Study of Mass Composition of Ultra-High Energy Cosmic Rays with AugerPrime using Machine Learning

von

Akash Parmar

Masterarbeit in Physik

vorgelegt der
Fakultät für Mathematik, Informatik und Naturwissenschaften
der
Rheinisch-Westfälischen Technischen Hochschule Aachen

im November des Jahres 2023

angefertigt am

III. Physikalischen Institut A

Erstgutachter und Betreuer

Prof. Dr. Thomas Hebbeker
III. Physikalisches Institut A
RWTH Aachen

Zweitgutachter

Prof. Dr. Christopher Wiebusch
III. Physikalisches Institut B
RWTH Aachen

ABSTRACT

The understanding of the mass composition of ultra-high energy cosmic rays offers a better understanding of the extreme events in the universe however it remains one of the biggest open questions in astroparticle physics. In the pursuit of an improved understanding of primary composition at the highest energies, the Pierre Auger Observatory is undergoing an upgrade known as AugerPrime. This observatory specializes in the study of cosmic rays, particularly focusing on the extensive air showers they generate within the Earth's atmosphere. This enhancement is primarily driven by the objective to deepen the understanding of cosmic rays and their interactions. The Scintillator Detectors installed on seventy-seven Water Cherenkov Detectors known as Pre-Production Array facilitate the development of new and improved methods to study cosmic rays.

In this thesis, previously developed observables based on the differences between signals at Water Cherenkov Detectors and Surface Scintillator Detectors are exploited using the Random Forest Regression machine learning method to understand the mass composition of ultra-high energy cosmic rays at the Pierre Auger Observatory. The data from the observatory were accompanied by the simulation data of the extensive air showers to train the machine learning model.

The results do not provide good discrimination of the mass composition of the ultra-high energy cosmic rays at the observatory but it is in acceptable agreement with the results from the Telescope Array located in Utah, U.S.A.

ZUSAMMENFASSUNG

Die Erforschung der Massenzusammensetzung ultrahochenergetischer kosmischer Strahlen ermöglicht tiefere Einblicke in die extremen Ereignisse des Universums. Diese Forschung stellt eine der zentralen Herausforderungen in der Astroteilchenphysik dar. Um ein detaillierteres Verständnis der primären Zusammensetzung bei höchsten Energien zu erlangen, erfährt das Pierre-Auger-Observatorium aktuell eine Erweiterung durch das AugerPrime-Upgrade.

Das Observatorium, das sich auf die Analyse kosmischer Strahlen spezialisiert hat, insbesondere auf die Untersuchung der ausgedehnten Luftschauer in der Erdatmosphäre, wird durch AugerPrime eine signifikante Weiterentwicklung erfahren. Die Hauptmotivation hinter diesem Upgrade ist es, unser Wissen über kosmische Strahlen und ihre Wechselwirkungen zu vertiefen. Ein Schlüsselelement des Upgrades sind die auf siebenundsiebzig Wasser-Tscherenkow-Detektoren installierten Szintillatordetektoren, das sogenannte Pre-Production Array. Diese ermöglichen die Entwicklung innovativer Methoden zur Analyse kosmischer Strahlen.

In dieser Dissertation wird die Massenzusammensetzung ultrahochenergetischer kosmischer Strahlen am Pierre-Auger-Observatorium untersucht. Dabei kommen neu entwickelte Beobachtungsmethoden zum Einsatz, die auf den Unterschieden zwischen den Signalen der Wasser-Tscherenkow-Detektoren und den Oberflächen-Szintillatordetektoren basieren. Unter Verwendung der Random-Forest-Regression, einer Methode des maschinellen Lernens, werden diese Beobachtungen analysiert. Ergänzt werden die Daten des Observatoriums durch Simulationsdaten der ausgedehnten Luftschauer, um das maschinelle Lernmodell zu trainieren.

Die Ergebnisse dieser Arbeit bieten zwar keine eindeutige Differenzierung der Massenzusammensetzung ultrahochenergetischer kosmischer Strahlen am Observatorium, zeigen jedoch eine akzeptable Übereinstimmung mit den Befunden des Telescope Array in Utah, USA.

CONTENTS

ACRONYMS

ADST Advanced Data Summary Tree.

AGN Active Galactic Nuclei.

AMIGA Auger Muon Detectors for the Infill Ground Array.

AMS Alpha Magnetic Spectrometer.

BDT Boosted Decision Trees.

CDAS Central Data Acquisition System.

CIC Constant Intensity Curve.

CMB Cosmic Microwave Background.

CORSIKA COsmic Ray SImulations for KAscade.

CR Cosmic Ray.

DNN Deep Neural Networks.

EAS Extensive Air Shower.

EBL Extra Galactic Background.

FADC Flash Analog-to-Digital Converter.

FD Fluorescence Detector.

FoV Field of View.

GZK Greisen–Zatsepin–Kuzmin.

ISOMAX Isotope Magnet Experiment.

KASCADE KArlsruhe Shower Core and Array DEtector.

LAT Large Area Telescope.

LDF Lateral Distribution Function.

LHC Large Hadron Collider.

MF Merit Factor.

MIP Minimum Ionizing Particles.

MoPS Multiplicity of Positive Steps.

MVA Multivariate Analysis.

NKG Nishimura-Kamata-Greisen.

PAO Pierre Auger Observatory.

PCA Principal Component Analysis.

PMT Photomultiplier Tubes.

PPA Pre-Production Array.

RF Random Forest.

RFECV Recursive Feature Elimination with Cross-Validation.

RMSE Root Mean Square Error.

ROC Receiver Operating Characteristic.

RoC Radius of Curvature.

SD Surface Detector.

SNR Supernova Remanant.

sPMT Small Photomultiplier Tubes.

SSD Surface Scintillator Detector.

TA Telescope Array.

THR Threshold.

ToF Time of Flight.

ToT Time over Threshold.

ToTD Time of Threshold Deconvoluted.

TSR Total Signal Ratio.

UB Unified Board.

UHECR Ultra-High Energy Cosmic Ray.

UMD Underground Muon Detector.

UUB Upgraded Unified Board.

VEM Vertical Equivalent Muons.

WCD Water Cherenkov Detector.

XGB eXtream Grdient Boost.

CHAPTER 1

INTRODUCTION

Understanding the Universe has been one of the fascination for humans. Astroparticle physics is a relatively newly developed field out of this long conquest. With all the developments in particle physics and astrophysics, this field has grown quickly in the past hundred years. This progress has opened the door to many new questions which are still open to be answered. One of the many unknown quests is to find ***"the composition of the ultra-high energy cosmic rays"***, which is the cosmic rays at the highest energies; this thesis is an attempt to answer this question. This is a particularly important question, as it can help us to understand its sources and acceleration mechanism, which are some of the most violent events in the universe.

There are many different observatories with different detection techniques deployed around the world to observe cosmic rays. One such observatory is the Pierre Auger Observatory in Argentina. Besides being the largest cosmic ray observatory in the world, it detects cosmic rays through multiple detection methods such as Surface detectors for the measurement of cosmic ray secondary particles (secondaries)[1] and Fluorescence Detectors to detect fluorescent light produced as the secondaries interact with the earth's atmosphere. The Fluorescence Detector at the Observatory currently has the highest shower-by-shower mass resolution of the cosmic ray primary particle, but it has a low duty cycle, as it can only measure during clear nights with very low background lights, which results in low statistics. Primarily Surface Detectors have Water Cherenkov Detectors(WCDs) and Surface Scintillator Detectors(SSDs) were installed on top of the WCDs as a part of an upgrade of the observatory, the so-called —*AugerPrime*. Since Surface Detectors have almost 100% duty cycle it is beneficial to explore the most properties of the cosmic rays using the Surface Detectors only which is why the work carried out in this thesis is only done using the data from SSDs and WCDs.

In Chapter 2 foundations of Ultra-High Energy Cosmic Rays have been laid.

[1]Secondary particles are produced when cosmic ray particles accelerate from sources and interact with Earth's atmosphere.

The composition of the UHECRs, which is the motivation for this thesis work is introduced in this chapter. The different detection methods with different intentions to study cosmic rays are briefly introduced. Also, the overview of the sources that produce UHECRs are given in this chapter.

In Chapter 3 we exclusively discuss the Pierre Auger Observatory. Since Pierre Auger Observatory utilizes multiple detection techniques, each of them is mentioned in this chapter. Chapter 4 follows from Chapter 3 as the reconstruction process of the events observed with the WCDs and SSDs is discussed in this chapter.

The simulation data are crucial in this thesis work. However, a deeper understanding of this topic is not necessary which is why it is mentioned in brief. Also, a discrepancy in the hadronic interaction model is discussed. The pre-processing of the data generated using this simulation software is carried out in chapter 6.

A detailed exploration of the observables at the Pierre Auger Observatory using the simulation software is executed in Chapter 7. The exploration of observable also involves training and testing multiple Random Forest regression machine learning models. A set of observables that performs best in the simulations training and testing are used with different simulated datasets in Chapter 8, where the trained model is applied to the data measured at the Pierre Auger Observatory to predict the mass composition of the Ultra-High Energy Cosmic Ray.

CHAPTER 2

LOW AND HIGH ENERGY COSMIC RAYS

2.1 Discovery of Cosmic Rays

In the early 20th century, there was a rising interest of many scientists in the phenomena of radioactivity after its discovery in 1896 by Henri Becquerel. One of the important instruments used for its study was the electroscope, which was also the instrument used over several balloon flights by Victor Hess in 1912 to study the change of radioactivity in the Earth's atmosphere as a function of altitude. With increasing height, he observed first a decrease and then a rapid increase in radiation, he concluded that particles from an extraterrestrial source ionize molecules as they enter our atmosphere[1]. This result was further confirmed through similar experiments conducted by W. Kolhörster at higher altitudes up to 9300m (see figure-2.1) and also by Robert Millikan performed further measurements using electroscopes above the below the surface of high altitude lakes in the Sierra Nevada and observed a continued but slower decrease in radiation with increase in depth. The term *Cosmic Rays* was coined by Millikan and Hess received the Nobel Prize in 136 for its discovery.

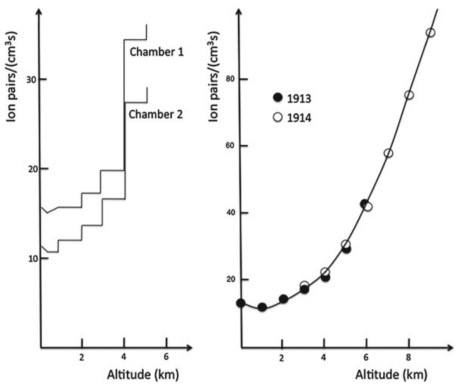


Figure 2.1: Change in ionization with altitude. Left: Results from two ion chambers by Hess(1912). Right: Results by Kolhörster (1913,1914) at higher altitude.[2]

2.2 Energy Spectrum of Cosmic Rays

The energy spectrum of cosmic rays, also referred to as flux, has a very wide range of energy, from 10^9 eV to 10^{20} eV. It is strongly dependent on its energy and diminishes strongly with an increase in the energy. At low energies (10^9 eV) we observe roughly 1,000 cosmic rays per m^2 per second, whereas at 10^{20} eV the rate falls to one particle per km^2 per century, making measurement of cosmic rays at the highest energies difficult.

The very low-energy cosmic rays (below $\sim$ 30 GeV)[1] traveling towards Earth are subject to *solar modulation*. The heliosphere is transparent for higher-energy particles. The predominant and most significant periodic variation in cosmic ray intensity is closely linked to the 11-year *solar cycle*. This recurring pattern finds compelling evidence in the historical record of sunspots dating back to the early 1600s and in the observed galactic CR intensity since the 1950s, monitored through a worldwide network of ground-based neutron detectors [31]. The modulation in CR flux is anti-correlated to the solar cycle. During the maximum solar activity (solar maximum) the intensity of cosmic rays reduces and vice-versa.

The flux is generally described by a broken power law of the form,

$$\frac{dN^4}{dE\ dA\ dt\ d\Omega} \propto E^{-\gamma} \tag{2.1}$$

where N is the number of observed cosmic rays at energy E within a solid angle Ω, Effective detection area A of an experiment, and exposition time t. γ is the spectral index that represents changes in the spectrum due to changes in physical phenomena such as propagation or source of the cosmic rays. To highlight features, the flux is usually scaled by energy; in Figure 2.1 a scaling factor of $E^-2.6$ is used, where the three most prominent visible features of the spectrum are,

Knee: Below 3×10^{15} the spectrum is relatively flat and steepens above, the γ changes from $\approx$ 2.7 to $\approx$ 3.1. This is often attributed to the maximum acceleration energy of the galactic Supernova Remanant (SNR) for lighter nuclei such as protons[2]. This has been confirmed by the *Karlsruhe Shower Core and Array Experiment* (KASCADE)[4].

Second Knee: A less distinctive, *second knee* at the energy $\sim 10^{17}$ eV where spectral index changes to $\gamma \sim 3.3$ as obsevered by KASKADE-Grande [5]. It is theorized that similar to the first knee, this is the maximum acceleration limit for heavier iron-like nuclei.

Ankle: At $\sim 5 \times 10^{18}$ eV hardening of the spectrum is observed, changing the spectral index to $\gamma \sim 2.6$. A favored explanation is the change of CR accelerators from galactic to extragalactic [6] and a pile-up effect from electron-positron pair production due to protons interacting with CMB photons [7].

[1] Not shown in the Figure 2.1

[2] More on the composition of CR in the next section.

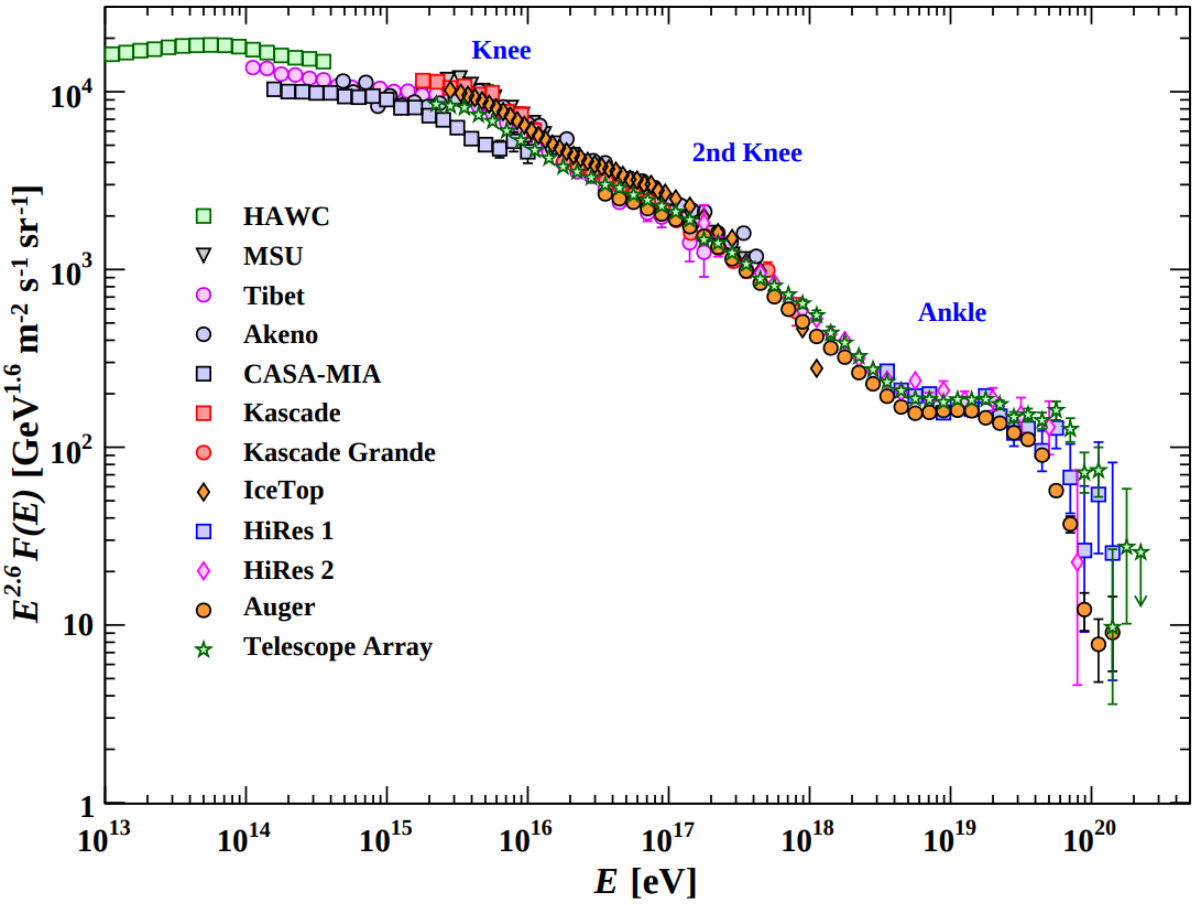


Figure 2.2: All particle flux of the cosmic rays as the function of its energy, obtained using various experiments which are denoted with different markers. To highlight the features the spectra are scaled by $E^{2.6}$. The verticle lines denote the statistical uncertainties which are only visible at the highest energies because the flux is falling steeply.[3]

Isotropic distribution of CRs in this energy range also supports its extragalactic origin.

Cut-off: At energies above $\sim 5 \times 10^{19}$ eV energy spectrum is strongly suppressed. The reason for this suppression is still unclear but the leading hypothesis is either GZK cut-off in which CR protons lose energy by interacting with CMB photons, photo-disintegration of heavier nuclei due to EBL[3] or CMB, or there is no natural accelerator of cosmic rays beyond this energy.

The most recent measurement of UHECRs at the Pierre Auger Observatory can be seen in the figure: 2.3 where features of Ankle and beyond are clearly visible. This is the region of interest for this thesis, and we would attempt to know the mass composition of this energy range.

[3]EBL is all the combined radiation due to stellar formation and AGN. It covers almost all the wavelengths of the electromagnetic spectrum, except microwave where CMB dominates.

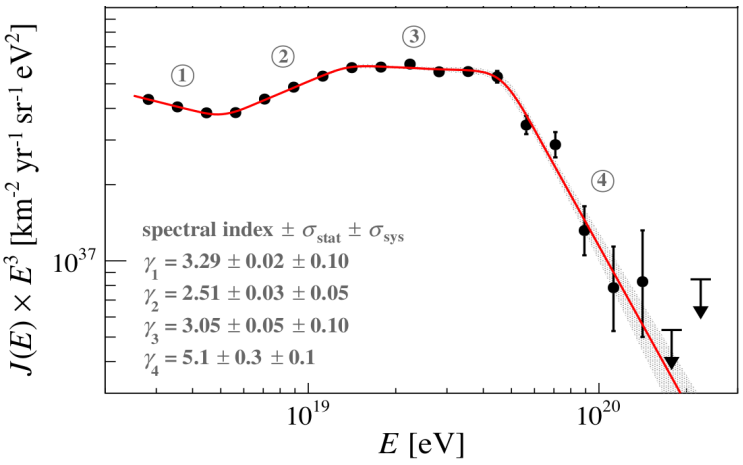


Figure 2.3: Spectrum of all particles measured at the Pierre Auger Observatory at and above the ankle. Numbers correspond to different spectral indices γ_i for different energy ranges.

2.3 Composition of Cosmic Rays

Due to high penetration power cosmic rays were generally believed to be γ radiation (In the early period penetrating power of relativistic charged particles was not understood). Millikan hypothesized that γ rays were produced in interstellar space when protons and electrons formed helium nuclei.

Using balloon-borne experiments, Jacob Clay made two voyages in 1927 and 1928 between Java and Geneva. He discovered that the ionization increased with latitude and proved that cosmic rays have to be mostly charged in nature as they interact with geomagnetic fields. In fact, the interaction of cosmic rays with geomagnetic fields led to the proof that cosmic rays are mostly positively charged which was observed by three independent experiments by Alvarez and Compton, Johnson, and Rossi. They discovered asymmetry in the cosmic rays close to the equator. The cosmic rays were coming from the West than from the East. This is generally known as the *East-West effect*.

In modern astrophysics, the mass composition of cosmic rays plays a crucial role. The relative abundance of cosmic rays allows us to get insights into the sources and acceleration mechanisms of the cosmic rays, shedding light on the energetic processes that occur in the supernova, pulsars, Active Galactic Nuclei, and other cosmic events. Also, with cosmic rays we can probe the physics at energies higher than current achievable limits by experiments in the laboratories. such as LHC, Geneva.

In recent decades, cosmic rays up to a few hundred TeV have been intensively studied with balloon- and satellite-borne experiments, known as *direct detection methods* (see Section 2.5). The observed cosmic rays contain $\sim$ 86% protons, $\sim$ 12% helium, $\sim$ 12% other heavier nuclei, and $\sim$ 2%[8]. The cosmic rays spectrum of different elements which is observed by different experiments is shown in figure 2.4.

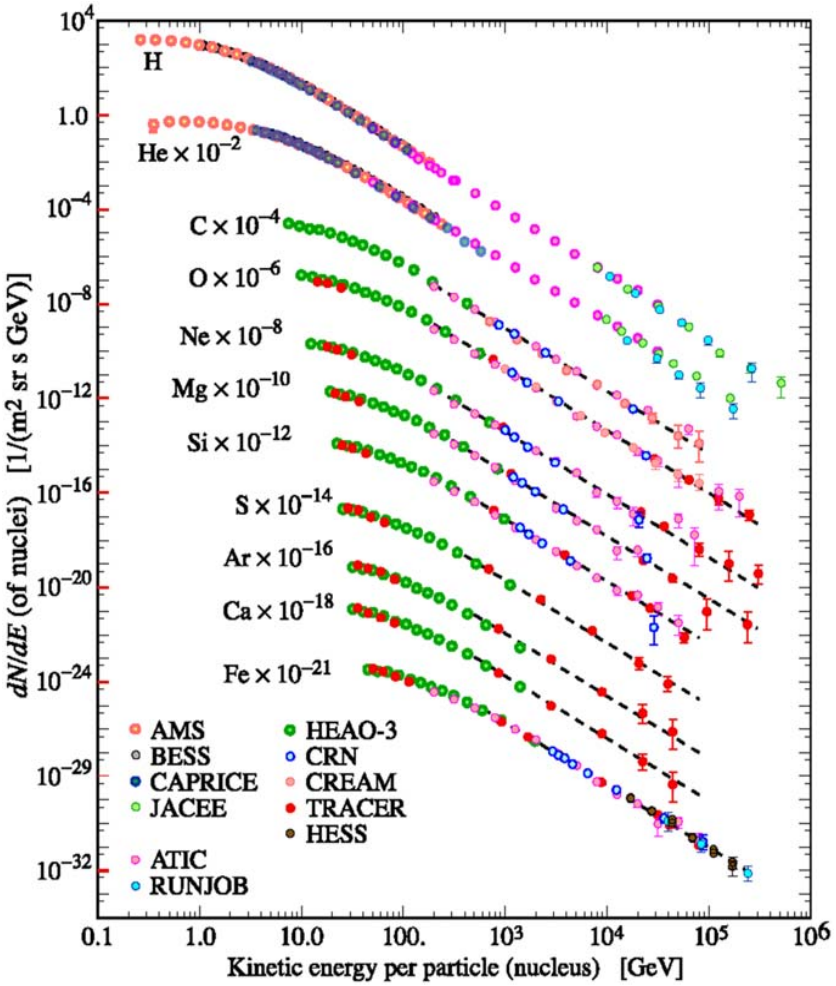


Figure 2.4: Differential spectrum of different elements as measured by different experiments. Fluxes are multiplied with different factors for better visibility. The fluxes have a constant power-law energy spectral index of ~ -2.7 and a constant abundance ratio independent of energy [9].

2.3.1 Composition of UHECRs

Above a few hundred TeV energy, the direct detection of the cosmic rays is not feasible, as it would require larger detectors to deposit the complete energy of a particle which is challenging economically and technically. Also, extremely low flux would require even larger detectors. In this case, we rely on indirect measurement of cosmic rays via EAS. The indirect detection is model-dependent, as the interaction processes in the EAS must be simulated, which depends on different particle interaction models [10, 11, 12]. These models are extrapolated from the experiments in the Earth-based laboratories, where particle energies are comparatively very low ($\sim$ 14 TeV at LHC, Geneva), this leads to uncertainties in the mass of the primary cosmic rays.

One way to indirectly determine the composition of cosmic rays involves deducing the mass of primary cosmic rays by examining variations and ratios among the three components of particle showers detected. The KArlsruhe Shower Core and

Array DEtector (KASCADE) experiment [13] effectively employed this approach, wherein they separately measured the electromagnetic and muonic components and measured all charged particles with its extension called KASCADE-Grande. To achieve this they utilized a combination of distinct detectors, each primarily sensitive to either the component of a shower. By analyzing the responses of these detectors, they reconstructed the quantities of particles: $N_{e/\gamma}$, N_μ, N_{ch}. The relationships between the electromagnetic and muonic component ratios, $N_{e/\gamma}/N_\mu$ (as observed in KASCADE), and the charged and muonic component ratios, N_{ch}/N_μ (as seen in KASCADE-Grande), are responsive to varying primary masses. This sensitivity arises because the primary particles having the same $N_{e/\gamma}$ or N_{ch} values will exhibit different N_μ values. Also, variations in N_{ch} or $N_{e\gamma}$ for events with the same primary mass tend to exhibit similar variations from shower-to-shower. Consequently, it becomes feasible to reconstruct the primary composition on an individual event basis. A parameter k is computed for various zenith intervals to facilitate this analysis:

$$k = \frac{log_{10}(N_{ch}/N_\mu)_{event} - log_{10}(N_{ch}/N_\mu)_H}{log_{10}(N_{ch}/N_\mu)_{Fe} - log_{10}(N_{ch}/N_\mu)_H} \tag{2.2}$$

Where the parameters $log_{10}(N_{ch}/N_\mu)_H$ and $log_{10}(N_{ch}/N_\mu)_{Fe}$ are determined through Monte Carlo simulations, and their values are based on the specific hadronic interaction model used. The parameter k is defined such that it converges to a value of approximately one when the primary particle is iron and zero when the primary particle is a proton. Depending on the value of k calculated for each event, we can categorize each event as having a lighter composition (H+He) or a heavier composition (C+Si+Fe).

The composition outcomes derived from the combined data of KASCADE and KASCADE-Grande using this approach are presented in Figure 2.5 for two different hadronic interaction models. These results confirm the initial knee structure occurring at $\sim 2 \times 10^{15}$ eV for lighter elements and the second knee for the heavier elements at around 10^{17} eV. Furthermore, the data also indicate a rise in the proportion of lighter elements following the second knee.

Measurement of the longitudinal parameter X_{max} provides another promising method to study the composition of cosmic rays, which will be mentioned in the section 2.4.1. This observable is effectively exploited at the Pierre Auger Observatory through its fluorescence detectors [15, 16] but it is not part of this thesis. The Fly's Eye experiment and the Telescope Array also use it extensively [17, 18].

Under the assumption of mixed particle composition, this distribution results from the combination of individual distributions, $f_i(X_{max})$, generated by various primary particles denoted as i with mass A_i. This combination involves multiplying each individual distribution by the fraction of the total flux attributed to that specific primary particle represented as p_i:

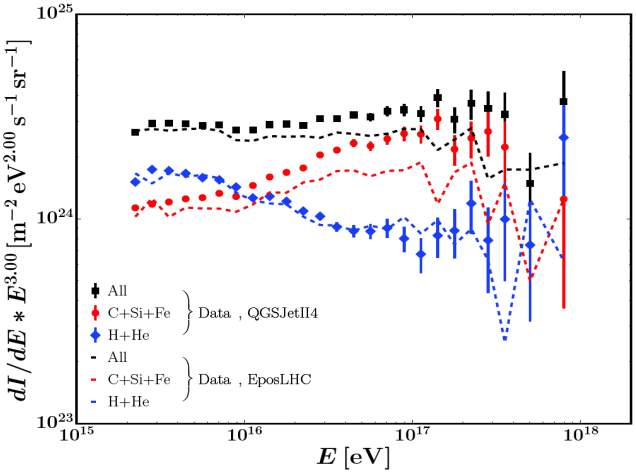


Figure 2.5: Cosmic ray spectrum as determined by KASCADE+KASCADE-Grande. In this representation, the black squares depict the measured spectrum for all particles, while the blue diamonds and red circles correspond to the components associated with lighter and heavier masses, respectively. Taken from [14].

$$f(X_{max}) \;=\; \sum_i p_i f_i(X_{max}) \tag{2.3}$$

The mean $<X_{max}>$ and standard deviation $\sigma(X_{max})$ of this are in linearity with the logarithm of mass of the primary particle, ln(A) [15]. Recent results on the $<X_{max}>$ and $\sigma(X_{max})$ from Pierre Auger Observatory are shown in the figure 2.6, where both values are a function of the energy of primaries which has been inferred to the mass of primaries with the selected hadronic model [19]. The results of these findings are in accordance with the results of KASCADE. The composition becomes less massive as we approach the ankle region, where a notable proportion consists of protons and as energy levels rise beyond the ankle, the composition progressively becomes heavier once more. As mentioned before a limiting factor of these results is low statistics in this energy range. One of the key motivations for upgrading the Pierre Auger Observatory to *AugerPrime* is the imperative need to improve these high-energy statistics. AugerPrime aims to provide information on both the mass composition and the origins of the reduced flux of UHECR. It also seeks to enhance our understanding of the hadronic interactions occurring in EAS at energy levels far surpassing those achievable in collider experiments.

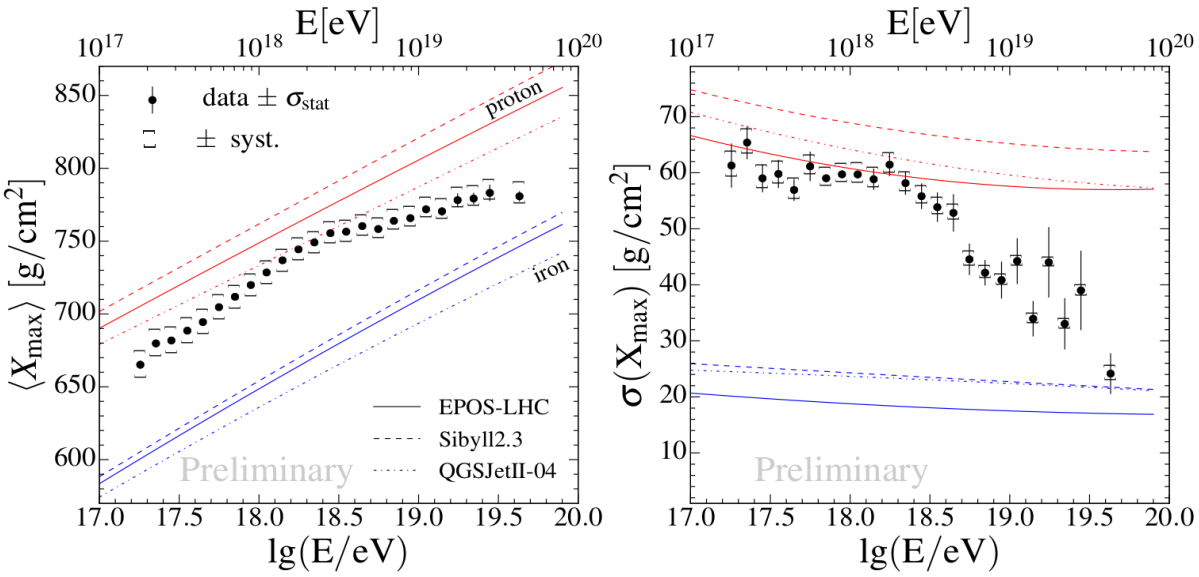


Figure 2.6: The $<X_{max}>$ (left) and $\sigma(X_{max})$ (right) observed at Pierre Auger Observatory, showing X_{max} getting deeper towards the ankle and shallow afterward indicating composition is lighter before ankle and heavier after ankle. Taken from [19].

2.4 Extensive Air Showers

As mentioned before, cosmic rays at the highest energies are not feasible with direct detection techniques; in this case, indirect methods are used. These indirect detection experiments are ground-based. Via nuclear interactions, the primaries usually interact with the molecules in the Earth's atmosphere at an altitude between 15 to 35 km [20], and they have enough energy to start a chain reaction via momentum transfer which creates a cascade of billions of particles that are called *cosmic ray secondary particle* or *secondaries*. This induced chain secondaries is known as *Extensive Air Shower* (EAS). A schematic view of an EAS is shown in Figure 2.7.

The first interaction, X_0 of a constant and isotropic flux of cosmic rays with the Earth's atmosphere (N_2, O_2, Ar, etc.) strongly depends on the mass and energy of the cosmic ray. In the first interaction, most of the particles produced are pions (π^+,π^- and π^0), and kaons(K^+,K^- and K^0) in smaller fractions. The charged and the neutral pions are produced roughly in equal numbers. Neutral pions predominantly decay into pairs of photons which leads to the generation of electrons and positrons through *pair-production*. These newly generated particles, in turn, produce additional photons through the *bremsstrahlung process*, creating a cyclical alternation of these two processes. The last two mentioned processes repeat in alternation, leading to $e^{\pm}$ and γ as the majority content of the EAS which is known as *electromagenetic component*.

When charged pions or charged kaons decay before the next interaction, they produce muons, which are highly likely to survive until they reach the Earth's

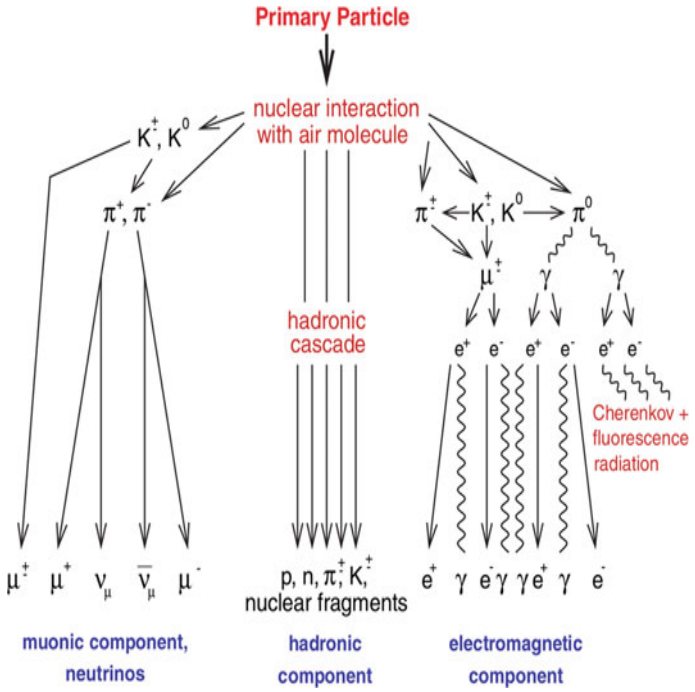


Figure 2.7: The development of Extensive Air Shower as induced by cosmic ray. The secondaries in EAS can be categorized into three components: the muonic components(including neutrinos), the hadronic, and the electromagnetic components. Taken from [21].

surface which is termed as *muonic component* of EAS. If a muon decays in the air before reaching the surface, it decays into an electron ($\mu^{\pm} \rightarrow e^{\pm} + (\nu_e/\overline{\nu}_e) + (\overline{\nu}_\mu/\nu_\mu)$), thereby contributing to the electromagnetic component. In the decay process of charged pions and kaons, along with muons, neutrinos are also produced which almost always remain undetected and are considered part of the missing energy.

High-energy pions and kaons with extended lifetimes, tend to interact with other atmospheric nuclei before undergoing decay which consequently produces more pions. The other fragments of the primary cosmic rays particle and also protons and neutrons are generated during these interactions further interact to produce more hadronic particles which is known as the *hadronic component*. Most of these hadronic particles ultimately interact or decay, contributing to the electromagnetic and muonic components.

As the EAS traverses the atmosphere, an increasing number of particles are generated, which share the initial energy of the primary cosmic ray. Over generations of interactions, these secondaries will reach an energy level that is insufficient to further induce a nuclear interaction. This reduction in energy results in more absorption of particles in the atmosphere compared to their production. The distance at which this transition occurs in the atmosphere is called the depth of *shower maximum*, X_{max}. This is named as such because this altitude has the highest number of particles produced due to the shower. X_{max} can provide valuable insight into our understanding of UHECR as it is directly correlated with

the energy and mass of the primary cosmic ray. Measurement of X_{max} is one of the main purposes of Fluorescence Detector at the PAO. Although X_{max} is an important variable in determining the mass of the primaries it is not used in this thesis and is considered as a part of possible future work.

2.4.1 Characterization of an EAS

A basic illustration of an Extensive Air Shower (EAS) in Figure 2.8. As visible in the figure, from the point of first interaction the vector along which a shower progresses is known as the *shower axis* and its central region as the *shower core*. These are generally the references to define other geometric features of an air shower. The shower axis is determined by the position of the shower core and the angle at which the primary particle enters the atmosphere, i.e., its zenith angle (θ) and azimuth angle (ϕ). The majority of the hadronic component develops along the shower axis. At any given moment during its evolution, the EAS appears as a dense disc of particles with a thickness of a few meters referred to as the *shower front*. The randomness in the interaction of particles provides the thickness and the curve, with the shower axis as the average direction. Consequently, the shower front is narrow near the core and thicker at further distances.

Typically, an important parameter for describing a hadron-induced air shower is the *slant depth* X, which quantifies the volume of atmosphere the shower has traversed up to a specific point in its development. For vertical showers, this quantity is called X_v, which is computed by integrating the air density from the top of the atmosphere to an altitude h

$$X_v(h) \;=\; \int_{\infty}^{h} \rho(h')\; dh', \tag{2.4}$$

it has a unit of g/cm^2. Scaling the vertical depth with the zenith angle θ of an EAS, the slant depth can be obtained as

$$X \;=\; \frac{X_v}{cos\ \theta}, \tag{2.5}$$

where $\theta = 0$ if the shower axis is exactly above the detector.

The crossing of the whole atmosphere is similar to $\sim$1000 $g \cdot cm^{-2}$ and the X_{max} value for hadron-induced showers at the highest energies typically falls within the range of 700 to 850 $g \cdot cm^{-2}$), indicating that these showers tend to reach their maximum at a shallower depth in the atmosphere. Detecting the shower closer to its maximum point enables more precise measurements since the number of secondaries is at its peak. Consequently, cosmic ray experiments are frequently conducted in elevated areas such as mountain regions or high plateaus, as exemplified by the Pierre Auger Observatory, positioned at an atmospheric depth of $\sim$870 $g \cdot cm^{-2}$ [20].

The longitudinal development profile depicts the evolution of the number of particles, *N(X)*, as the EAS progresses. This is indicated in Figure 2.8 by the

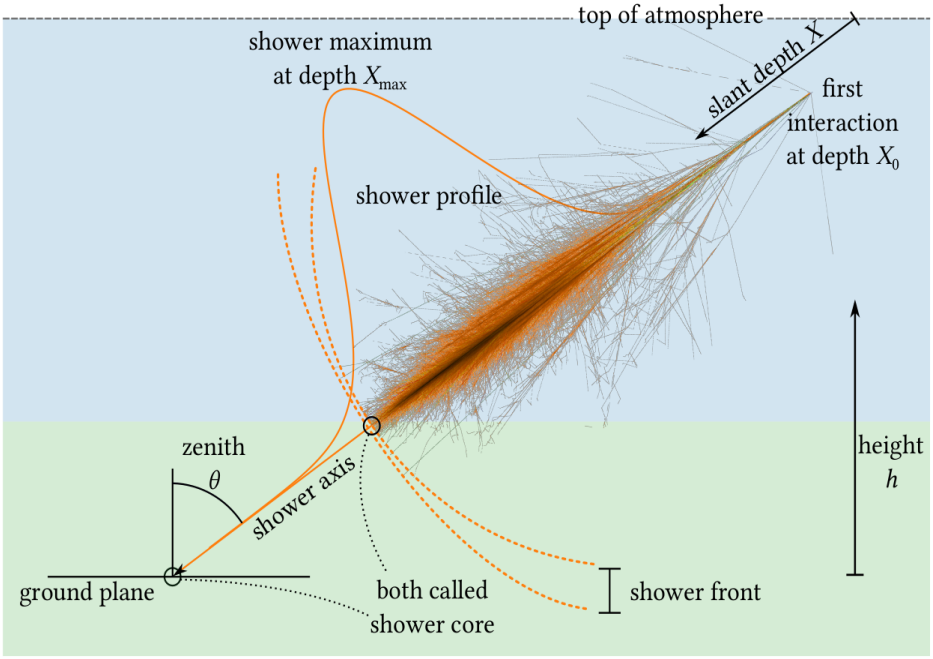


Figure 2.8: A schematic diagram of an EAS. A cosmic ray entering the Earth's atmosphere at an inclined angle produces an EAS. Taken from [22].

orange curve above the shower. This profile also called the *shower profile*, can be effectively characterized using the Gaisser-Hillas function [23],

$$N(X) = N_{max}\left(\frac{X - X_1}{X_{max} - X_1}\right)^{\frac{X_{max}-X_1}{\lambda}} exp\left(\frac{X_{max} - X_1}{\lambda}\right) \tag{2.6}$$

where X_1 and λ are fitting parameters depending on the mass and energy of the primary particle and N_{max} is the maximum number of particles at X_{max}. The energy contained in an EAS can be obtained by integrating its shower profile.

We do not directly measure all the particles in the X_{max} but only when they reach the ground, this is called the *shower footprint*. Ground-based cosmic-ray observatories mainly exploit the arrival time and density of the particle to reconstruct and classify EAS. The arrival time of the shower footprint at different positions on the ground can be used to reconstruct the shower geometry. With the available geometry information, the density of the arriving particles can be correlated with their radially outward distance from the shower core on the ground to derive the Lateral Distribution Function (LDF) and with this, the shower energy can be reconstructed. Typically, the LDF is fitted with an NKG(Nishimura-Kamata-Greisen)-like [24, 25] function. NKG function can be expressed by the particle density $\rho(r)$ as a function of distance from the shower axis r,

$$\rho_e(r) = \frac{N_e}{2\pi R_M^2} C(s)\left(\frac{r}{R_M}\right)^{(s-2)}\left(\frac{r}{R_M} + 1\right)^{(s-4.5)} \tag{2.7}$$

where N_e is the total number of electrons and R_M is the Moliere radius. The parameter s is the *shower age* which is defined as,

$$s \equiv \frac{3X}{X + 2X_{max}} \tag{2.8}$$

The detection of a EAS using the combination of SSDs and WCDs at Pierre Auger Observatory can distinguish muonic and hadronic components from electromagnetic components (See sections 3.1, 3.3.1 and chapter **??**). Crucially, all three components of an EAS evolve differently, leading to different lateral particle density distributions and, consequently, different shower footprints. Given that the relative fraction of particles in each component correlated with the composition of the cosmic ray primaries, distinguishing the footprints of these components allows for the reconstruction of the initial particle type using a ground array, a key aspect addressed in this thesis.

2.5 Detection of Cosmic rays

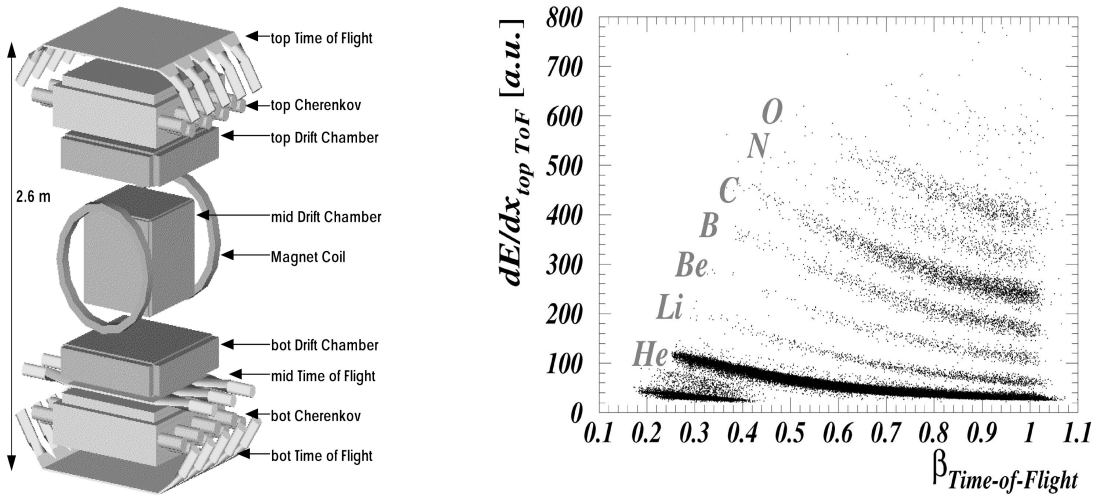
Numerous experiments with different techniques have been performed in the past to detect cosmic rays [26]. Cosmic ray experiments can be classified mainly into two categories: *direct methods* and *indirect methods*.

Instruments designed for direct detection typically incorporate a calorimeter that registers ionization upon interaction with a cosmic ray. However, due to the sharply declining flux of high-energy CRs and the inherent limitations of these instruments in terms of size, they are not well-suited for the detection of the most energetic extragalactic CRs. Nevertheless, they do offer precise insights into the characteristics and origins of cosmic rays up to energies of approximately ∼10 TeV. For instance, the balloon-borne ISOMAX experiment uses a complex multi-detector method which can be seen in Figure 2.9a. Under the effect of the magnetic field inside the detector, the particles follow a helical trajectory with a radius, R_{gyro}, which is related to momentum by the Lorentz force:

$$\gamma mv \;=\; p \;=\; BzeR_{gyro} \tag{2.9}$$

where B is applied magnetic field, z, p and e are mass, momentum, and charge of the particle, respectively. The results from this experiment are shown in Figure 2.9b where the separation between different particles is clearly visible. The same Time of Flight (ToF) principle is applied in the AMS experiment [27], which operates aboard the International Space Station (ISS). The Physics Department of RWTH Aachen University is actively involved in this collaborative effort, and at present, its successor, AMS-2, is in operation.

Another pioneering experiment is the Fermi-LAT instrument onboard the Fermi satellite. This is currently the most sensitive instrument for cosmic ray energies between ∼100 MeV to ∼100 GeV. The detection method relies on the conversion of incoming photons into an electron-positron e^+e^- pair. This detector consists



(a) (b)

Figure 2.9: (left) The schematic of ISOMAX balloon experiment. (right) Measured charge separation using the ToF method. For the clear identification a threshold was set to the velocity, $\beta = 0.4$. Obtained from [28].

of a series of silicon strip detectors arranged with alternating orientations in the x and y axes. Incoming photons interact with thin tungsten foils, which are interleaved between the silicon layers. The tracking of charged particles enables the reconstruction of the initial photon's direction. Beneath the tracker, an electromagnetic calorimeter is employed to measure the total energy of the e^+e^- pair. Some possible candidates of SNRs observed with Fermi-LAT are shown in Figure 2.10.

Direct observations are capable of measuring primary energies up to approximately 1 PeV. Beyond this energy range, observations are conducted by utilizing the EAS generated in the atmosphere due to interactions with cosmic rays.

2.5.1 Detection of UHECRs

The significant reduction in the cosmic ray flux (Φ) as a function of energy, which can be roughly approximated as $\Phi \propto E^{-3}$, presents a substantial challenge in terms of the required dimensions and operational duration of experimental setups when investigating high-energy phenomena. For the measurement of UHECRs we deploy instruments to detect EAS. As we have seen in Section 2.4.1 EAS is a feature-rich phenomenon and for this different detection methods are used to detect different parts of the shower. Along with the shower, it is necessary to know the weather conditions at the detector location, as it can fluctuate development of EAS. Presently, there are four main techniques used for measuring EAS.

1. Fluorescent Technique

As charged particles within an EAS traverse the atmosphere, they cause ion-

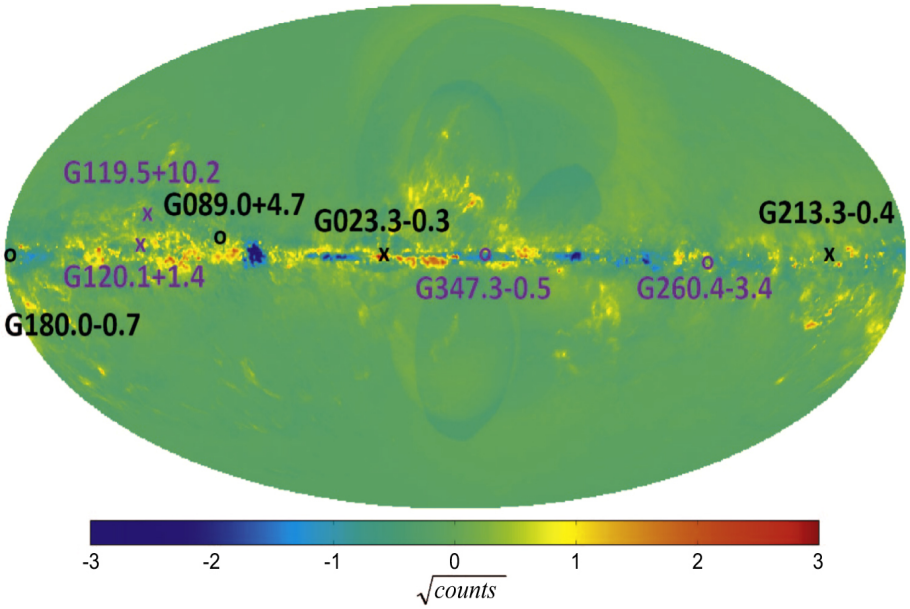


Figure 2.10: Some SNRs as observed by the Fermi-LAT. The selected SNRs represent all combinations of hard (purple) and soft (black), point-like (x), and extended (o) sources. The diffused γ-radiation can also be noticed as background [29].

ization of the atmospheric gas molecules, predominantly nitrogen. Subsequently, as these molecules de-excite, visible and ultraviolet radiations in the wavelength range of 300-400 nm are emitted, giving rise to what is known as the fluorescence light associated with the shower [30]. A schematic of the fluorescent technique is shown in Figure 2.11, together with the other detection methods.

It is important to note that the number of fluorescence photons emitted is relatively low, typically on the order of a few photons per electron per meter in air. Consequently, the fluorescence technique is primarily applicable at high energies ($E > 10^{17} eV$). Unlike Cherenkov photons (discussed below), this method is not directional, making it suitable for fortuitous observations.

Fluorescence due EAS happens isotropically, indicating that the telescope can effectively capture the shower's light emissions even from a few kilometers away, without the necessity of being directly underneath it. The Fluorescence detectors generally have a large field of view, which provides a large aperture.

The amount of fluorescent light emitted is proportional to the energy of the primary particle and the light emitted at a particular moment is proportional to the energy along the shower axis at that time, through this property the longitudinal profile of a shower is observed. Consequently, X_{max} is obtained, which is an important variable directly associated with the composition of the primary particle.

As mentioned weather conditions play an important role. Detecting fluorescence light requires conditions of darkness, which means that it can only be utilized during nighttime when there is an absence of moonlight and minimal light pollution. Additionally, the presence of clouds obstructs light, which brings a large

accuracy in the measurement of the longitudinal shower profile. This reduces the operating window of the fluorescence detectors limiting to the duty cycle of approximately 15%..

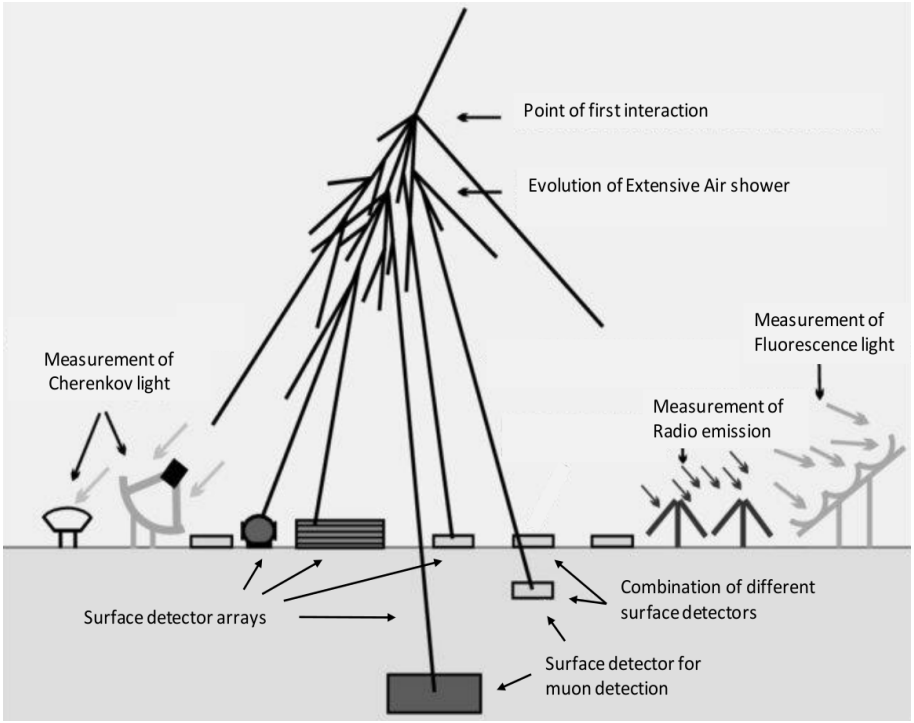


Figure 2.11: A sketch showing all four different EAS techniques discussed above. The Cherenkov detectors are not used in the hybrid mode as depicted here. A combination of different surface detectors (surface + underground) to detect muon can also be seen. Obtained from [31].

2. Surface Detector Array

Arrays of surface detectors have been in use ever since the discovery of Extensive Air Showers. This methodology is employed in experiments like KASKADE and its successor, KASKADE-Grande, as well as in the Telescope Array (Telescope Array) and the Pierre Auger Observatory (Pierre Auger Observatory). By employing these arrays, one can conduct a fitting analysis of the lateral distribution function of the shower, which, in turn, facilitates the estimation of the shower size and, consequently, the energy contained within it.

The minimum detectable energy with the surface array is determined by the spacing between the detectors, while the total area encompassed by all the detectors establishes the upper limit for the observable energy by the array. For example, the surface array at the Telescope Array is optimized for energies greater than $10^{18.8}$ eV with a detector distance of 1.2 km [32], covering an area of 762 km^2, which facilitates to detection of UHECR with reasonable statistics.

Various surface detectors are typically specialized in the detection of charged particles, but they are unable to distinguish between different types of charged particles. This limitation can be addressed by combining different types of detectors at each station within the surface detector array, a simple payout of detec-

tors can be seen in Figure 2.11. While Water Cherenkov Detectors and Surface Scintillator Detectors are incapable of directly segregating the electromagnetic and muonic components, they exhibit varying interactions with different kinds of charged particles. Scintillators (SSD), due to their thin composition, demonstrate higher sensitivity to electrons compared to muons, as the latter interact less with matter. However, electrons are absorbed within the water tanks, leading to the predominance of muon signals in WCD.

The underground detector array such as AMIGA [33] at the Pierre Auger Observatory is generally considered a type of surface detector. These detectors are spaced a few meters below the surface to detect muonic showers.

3. Cherenkov Technique

Numerous secondary particles within an Extensive Air Shower (EAS) travel faster than the speed of light in the surrounding medium, producing Cherenkov light[4]that is detectable. At sea level, the Cherenkov angle θ_c in air, corresponding to a speed $\beta = 1$, is approximately 1.3 degrees, while it reduces to about 1 degree at an elevation of 8 km above sea level. For Cherenkov emission, the energy thresholds are 21 MeV for a primary electron and 44 GeV for a primary muon. About half of this emission occurs within 20 meters of the shower axis (around 70 meters for a proton shower). Given that the intrinsic angular spread of charged particles in an electromagnetic shower is roughly 0.5 degrees, the opening of the light cone is mainly governed by the Cherenkov angle. Consequently, the ground area illuminated by Cherenkov photons from a 1 TeV shower, referred to as the "light pool," has a radius of approximately 120 meters, featuring a relatively uniform photon density per unit area [8].

These ground-reaching photons can be concentrated using a mirror directed onto a camera in the focal plane. A larger mirror facilitates the collection of more photons and lowers the energy threshold for detection. Imaging the diffused shower results in an elliptical image of the shower's development within the camera. The origin of the photons aligns with the principal axis of this ellipse. Considering the measured arrival times in each pixel and the geometric orientation of the ellipse, the direction of the shower can be reconstructed. The shower's energy is determined by the number of photons registered.

This method can be optimized by simultaneous observation of the shower using multiple Cherenkov telescopes. An example of such a telescope system is the High Energy Stereoscopic System (HESS) in Namibia [34].

[4]The Cherenkov radiation or Vavilov–Cherenkov radiation occurs when a charged particle moves through a medium faster than the speed of light in that medium. The light is emitted in a coherent cone along the direction of the particle at an angle such that

$$cos\ \theta_c\ =\ \frac{1}{n\beta} \tag{2.10}$$

where n is the refractive index of the medium and the threshold of the velocity is $\beta\ =\ 1/n$. The number of photons emitted is directly proportional to the energy of the particle.

4. Radio Technique

Radio signals are emitted from EAS mainly by two effects: *Geo-synchrotron effect* and *Askaryan effect*. The radio signal is emitted forward and can be picked up at ground level using antennas typically operating in the frequency range of approximately 10 MHz to 300 MHz. The EAS strongly depends on the Earth's magnetic field and the distance and orientation of the receiving antenna on the ground. However, radio detection has a potential duty cycle of nearly 100% and the capability to determine X_{max}, which can yield higher statistics compared to the Fluorescent technique. The initial studies on radio signals emitted by air showers were conducted using the LOPES detector [35], which was an integral component of the KASCADE detector. Building upon these early findings, the AERA detector has been established as part of the Pierre Auger Observatory [36], representing the first comprehensive radio array to operate in hybrid mode with other detection techniques.

2.6 Sources of Cosmic Rays

Cosmic rays originate from diverse regions of the Universe. Particles with energies below a few GeV arrive from the regular stars which can be described by the Hertzsprung-Russell diagram, such as our sun and they are strongly modulated by solar activities, while particles with energies of $\sim 10^{18}$ eV must undergo acceleration processes within our galaxy. As we transition to higher energies, the sources of cosmic rays tend to be extragalactic, although the precise threshold energy marking this transition remains somewhat indistinct and subject to debate, allowing for potential overlap [37].

The prevailing notion suggests that galactic nuclei are accelerated at shockwaves within Supernova Remanant (SNR) through the first-order Fermi mechanism. Subsequently, these particles are somehow released into the Interstellar Medium (ISM), where they temporarily reside before escaping the confines of our Galaxy. This scenario aligns with the SNR paradigm for the genesis of galactic cosmic rays [38]. However, it is worth noting that certain types of expanding supernova remnants may have limitations in terms of their capacity to accelerate protons beyond energies around 10^{15} eV [3]. Potential sources of protons with energies in this range encompass massive star clusters, which have the capability to reach energies in the peta-electronvolt (PeV) range, as well as regions near the Galactic Center [39, 40]. However, the question persists regarding the extent to which the acceleration mechanisms operating within SNRs can elevate particle energies to the PeV range and beyond. Unraveling the true capability of SNRs in accelerating galactic cosmic rays to these extreme energies remains an open and challenging research endeavor.

2.6.1 Sources of UHECRs

While the sources of most cosmic rays, up to energies of approximately 10^{18} eV, have been reasonably identified, particularly as Supernova remnants [41], pinpointing singular sources of Ultra-High Energy Cosmic Ray (UHECR) remains an elusive endeavor [42]. Nonetheless, there exist several astrophysical entities considered as potential sources.

In the quest to discover the origins of UHECRs, the Hillas criterion emerges as a valuable description, serving as a link between the dimensions and magnetic field of the accelerators [25]. An accelerator of size R generating a magnetic field B is constrained by the Larmor radius in terms of its acceleration capability. A particle with energy surpassing the threshold

$$E_{max} \;=\; q\beta RB\Gamma \tag{2.11}$$

will break free from the accelerator, where q represents the particle's charge, β indicates the speed of the shock fronts propelling the particle, measured in natural units, and Γ denotes the Lorentz factor of the source. Figure 2.12 illustrates a range of potential UHECR sources based on their distinctive sizes and magnetic-field strengths, with a focus on three sources that we will highlight below.

Three prime candidates for sources of UHECRs stand out. AGNs frequently hosting super-massive black holes emitting relativistic jets, exhibit great promise as accelerators for UHECRs. For instance, Centaurus A, a radio galaxy harboring an AGN, positioned approximately 5 Mpc away, emerges as a compelling prospect as a single source of UHECRs. The Auger Collaboration first reported a correlation between the arrival directions of UHECRs with energies exceeding 6×10^{19} eV and AGNs within a 75 Mpc vicinity, although this correlation necessitated thorough reevaluation as more data accumulation [43, 44]. AGNs, characterized by substantial size and relatively modest magnetic fields (approximately six orders of magnitude smaller than the magnetic field of the Earth), could conceivably retain cosmic rays over extended periods, gradually accelerating them over millions of years until they gain high enough energy to escape. In particular, if an AGN is classified as a blazar, signifying that its jet axis aligns with Earth, it becomes an even more favorable source candidate due to its increased brightness and reduced magnetic field deflection of the cosmic rays [45]. Additionally, the signature of UHECR would manifest in the gamma-ray spectrum of the blazar. Although the blazars in the vicinity of the Milkyway might not suffice to account for the observed UHECRs, they remain plausible candidates in a scenario involving multiple contributing accelerators.

Tidal Disruption Events (TDE), fleeting occurrences where a black hole disrupts a star in its vicinity, typically the black hole in the center of a galaxy, have recently surfaced as conceivable sources for UHECRs [46]. TDEs are brief (in astrophysical time scale) and luminous events, during which nearly half of the star's mass is consumed by the black hole. The remaining material is ejected either as relativistic jets or released via thermal processes in the accretion disk. These

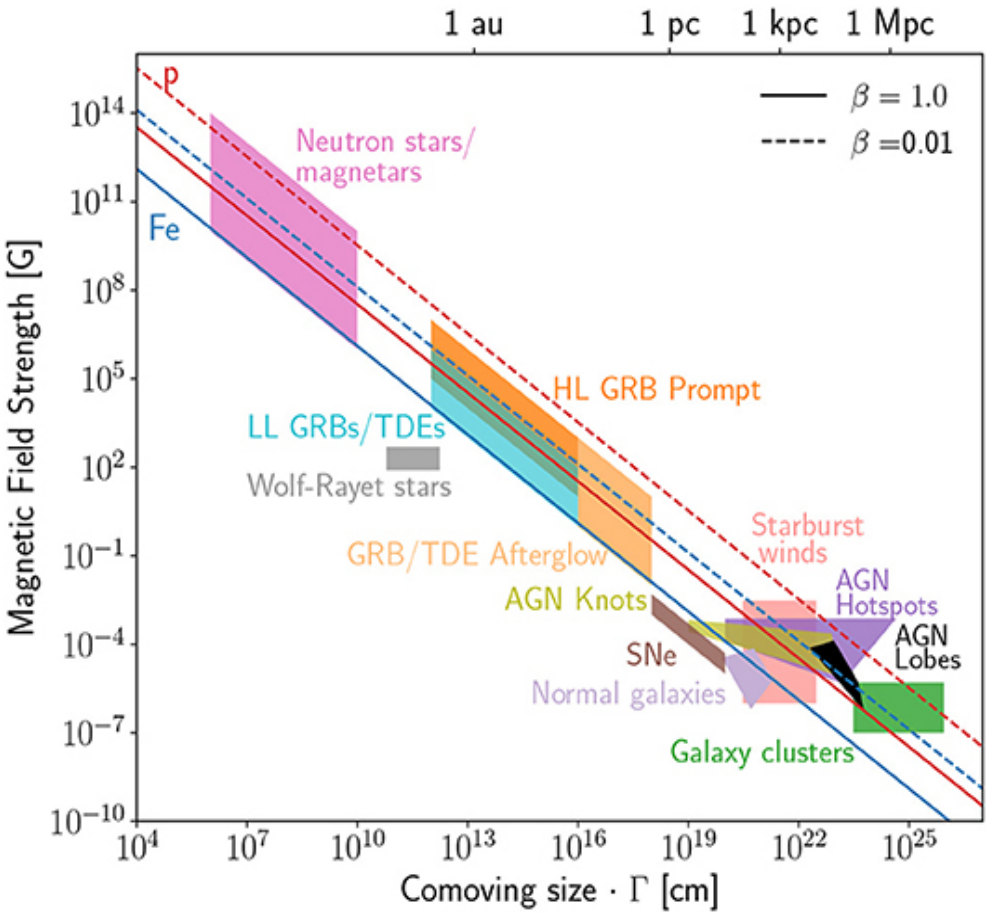


Figure 2.12: Hillas Diagram. The potential origins of UHECRs are illustrated based on their characteristic size (L) and the accompanying magnetic field (B). The boundaries for accelerating protons and ionized iron nuclei to energies of 10^{20} eV are denoted by solid (ultra-relativistic) and dashed (non-relativistic) lines. Obtained from [42].

blazar-like objects or jetted fulfill the prerequisites for accelerating protons to energies of 10^{20} eV [47]. Nevertheless, knowledge of TDEs, let alone jetted TDEs, remains limited, with only a handful known, and the nearest candidate positioned at a distance of approximately 150 Mpc [48].

Fast-rotating jetted neutron stars, known as pulsars, or in the case of extremely powerful magnetic fields, magnetars, represent remarkably compact source candidates for UHECRs, with diameters of just a few kilometers. These celestial objects boast magnetic fields range is approximately 10^{14} - 10^{15} Gauss [49], displaying a strength so formidable that its magnetic energy density outstrips even the heaviest elements by multiple orders in magnitude. Being remnants of massive stars, they contain abundant heavy elements, allowing them to accelerate heavy nuclei to extraordinary energies [50]. Subsequent to the supernova event leading to pulsar formation, a substantial portion of the pulsar's matter and energy is rapidly transferred to its surroundings within a matter of months. In this process, particles are propelled as the pulsar wind. These particles can then escape as UHECRs.

In the quest for UHECR sources, the field of multi-messenger astronomy has

assumed a leading role. This approach involves the simultaneous exploration of highly energetic astrophysical sources through neutrinos, gravitational waves, gamma rays, and cosmic rays [42, 51, 52]. The presence or the absence of one or more of these multi-messengers serves as an indicator to deduce the mechanisms accountable for UHECR production.

CHAPTER 3

THE PIERRE AUGER OBSERVATORY

The Pierre Auger Observatory is the world's largest array of detectors to measure Extensive Air Shower due to UHECRs, located at the Malargüe in the province of Mendoza, Argentina. It was named after the French physicist, Pierre Auger, who is credited with the discovery of Extensive Air Shower. The construction of the observatory commenced in 2000, and has been collecting data since 2004; it achieved full operational status in 2008.

The observatory covers an area of approximately 3000 km^2 with 1600 *Water Cherenkov Detectors (WCDs)* which comprises the *Surface Detector (SD)* and four stations to detect fluorescence emission (see Figure 3.1). The location is a semi-flat desert at an average altitude of ∼1400 m (∼870 $g \cdot cm^{-2}$) providing favorable conditions to measure X_{max}. Additionally, less light pollution due to low population density and low-temperature desert conditions makes it perfect for fluorescence detection.

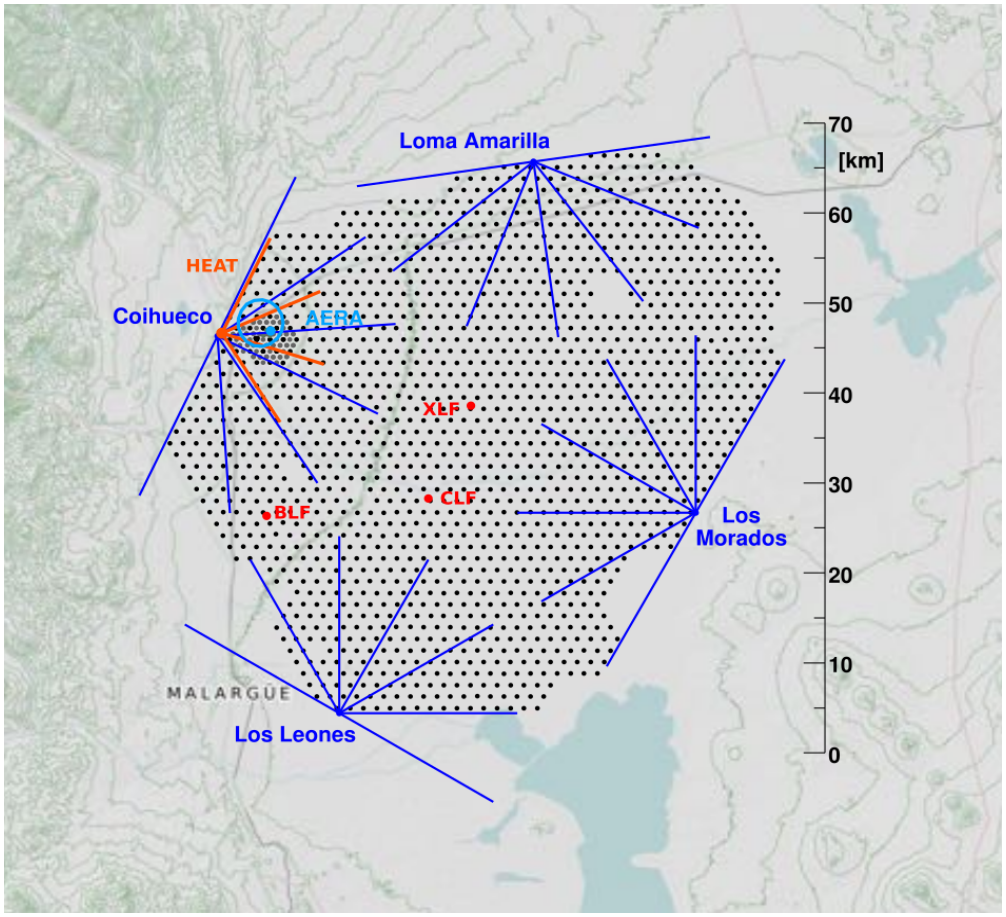


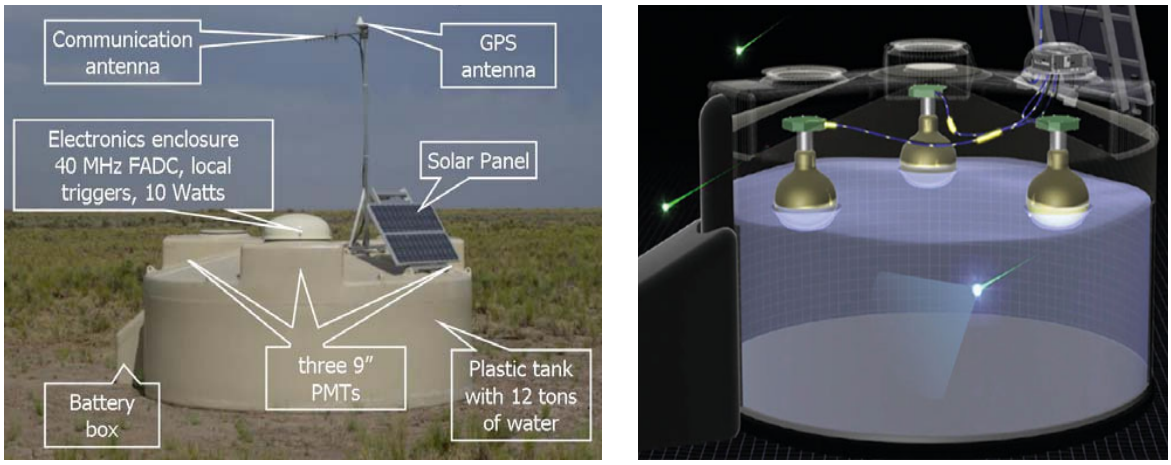
Figure 3.1: A schematic map of the Pierre Auger Observatory. The grey points represent the SDs. The four Fluorescence Detectors are marked in blue along with the names of their location. The length of the blue lines signifies the azimuthal field of view of FDs. Atmospheric monitoring stations, XLF, CLF, and BLF are marked in red. Additional radio detection stations (AERA) and enhancement to FD (HEAT) which detects fluorescent light higher in the atmosphere can be seen in the top-left corner. Obtained from [53].

3.1 Water Cherenkov Detectors

The *Water Cherenkov Detector (WCD)* are the main section of the Surface Detector at the Pierre Auger Observatory. The installation of SSDs as part of the SD is discussed in section 3.3.1. The 1600 WCDs are arranged hexagonally such that each station is at a distance of 1500 m from its neighboring detectors. These are re as SD-1500. As elaborated within the section dedicated to SSD detection of UHECRs, it becomes apparent that the separation between detectors plays a pivotal role in establishing the minimum detectable energy threshold at the observatory. In this context, the spatial arrangement of WCDs is efficient in identifying hadron-induced showers characterized by energies above $\sim$3 EeV and zenith angles below 60°.

The WCDs are cylindrical tanks, possessing a radius of 1.8 meters and a height of 1.2 meters, filled with highly purified water, providing an active detector volume

of 12,000 liters. Each WCD is equipped with three *XP1805 Photomultiplier Tubes* PMTs, each having a diameter of 9 inches, strategically positioned to monitor the Cherenkov light generated when particles from an air-shower traverse the water [54, 55]. These PMTs are arranged symmetrically in the water tank, located at a distance of 1.2 meters from its center and oriented downward. A photograph of a station and an animated view of a particle passing through a WCD can be seen in Figure 3.2. The innermost layer of the water tank is coated with an opaque, highly reflective Tyvek material (DuPont Tyvek 1025-BL). The high-purity water is vital for minimizing Cherenkov light attenuation and ensuring the long-term stability of both the water and liner properties during the extended operational lifespan of the Observatory. Temperature variation and light and bacterial contamination are controlled by the other four layers between Tyvek and the tank: a Carbon black LDPE (Low-Density Polyethylene) layer between two layers of clear LDPE, a layer of titanium dioxide pigmented in LDPE. The Surface Detector (SD) station, characterized by its resilience to environmental factors, achieves an impressively high duty cycle of nearly 100%, enabling the continuous observation of EAS.



(a) **(b)**

Figure 3.2: **(a)** A picture of WCD with its components labeled; from [56]. **(b)** An animated picture of a particle passing through water containment of a WCD, producing Cherenkov radiation. Three PMTs are connected by wires (dark blue) connected to UB and the current carried by wires is shown in yellow; from [57].

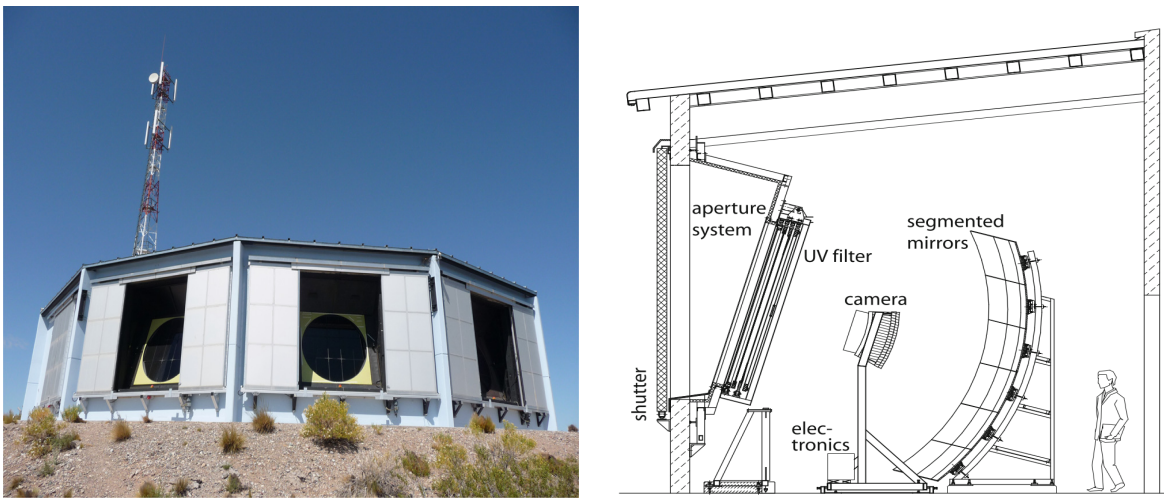
The temporal and geographical coordinates of each SD station are precisely determined using a Motorola GPS unit, which features a time precision of ∼8 nanoseconds. Communication between the SD stations and the Central Data Acquisition System (CDAS) is established through a Wireless-LAN connection via an integrated antenna [58]. Each SD station operates independently of the others and its power requirements are satisfied by a solar power system, providing an average power output of 10 W to operate the PMTs and associated electronics. The electronics are consolidated into a *Unified Board (UB)* housing a processor, power controller, GPS receiver, and radio transceiver [59].

To expand the lower limit of observable energy, an auxiliary array, denoted as

the *Infill array* or SD-750, was established in 2011. Comprising 71 stations, these detectors are placed at intervals of 750 m, thereby reducing the energy threshold to $\sim 3 \times 10^{17}$ eV [60]. Furthermore, an additional hexagonal array was introduced in 2013, featuring a more compact spacing of 433 m, with the explicit aim of further lowering the energy threshold to ∼40 PeV [61, 58].

3.2 Fluorescence Detectors

The Fluorescence Detector (FD) constitutes a critical component of the instrumentation, housing 24 telescopes stationed at four distinct FD sites, denoted as *Eyes*: Coihueco, Los Morados, Loma Amarilla, and Los Leones, as illustrated in Figure 3.1. Within this expanse, the low-energy extension HEAT directly monitors the upper atmosphere directly above a 66-detector sub-array of the Surface Detector. The spatial arrangement features detector stations with a spacing of 750 meters and an elevated observation level, enabling measurements of events reaching down to nearly PeV energies. Each fluorescence telescope has a UV filter window, a segmented spherical mirror with an area of 13 m^2, and a camera equipped with 440 PMTs. These telescopes are designed to provide a fixed Field of View (FoV) of $30° \times 30°$, with each PMT dedicated to observing a specific ∼1.5° section of the sky.



(a) **(b)**

Figure 3.3: **(a)** A FD where three out of four telescopes are open during the day for maintenance; from [62]. **(b)** A schematic of a fluorescence telescope with its parts labeled and human for the size comparison; obtained from [63].

The operation of the FD involves a multi-step process: initially, the background light is eliminated from the UV fluorescence light of the air shower using the filter window. Subsequently, the mirror collects and directs the light onto the camera, positioned in the focal plane. Each PMT records the UV light within its designated FoV with a timing resolution of 100 ns, generating an intensity and timing profile

of the shower. This data is subsequently utilized to reconstruct the shower's geometry and profile [63].

The FD mandates optimal conditions for data acquisition, requiring exceptionally dark, cloudless nights, resulting in a duty cycle of only 14–15%. Consequently, it offers relatively limited event statistics, particularly at the highest energies. In contrast, the SD have a significantly higher duty cycle (as discussed in section **??**), accumulating a more substantial event dataset at these elevated energy levels. Nevertheless, the composition resolution of SD-only analyses does not match the precision achieved by the FD. To address this limitation, ongoing upgrades such as AugerPrime aim to enhance the SD's capacity for measuring composition on a shower-by-shower basis, presenting an opportunity to significantly augment composition information at the highest energy levels. This thesis is motivated by this upgrade of the facility.

3.3 The AugerPrime Upgrade

Several of the previously addressed inquiries and findings are constrained by the 15% limited duty cycle of Fluorescence Detector. This is insufficient to provide enough data, particularly for in-depth investigations into parameters such as mass composition at the highest energy levels. To circumvent this constraint, the ongoing upgrade to the Surface Detector has been proposed, offering a 100% duty cycle independent of the FD. The primary motivation for this upgrade focuses on disentangling various components of extensive air showers, primarily the muonic and electromagnetic components. This is achieved by introducing an additional detector alongside the WCDs. The distinctive responses of both detectors to the different shower components are leveraged to discriminate between them, thereby enhancing the capacity to perform mass composition reconstructions at the most elevated energy levels. The ensuing sections provide a concise overview of the key components comprising the *AugerPrime* upgrade. All the upgrades of the observatory are discussed below and visualized in Figure 3.5.

3.3.1 Surface Scintillator Detectors

With this upgrade, we are trying to improve the observatory to discriminate the mass composition of the UHECRs, which is the main aim of this thesis. A Surface Scintillator Detector (SSD) is mounted on top of each WCD. This introduces a new response to the EASs which is characterized by a ratio of measured electromagnetic to muonic components that is nearly double that of the corresponding WCD. Figure 3.4**(b)** shows an SD station where an SSD is installed.

Each SSD has an area of $\sim$4 m^2 which is divided into two modules, each module composed of 24 plastic scintillator bars. Each scintillator bar has a length of 1.6 m, a width of 5 cm, and a thickness of 1 cm, and is made up of polystyrene (Polystyrene Dow Styron 663 W) mixed with PPO (1%) and POPOP (0.03%) as wavelength-shifting dopants [53, 64]. Every bar has two holes on the longer side

through which the 1 mm diameter wavelength-shifting optical fiber passes [65]. The emission spectrum of the scintillator bars is in the range of 330 to 480 nm; this matches the absorption spectrum of the fibers. The fibers collect light from the bars and pass it on to a 15-inch PMT (Hamamatsu R9420 [66]) located in the gap between two modules of the SSD. This arrangement is visualized in the Figure 3.4. To resist external effects and provide more stability all this is covered in polystyrene panels and a waterproof aluminum enclosure on top of that.

The WCD initiates the SSD triggers, leading to the acquisition of signals from the SSD every time an event is triggered by the WCD, irrespective of the signal magnitude within the SSD. The SSD quantifies the signal in terms of *Minimum Ionizing Particles (MIP)*, where 1 MIP represents the passage of a single particle moving vertically through the SSD.

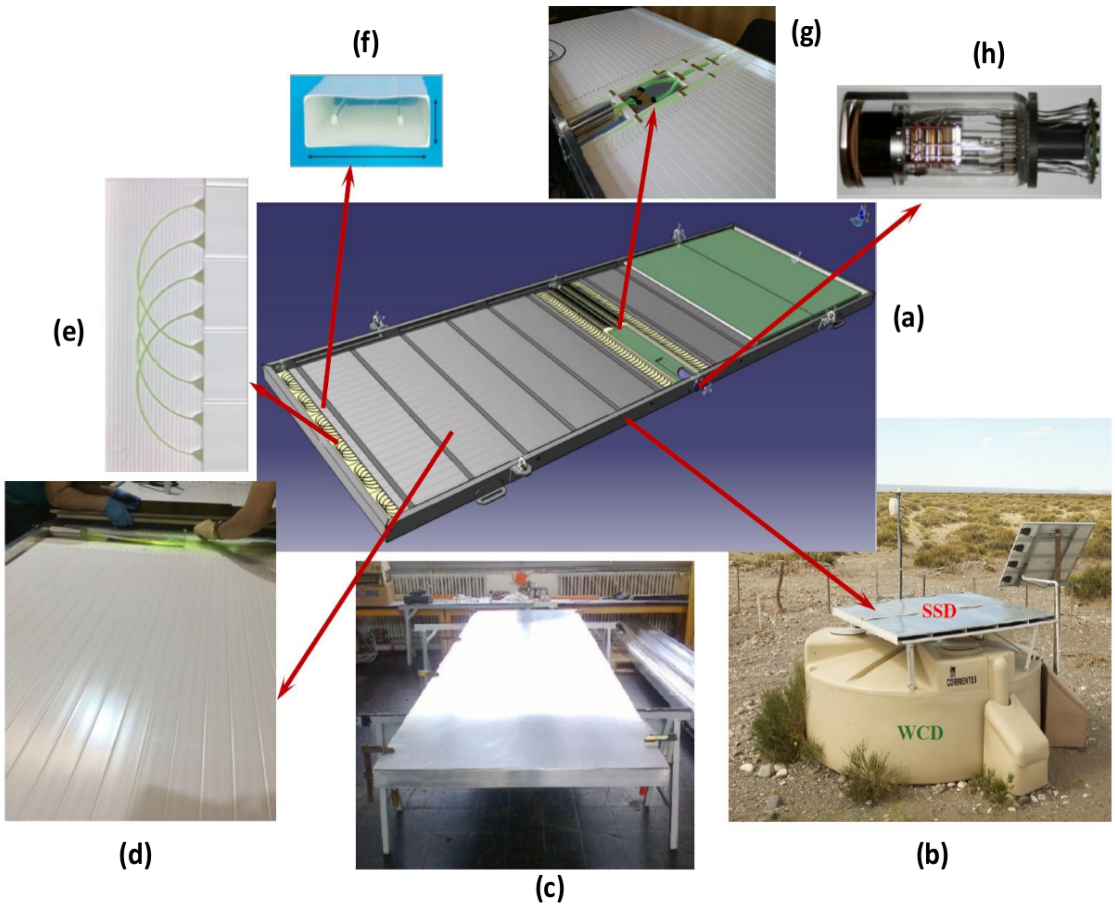


Figure 3.4: **(a)** A schematic of Surface Scintillator Detector. **(b)** A photograph of installed SSD on a WCD. **(c)** Aluminum enclosure for a SSD. **(d)** Scitillator bars covered in the reflective layer of TiO_2, giving it opaque white color. **(e)** U-shaped layout of optical fibers to maximize light yield. **(f)** A side view of a scintillating bar with horizontal and vertical dimensions are 5 cm and 1 cm, respectively. **(g)** Optical fibers coming out bars in between both sections of SSD, connecting to a PMT. **(h)** A Hamamatsu R9420 PMT. Referred from [67].

The production and testing of 1,500 scintillators were distributed across six

different facilities, with RWTH Aachen University being one of them. During the course of this project, 135 detectors were assembled and subjected to testing procedures that covered the period from March 2018 to October 2019 [68]. Seventy-seven SSDs have been operational since March 2019, forming the *SSD Pre-Production Array (PPA)*, of which the data will be presented in this thesis at a later stage.

3.3.2 Small Photo-Multiplier Tube (sPMT)

Incorporated alongside the three PMTs within the WCDs, a fourth smaller PMT (sPMT, Hamamatsu R8619 [155]) is placed in the center of the tank to expand the dynamic range. The smaller active surface area of the sPMT results in a decrease in light collection, enabling the detection of larger signals. This feature proves especially advantageous when capturing signals in close proximity to the shower core, where the other PMTs in the WCDs tend to saturate due to the presence of large signals.

3.3.3 Upgraded Unified Board (UUB)

The availability of only six readout channels of the existing Unified Board was insufficient to accommodate readout from new SSD and sPMT as they were pre-occupied with the three PMTs from WCD. To facilitate this enhancement, a novel electronic system, denoted *Upgraded Unified Board (UUB)*, has been introduced, which has a total of 10 input channels. Of these, six are designated for conventional PMT of WCD, one channel is assigned for a sPMT, and two input channels for the high and low gain modes of the scintillator PMT. The 10^{th} channel is reserved as a spare, allowing for potential future extensions. Simultaneously, the sampling frequency is increased from the pre-existing 40 MHz to 120 MHz, facilitated by a 12-bit *Flash Analog-to-Digital Converter (FADC)* [69].

3.3.4 Increased FD Uptime

An additional upgrade of Fluorescence Detector, intended to increase its duty cycle from 15% to 30%. This enhancement is attained through a reduction in the FD PMT gains by a factor of 10 This is achieved by decreasing the supplied high voltage, making PMTs susceptible to higher light fluxes, thereby facilitating data collection during periods characterized by a significant moon fraction and closer proximity to sunrise and sunset [53].

3.3.5 Radio Detector (RD)

To complement the SSDs, a set of small radio antennas will be installed at each of the array stations. Air showers exhibit a substantial radio emission footprint at zenith angles exceeding 60°. Although the SSD and WCD detectors collectively

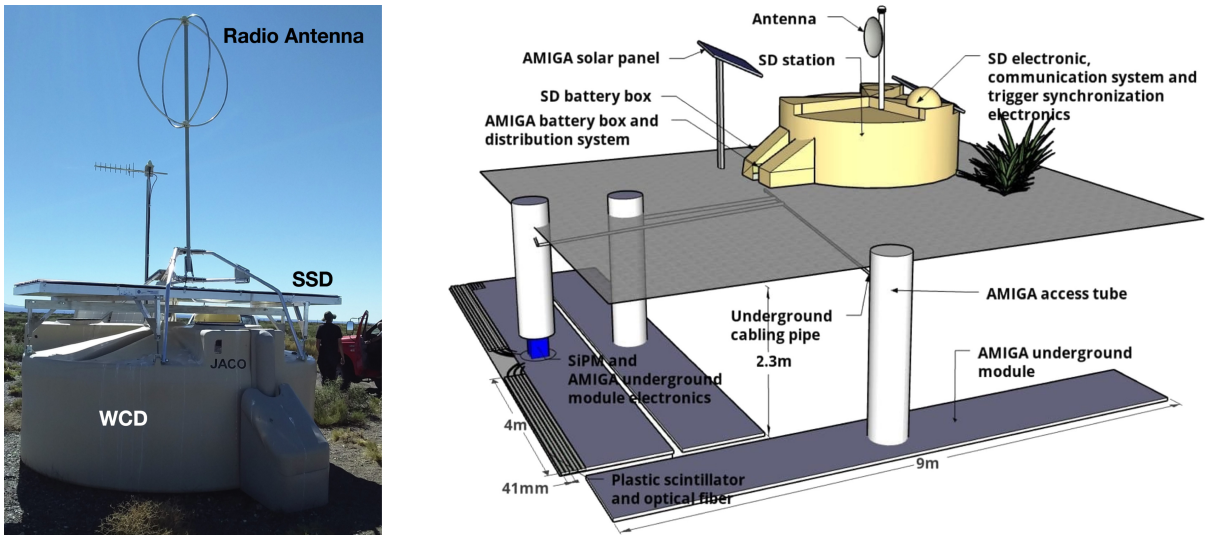


Figure 3.5: Left: A SD station with Radio Detector, Water Cherenkov Detector and Surface Scintillator Detector. **Left:** A schematic of Underground Muon Detector at an SD station with parts labeled within figure. The images are obtained from [70].

offer sensitivity to vertical air showers, the incorporation of WCDs alongside radio antennas will broaden the scope to measure more inclined air showers [71]. Consequently, this expansion facilitates additional investigations into mass composition at higher zenith angles.

3.3.6 Underground Muon Detector (UMD)

Within the SD-750 array, the deployment of Underground Muon Detector (UMD) is planned, and strategically placed at a depth of 2.3 meters below the ground [30]. These detectors, known as Auger Muon Detectors for the Infill Ground Array AMIGA, will be placed in the vicinity of 61 WCD stations within the array. Their positioning makes UMD allow measurement of muonic components while electronic components absorb in the soil. This advantageous depth allows for direct calibration of the combined WCD and SSD reconstruction of the muon content [72, 73]. Furthermore, this integration will enhance the composition sensitivity of the energy spectrum, particularly in the region of the ankle of cosmic rays.

CHAPTER 4

SURFACE DETECTOR RECONSTRUCTION

4.1 Calibration

The electronics utilized in SD are equipped with six 40 MHz 10-bit Flash Analog-to-Digital Converter (FADC), which digitize signals from the three PMTs. These signals are processed through a *low gain* (LG) channel and an amplified channel *high gain* (HG). The LG is read from the anode, while the HG is read from the last dynode. The latter is inverted and subsequently amplified by a factor of 32. The combined operation of these two channels effectively spans a dynamic range capable of accurately measuring particle flux in proximity to the shower core ($\sim$1000 $particle(s)/\mu s$) and at greater distances from it ($\sim$1 $particle(s)/\mu s$). Each of these FADC channels records a signal trace encompassing 768 bins for each PMT, with each bin boasting a time width of 25 ns, in light of the 40 MHz sampling rate. The recorded signals are temporarily stored in buffer memory, and if they meet the specified trigger criteria (elaborated in the subsequent section), the signals are then transmitted to the Central Data Acquisition System (CDAS).

To achieve consistency in data interpretation, it is necessary to convert the signals obtained from each detector station into units of *Vertical Equivalent Muons (VEM)*. A VEM is defined as the charge deposited within a station by a vertically incident and centrally traversing muon, referred to as a *vertical and central through-going (VCT)* muon. Employing the concept of VEM establishes a standardized measurement unit for assessing signals generated by particles traversing the detector stations. Although the SD does not possess the capability to exclusively isolate and measure VCT muons but can be calibrated the response of PMTs to random atmospheric muons in conjunction with reference measurements obtained from a specialized muon telescope, which captures data related to VCT muons. This approach ensures that the data collected from the various detector stations can be consistently evaluated and interpreted.

The SD calibration procedure relies predominantly on two key calibration histograms: the charge histogram (constructed for each PMT, as well as their collec-

tive sum) and the pulse height histogram (likewise produced for each PMT and their aggregate sum). Determining the charge corresponding to a VEM involves the integration of amplitudes over the signal pulse duration. As depicted in Figure 4.1, the Cherenkov light generated by atmospheric muons yields a visible peak in the charge histogram. With an ample number of events, this peak consistently aligns with $Q_{VEM}^{peak} \sim 1.09 \pm 0.02$ VEM. This peak, in conjunction with the reference measurement of $Q_{VEM} = 1$ VEM obtained from the muon telescope (as depicted in Figure 4.1), facilitates the derivation of the calibration value essential for rescaling Q_{VEM}^{peak} for the collective output of all three PMTs. Consequently, this procedure enables the expression of charge measurements in terms of VEM. These calibration histograms are generated at one-minute intervals, ensuring that in the event of a trigger within the array, there is access to calibration data from the preceding minute, thereby guaranteeing a high degree of calibration accuracy.

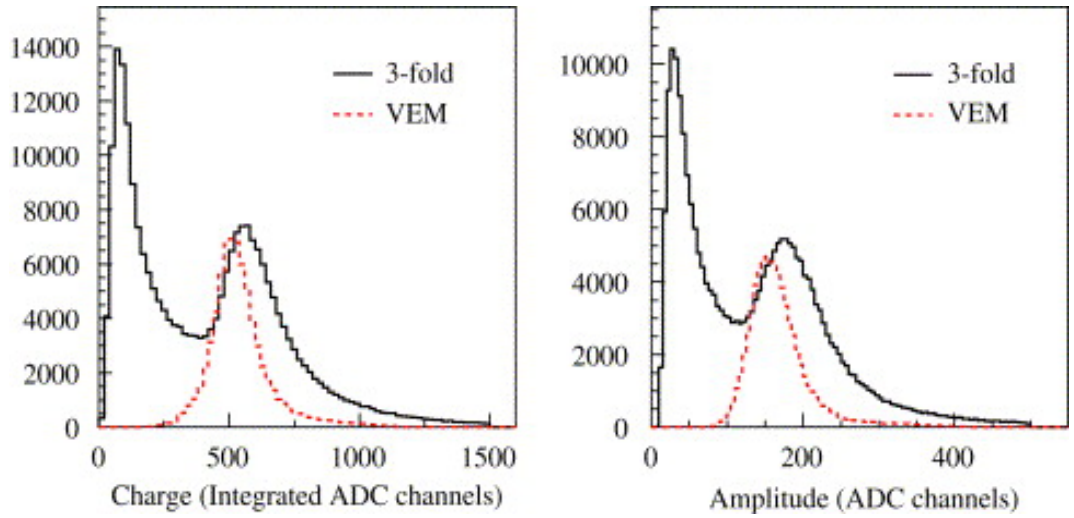


Figure 4.1: Left: The presented histogram of charge (black) results from the summation of responses obtained from all three PMT. The dashed red line corresponds to a reference measurement performed using an external muon telescope. The first peak within the charge histogram primarily corresponds to noise originating from low-energy particles, whereas the second peak is due to VEM. **Right:** The pulse height histogram (black) is employed for calibrating the I_{VEM}^{peak}, which establishes trigger thresholds. This calibration process is conducted using the red dashed line from the reference measurement acquired from the muon telescope. Obtained from [74].

Every triggered event from the SD necessitates that the signal trace is quantified as the current registered by the PMT and expressed in ADC counts, over a specific threshold. Consequently, it is necessary to calibrate the station current to a reference unit, I_{VEM}. This calibration process uses the technique for VEM calibration. Atmospheric muons generate a peak current, denoted as I_{VEM}^{peak}, within the pulse height histogram. Calibration of I_{VEM}^{peak} is accomplished by referencing it to the muon telescope's I_{VEM}. For further details on the calibration of SD see [74].

4.2 Trigger System

Various hardware and software triggers are employed to detect air shower events and discriminate them from background signals generated by low-energy showers. The trigger sequence encompasses single-station-level triggers, central data station triggers, physics event selection, and quality triggers. This trigger system is sequential in five levels, from T1 to T5.

Single Station Trigger

The first two levels are T1 and T2, which function at individual SD stations. At this stage, four distinct algorithms come into play: *Threshold (THR)*, *Time over Threshold (ToT)*, *Time of Threshold Deconvoluted (ToTD)*, and *Multiplicity of Positive Steps (MoPS)* [75, 76]. The Threshold trigger is tailored to address signals originating in close proximity to the shower core, while the remaining three algorithms are finely tuned to accommodate low-energy particles, specifically those originating far from the shower core. T1 and T2 trigger differs only at the THR triggers while other triggers have similar requirements.

- *THR:* All three PMTs to measure a signal above 1.75 times the peak signal in terms of VEM (I^{peak}_{VEM}) in coincidence for T1 and 3.2 I^{peak}_{VEM} for T2. This trigger is specifically designed to detect the extensive, inclined events characterized by the predominance of the muonic component. If the station has less than three PMTs then the threshold value is increased to reduce the random coincidences.

- *ToT:* This requires the presence of a signal surpassing the 0.2 I^{peak}_{VEM} in a minimum of two PMTs within a time window of 120 bins (3 μs), and this signal must last longer than 13 bins (325 ns). In instances where a SD station lacks three functional PMTs, the trigger is applied to the available PMTs, which could be either one or two.

- *ToTD:* This is an enhancement of ToT trigger mechanism. When a charged particle traverses the water Cherenkov tank, the emitted Cherenkov light is produced within a time interval smaller than the temporal resolution of the FADC. However, due to multiple reflections of this light within the tank liner, a portion of the signal experiences a substantial delay, yielding an exponential tail of approximately 70 ns (associated with the light's decay time in the tank). The deconvolution algorithm applied to the trace effectively mitigates the impact of this exponential tail, leading to the transformation of the signal into a prominent peak spanning one or two time bins. In scenarios where multiple particles traverse the tank, as is common in extensive air showers, the deconvoluted trace manifests as a series of peaks. After the trace's deconvolution, the ToT trigger is subsequently applied.

- *MoPS:* This trigger is designed to detect consecutive increments in the FADC

counts. It requires the presence of four successive positive steps in ADC counts, each ranging from 4 to 30 ADC counts, and occurring concurrently in at least two PMTs within a 3 μs time frame. This specific trigger operates using ADC counts, distinguishing it from others that employ VEM calibrated values.

ToTD and MoPS find their specialized applications in studies focused on photons and neutrinos, therefore these triggers do not pertain to this thesis but the data used during this thesis work has passed through this trigger as it was used before for the search of photons in UHECRs [77]. Nevertheless, these additional triggers are no hindrance to this thesis.

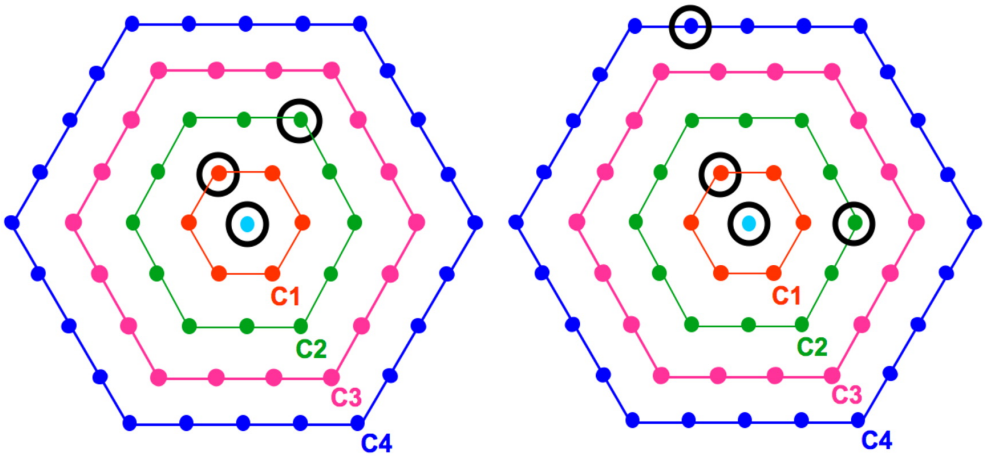


Figure 4.2: Illustration of possible T3 configurations. The different station levels are shown with hexagonal shapes, with respect to the station denoted in sky-blue color. The levels C1, C2, C3, and C4 are at a distance of 1.5, 3, 4.5, and 6 km, respectively [78].

Multi-Station Trigger

If either of the T2 triggers is satisfied then the station level trigger information is sent to the CDAS. The data goes through three triggers T3, T4, and T5 to consider the measured signals as an event.

- *T3:* This trigger is based on signal coincidence which requires the satisfaction of certain spatial conditions, wherein a minimum of three stations must meet the ToT condition, and among these stations, two must be immediate neighbors. Following this requirement, the time difference between the signals recorded by the stations is subsequently examined. Each T2 signal is expected to fall within a time window of $(6 + 5C_n)$ μs from the first triggered signal, where C_n denotes the degree of neighborliness between the stations (see Figure 4.2). The $2C_n$ refers to direct neighbors, and $3C_2$ for the third station is at least as close as the C_2 hexagon. This type of T3 trigger, which derives from the ToT trigger, predominantly yields physics events. Furthermore, it is essential to underscore that this trigger exhibits heightened efficiency when

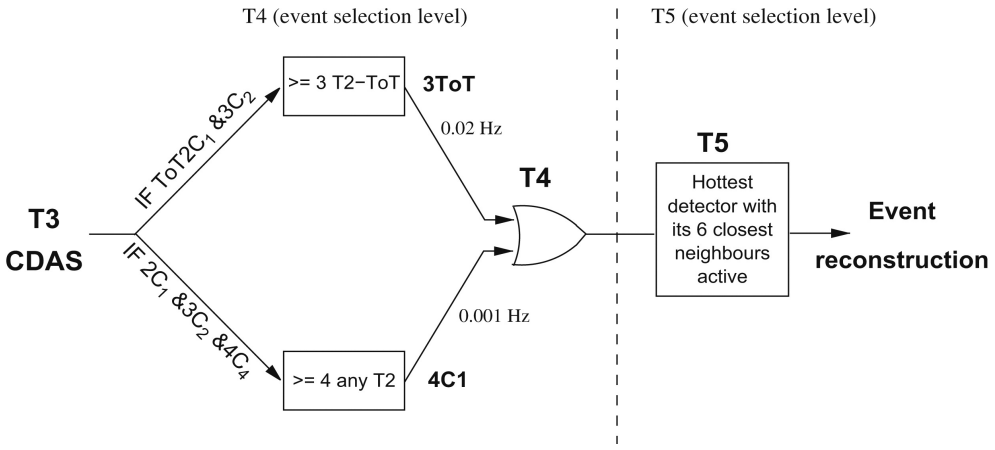


Figure 4.3: Schematic of the station level trigger hierarchy, from T3 to T5. [78].

applied to air showers with zenith angles below 60 degrees. Consequently, it has been aptly denominated as $ToT2C_1$ & $3C_2$.

The second T3 trigger has a more relaxed approach. It mandates that all stations should possess either ToT-T2 or THR-T2 triggers, as illustrated in Figure 4.3, and necessitates a 4-fold coincidence. In particular, the first two neighboring stations of a selected station should align in a manner similar to the arrangement in $ToT2C_1$ & $3C_2$, while the fourth station can be located as far as the 4^{th} neighbors. This trigger is denoted as $2C_1$ & $3C_2$ & $4C_4$ and is designed primarily to capture inclined showers that exhibit a larger ground footprint.

Upon successful passage of a T3 trigger event, the FADC traces from the stations are subsequently transmitted to the CDAS. Subsequently, after the acquisition of the data, the processing of T4 and T5 begins.

- *T4:* To ensure the selection of real EAS events, a physics trigger denoted as T4 is introduced. T4 triggers encompass two distinct categories: 3ToT and $4C_1$. The 3ToT trigger necessitates the arrival times in a minimum of three additional stations, forming a triangular configuration, which has successfully passed the T2-ToT criterion and can be accurately fitted to a planar wavefront traveling at the speed of light. This trigger is effective in identifying events characterized by zenith angles less than 60°. The second variant of the T4 trigger, $4C_1$, extends its selection criteria to include events featuring four or more stations (see Figure 4.3). These stations should have fulfilled any T2 level trigger conditions and be capable of fitting with a planar wave. This trigger reduces the random coincidence probability to 2%.

- *T5:* This trigger serves the function of removing events located near the boundaries of the array. These events are susceptible to missing station data within their footprint, leading to an incorrect reconstruction process. To

mitigate this issue, T5 employs a stringent criterion for event selection, mandating that the station with the highest recorded signal must have all six of its immediate neighboring stations, denoted as C_1, operational at the time of measurement. This ensures that the high-energy event footprint is close to the border but within the vicinity of the array, and can be retained and accurately reconstructed. It also protects against potential biases arising from incomplete event data [78].

4.3 The Surface Detector Shower Reconstruction

Upon initiation by a T4 or T5 trigger, the subsequent phase involves the reconstruction of the shower geometry and energy. The main objective of the SD reconstruction process is to determine the primary energy and its arrival direction. For further understanding of SD shower reconstruction, readers are referred to [79].

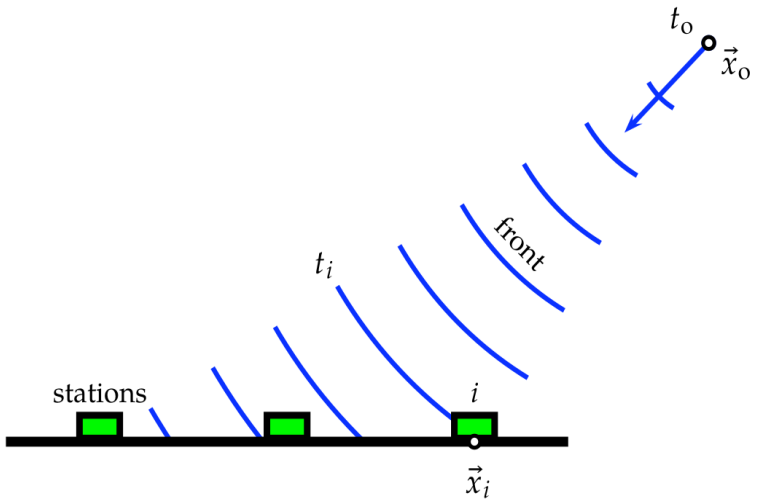


Figure 4.4: Caption [79].

Shower Geometry

Shower geometry is reconstructed by determining the landing position of the shower core and precisely measuring signal timing. First, the arrival direction of the shower is approximated assuming a spherical shower front moving at the speed of light:

$$c\,(t_i - t_0) = |\overrightarrow{x_0} - \overrightarrow{x_i}| \tag{4.1}$$

Here, $\overrightarrow{x_i}$ is the arrival of shower front at the station i at the time t_i, and $\overrightarrow{x_0}$ and t_0 are the starting position and time of the shower development. An illustration of the shower plane can be seen in Figure 4.4. A more precise estimation of the shower front by assuming it is spherical, with radius R. R is determined from the estimated shower core arrival time at the ground [80].

Lateral Distribution Function

The determination of particle density at ground level, and consequently the recorded signal, relies on the application of the *Lateral Distribution Function (LDF)*. This function reconstructs both the energy and composition of the primary cosmic particle, as illustrated in Figure 4.5. It is essential to note that the LDF employed in this thesis is *Nishimura-Kamata-Greisen (NKG)* type function:

$$S(r) = S(r_{opt}) f_{NKG}(r), \tag{4.2}$$

$$f_{NKG}(r) = \left(\frac{r}{r_{opt}}\right)^{\beta} \left(\frac{r + r_{scale}}{r_{opt} + r_{scale}}\right)^{\beta+\gamma} \tag{4.3}$$

Where β and γ are the parameters of LDF which describes its steepness and also depends on the zenith angle. $S(r_{opt})$ is connected with the shower size, a parameter derived from an optimized distance denoted as r_{opt}. The selection of this distance is optimized to ensure a precise and reliable determination of the shower size. Minimal alterations in the LDF are anticipated from slope fluctuations at this point. It's essential to note that r_{opt} primarily depends on the specific characteristics of the detector geometry, with a standardized value of $r_{opt} = 1000$ m applicable to the SD-1500 array while differing values are assigned to other arrays within the observatory. Similarly, r_{scale} is estimated to be 700 m. Refer [81] for additional details on the LDF in SD-750 and SD-433.

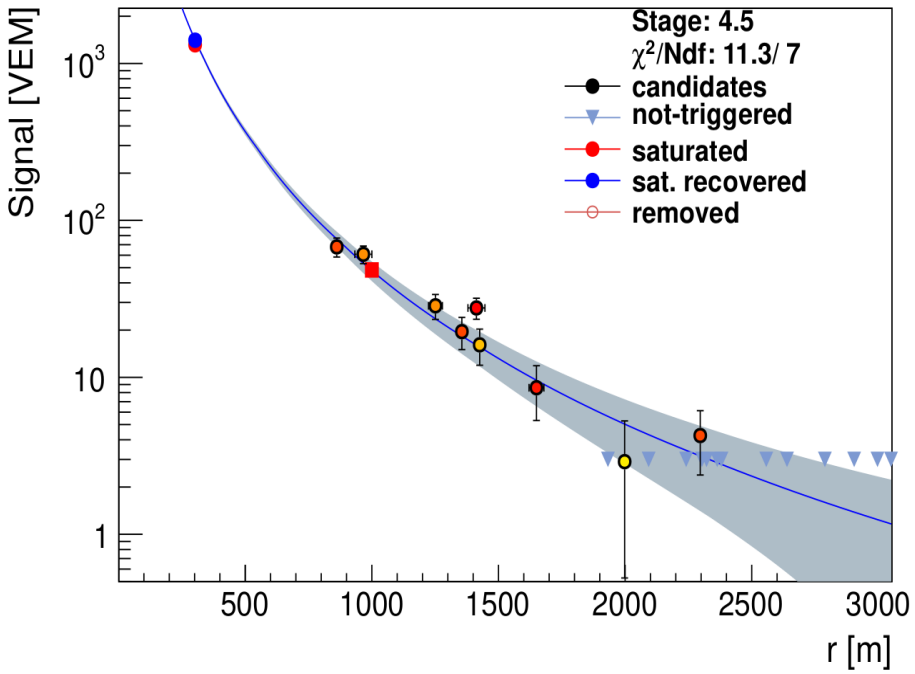


Figure 4.5: An illustrative reconstruction extracted from offline data showcases the lateral distribution (LDF) of a shower event, highlighting the distinct treatment methods applied at various stations. Sourced from [82].

Shower Energy Estimation

The energy of the primary particle can be estimated with shower size S_{1000} or $S(r_{opt} = 1000)$. The shower size is influenced by the zenith angle (θ) EAS, primarily due to the attenuation of the electromagnetic component in the atmosphere. To reduce this dependence on the zenith angle and obtain a shower size estimator that is independent of the zenith angle, a correction is introduced and applied to S_{1000}.

$$S_{38} = \frac{S_{1000}}{f_{CIC}(\theta)} \tag{4.4}$$

Where, $f_{CIC}(\theta)$ is a function that uses Constant Intensity Curve (CIC) [83, 84]. It is a polynomial of the third order:

$$f_{CIC}(x) = 1 \ + \ ax \ + \ bx^2 \ + \ cx^3 \tag{4.5}$$

where,

$$x = cos^2(\theta) \ - \ cos^2(38^\circ) \ and,$$
$$a \ = \ 0.980 \pm 0.004, \quad b \ = \ -1.68 \pm 0.01, \quad c \ = \ -1.30 \pm 0.45$$

Here, the determination of the free parameters a, b, and c is achieved through the fitting process, involving the event intensities measured at different zenith angles. The fitting procedure operates under the presumption of an isotropic flux in $cos^2(\theta)$. A newly introduced variable denoted as S_{38} may be regarded as the theoretical shower size of an event, signifying the shower's characteristics when it approaches Earth at a zenith angle of 38°. From S_{38} energy can be estimated:

$$E_{SD} \ = \ A \left(\frac{S_{38}}{VEM} \right)^B \ EeV \tag{4.6}$$

With $A = (0.178 \pm 0.003)\ EeV$ and $B = 1.042 \pm 0.005$ are the values obtained from FD hybrid events [85], providing weather information.

CHAPTER 5

SIMULATION OF EXTENSIVE AIR SHOWER AND PIERRE AUGER OBSERVATORY

As mentioned in Chapter 2 development of Extensive Air Showers is a complex process that depends on the interactions of particles, producing in billions and atmospheric conditions during interaction. Since the particle interaction that occurs in EAS is currently beyond the reach of human-made accelerators, Monte Carlo simulations of EAS are generated under certain conditions, such as energy, direction, type of cosmic ray primary particle and weather conditions. A complete simulation requires a simulation of the shower as well as the response of detectors. Later is discussed subsequently in this Chapter.

5.1 CORSIKA

CORSIKA stands for "*COsmic Ray SImulations for KAscade*" is a sophisticated state-of-the-art software program designed to simulate Extensive Air Shower initiated by high-energy cosmic ray particles. It accommodates a broad range of primary particles, including protons, light nuclei up to iron, photons, and various others, up to the highest energy of 100 EeV. CORSIKA can not only estimate the average properties of the resulting showers but also capture the inherent uncertainties around these values. Based on the input parameters, CORSIKA returns information about secondaries in the shower such as energy, particle type, time of arrival, and direction.

The simulation process involves the tracking of particles as they traverse the atmosphere, interacting with air nuclei and the decay of the unstable particles. To describe hadronic interactions at high energies, CORSIKA provides several interaction models: VENUS, QGSJET, and DPMJET, SIBYLL, EPOS. For interactions at lower energies, CORSIKA offers the GHEISHA, FLUKA, and UrQMD models. This study employs EPOS-LHC, QGSJet-II-04, and SIBYLL2.3 as the specific versions of relevant models. For further information on these interaction

models refer [77].

Within the Pierre Auger Collaboration, *Napoli* and *Praha* are the two libraries available for CORSIKA files, which only differ slightly from each other [86, 87]. CORSIKA files used during this thesis work are produced and stored by the members of the Pierre Auger Collaboration. The simulation output files are used as input for Offline, which simulates detectors and their response to air showers. This is discussed later in this chapter.

5.1.1 Muon Deficiency

Recent measurements from the Pierre Auger Observatory have demonstrated a deficiency of muons in the simulation of EAS in comparison with the measured data [88]. Further investigations reported the existence of this deficiency for the primary particle of energy above 10 PeV [89]. Muon deficiency for the highest energy particle can be seen in Figure 5.1. By three interaction models with the measurements, SIBYLL-2.3c shows the least difference in the muon numbers. The indifference between models and data increases with the energy. This can be seen in Figure 5.1.

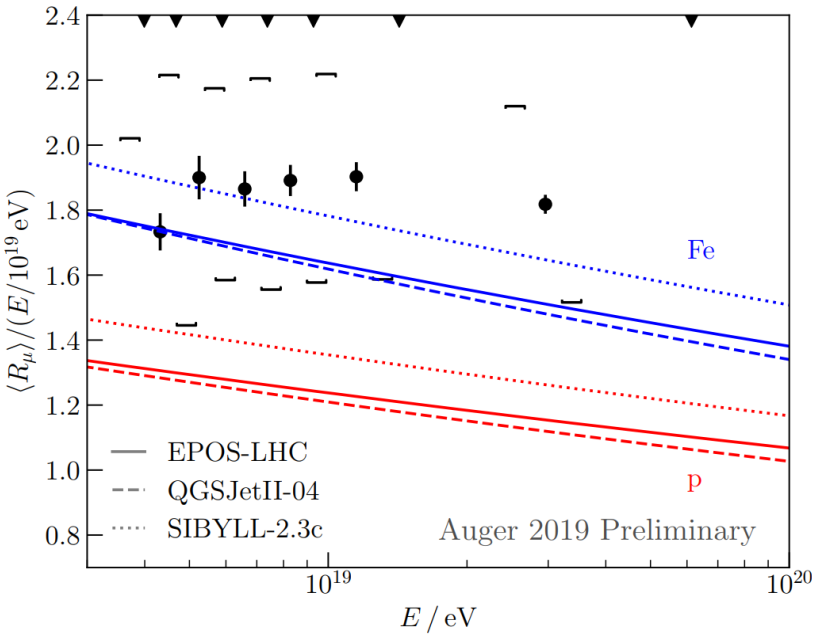


Figure 5.1: Average number of muons as a function of the energy of the primary cosmic ray particle. R_μ is normalized number of muons to that of 10^{19} eV for proton shower. It is also divided by energy to reduce its energy dependency. Taken From [90].

More recent results using Auger hybrid data found constancy between hadronic model predictions and composition of X_{max} [91]. This suggests that any fluctuations in muon content within the simulations are primarily attributed to a slight deficit observed at the initial phases of the shower, which is accumulated over the progress of the shower. The results imply that the discrepancies do not emerge as

a pronounced deviation at the onset of the first interaction, hinting toward known physics processes in the interactions. Although, further analysis of more data is required, which is facilitated by the AugerPrime.

5.2 $\overline{\text{Off}}$line

$\overline{\text{Off}}$line software for the Pierre Auger Observatory is a universal framework for event reconstruction and simulation. The framework is designed in C++, providing Object Oriented nature of language that allows quick integration of the upgrade of the observatory to AugerPrime within the framework. The CORSIKA output files are provided into $\overline{\text{Off}}$line framework to generate detector response for each shower. The framework is flexible enough to take input from other EAS simulation software mentioned above. The output is in the Advanced Data Summary Tree (ADST) format [92] for the end-user that can be handled with the ROOT framework [93].

The $\overline{\text{Off}}$line framework is configured in three parts. *Detector description* - providing detailed information about each detector as well as up-to-date atmospheric conditions. All final information is stored in *Event data* which includes information on the detectors and reconstructed shower. This process is intermediated by *algorithms*, providing an I/O interface for the data flow through different modules. The connections between different parts are illustrated in Figure 5.2.

For this thesis a beta version (trunk) of $\overline{\text{Off}}$line has been used since the scintillator detectors were not included in any tagged version of $\overline{\text{Off}}$line.

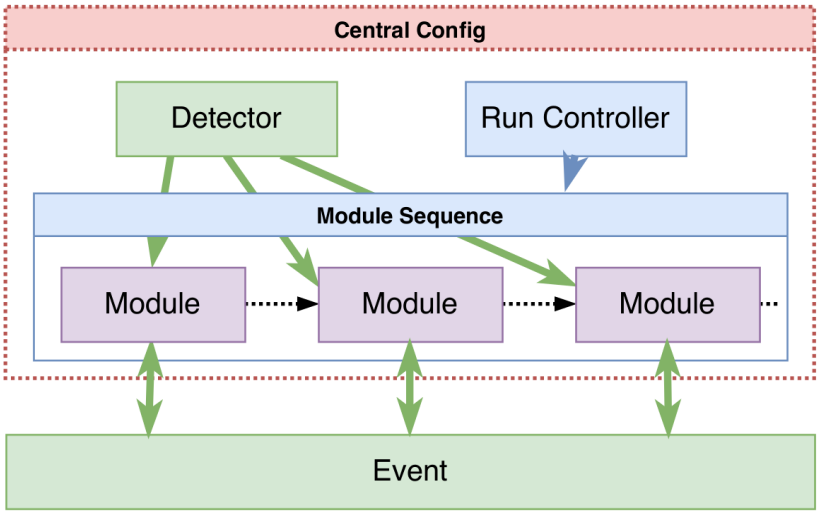


Figure 5.2: A schematic of $\overline{\text{Off}}$line. The algorithm modules and detector configuration are controlled by `CentralConfig`, where `RunContoller` manages the sequencing of modules. Sourced from [94].

CHAPTER 6

PRE-PROCESSING OF SIMULATED SHOWERS

Compared to the other detection methods, the installation of SSDs at Pierre Auger Observatory is a different approach to studying the Extensive Air Showers. This demands the development of new analysis techniques. This work examines simulated data about protons-, helium-, oxygen-, and iron-induced showers. These simulations are derived from the response characteristics of Water Cherenkov Detector (WCD) and Surface Scintillator Detector (SSD). The early analysis is based on [77], in which analysis method was developed for the search of photon-induced UHECRs and it is recommended to the reader for further details, especially discussion of this chapter.

In this chapter, we discuss different conditions applied to the simulated datasets, making sure that the analysis is performed on reliable data. These conditions are applied based on different detector responses and shower reconstruction and it is applied later to observed data (discussed in Chapter 8).

The analysis of simulated and observed data is performed using the High-Performance Computing facility at the RWTH Aachen University.

6.1 $\overline{\text{Off}}\underline{\text{line}}$ Configuration and Simulated Data

Numerous showers were simulated using $\overline{\text{Off}}\underline{\text{line}}$ with different primary particle characteristics such as type, energy, and zenith angle. In this work, generated showers are based on the EPOS-LHC hadronic interaction model. The trunk version of $\overline{\text{Off}}\underline{\text{line}}$ was employed due to the absence of an officially released $\overline{\text{Off}}\underline{\text{line}}$ version containing a functioning simulation of the SSD. The specific trunk version used for each simulated dataset will be explicitly mentioned later in this section. Along with this the $\overline{\text{Off}}\underline{\text{line}}$ itself has been configured. From many $\overline{\text{Off}}\underline{\text{line}}$ settings, there are two points to focus on.

The $\overline{\text{Off}}\underline{\text{line}}$ As mentioned in Section 3.3.3, installation of Upgraded Unified Board (UUB) is part of the AugerPrime upgrade. Although, during the collection of data used in this thesis, there were SSDs operational in the field with a UB.

Datasets	A0	A1	A2	A3
Primary Particle	proton (P)	Helium (He)	Oxygen (O)	Iron (Fe)
Energy [EeV]	18 - 20.2			
θ[°]	0 - 65			
ϕ [rad]	0 - 2π			
Corsika Library	Napoli			
Corsika Files	$\sim 10^5$			
Offline Sequence	SdSimulationReconstructionUpgrade			
Offline Version	Trunk rev 32846			
Detectors	SSD + WCD			
Stations List	SIdealUpgradedUBStationList			
Electronics	*UB*			
Energy Spectrum Slope (before selection)	E^{-1}			
Generated Events	98254	34688	38626	38811
Selected Events	51347	17573	20209	20473

Table 6.1: A comprehensive description of the simulated datasets employed in the preliminary analysis in this chapter.

Since, UB cannot accommodate all channels from PMTs from WCD and SSD, only two of the three PMTs from WCD are selected. This leads to early saturation of the detectors with UB compared to UUB. However, this difference is not too large. The analysis of the comparison between different electronics is discussed in detail in [77].

The triggers ToTD and MoPS mentioned in 4.2 are developed for photon analysis. The photon-induced shower has a smaller muonic component, consequently the lower triggered detectors and hence the lower signal. These triggers are used to overcome this problem. In the case of hadron-induced showers, these triggers are not required. They are still used in this work as the data is inherited from [77]. Therefore due to these triggers, we obtained more triggered stations which are typically situated at the periphery of the particle shower. Consequently, they frequently fail to register a signal exceeding the 1 MIP threshold in the SSDs, which is an established criterion for quality assessment (as detailed in Section 6.2). Thus, including or excluding these additional triggers the number of selected events remains almost the same. A complete analysis of these triggers is carried out in [77].

The detailed description of Extensive Air Showers simulated with CORSIKA can be seen in Table 6.1. The produced data sets are used as input for the Offline to reconstruct showers along with detector response. The CORSIKA files are

obtained from another work [86, 87].

The generated events described in Table 6.1 are those events that effectively initiated the triggering of the array and subsequently underwent successful reconstruction and the selected events refer to events that remained after the application of quality cuts. The data sets exhibit an energy spectrum characterized by an E^{-1} power-law distribution prior to implementing any selection criteria, this is in alignment with the anticipated behavior originating from astrophysical phenomena [95].

6.2 Event Selection

To avoid miscalculations and unintended deviations from the results, it is important to select only those events which are well reconstructed. To achieve this quality cuts are applied which are described below.

- Event Reconstruction: To ensure the accurate determination of the curvature of the shower only fully reconstructed events are considered which are characterized by a `SdRecLevel = 4`.

- Number of Stations: To enhance the accuracy of estimating the variables related to Surface Detector, a set of criteria must be met. Specifically, an event should have a minimum of three unsaturated SSDs and WCDs, with no rejection flags applied to those respective stations. This criterion applies independently to each detector type, meaning that the three SSDs and WCDs do not necessarily need to be at the same stations.

- Minimum SSD signal: SSD must exhibit a signal intensity exceeding that of 1 MIP. It is worth mentioning that SSDs are triggered only through the trigger of WCDs, i.e., SSDs do not possess self-triggering capabilities. Consequently, there are instances when the signal recorded by SSDs is exceptionally low, which has no physical significance, and can be attributed to electronic background noise or baseline fluctuations.

- Zenith angle (θ): Showers with zenith angle exceeding 55° are excluded from consideration due to the inability of the Lateral Distribution Function fit to accurately describe the showers at greater angles.

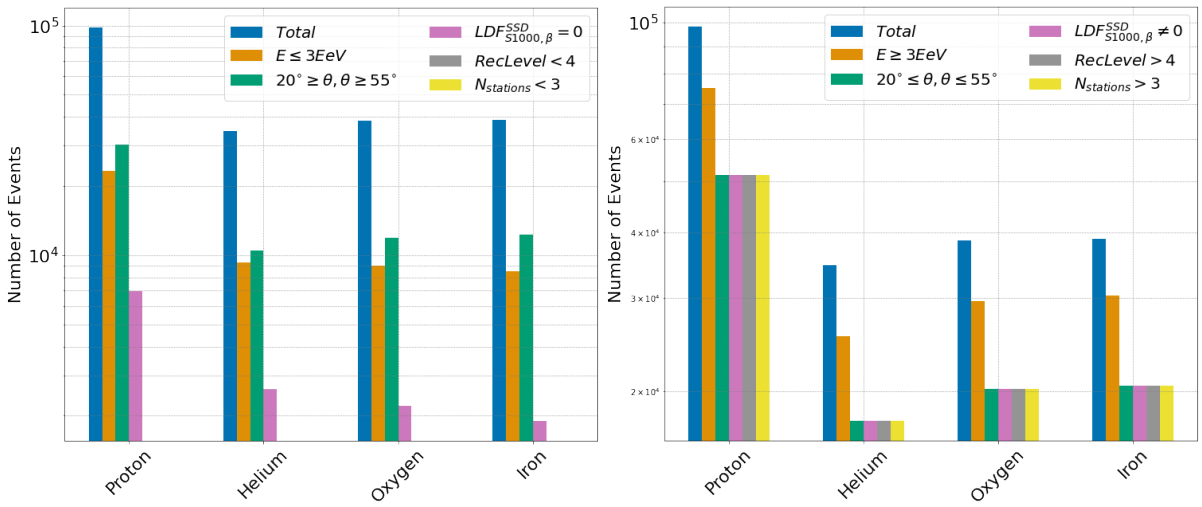
- Vertical showers: Vertical cosmic ray showers, particularly those initiated by protons, exhibit a deeper shower maximum, X_{max} and some showers could occur below the Earth's surface at the Pierre Auger Observatory. This introduces a potential source of bias, as the array detects the shower prior to its reaching its maximum development. Consequently, such vertical showers have been excluded from the analysis.

- LDF SSD fit: Since we are especially investigating the SSDs and thoroughly assessing its potential. This criterion restricts the consideration of events

to those where it is feasible to determine the LDF from the SSD. Despite implementing the above-mentioned measures, there are instances where the fitting process encounters difficulties. Consequently, to ensure that only events with a successfully fitted SSD LDF are included in the analysis, events with values of S_{1000} or β equal to zero are deliberately excluded from consideration. This rigorous approach ensures that the selected events meet the criteria for accurate SSD LDF determination.

- Low energy events: Events with reconstructed energies below 3 EeV are excluded from the analysis. This energy threshold corresponds to the point at which the observatory attains full efficiency in detecting hadronic showers.

Figures 6.1a and 6.2 give an overview of the number of events that are required to be removed by each selection cut. This is considered for showers induced by protons, helium, oxygen, and iron, that is, hadron-induced showers. Also, the number of events that survived after each selection. The cuts are applied one after the other, so the figure 6.1b should be read from left to right.



(a) **(b)**

Figure 6.1: **(a)** The number of unfit hadron-induced events according to different selection cuts. **(b)** Number of events that survived after each quality cuts in order of left to right. The total events generated are also visualized in blue.

Most of the unit events either have low energy or the shower inclination is too large or too small. Only a fraction of events are not suitable to fit the LDF SSD. All the reconstructed events can reconstruct the curvature and have at least three unsaturated SSDs and WCDs. The fraction and number of unfit events according to each criterion are summarized in Table 6.2.

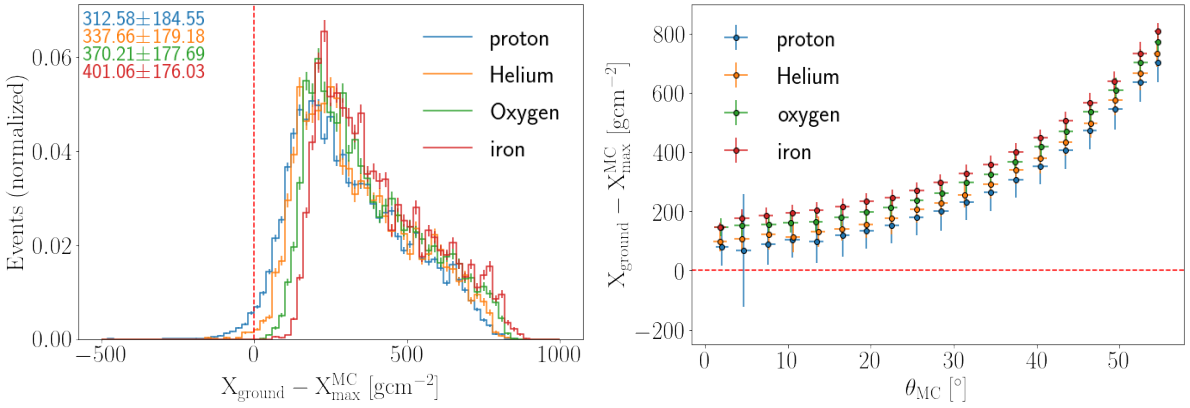
6.2.1 Cut on Vertical Showers

As previously discussed, the application of a selection criterion involves restricting events with small θ angles, primarily intended to diminish the occurrence

Particle type	Proton	Helium	Oxygen	Iron
Generated events	98254	34688	38626	38811
$E \leq 3\ EeV$	23232	9257	8996	8486
$\theta \leq 20°$, $\theta \geq 55°$,	30423	10443	11920	12300
$LDF^{SSD}_{S1000,\beta}$	6991	2612	2210	1897
`SdRecLevel < 4`	0	0	0	0
$\mathbf{N}_{stations} < \mathbf{3}$	0	0	0	0

Table 6.2: The number of events needed to be excluded according to different selection cuts. This table is in agreement with the figure 6.1a.

of air showers with a shower maximum (X_{max}) below the Earth's surface. This is visible in proton and helium high-energy showers, as they have larger penetration depth. Although X_{max} is not a parameter considered for the analysis in this thesis, it remains a valuable tool for discriminating between various primary particles. Therefore, a true Monte Carlo shower maximum (X^{MC}_{max}) is employed for implementing selection criteria based on the zenith angle. This can be seen in Figure



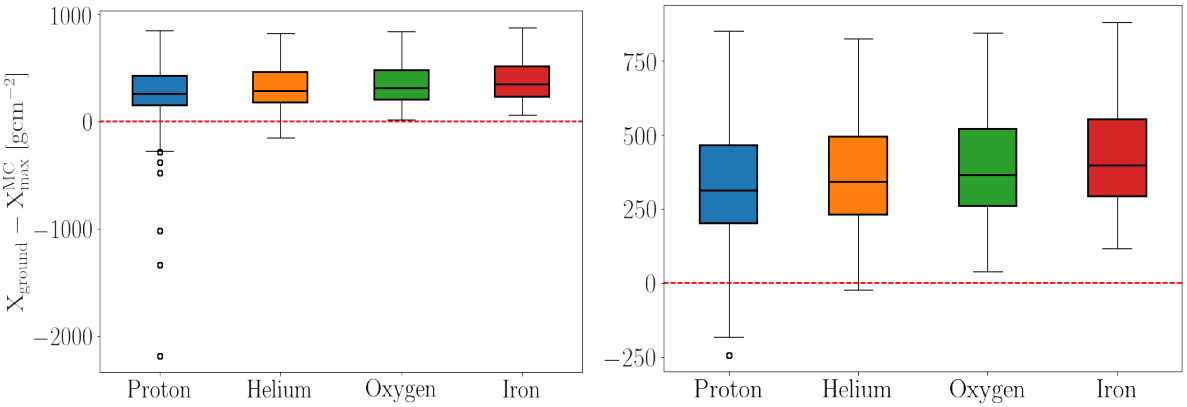
(a) **(b)**

Figure 6.2: **(a)** The distances between the location of X^{MC}_{max} and the Earth's surface X_{ground} are presented for proton, helium, oxygen, and iron showers in the units of $g \cdot cm^{-2}$. To account for the shower's zenith angle, X_{ground} is appropriately adjusted. The vertical dashed red line demarcates the observation level of the observatory. The top-left values along the boundary indicate the mean values and standard deviations for the distributions associated with each respective particle type. Negative values on the underground side of the X^{MC}_{max} axis represent cases where the shower maximum occurs below the Earth's surface. **(b)** The same distance is shown as a function of the true Monte Carlo zenith angle, θ_{MC}. The events with X^{MC}_{max} below above the ground are separated by the horizontal red line.

In the proton data set, the mean value of X_{max} is approximately 800 $g \cdot cm^{-2}$. In the context of a completely vertical shower, the altitude of the Malargue site is

situated at X_{ground} ~870 $g \cdot cm^{-2}$, which is in good agreement with most of the proton showers and also showers due to other hadrons. As discussed in Chapter 2.4.1, a correction is implemented for the inclined showers which adjust the ground altitude, denoted as $X_{ground}(\theta)$, see equation 2.5. This correction accounts for the angle of inclination (θ) and ensures that the observed X_{ground} aligns appropriately with the zenith angle.

For different hadron-induced showers, 6.2a shows the heights at which showers have developed to their maximum according to the location of the Observatory. Also, the relationship between the average distance from the surface to the X_{max} and the true zenith angle (θ_{MC}) is illustrated in Figure 6.2b. Only ~2.7% of the proton and ~0.5% helium showers have fully developed below the observation level. As these showers are in very low amounts, their absence of presence in our analysis would not cause any significant deviations, as such outliers are efficiently handled in our analysis technique, which is described in the subsequent chapter. Although, without applying a cut at the zenith angle (Fig. 6.3a) there are few outlying data points in the proton showers while with a cut at $\theta_{MC} = 20°$ (Fig. 6.3b) the range of X_{max} is quite stringent. Consequently, applying a selection criterion of 20° at the zenith angle is the beneficiary in the analysis.



(a) **(b)**

Figure 6.3: The box plots of the distances between the location of X_{max}^{MC} and the Earth's surface X_{ground} are presented in the units of $g \cdot cm^{-2}$. Different primary particles are labeled on the x-axis and are shown in different colors. The color section for each particle in the plot represents 50% of the data and the black line between the colored boxes delimitates the median of the data sets. The black verticle line with whiskers at both ends of the boxes contains the upper 25% range and the lower 25% range of the data sets. The black points outside this complete range are called outliers. The vertical dashed red line demarcates the observation level of the observatory. Values below this line have their X_{max}^{MC} below the ground. **(a)** Box plot for the showers that have θ_{MC} below 55°. Certain showers have developed much deeper in the ground which are considered as outliers. **(b)** Box plot for the showers that have θ_{MC} below 55° and above 20°.

CHAPTER 7

ANALYSIS OF SIMULATED SHOWERS

In this chapter, we will investigate the effectiveness of AugerPrime in differentiating the composition of primary cosmic rays. To achieve this, we will analyze various observables at the Pierre Auger Observatory, utilizing simulated showers as described in the previous chapter.

Characterization of Extensive Air Showers can be done separately with the Water Cherenkov Detector (WCD) and Surface Scintillator Detector (SSD) and also with their combination. As discussed before, both detector types respond differently to the electromagnetic and muonic components of the showers, which are produced in different amounts based on the primaries that induced the showers. For this reason, we emphasize the combined response of both detectors. Initially, a detailed exploration of the relevant observables is conducted, followed by the application of Multivariate Analysis (MVA) on these variables to investigate the mass composition of the simulated showers of UHECRs. The methods developed in this chapter are applied to the observed data in the subsequent chapters.

7.1 Observables for hadron-induced shower discrimination

In Chapter 6, quality criteria were applied to the simulated showers to obtain a refined dataset. This section provides an overview of the initial method employed in selecting the most relevant observable for this work, subsequently proceeding to conduct an in-depth analysis of these chosen variables. These variables of different particles are studied through their distribution, zenith angle, and energy dependency. Primarily our focus is on the AugerPrime variables but other variables optimized for WCD are also used.

7.1.1 Variable selection method

The Merit Factor (MF) is used as a first method to select a variable by understanding discrimination between the distribution of different particles. The MF is defined as

$$MF = \frac{|\langle X_1 \rangle - \langle X_2 \rangle|}{\sqrt{\sigma^2_{X_1} + \sigma^2_{X_2}}} \tag{7.1}$$

where X represents a mean variable and σ is the standard deviation of a variable. The differentiation between primary particle types is denoted by numbers 1 and 2. Proton and iron particles are mostly used for the investigation as they are more distinct compared to other particles. The distribution of helium and oxygen particles falls intermediate of the proton and iron. However, all four particle's data is compared wherever necessary. Several variables are studied using the merit factor, only the necessary variables are presented in this thesis, which is used for the analysis.

Additionally, the RFECV is also used and run random forest including those variables and compare them with the variables only selected with merit factor. Also, include the person correlation test for the better understanding of the relation between observables.

7.1.2 Total Signal Ratio

The composition of extensive air showers (EAS) is influenced by the nature of the primary particle. This characteristic can be exploited by employing Water Cherenkov Detectors and Surface Scintillator Detectors, as they exhibit distinct responses to the electromagnetic and muonic components of the showers (as detailed in section 3.3.1).

Within this section, we delve into the signals registered by the Water Cherenkov Detectors and Surface Scintillator Detectors, and explore various signal combinations to derive the intended variables. However, this has been carried out extensively in a previous work [77] and here we only summarize this detailed study. The signals in the WCDs and SSDs linearly increase with increasing energy and are mostly constant with the zenith angle, this can be seen in Figure 7.1. However, a slight decrease can be noticed in the SSD signal at the larger angle in the case of a hadronic shower. Moreover, the signals in both detector types are linearly correlated to each other but the WCDs saturate faster than SSDs this linearity is deviated. This is observed in the hottest stations that are closest to the core of the shower, where the WCDs get saturated more often. The three hottest stations collectively contribute to more than 80% of the signal on average. The omission of saturated stations, which are generally the hottest ones, leads to a reduction in the overall signal strength. This issue is discussed again in the chapter about the future.

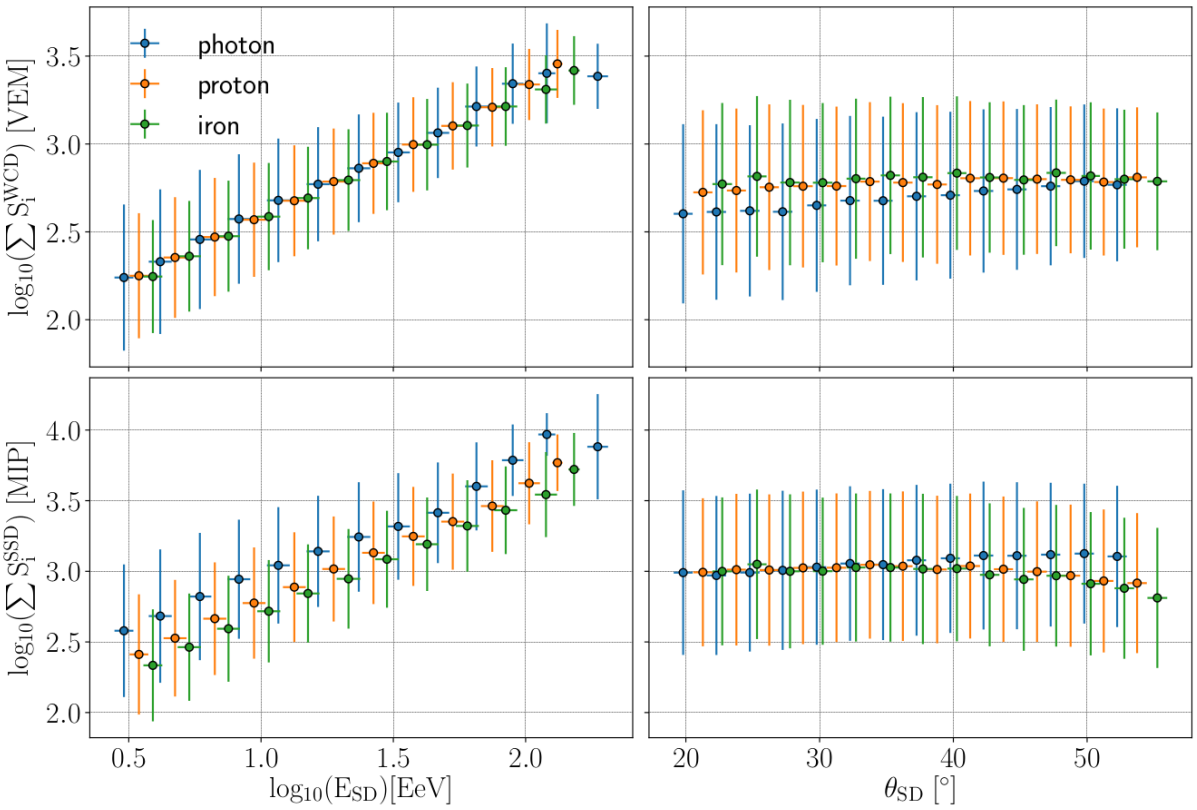


Figure 7.1: The average total signals from WCDs and SSDs are depicted in response to changes in reconstructed energy (left) and zenith angle (right) for the simulated events. Sourced from [77].

The Total Signal Ratio (TSR) for an event is defined as the ratio of the total signal measure by all the triggered SSDs to that of all the triggered WCDs. It is given in the unit of *MIP/VEM*. As mentioned, these detector types have different responses to electromagnetic and muonic components, consequently, the ratio of their detected signals provides cumulative information on these different components. The TSR is given as,

$$TSR = \frac{\sum S_i^{SSD}}{\sum S_i^{WCD}} \left[\frac{MIP}{VEM}\right] \tag{7.2}$$

Figure 7.2 illustrates the distributions of the TSR of the proton and iron simulated data sets. The distribution of helium and oxygen are intermediary and their mean and standard deviation can be seen in table 7.1a. As anticipated, the proton distribution exhibits its peak at higher values, reflecting the smaller presence of their muonic component. Subsequently, the helium distribution follows, succeeded by the oxygen distribution, and lastly, the iron distribution which has the largest muonic component.

In Figure 7.4, the correlations of the TSR with the reconstructed energy and zenith angle. The TSR is constant with the energy but has a strong sensitivity to the zenith angle, wherein TSR experiences a reduction of approximately

Particle	Mean with σ
Proton	1.11±0.17
Helium	1.07±0.17
Oxygen	1.03±0.17
Iron	0.99±0.16

(a)

Particle - Particle	Merit Factor
Proton - Helium	0.14
Helium - Oxygen	0.19
Oxygen - Iron	0.20
Proton - Iron	0.52

(b)

Table 7.1: (a) The mean with standard deviation (σ) of TSR distribution for all particles. The mean value decreases with the increase in mass of the particle. **(b)** The MF for the distribution of two consecutive particles in the context of their mass. Also, the MF given for the iron with proton distribution, which can be seen in Figure 7.2.

30% as the zenith angle increases from 20° to 55°. This observed trend aligns with the anticipated electromagnetic to muonic component ratio at larger zenith angles, where the electromagnetic component has already undergone substantial absorption, possibly even completely absorbed. A consequence of the decrease in TSR or the electromagnetic component with increasing θ_{SD} is the corresponding reduction in the number of detectors associated with events characterized by a higher number of detectors. This is illustrated in Figure 7.3, where the TSR is represented alongside the numbers of SSD and WCD detectors.

Another approach to examining station dependency involves quantifying the extent to which the TSR undergoes variations in response to the absence of SSDs and WCDs. This method offers a different perspective on station influence and provides insights into the impact of station removal on TSR. This has been studied with details in the [77]. From this study, it can be concluded that removing the saturated detectors from the analysis gives better discrimination between the TSR of the showers induced by different particles.

7.1.3 Lateral Distribution Function

Lateral Distribution Function (LDF) describes the particle density in a shower as a function of distance from the shower axis. As described in the section 4.3 it is parameterized by $S(r_{opt} = 1000m)$, β, and γ. $S(r_{opt} = 1000m)$ is the signal at distance r_{opt}, and β and γ are the slope parameters. For the Pierre Auger Observatory, r_{opt} is optimized to 1000 meters, such that fluctuation in signal with respect to the slope parameter, β, is smallest. The distribution of S(1000 m) or S_{1000}, β, and γ for SSDs and WCDs is shown in Figure 7.5. However, there are no clear differences in these parameters for different particle types can be seen. Their merit factor is also shown in Table 7.2.

Similar to Total Signal Ratio (TSR), the ratio of S_{1000} ratio is also obtained by dividing S_{1000}^{SSD} with S_{1000}^{WCD}, which is shown in Figure 7.6 and its dependency on the reconstructed energy and the zenith angle is shown in Figure 7.7. In contrast to the TSR, there is a pronounced energy dependency, where S_{1000} ratio increases with E_{SD}, irrespective of the primary particle type. Additionally, the ratio demonstrates

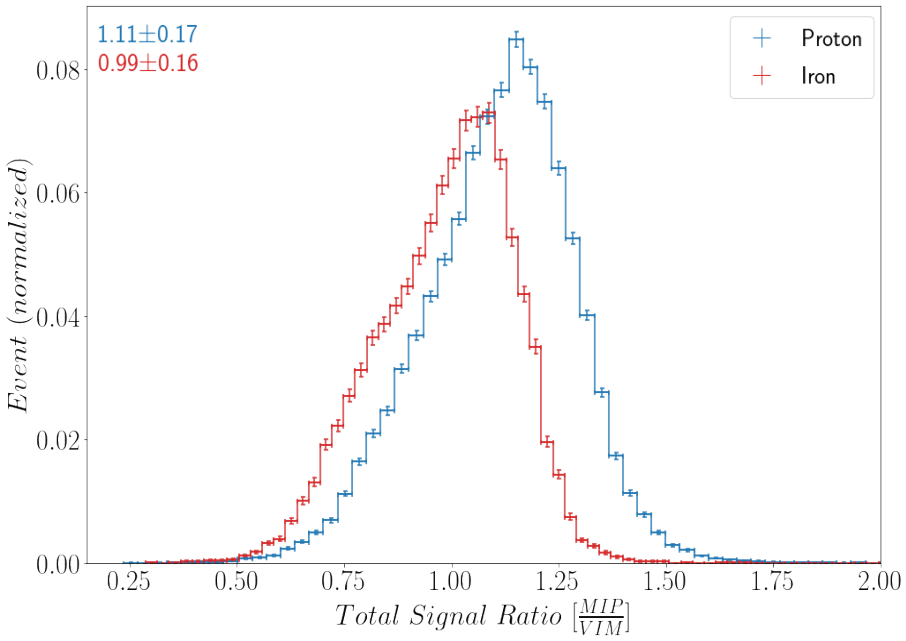


Figure 7.2: The distributions of Total Signal Ratio were generated by dividing the total signal obtained from SSDs by that from WCDs within simulated datasets. The values presented in the upper-left corner of the figure denote the mean and standard deviation for each respective distribution. The distribution of proton is shown with blue while iron with red. These distributions have been normalized to account for the total number of events.

a noticeable inverse dependence on the zenith angle. The merit factor between hadronic particles for S_{1000} ratio is shown in Table 7.2.

7.1.4 Number of Station

The number of triggered stations is significantly dependent on the nature of the primary particle initiating the shower. For a particular energy, iron-induced showers trigger a greater number of stations compared to proton-induced showers. The showers induced by helium and oxygen particles trigger more stations than those triggered by protons but fewer than those by iron particles. However, this observable is vulnerable to fluctuation due to the geometry of showers, trigger criteria, aging of detectors, and 1MIP signal cut on SSDs. These factors mainly affect the detectors on the edge of the shower. The distribution of the number of SSD ($N^{SSD}_{Stations}$) and WCD ($N^{WCD}_{Stations}$) detectors selected in each shower is shown in Figure 7.8 and its dependency on reconstructed energy and zenith angle in Figure 7.9.

The ratio of several stations is also studied, its distribution is shown in Figure 7.10c along with its correlation with reconstructed energy and zenith angle. The events where the number of SSDs are larger than WCDs detectors is due to the removal of saturated WCDs.

Table 7.3 shows the merit factor for several Surface Scintillator Detector (SSD)

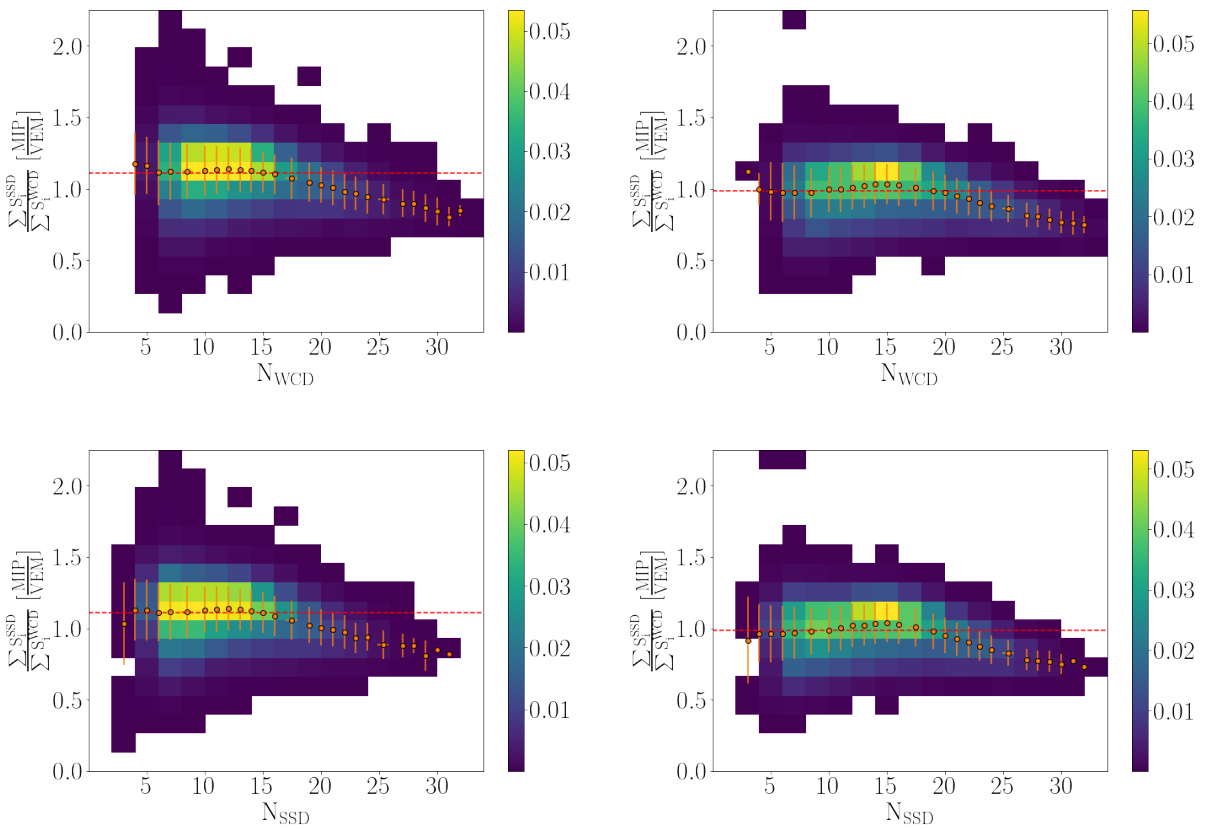


Figure 7.3: Density plots depict the correlation between the number of chosen detectors and their corresponding TSR. The left panels portray simulated proton-induced showers, while the right panels represent iron-induced showers. The upper panel visualizes correlations with the selected WCDs, while the lower panels illustrate the same correlations with the SSD. A dashed red line demarcates the average TSR value within the corresponding simulated dataset. The orange circles and vertical bars indicate the average TSR and the standard deviation associated with the specific number of detectors.

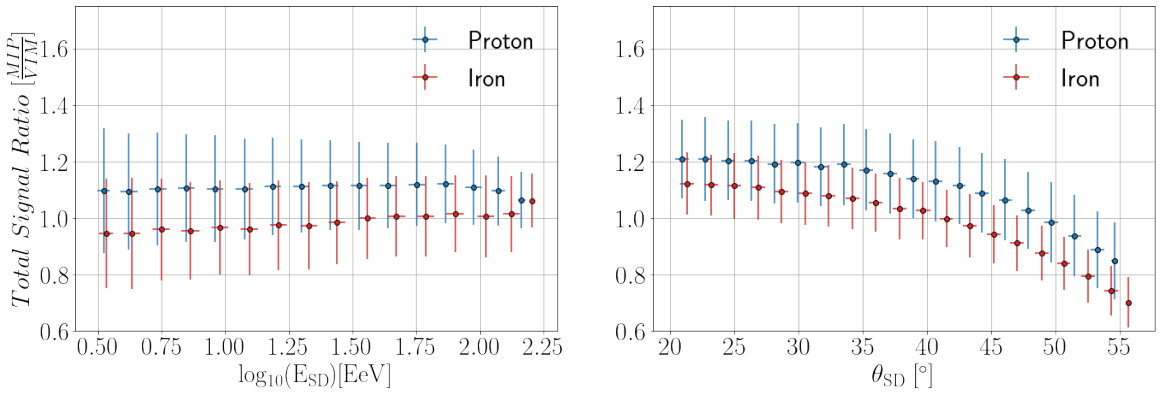
detectors and WCD detectors. Also, the merit factor of their ratio is shown in the table.

7.1.5 Reconstructed Energy

7.1.6 Time over Distance

The risetime ($t_{1/2}$) is a parameter denoting the time period required for the signal registered by a WCD detector transition from 10% to 50% of the total signal magnitude. The specific value of the $t_{1/2}$ is dependent upon the radial distance r of the detector from the shower core. Stations positioned at greater distances from the shower core will observe a broader temporal distribution in the arrival of particles, leading to higher values of $t_{1/2}$. Conversely, stations located near the shower core will record lower $t_{1/2}$ values.

Using the $t_{1/2}$ and distance of the stations from the shower core a new variable



(a) **(b)**

Figure 7.4: The TSR presented with reconstructed energy **(a)** and zenith angle **(b)**. The error bars denote the standard deviation of distributions. A horizontal shift was introduced for iron distributions by multiplying E_{SD} and θ_{SD} with 1.02. This is executed only for the clear reading of the plots. The true values of the energy and zenith angle can be read from the proton distribution.

Particle -Particle	**Merit Factor**						
	S_{1000}^{WCD}	S_{1000}^{SSD}	β^{WCD}	β^{SSD}	γ^{WCD}	γ^{SSD}	$S_{1000}^{SSD}/S_{1000}^{WCD}$
Proton-Iron	0.30	0.22	0.12	0.08	0.06	0.09	0.46
Proton-Helium	0.08	0.08	0.08	0.06	0.02	0.02	0.16
Helium-Oxygen	0.09	0.06	0.03	0.02	0.00	0.01	0.13
Oxygen-Iron	0.08	0.07	0.00	0.01	0.01	0.02	0.17

Table 7.2: Merit factor of LDF parameters from simulated showers of the SSD and WCD. Also, the MF for S_{1000} ratio is shown here. The MF calculated between two consecutive hadronic particles based on increasing order of their atomic mass and also between iron and proton.

has been adopted from [96]. This new variable known as "Time over Distance" ($\overline{\text{ToD}}$) is given as,

$$\overline{ToD} = \left\langle \frac{t_{1/2}}{r} \right\rangle = \frac{1}{N} \sum_{i=1}^{N} \frac{t_{1/2}}{r_i} \tag{7.3}$$

Here, the summation carried from 1 to the total number of selected detectors, N. A drawback of this variable is that the risetime is not linear to the detector distances beyond 2000 meters [96]. However, this deviation is not too large this observable is still used in our preliminary analysis.

The distribution of the variable $\overline{\text{ToD}}$ is shown in Figure 7.11. From the distribution, no difference is found between the proton- and iron-induced showers. However, it is still used in our analysis as it is selected by other observable selec-

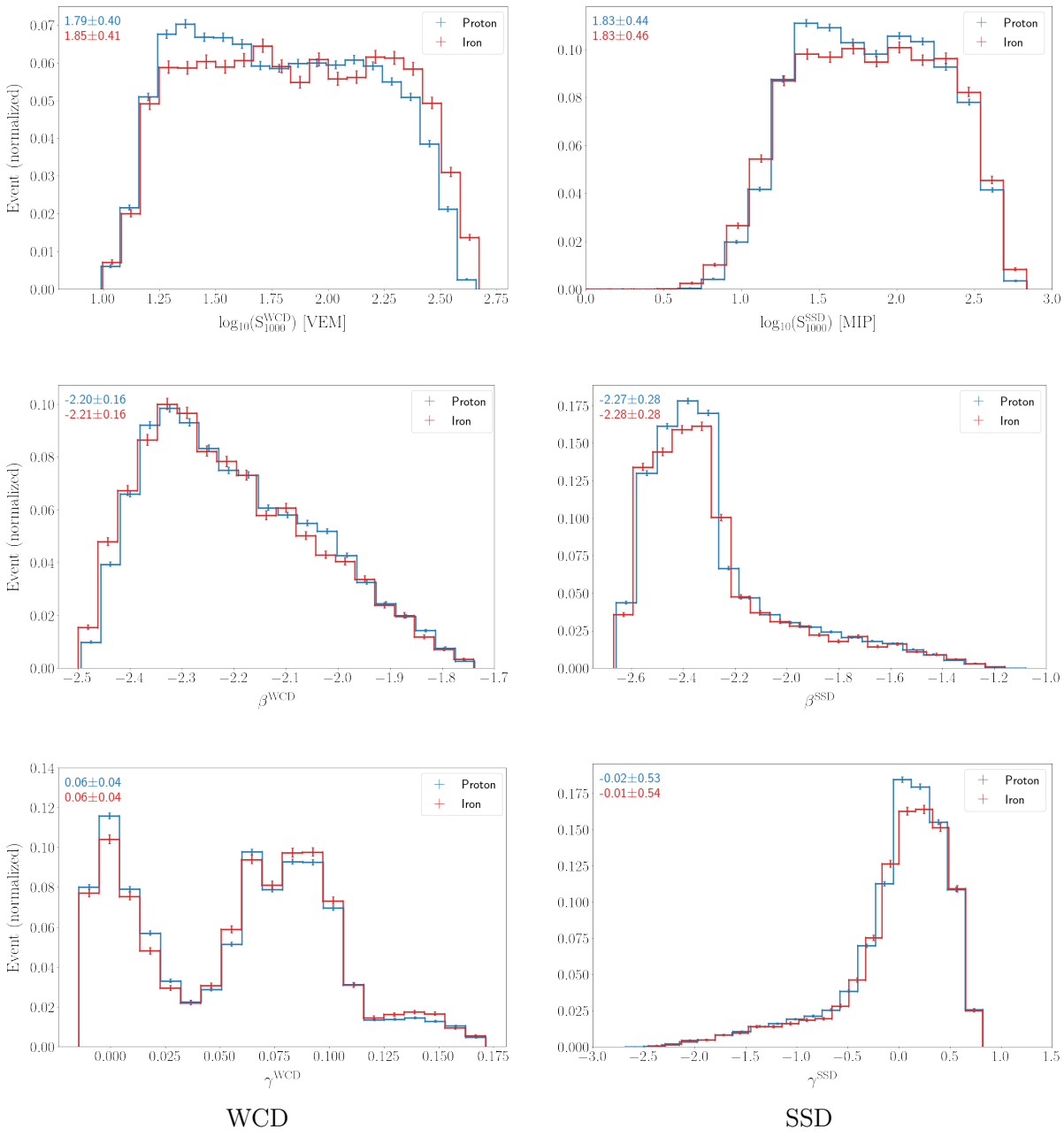


Figure 7.5: The normalized distribution of S_{1000} (top), β (middle), γ (bottom) for WCD (left) and SSD (right). The mean and standard deviation of distributions are shown in the top left corner of each plot, which are color-coded according to their particle type, namely proton and iron.

tion methods, which will be discussed later in this chapter. $\overline{\text{ToD}}$ shows a strong correlation with reconstructed energy and remains constant with the change in the zenith angle, this can be seen in Figure 7.11. The correlation with energy comes from the observation that primary particles possessing higher energy levels result in the trigger of detectors situated at greater distances from the shower core, consequently inducing a longer risetime in these detectors.

The MF of $\overline{\text{ToD}}$ for showers induced by adronic particles in consideration is

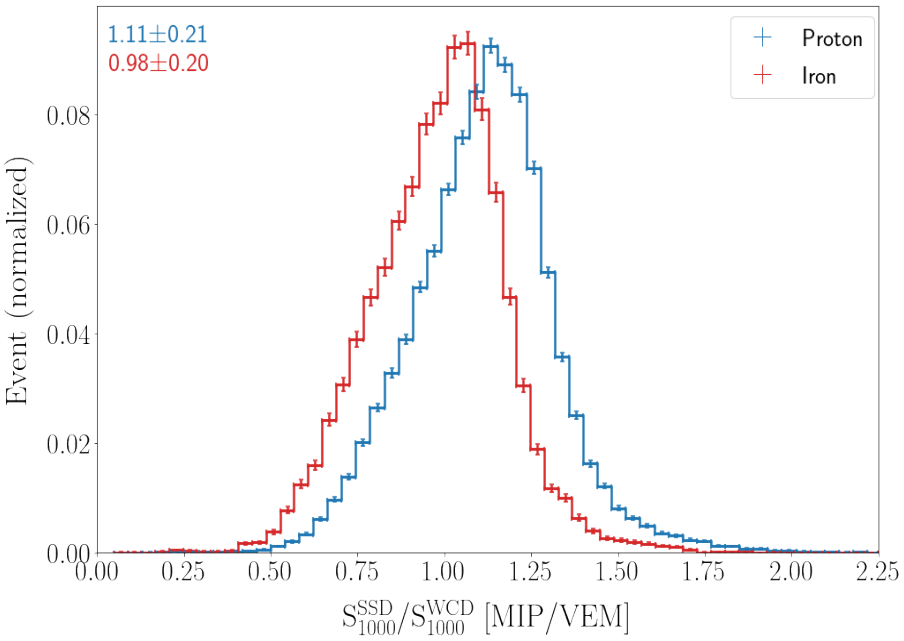


Figure 7.6: The distributions of $S(r_{opt} = 1000\ m)$ ratio were generated by dividing the summation of S_{1000} signal obtained from SSDs by that from WCDs for each event within simulated datasets. The values presented in the upper-left corner of the figure denote the mean and standard deviation for each respective distribution. The distribution of proton is shown with blue while iron with red. These distributions have been normalized to account for the total number of events.

given in Table 7.4.

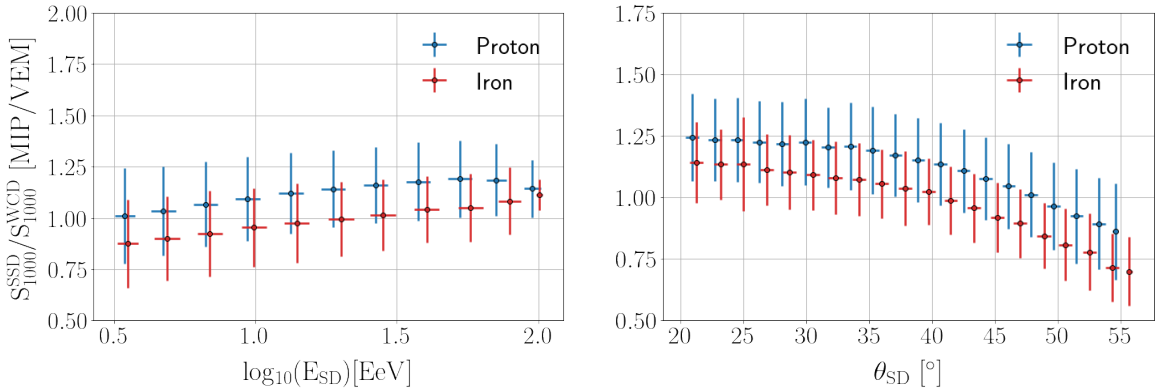
7.1.7 Observable S_b

The parameter S_b exhibits sensitivity to the collective impact of muonic and electromagnetic components on the lateral distribution function within air showers initiated by different hadronic particles [97]. S_b is defined as

$$S_b = \sum_{i=1}^{N} \left[S_i \left(\frac{r_i}{r_0} \right)^b \right] \tag{7.4}$$

Where r_0 is the reference distance set to 1000 m, S_i is the signal measured at the i^{th} detector located at the r_i distance from the shower axis. The summation runs over N numbers of triggered stations in an event. The free parameter b is optimized to 3 for the hadronic showers.

In the parameter S_b the signal in the detector is scaled by the ratio r_i/r_0, the exponent 3 enhances this scaling factor if it is larger than 1 and diminished if it is smaller than 1. Consequently, the signal at the farther detector signal is enhanced, and for detectors near the shower core, the signal is reduced. Thus, S_b emphasizes the detectors at the edge of the shower.



(a) **(b)**

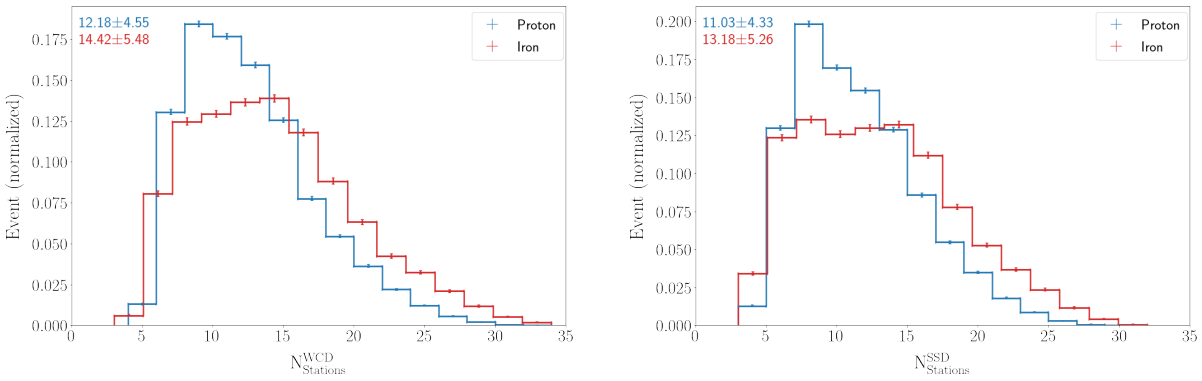
Figure 7.7: The $S(r_{opt} = 1000\ m)$ ratio presented with reconstructed energy **(a)** and zenith angle **(b)**. The error bars denote the standard deviation of distributions. A horizontal shift was introduced for iron distributions by multiplying E_{SD} and θ_{SD} with 1.02. This is executed only for the clear reading of the plots. The true values of the energy and zenith angle can be read from the proton distribution.

Particle-Particle	$N^{SSD}_{Stations}$	$N^{WCD}_{Stations}$	$N^{SSD}_{Stations}/N^{WCD}_{Stations}$
Proton-Iron	0.32	0.31	0.90
Proton-Helium	0.16	0.16	0.90
Helium-Oxygen	0.10	0.10	0.87
Oxygen-Iron	0.05	0.06	0.86

Table 7.3: Merit factor for number of SSD and WCD detectors and their ratio. The MF calculated between two consecutive hadronic particles based on increasing order of their atomic mass and also between iron and proton.

The parameter S_b is employed to modulate the signal recorded by detectors, utilizing the ratio r_i/r_0 as a scaling factor. The inclusion of an exponent 3 serves to amplify this scaling effect when the ratio exceeds 1 and to diminish it when the ratio is less than 1. Consequently, detectors located at a greater distance from the shower core exhibit an enhanced signal, while detectors near the core experience a reduction in signal strength. As a result, S_b accentuates the role of detectors situated at the edge of the shower. In this section, we discuss S_b for WCD, SSD and their ratio.

The distribution of S_b^{WCD} for iron- and proton-induced showers is shown in Figure 7.12 along with its reconstructed energy and zenith angle dependency. Similarly, the plots are also shown for S_b^{SSD} in Figure 7.13. In S_b^{WCD} and S_b^{SSD} distributions, iron dominates at higher values as the shower is larger and triggers more stations. Also, both S_b^{WCD} and S_b^{SSD} show a high dependency on energy, as the higher the energy, the more stations will be triggered beyond 1000 m. However, there is no zenith angle correlation is observed. The correlation between S_b^{WCD}



(a) **(b)**

Figure 7.8: The presented distributions of the number of SSD detectors **(a)** and number of WCD detectors **(b)**, for simulated showers initiated by protons and iron nuclei. The statistics, including the mean and standard deviation, are indicated in the top left corner for reference.

Particle-Particle	$\overline{\text{ToD}}$
Proton-Iron	0.09
Proton-Helium	0.04
Helium-Oxygen	0.03
Oxygen-Iron	0.03

Table 7.4: Merit factor for $\overline{\text{ToD}}$. The MF calculated between two consecutive hadronic particles based on increasing order of their atomic mass and also between iron and proton.

and S_b^{SSD} is shown in Figure 7.14. The MF of the observables S_b^{WCD}, S_b^{SSD} and their ratio for all the hadronic particles is shown in Table 7.5.

Conversely, the ratio S_b^{SSD}/S_b^{WCD} remains almost constant with energy but dependent on the zenith angle. This can be seen in Figure 7.15 along with its distribution.

7.1.8 Radius of Curvature

Measurement of the radius of curvature of the shower provides insight into the curvature of the shower front. This curvature stems from the fact that particles at a specific lateral distance from the shower axis take longer to reach the ground compared to those near the axis. The progressive time delay of particles, increasing with their distance from the shower axis, contributes to the curvature observed in the shower front. The curvature, or the radius of curvature, is more pronounced when outer particles experience greater delays. Consequently, showers characterized by a smaller X_{max} exhibit a more substantial radius of curvature. This implies that proton-induced showers, develop at greater atmospheric depths,

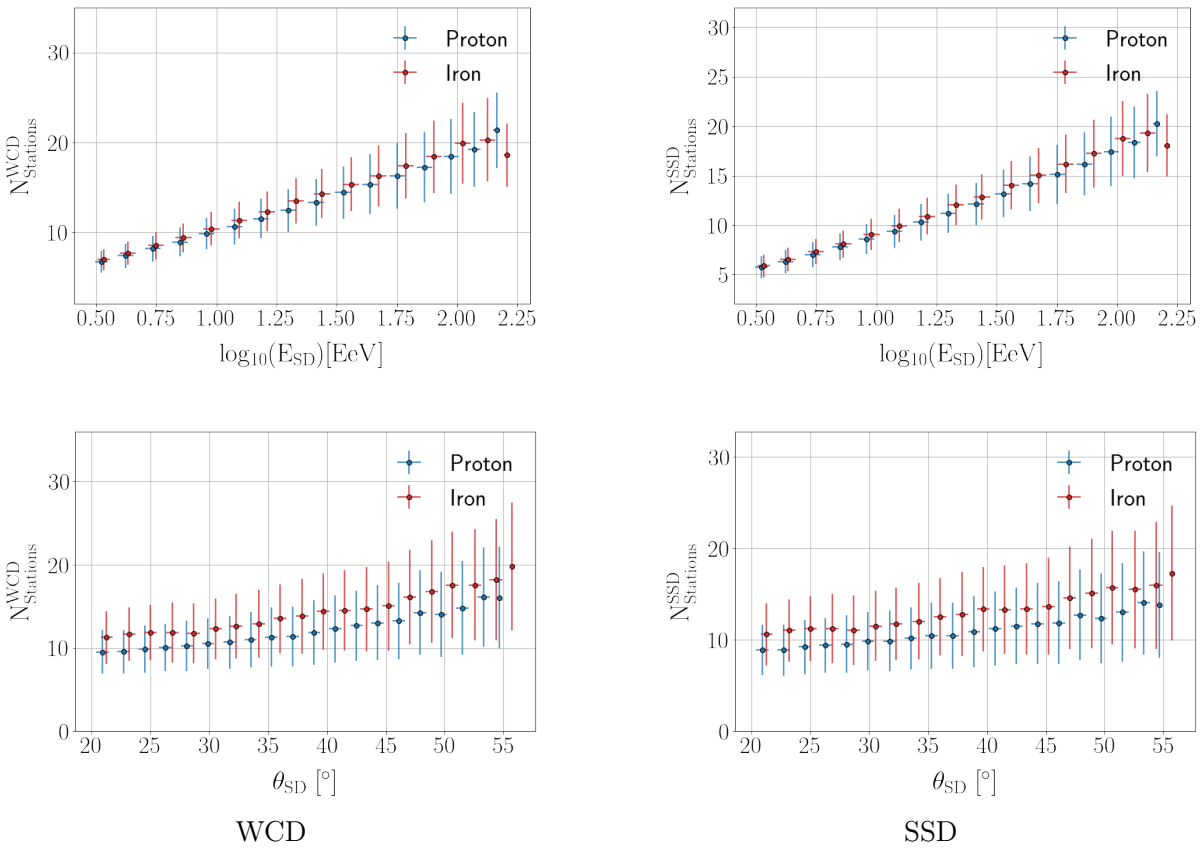


Figure 7.9: The presented distributions of the number of SSD **(a)** stations and a number of WCD **(b)** stations, for simulated showers initiated by protons and iron nuclei. The statistics, including the mean and standard deviation, are indicated in the top left corner for reference.

as they display a smaller radius of curvature compared to iron-induced showers. This can be seen in Figure 7.16. The higher energy particle produces larger showers and larger energy leads to X_{max} at the higher altitudes, thus the larger radius of curvature. The correlation of the radius of curvature with reconstructed energy and zenith angle is also shown figure which shows its dependency on both the observable. The MF for the radius of curvature for all hadronic particles is shown in Table 7.6.

The Offline Framework incorporates the fitting for the radius of curvature and it is described in [77].

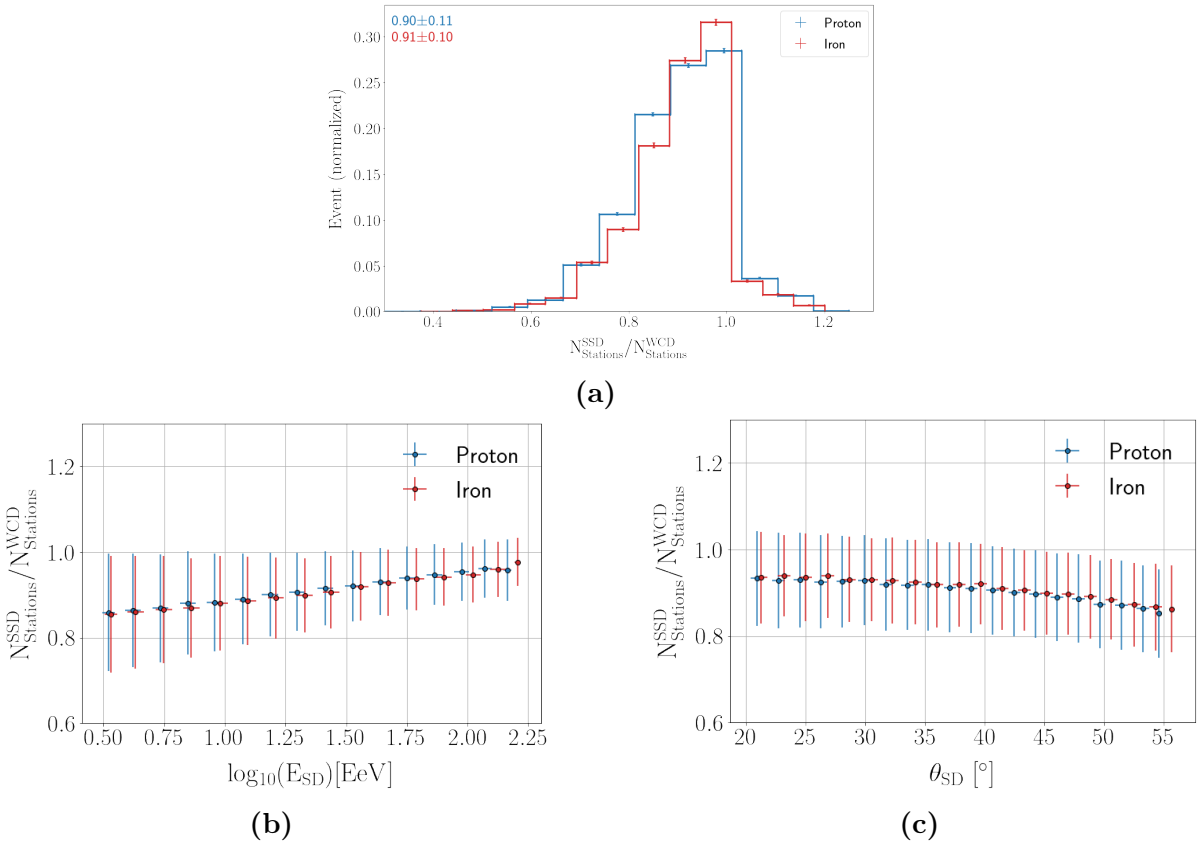


Figure 7.10: The ratio of number of detectors presented with reconstructed energy **(a)** and zenith angle **(b)**. The error bars denote the standard deviation of distributions. A horizontal shift was introduced for iron distributions by multiplying E_{SD} and θ_{SD} with 1.02.

Particle-Particle	S_b^{WCD}	S_b^{SSD}	S_b^{WCD}/S_b^{WCD}
Proton-Iron	0.18	0.12	0.34
Proton-Helium	0.05	0.06	0.11
Helium-Oxygen	0.06	0.04	0.10
Oxygen-Iron	0.06	0.03	0.13

Table 7.5: Merit factor for S_b^{WCD}, S_b^{SSD} and their ratio. The MF calculated between two consecutive hadronic particles based on increasing order of their atomic mass and also between iron and proton.

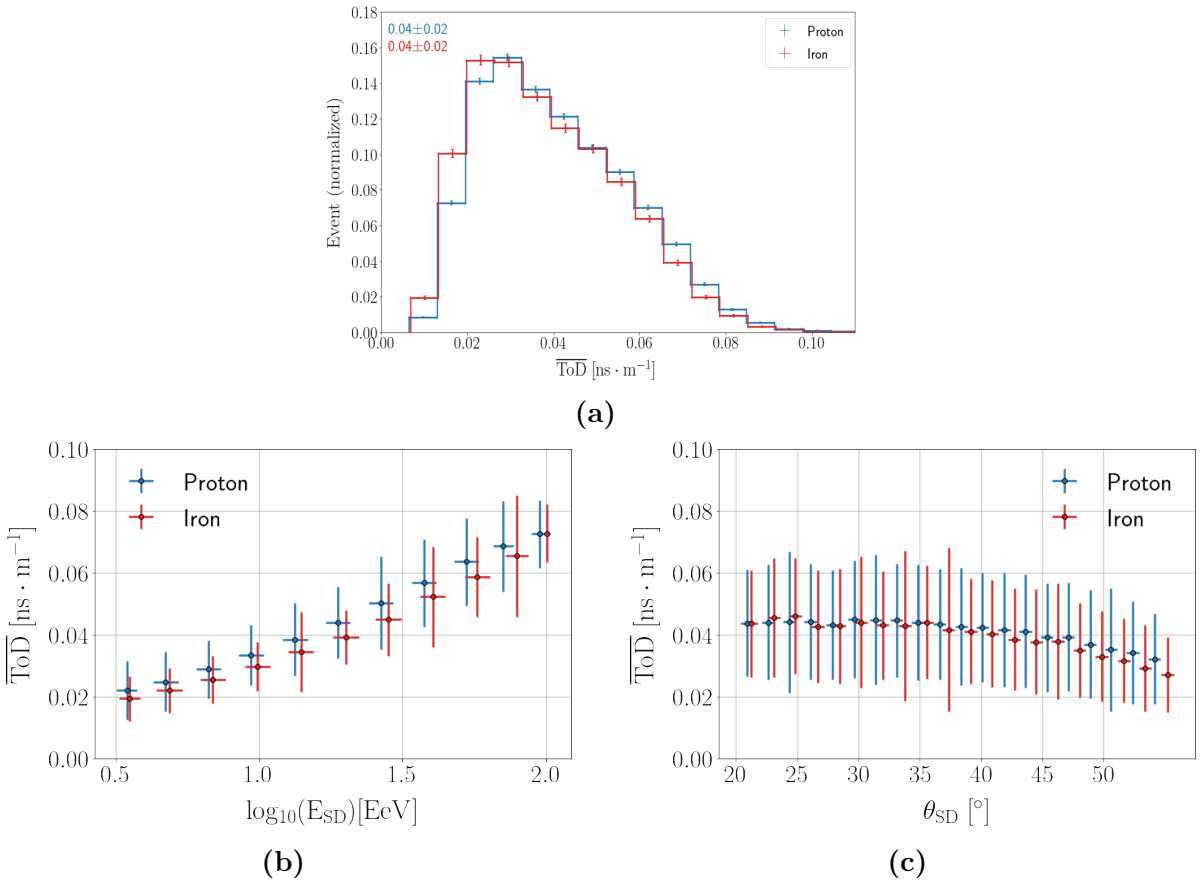


Figure 7.11: **(a)**The presented distributions of the $\overline{\text{ToD}}$, for simulated showers initiated by protons and iron nuclei. The statistics, including the mean and standard deviation, are indicated in the top left corner for reference. The correlation plots illustrating the relationships between the $\overline{\text{ToD}}$ with the reconstructed energy and zenith angle are shown in **(b)** and **(c)**, respectively.

Particle-Particle	Radius of Curvature
Proton-Iron	0.38
Proton-Helium	0.10
Helium-Oxygen	0.16
Oxygen-Iron	0.11

Table 7.6: Merit factor for Radius of Curvature. The MF calculated between two consecutive hadronic particles based on increasing order of their atomic mass and also between iron and proton.

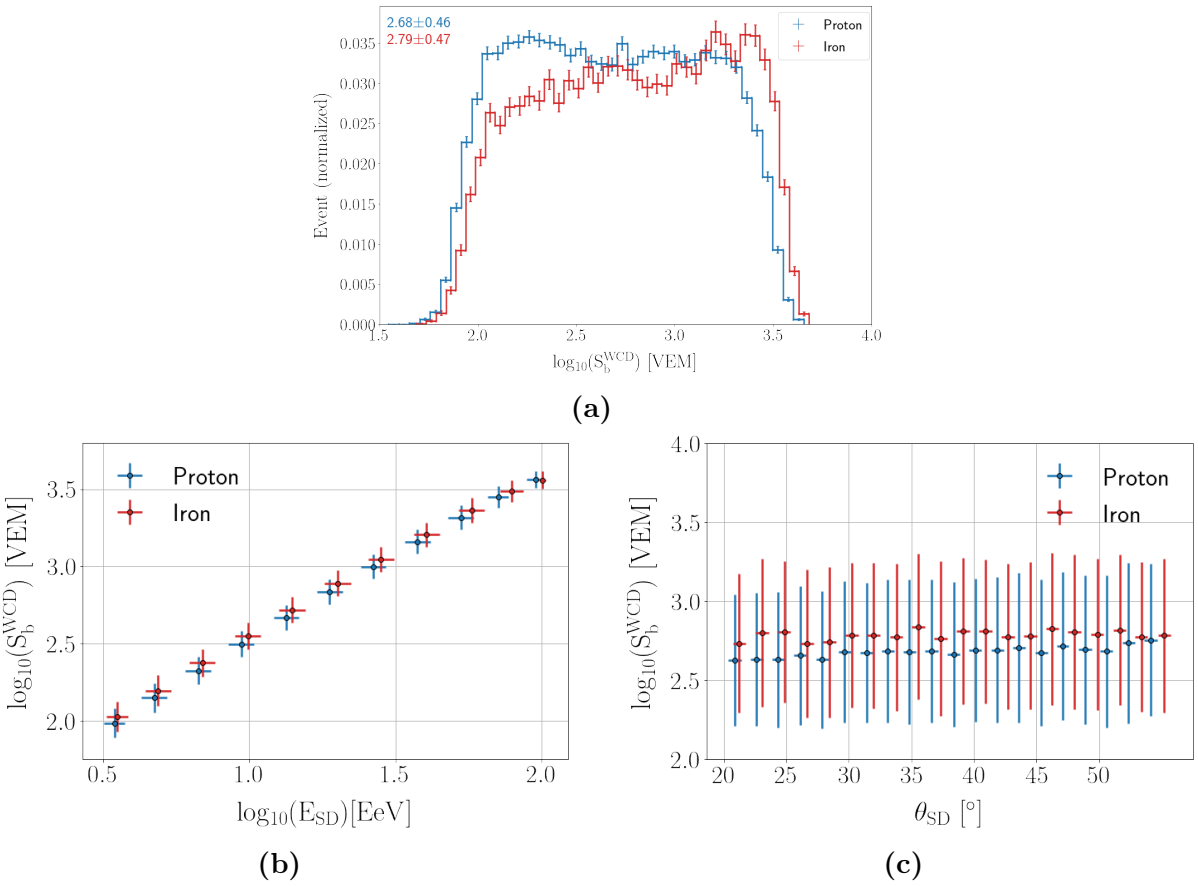


Figure 7.12: **(a)**The presented distributions of S_b^{WCD}, for simulated showers initiated by protons and iron nuclei. The statistics, including the mean and standard deviation, are indicated in the top left corner for reference. The correlation plots illustrating the relationships between S_b^{WCD} and the reconstructed energy and zenith angle are shown in **(b)** and **(c)**, respectively.

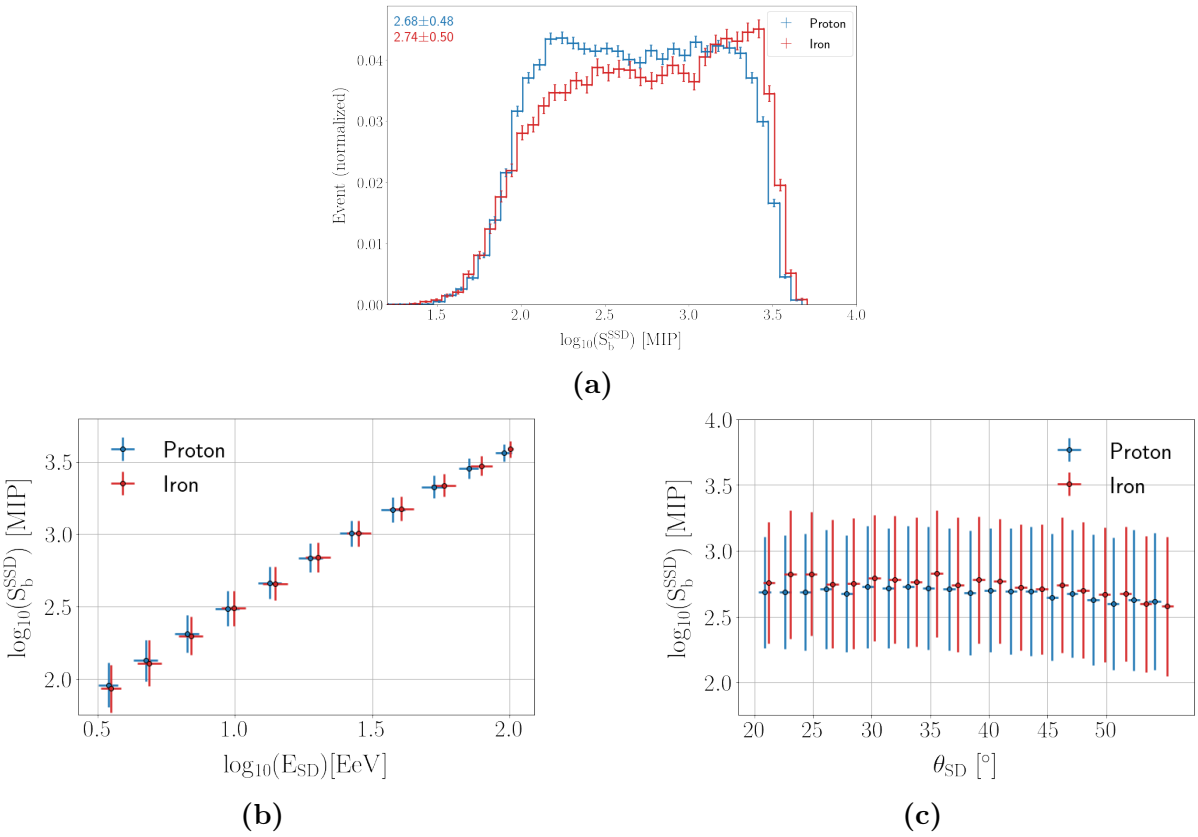


Figure 7.13: **(a)**The presented distributions of S_b^{SSD}, for simulated showers initiated by protons and iron nuclei. The statistics, including the mean and standard deviation, are indicated in the top left corner for reference. The correlation plots illustrating the relationships between S_b^{SSD} and the reconstructed energy and zenith angle are shown in **(b)** and **(c)**, respectively.

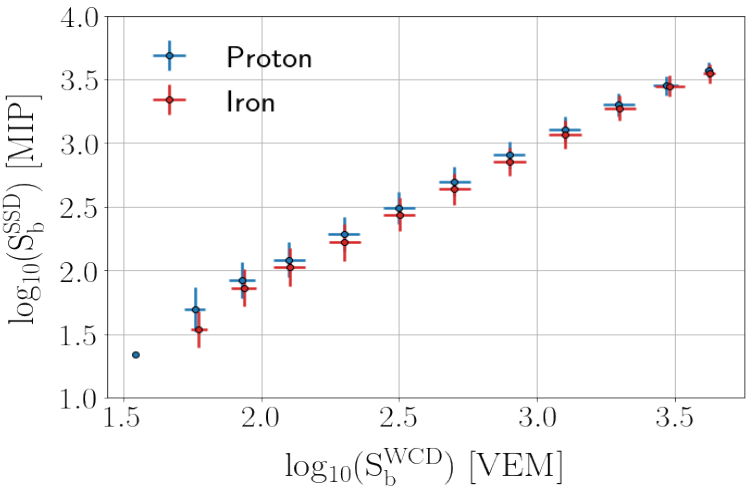


Figure 7.14: Plot Showing correction between S_b^{SSD} and S_b^{WCD} for proton- and iron-induced showers.

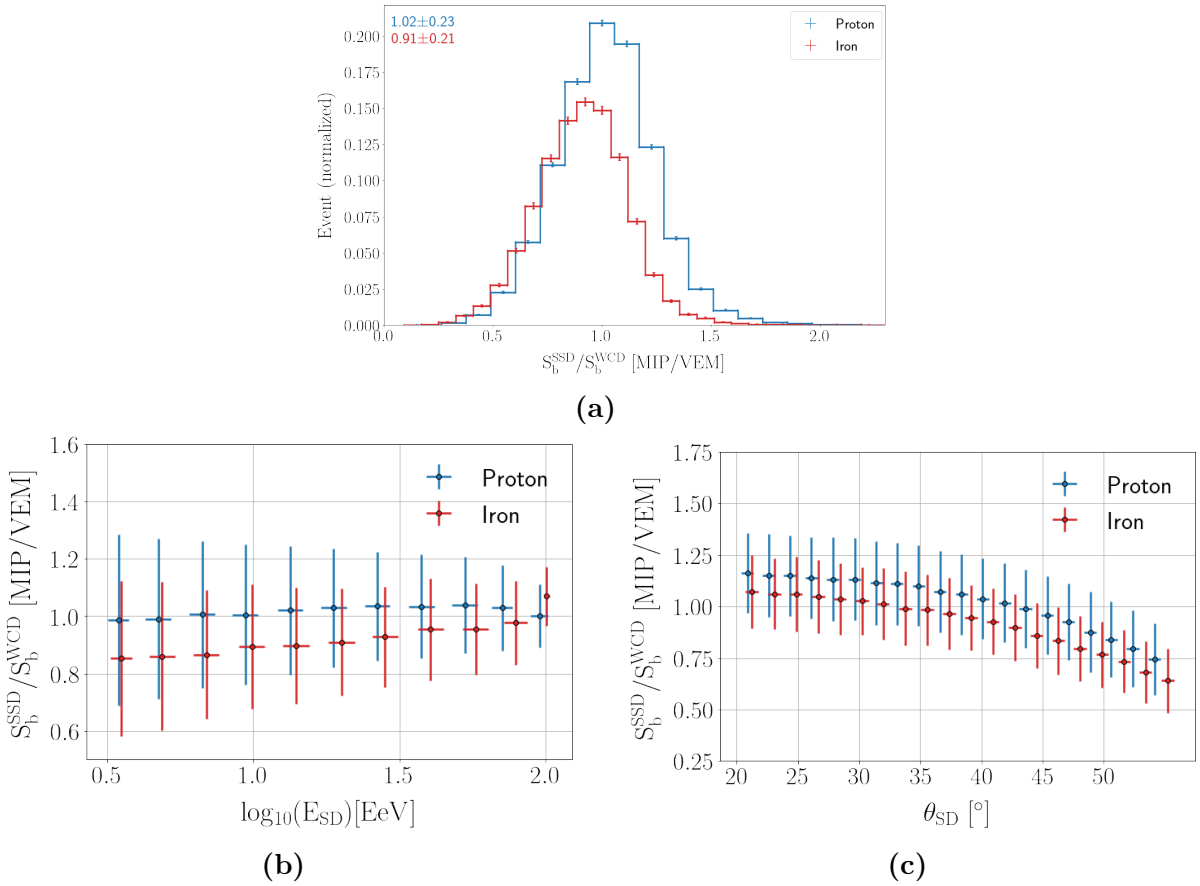


Figure 7.15: **(a)**The distributions of the ratio of S_b^{SSD} and S_b^{WCD}, for simulated showers initiated by protons and iron nuclei. The statistics, including the mean and standard deviation, are indicated in the top left corner for reference. **(b)** and **(c)** show the correlation between this ratio and the reconstructed energy and zenith angle, respectively.

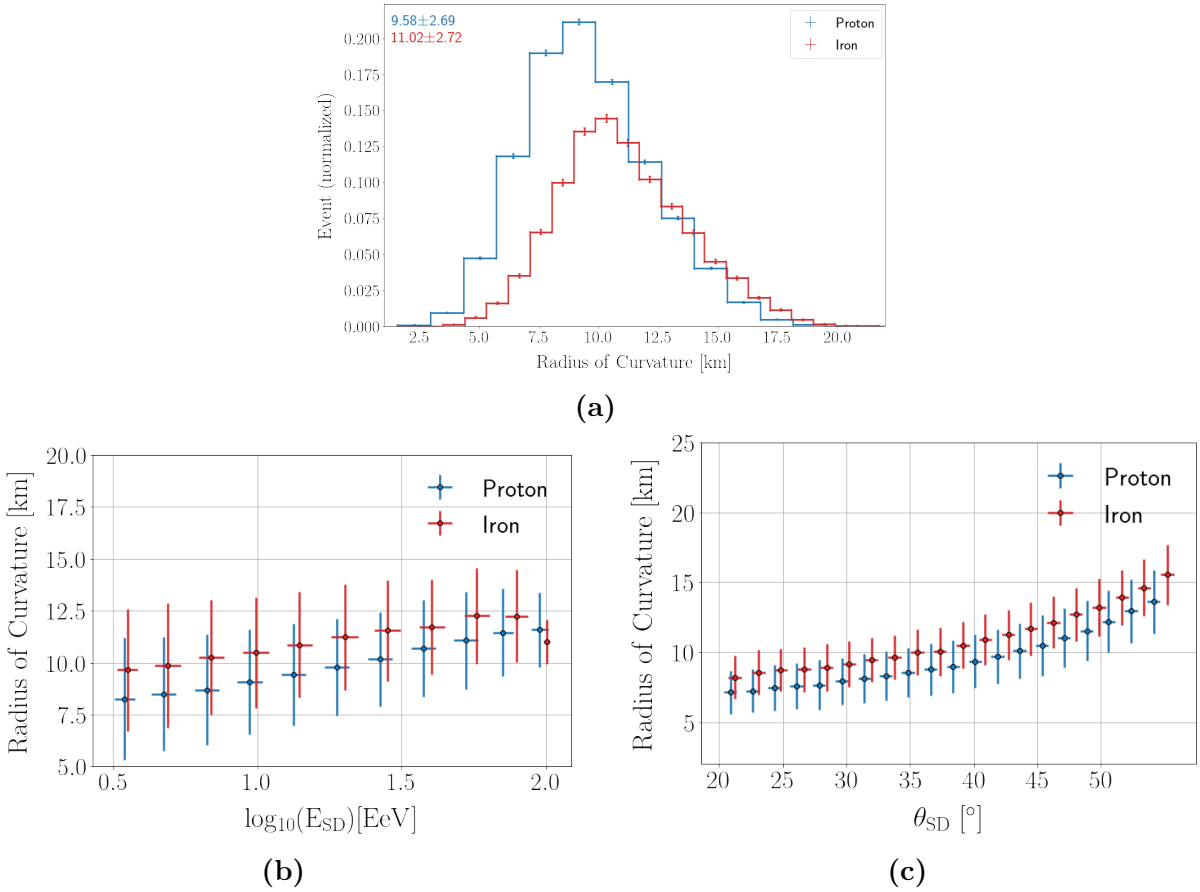


Figure 7.16: **(a)**The presented distributions of the radius of curvature, for simulated showers initiated by protons and iron nuclei. The statistics, including the mean and standard deviation, are indicated in the top left corner for reference. The correlation plots illustrating the relationships between the radius of curvature with the reconstructed energy and zenith angle are shown in **(b)** and **(c)**, respectively.

7.2 Multivariate Analysis

Multivariate Analysis (MVA) involves the combination of multiple variables to generate a unified outcome, making it a pivotal branch of Statistics with wide-ranging applications. Various techniques, ranging from traditional approaches like Principal Component Analysis (PCA) to modern machine learning algorithms, such as Boosted Decision Trees (BDT), Random Forest (RF), eXtream Grdient Boost (XGB) and Deep Neural Networks (DNN), are employed within this domain. This section delves into an MVA specifically applied to discriminate between proton, helium, oxygen, and iron. The chosen approach utilizes a Random Forest implemented in *Python*[1]. Following a concise introduction to the Random Forest methodology, the optimal combinations of observables are systematically assessed. In the evaluation potential uncertainties and opportunities for enhancements are considered.

7.2.1 Random Forest

Random Forest [98] is an ensemble learning technique extensively applied in scientific research for tasks such as classification and regression. Ensemble techniques involve the creation of multiple models instead of individual models, and with this prediction, performance is highly improved. In the context of Random Forest, the ensemble is comprised of multiple decision trees built during the training phase. Each tree is constructed using a randomly selected subset of training data which is called bagging or bootstrapping, this introduces variability and reduces overfitting, making models more versatile and resilient and capable of handling complex relationships within high-dimensional data. The final prediction is then determined by aggregating the outputs of these individual trees. This ensemble approach enhances the algorithm's generalization capabilities, robustness, and accuracy.

7.2.2 Random Forest Library in Python

In this thesis, the implementation utilizes the `RandomForestRegressor` library from the Python Scikit-learn package, as detailed in [99]. While Python is the primary choice, it is noteworthy that Random Forest algorithms can also be implemented using C/C++ or R. The hyper-parameter configurations utilized during the training of the RF in this chapter are as below.

```
sklearn.ensemble.RandomForestRegressor(n_estimators=800, *,
criterion='squared_error', max_depth=None, min_samples_split=2,
min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=1.0,
max_leaf_nodes=None, min_impurity_decrease=0.0, bootstrap=True,
oob_score=True, n_jobs=-1, random_state=101, verbose=0,
warm_start=False,ccp_alpha=0.0, max_samples=None)
```

[1]Python version 3.6.

To optimize the analysis, the `n_estimators`, `oob_score`, `n_jobs`, and `random_state` are the parameters that have changed from the default values. More details on these parameters can be found in [100].

7.2.3 Random Forest Implimentation

In the application of RF, the datasets are randomly partitioned into training and testing subsets. The datasets are detailed in Section 6.1. Approximately two-thirds of the datasets are allocated for training, while the remaining one-third is designated for testing. The training datasets of proton, helium, oxygen, and iron datasets are combined, and a similar grouping is formed for the testing sets. Before the combination of datasets, events are annotated based on the natural logarithm of the atomic mass of the primary particle, ln(A). Specifically, a proton is denoted as 0.0, helium as 1.39, oxygen as 2.77, and iron as 4.02.

In the previous subsection, the parameters for training sets were outlined. These sets are employed to train a model to classify based on the ln(A). This process involves utilizing various combinations of the observables described in the earlier chapter, with six observables at a time as training features. The criteria for selecting these features, along with a comprehensive discussion of their selection process are elaborated later in this chapter.

The decision to select six features at a time is decided based on the Recursive Feature Elimination with Cross-Validation (RFECV) feature selection method [101]. The RFECV method in Scikit-learn is a technique used for feature selection in machine learning. It combines the principles of recursive feature elimination (RFE) with cross-validated selection of the best number of features. RFECV works by fitting a model multiple times, each time removing the least important features based on specific criteria, such as feature weights in the models. It uses cross-validation to evaluate the performance of the model with different subsets of features, ensuring robustness and preventing overfitting. This process continues until the optimal subset of features is identified, offering a balance between model complexity and performance. In our analysis, the optimization of feature selection through the RFECV method is illustrated in Figure 7.17. This process involved an initial training and testing dataset comprising 17 variables. It was observed that the cross-validation score did not show significant improvement when more than six features were utilized.

The trained model is applied to the testing set to evaluate the performance of the RF model. This curve is built for each particle separately. The predicted data of the testing set for a particle in consideration is labeled as a signal and the average of the other three particles is taken as background rejection. For example, predictions of the proton are treated as a signal then the average of helium, oxygen, and iron predictions are background to the proton signal. Similarly, such calculations are carried out separately for datasets of helium, oxygen, and iron predictions. The performance evaluation of signal and background utilizes Receiver Operating Characteristic (ROC) curve. An ROC curve visualizes background re-

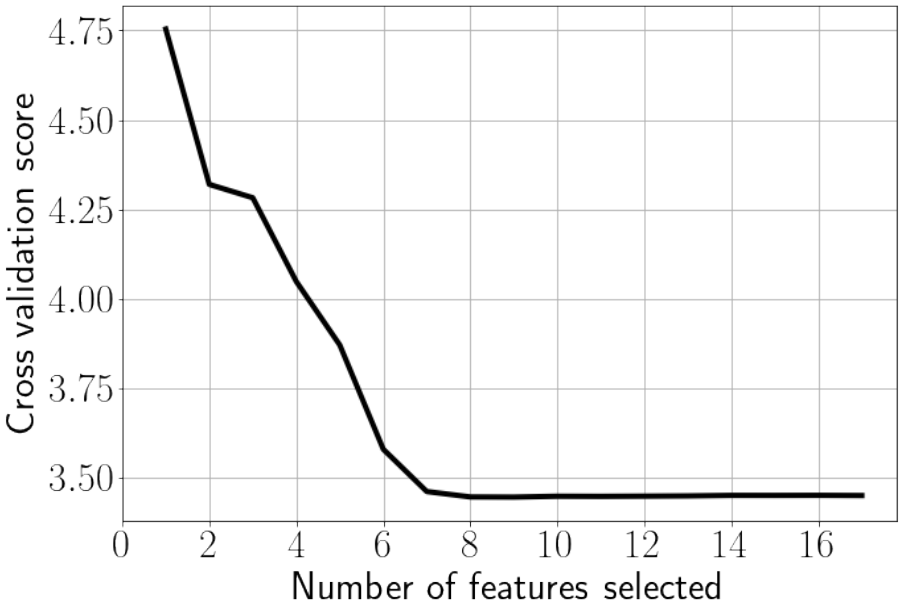


Figure 7.17: The cross-validation score obtained for a different number of selected features using the RFECV method. The lower cross-validation score provides better results.

jection as a function of signal. For example, if a threshold is set at a specific value x within the range of 0 to 4.02 in the RF predictions. The proportion of proton events exceeding this threshold x is defined as the signal efficiency and the proportion of helium, oxygen, and iron events that fall below this threshold constitutes the background rejection. To construct the ROC curve, successive thresholds are applied in descending order from 4.02 to 0. This methodical approach initiates the curve at a point representing 0% signal efficiency and 100% background rejection. In this chapter, ROC curves are plotted for proton- and iron-induced events only. The ROC curves are ranging from 0 to 1, representing 0% and 100% respectively.

In an optimal case, the ROC curve would be a straightforward horizontal line positioned at a value of 1. This configuration represents the achievement of 100% background rejection along with 100% signal efficiency.

Besides ROC curves, Root Mean Square Error (RMSE) [102] and MF are calculated for the output of each set of six observables. RMSE gives the spread of the predicted value compared to the actual value. It is defined as,

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y_i})^2} \tag{7.5}$$

Here, y_i and $\hat{y_i}$ are the true value and the predicted value and summation of their square runs for the total number of events, N. The lesser the RMSE score, the lesser will be the value deviation of the predicted value from the real value.

The errors associated with RMSE and MF are determined using the bootstrap method. This statistical approach involves the random re-sampling of the proton, helium, oxygen, and iron distributions, derived new from the initial RF predic-

tions, with allowance for repeated entries. This process is iteratively conducted n times, facilitating the extraction of a distribution for a specific feature. The statistical uncertainty associated with this feature is then inferred from the standard deviation of the obtained distribution.

The primary source of systematic uncertainties comes from various hadronic interaction models used for the Monte Carlo simulations. These uncertainties are challenging to quantify, however, we have used generated events with EPOS-LHC hadronic-interaction mode. The comparison of different hadron models shows no significant effect on the MVA, which is briefly studied in [77].

7.2.4 Selection of Observables with Random Forest

In the previous section, seventeen distinct observables were explored mainly focused on the differentiation between protons and iron. Subsequently, a variety of these observables were evaluated with the Random Forest algorithm to test their efficacy in discrimination. The comparison of these different observable combinations aims to evaluate observables developed with AugerPrime. Specifically, AugerPrime observables are compared with observables derived with WCDs only.

Along with the observables described in Section 7.1, this study also incorporates the reconstructed energy, denoted as E_{SD}, and the zenith angle, θ_{SD}, into the Random Forest analyses.

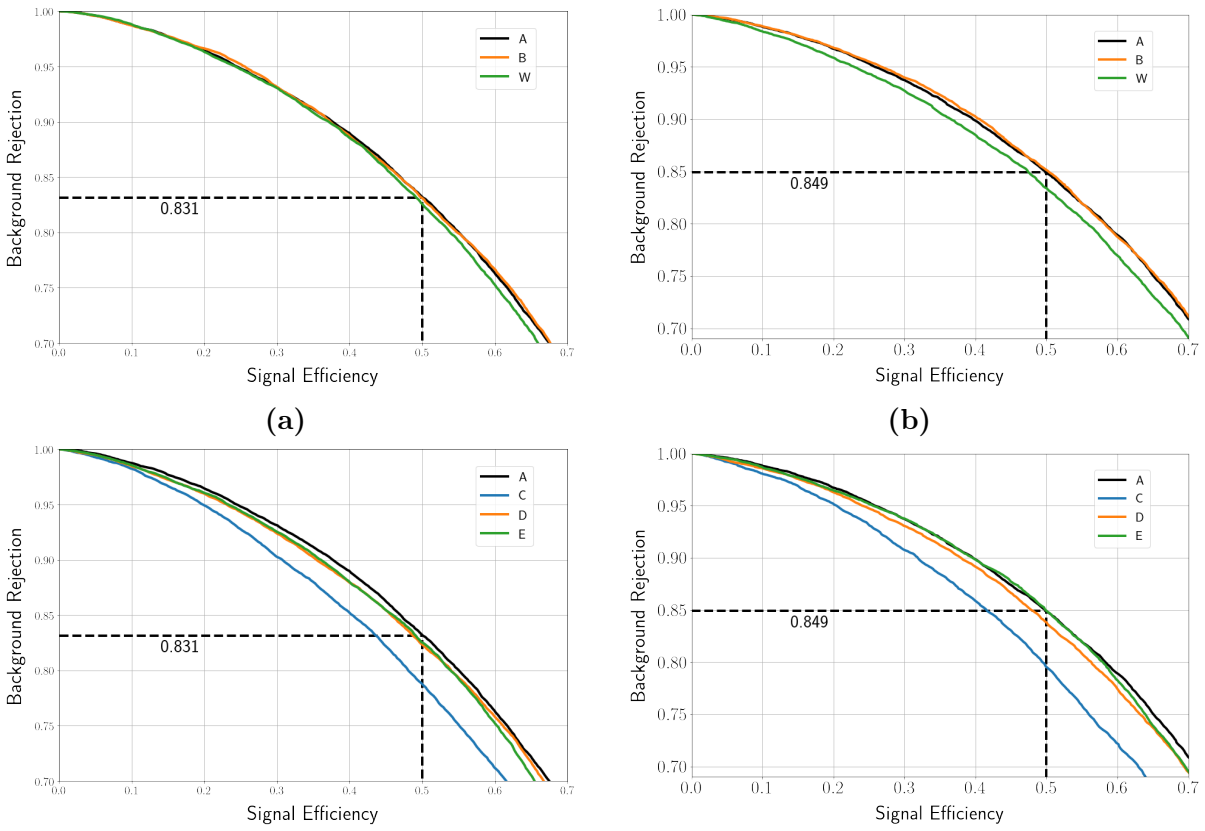
Table 7.7 describes different combinations of variables for proton, helium, oxygen, and iron datasets. For each combination, MF and RMSE are also shown.

Approach	Selected Observables	Merit Factor	RMSE
A	E_{SD}, θ_{SD}, TSR, S_{1000} Ratio, S_b^{WCD}, RoC	1.18±0.003	0.89±0.002
B	E_{SD}, θ_{SD}, TSR, S_{750} Ratio, S_b^{WCD}, RoC	1.17±0.003	0.89±0.002
C	E_{SD}, θ_{SD}, $\sum S_i^{SSD}$, $\sum S_i^{WCD}$, S_{1000}^{SSD}, S_{1000}^{WCD}	0.92±0.008	1.00±0.005
D	E_{SD}, θ_{SD}, $\sum S_i^{SSD}$, $\sum S_i^{WCD}$, S_b^{WCD}, RoC	1.14±0.007	0.91±0.004
E	E_{SD}, θ_{SD}, S_{1000}^{SSD}, S_{1000}^{WCD}, RoC, S_b^{WCD}	1.11±0.006	0.91±0.004
F	E_{SD}, θ_{SD}, TSR, S_{1000}^{SSD}, S_bWCD, RoC	0.99±0.005	0.98±0.004
G	E_{SD}, θ_{SD}, TSR, S_{1000}^{WCD}, S_bWCD, RoC	1.00±0.009	0.96±0.004
H	E_{SD}, θ_{SD}, TSR, $\overline{\text{ToD}}$, S_bWCD, RoC	1.02±0.009	0.96±0.005
I	E_{SD}, θ_{SD}, TSR, S_{1000} Ratio, S_b Ratio, $\overline{\text{ToD}}$	0.94±0.007	0.99±0.003
J	E_{SD}, θ_{SD}, TSR, S_{1000} Ratio, S_b^{SSD}, RoC	1.05±0.007	0.95±0.004
K	E_{SD}, θ_{SD}, TSR, S_{1000} Ratio, S_b Ratio, RoC	1.07±0.008	0.94±0.003
L	E_{SSD}, θ_{SD}, RoC, $\sum S_i^{SSD}$, S_{1000}^{SSD}, $N_{station}^{SSD}$	0.85±0.004	1.04±0.004
M	E_{WCD}, θ_{SD}, RoC, $\sum S_i^{WCD}$, S_{1000}^{WCD}, $N_{station}^{WCD}$	0.94±0.007	0.97±0.004
N	E_{SD}, θ_{SD}, TSR, S_{1000} Ratio, S_bWCD, $N_{station}^{SSD}$	1.15±0.005	0.90±0.004
O	E_{SD}, θ_{SD}, TSR, S_{1000} Ratio, S_bWCD, $N_{station}^{WCD}$	1.13±0.005	0.91±0.003
P	E_{SD}, θ_{SD}, TSR, S_{1000} Ratio, $N_{station}^{WCD}$, $N_{station}^{SSD}$	1.02±0.006	0.96±0.004
Q	E_{SSD}, θ_{SD}, TSR, S_{1000} Ratio, S_b^{WCD}, RoC	1.17±0.007	0.89±0.004
R	E_{WCD}, θ_{SD}, TSR, S_{1000} Ratio, S_b^{WCD}, RoC	1.16±0.007	0.89±0.003
S	TSR, S_{1000} Ratio, S_b^{WCD}, RoC	0.67±0.008	1.12±0.005
T	θ_{SD}, TSR, S_{1000} Ratio, S_b^{WCD}, RoC	1.12±0.005	0.91±0.004
U	E_{SD}, TSR, S_{1000} Ratio, S_b^{WCD}, RoC	0.74±0.006	1.08±0.003
V	E_{SD}, θ_{SD}, S_{1000} Ratio, S_b^{WCD}, RoC	1.10±0.008	0.92±0.002
W	E_{SD}, θ_{SD}, TSR, S_b^{WCD}, RoC	1.07±0.003	0.93±0.004
X	E_{SD}, θ_{SD}, TSR, S_{1000} Ratio, S_b^{WCD}	1.07±0.004	0.93±0.003
Y	E_{SD}, θ_{SD}, TSR, S_{1000} Ratio, RoC	1.09±0.007	0.92±0.003

Table 7.7: Various combinations of observables are summarized in this table. Each method is identified by a capital letter, wherein approach A signifies the ultimately chosen method. The MF and RMSE values for each combination are presented with their standard deviation as uncertainty.

The final selection of observables is established in approach A with a combination of the reconstructed energy, zenith angle, the radius of curvature, the S_b^{WCD}, S(r=1000m) ratio, and the TSR. This specific selection, however, should be noted as not exclusive in terms of efficiency, as alternative combinations have demonstrated comparable MF.

Further in this Section, we discuss ROC curves for the combinations mentioned in Table 7.7. We compare these variable combinations with Approach-A. For the compare ROC curves for proton and iron are presented.

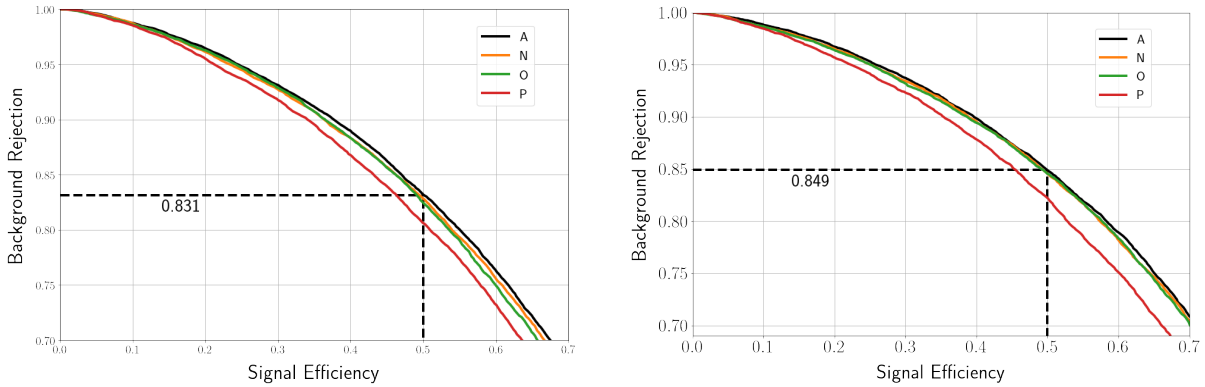


(a) **(b)**

(c) Proton **(d)** Iron

Figure 7.18: The ROC curves of various RF outputs, comparing signals determined at different distances and total signals observables for SSDs and WCDs with the Approach-A. A dashed vertical line represents 50% signal efficiency and a horizontal line indicates a corresponding background rejection Approach-A. The ROC curves for proton are on the left and for iron are on the right.

Figure 7.18a and 7.18b compare the signal ratio at the distances 1000 m and 750 m and in Figure 7.18c and 7.18d compare the signals determined SSDs and WCDs separately at these distances. It is noticeable that Approach-B performs slightly better than Approach-A. However, it is discussed in the previous section that signals at 1000 m are optimized for observatories like the Pierre Auger Observatory. For this reason, signals determined at 1000 m will remain our choice of variable compared to signals at 750 m. Also, in Figure 7.18c and 7.18d approaches developed with total signal measured separately by SSDs and WCDs.



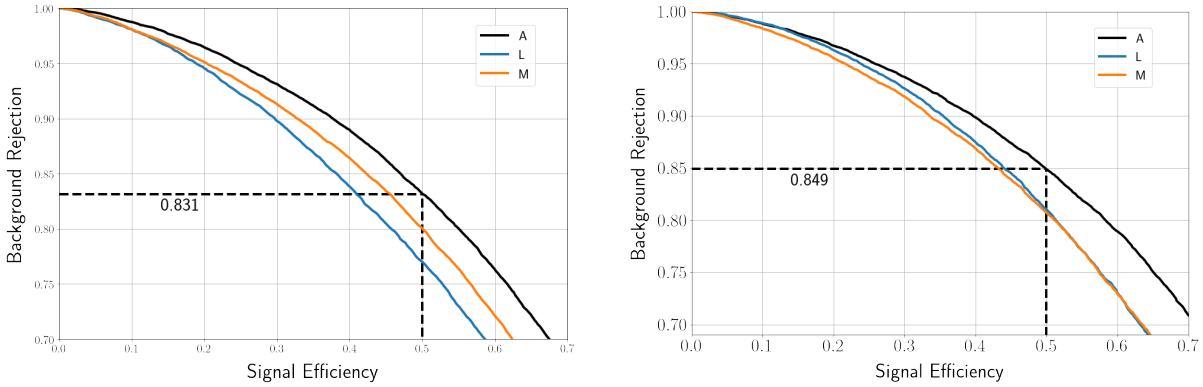
(a) Proton **(b)** Iron

Figure 7.19: The ROC curves of various RF outputs, comparing SSD and WCD number of stations with the Approach-A. A dashed vertical line represents 50% signal efficiency and a horizontal line indicates a corresponding background rejection in Approach-A. ROC curves for proton are on the left and for iron are on the right.

The number of selected SSD and WCD stations are examined in the Approach-N, -O, and -P. The RoC is replaced with the number of stations observable in Approach-N and -O. Also, in Approach-P both numbers of selected stations are included as two separate variables replacing RoC and S_b^{WCD}. This ROC curves for all three approaches are compared with Approach-A in Figure 7.19. The reduction in the accuracy of Approach-P comes from the absence of S_b^{WCD} observable while the other two approaches from similar performance as Approach-A. However, neither of these approaches provides an improvement in the hadron-induced shower discrimination. Moreover, the radius of curvature is less likely to be affected by trigger fluctuations. Conclusively, the radius of curvature is preferred over the number of selected stations. Despite being rejected the number of stations are used as observables in Approach-L and -M.

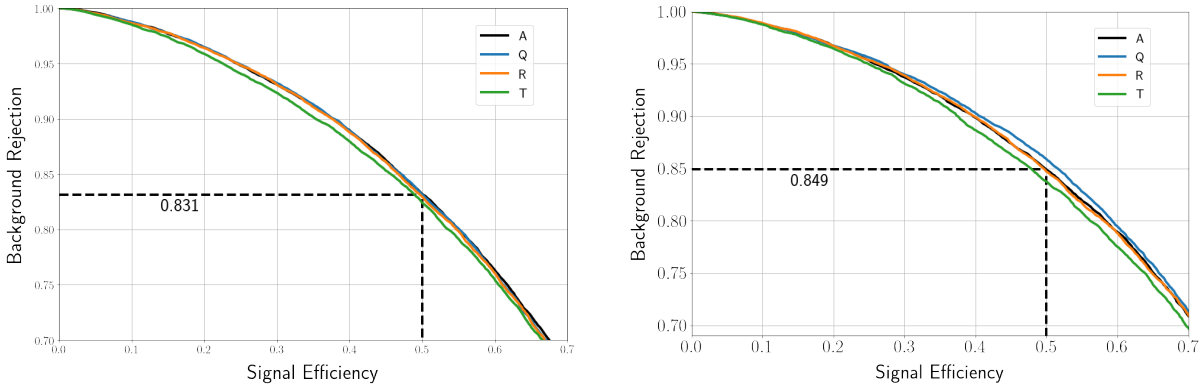
The observables derived separately from SSDs and WCDs are also studied with Approach-L and Approach-M, respectively. These observables are combined with the zenith angle and radius of curvature. The WCD observables provide better discrimination than the SSD observables, this is stated from their respective MF, which is $\sim$0.94 and $\sim$0.85, respectively. These two approaches are compared with Approach-A in Figure 7.20. However, these combinations are discarded as they do not provide better discrimination compared to Approach-A.

The energy of the primary particle derived separately with the SSDs and WCDs has been studied exclusively with Approach-Q and Approach-R. The ROC curves for energy approaches are shown in Figure 7.21. ROC curve for Approach-T which does not contain reconstructed energy observable is also displayed. For the iron discrimination, the Approach-Q with SSD reconstructed energy performs better than Approach-A and Approach-R. However, the MF for proton-iron discrimination is comparatively small. However, this approach is overlooked and Approach-A is still preferred as the E_{SSD} variable is not available in the files containing ob-



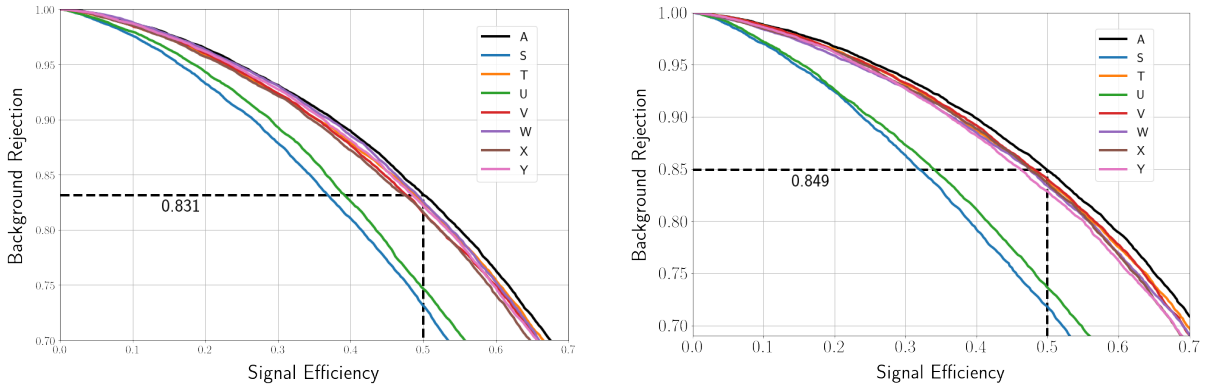
(a) Proton **(b)** Iron

Figure 7.20: The ROC curves of various RF outputs, comparing observables defined separately for SSD and WCD with the Approach-A. A dashed vertical line represents 50% signal efficiency and a horizontal line indicates a corresponding background rejection in Approach-A. ROC curves for proton are on the left and for iron are on the right.



(a) Proton **(b)** Iron

Figure 7.21: The ROC curves of various RF outputs, comparing energy reconstructed separately for SSD and WCD with the Approach-A. A dashed vertical line represents 50% signal efficiency and a horizontal line indicates a corresponding background rejection in Approach-A. ROC curves for proton are on the left and for iron are on the right.



(a) Proton **(b)** Iron

Figure 7.22: The ROC curves of various RF outputs, comparing energy reconstructed separately for SSD and WCD with the Approach-A. A dashed vertical line represents 50% signal efficiency and a horizontal line indicates a corresponding background rejection in Approach-A. ROC curves for proton are on the left and for iron are on the right.

served AugerPrime data which is described in the next chapter. This limitation does not come from the observatory.

Variable combinations based on Approach-A are defined in Approach-S to Approach-Y such that each combination contains one variable less from Approach-A, except Approach-S, where the zenith angle and reconstructed energy are excluded. ROC curve for these seven approaches are shown in Figure 7.22. It can be noticed that Approach-S and -U are the most affected by the removal of the zenith angle, signifying the zenith angle is the most important variable in our analysis.

From the values of MF, RMSE and ROC Approach-A has been selected as the choice to train a random forest regression model. In the next chapter, this model is used to predict the primary particle of the UHECRs measured at the Pierre Auger Observatory. Also, for a deeper understanding combination of variables in Approach-A is explored further in this chapter.

Another method of interpreting the variables can be obtained by determining the Pearson correlation coefficient between the variables, which is shown in Figure 7.23. In Figure, variables from Approach-A are plotted where the correlation for iron is below diagonal and above are for proton. There is a linear correlation between S_b^{WCD} and E_{SD} however, reconstructed energy and zenith angle are used for the characterization of showers and not for discrimination sensitivity.

Considering only the sensitive variables the TSR and S_{1000} Ratio display a strong correlation for proton and iron events and TSR's correlation with S_b^{WCD} is minimum. The radius of curvature and S_b^{WCD} have a negative correlation of approximately 50% for both types of hadronic events. Also, the radius of curvature has nearly 42% correlation with the TSR and 27% correlation with S_{1000} Ratio. S_{1000} Ratio is similarly correlated with S_b^{WCD} at around 27%.

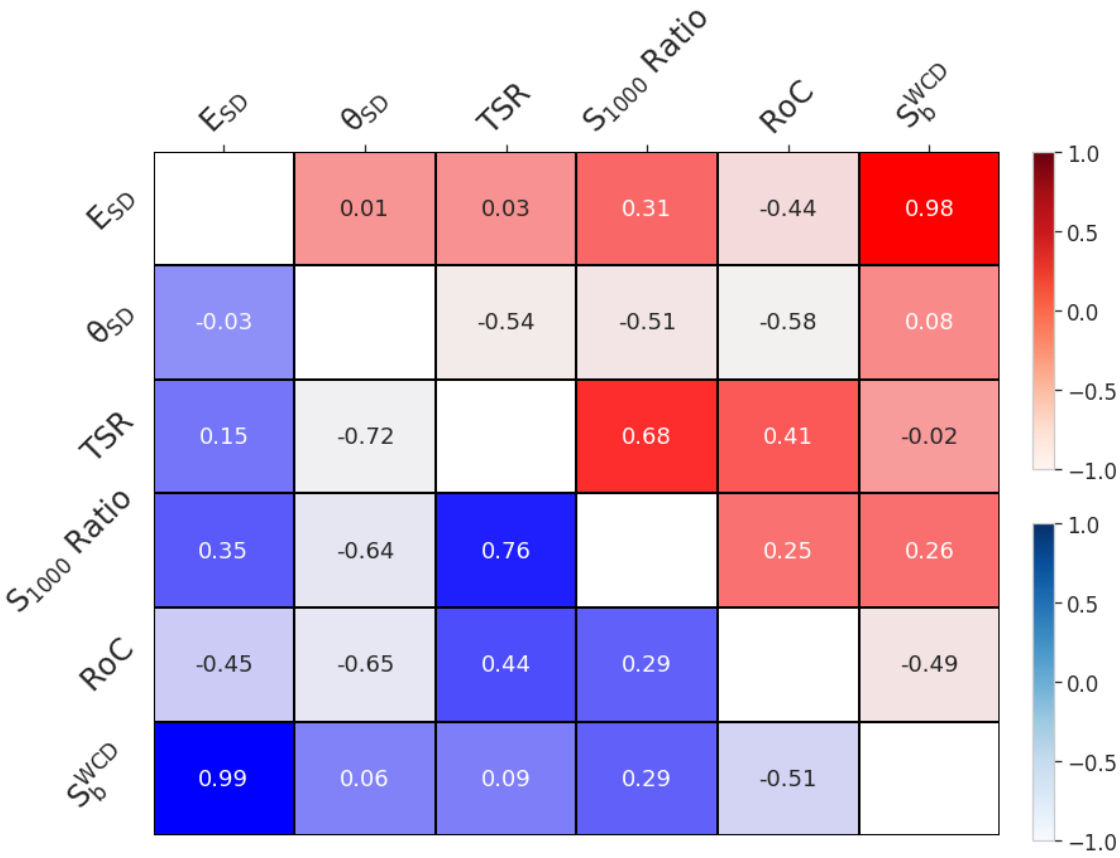


Figure 7.23: Heatmap representing Pearson correlation coefficient for variables in Approach-A. Coefficients for proton-induced events are below empty squares in the diagonal while above the diagonal represent coefficient values for iron-induced events.

7.2.5 Predictions with Random Forest

It was highlighted above that the RF regression model was developed to categorize specific particles by assigning a distinct value to the natural logarithm of their atomic number, ln(A). Specifically, this model classifies proton showers as 0, helium showers as 1.39, oxygen showers as 2.77, and iron showers as 4.02. The predictions of the testing dataset obtained with the trained model based on Approach-A are shown in Figure 7.24.

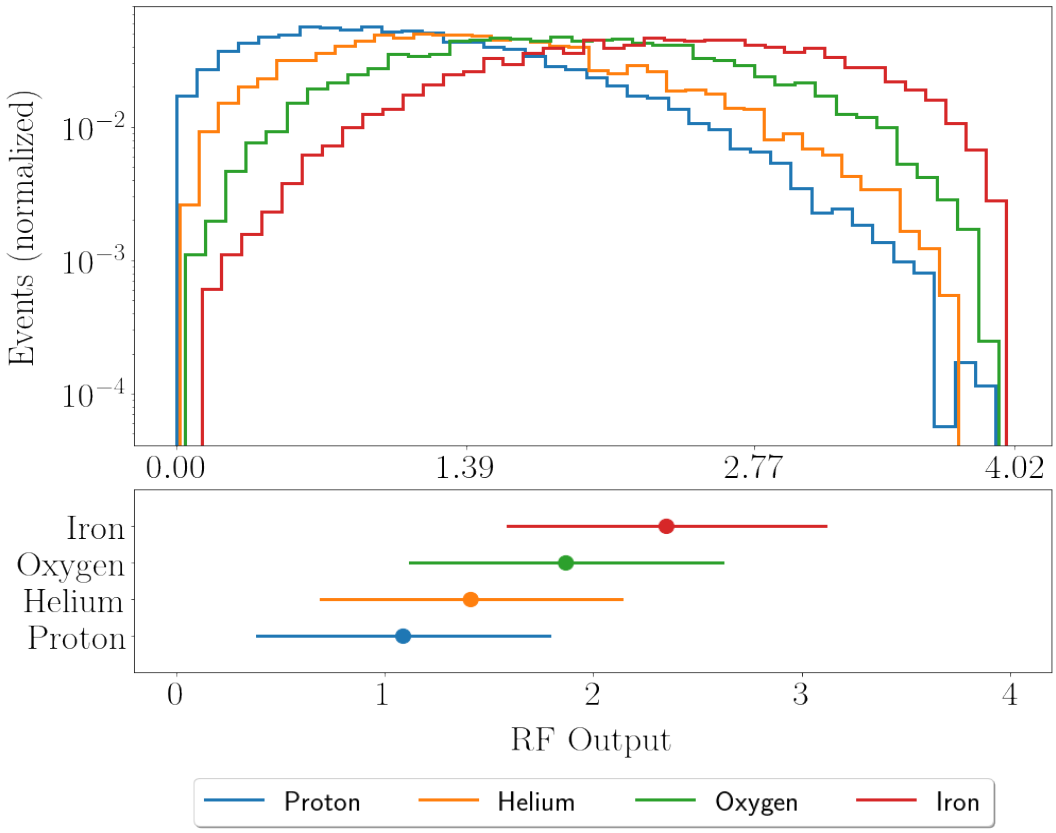


Figure 7.24: Upper histogram shows predicted outcomes of the testing datasets for all hadronic particle showers. In the training phase, protons were labeled as 0, helium as 1.39, oxygen as 2.77, and iron as 4.02. The mean and the standard deviation for the distribution of each particle type are shown in the below plot.

In the figure, the distribution of each particle overlaps the distribution of other particles and this overlapping is more pronounced for the helium and oxygen showers. Some distinctions can be made between proton and iron showers near the true value for both particles. This can justified by calculating MF as shown in Table 7.8. The mean of predictions for each particle is highly deviated from their true value signifying a large amount of misidentification. However, this is still helpful in understanding the trend of the observed showers. This is the aim of the next chapter. The poor prediction of hadronic showers stems from their complex nature and high similarity between them.

In Figure 7.25 RF output is compared with each of the six variables and the

Particle	Merit Factor for prediction
Proton	0.677
Helium	0.131
Oxygen	0.312
Iron	0.828

Table 7.8: Merit factor of the predicted data for each particle type showers where the MF is calculated by differentiating each particle with the other three particles combined.

Particle	Mean with σ $E_{SD} > 50\ EeV$	Mean with σ $E_{SD} = All$
Proton	1.01±0.721	1.08±0.698
Helium	1.36±0.740	1.41±0.720
Oxygen	1.92±0.777	1.86±0.749
Iron	2.43±0.780	2.35±0.759

Table 7.9: Mean with standard deviation for RF output comparison of Approach-A with particles of all reconstructed energy and $E_{SD} > 50\ EeV$.

ambiguity in the prediction can be seen here. However, with the median value of distributions of each particle, slight trends are visible with most of the variables except the zenith angle which is spread evenly for all particles. With the higher values of S_b^{WCD} and E_{SD} the iron and oxygen showers are predicted with less ambiguity. The similar trend in both distributions also confirms their correlation mentioned in 7.23. A similarly inverse trend is visible for TSR and S_{1000} Ratio.

Influence of reconstructed energy on Random Forest performance

The small effect of reconstructed energy on the predictions is already discussed above. In this section, we explore this trend by training models in different energy ranges. In Figure 7.26, ROC curves are plotted for training and testing sets for three energy ranges, 3-10 EeV, 10-50 EeV, and all particles with energy above 50 EeV. For comparison all energy ROC curve is also displayed. From the figure, it can be concluded that the predictions become better for proton and iron particles with the increase in energy.

The prediction with the energy above 50 EeV performs slightly better than Approach-A and provides some isolation near the true value of proton and iron. This is visualized in Figure 7.27 along with the RF output of all particles. The mean and the standard deviation of this prediction are compared with Approach-A in Table 7.9. One can also notice that predictions with different energy segment shows no visible difference between proton and helium.

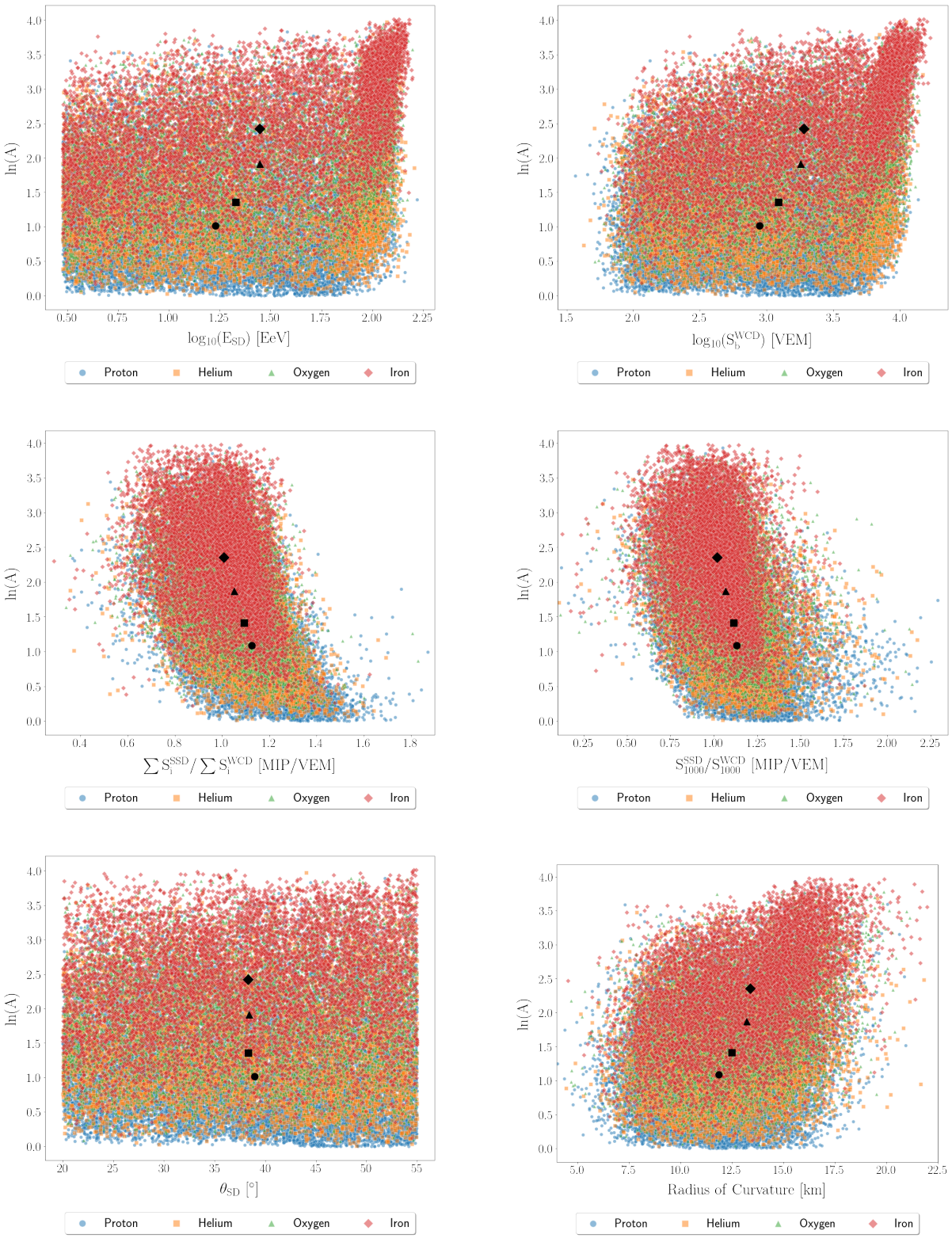
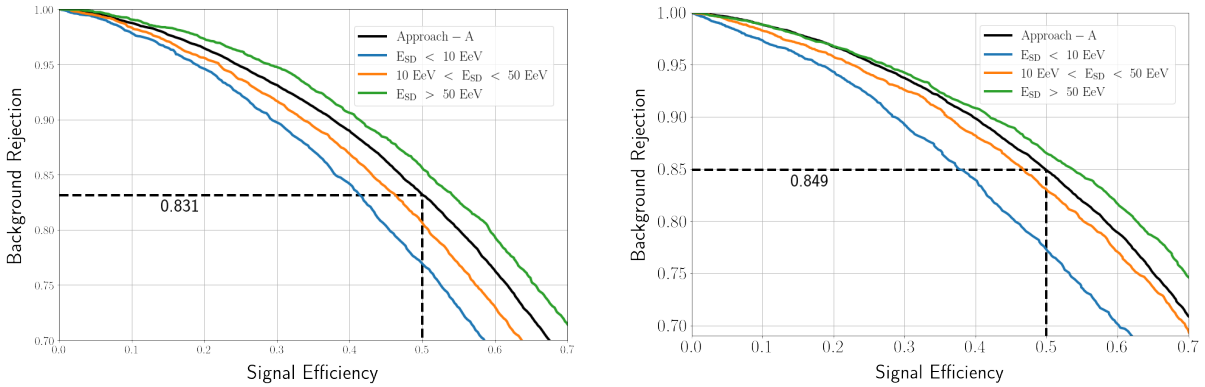


Figure 7.25: Correlation distributions are presented to illustrate the relationship between the predictions of the RF output and the value of each observable for the respective event. The median point for both the x and y axes is indicated by a black marker. Furthermore, the markers used to represent different particles are distinguished by their shapes. In this context, circles, squares, triangles, and rhombuses correspond to protons, helium, oxygen, and iron, respectively.



(a) Proton **(b)** Iron

Figure 7.26: The ROC curves of RF outputs in different energy ranges, comparing with the Approach-A. A dashed vertical line represents 50% signal efficiency and a horizontal line indicates a corresponding background rejection in Approach-A. ROC curves for proton are on the left and for iron are on the right.

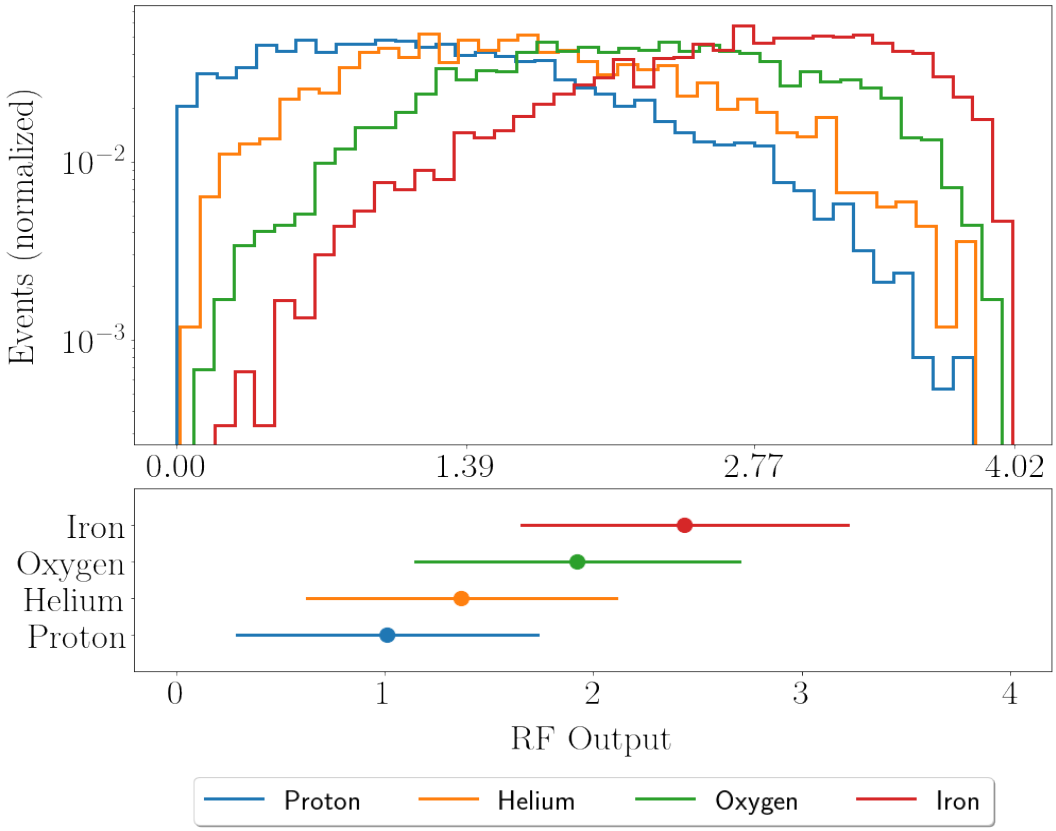
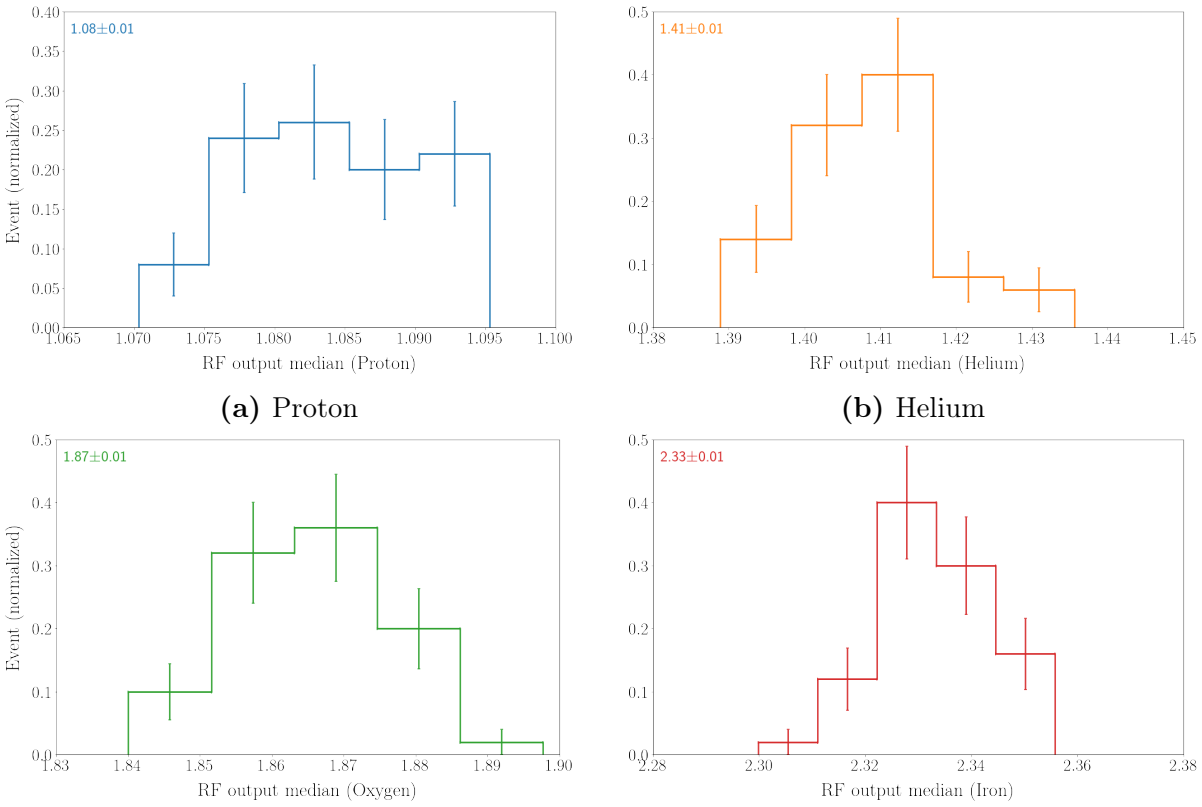


Figure 7.27: Upper distribution at $E_{SD} > 50$/EeV shows predicted outcomes of the testing datasets for all hadronic particle showers. In the training phase, protons were labeled as 0, helium as 1.39, oxygen as 2.77, and iron as 4.02. The mean and the standard deviation for the distribution of each particle type are shown in the below plot.



(a) Proton **(b)** Helium

(c) Oxygen **(d)** Iron

Figure 7.28: For all particle types, the median of several outputs obtained with the bootstrap method is presented. With the median value, the author presents statistical uncertainty associated with the random forest regressor. The mean of all median values for each particle type is presented in the top-left corner of each plot.

Statistical uncertainty

Similarly to the rest of this chapter, the bootstrap method is deployed to assess the uncertainty. Using the bootstrap method multiple randomly selected training and testing datasets are obtained and the median value of the output from all testing sets is shown in Figure 7.28 for all particles separately as statistical uncertainty.

7.3 Chapter Remarks and Conclusion

In our analysis, the EPOS-LHC hadron interaction model is exclusively selected. However, there are various models available that can be deployed for the simulation of EASs. In particular, the results of QGSJet-II-04 and SIBYLL2.3 are also compared in [77]. In that study, the author found no major differences between these models for photon and proton discrimination, and similar results are assumed for this work.

The incorporation of scintillators into the Surface Detector array at the Pierre Auger Observatory has yielded novel insights into the nature of air showers. These scintillators capture data at a consistent relative distance from the shower core and the same developmental stage of the shower as the WCDs. This alignment not only facilitates the enhancement of traditional WCD observations using the Scintillator Surface Detectors (SSDs) but also presents the opportunity to combine the data from both detector types, leading to better understandings of air shower phenomena.

The development of AugerPrime observables is dependent on a combination of factors: the distances between stations, the signals they received, and the predictions derived from the LDFs. These developed observables do not provide good discrimination for hadron-induced showers. However, these observables surpass the capabilities of variables confined to a single detector type. An exception is observed in the case of the observable S_b. In this instance, no discernible progress was achieved in comparison to the traditional WCD determination methods.

In the tested SD observables, the ratio of the total signal from the SSD to that from the WCD exhibited the highest sensitivity to hadron-induced showers. This observable achieved a distinction factor of $\sim$0.5 in differentiating proton from iron events. Additionally, the TSR demonstrated considerable stability at varying energy levels, with only minor fluctuations at larger zenith angles. The S_{1000}, calculated from the ratio of two LDFs also provides effective separation capabilities. The enhanced sensitivity of the two AugerPrime observables can be attributed to their indirect association with the ratio of electron to muon count. The scintillators exhibit greater responsiveness to the electromagnetic component of the shower, whereas the WCD signals are predominantly influenced by the muonic component. Consequently, a ratio of these disparate signals correlates to the electron-to-muon number ratio. Proton-induced showers, characterized by a reduced muon count, yield the highest ratios in these events. Conversely, iron-induced showers, with the highest muon count for a given primary energy, represent the lower boundary for these ratios.

In the study, a MVA employing the Random Forest method was designed to distinguish between proton, helium, oxygen, and iron events. This analysis involved testing various combinations of observables, ultimately determining a combination of six key input variables for optimal discrimination. These variables include Total Signal Ratio (TSR), S_{1000} ratio, S_b^{WCD}, the radius of curvature of showers, and the SD reconstructed energy and zenith angle. The effectiveness of

the MVA is highlighted by its capacity to attain approximately 34% background rejection at a 50% signal efficiency rate for helium events, the lowest rate among all tested events. Conversely, it achieves around 85% background rejection at the same efficiency level for iron events, representing the highest rate observed. Furthermore, the results from the Random Forest model demonstrate a Merit Factor of 0.82 for iron events and 0.13 for helium events, indicating the maximum and minimum Merit Factors, respectively.

CHAPTER 8

THE AUGERPRIME ANALYSIS

The mass discrimination of the UHECR simulated showers with AugerPrime has been conducted extensively. Total Signal Ratio and S_{1000} ratio are the two AugerPrime observables were developed with the Water Cherenkov Detector and Surface Scintillator Detector. These two observables along with E_{SD}, θ_{SD}, S_b^{WCD} and radius of curvature were used to train and test Random Forest regression model.

In the present study, the developed machine learning model has been employed to analyze field data, prioritizing the AugerPrime observables. This process involves a detailed comparison between simulations and the measured data. The events observed at the observatory have been evaluated using a Random Forest model for mass discriminating of hadron-induced events.

The data utilized in this analysis were obtained between March 2019 and December 2021 using the Pre-Production Array (PPA), which refers to the initial phase of SSD installation. As a part of PPA the newly installed scintillators were integrated with the existing electronics (UB). The PPA was specifically designed to assess the performance of the SSDs on a larger scale, thereby providing valuable insights into their operational efficiency.

The deployment of the scintillators led to the disconnection of one PMT in a WCD from the pre-existing electronics infrastructure (UB). This disconnection resulted in discrepancies at low signal levels between collected data and simulations. In standard Offline simulations, each WCD is typically equipped with three PMTs, however, this is not the configuration of the observatory setup. These discrepancies, as well as the impact of aging effects on WCDs contributing to low signal disagreements, are comprehensively discussed in [77]. To overcome these issues, a revised signal threshold of 5 VEM was implemented in both the simulations and the PPA data. Applying this correction and the selection cuts, 636 events were suitable for the application ofMVA.

In this chapter, the simulation data presented differs from that utilized in Chapter 7. Section 8.1 details both the simulated data and the corresponding

measured data. Subsequently, there is a brief description of the reconstruction of events from the PPA. The completion of this analysis involves the evaluation of these events using RF, which facilitates the mass composition of observed UHECRs.

8.1 Simulated and Pre-Production Array data sets

In this section, a different approach has been employed than the previous chapters, as it encompasses an examination of both simulations and field data. The data sets utilized in this analysis are comprehensively summarized herein. As mentioned above the simulated data sets from previous chapters are not incorporated into this current analysis.

The construction of the datasets, derived from events observed by the SSDs and WCDs of the Pierre Auger Observatory, is detailed in Table 8.1. This dataset incorporated a 5 VEM signal threshold at the WCDs to resolve discrepancies observed in low signal measurements.

Data sets	**PPA**
Offline Sequence for Reconstruction	SdSSDData Reconstruction
Offline Version	Trunk rev 33552
Detectors	WCD+SSD
ToTd and MoPS?	No
Data Period	22^{nd} March 2019 to 31^{st} December 2021
Collected Events (within PPA)	13058
Selected Events	636
Signal Threshold of 5 VEM	Yes

Table 8.1: Description of the data sets built from events measured with the Pierre Auger Observatory.

The data sets simulated are described in Table 8.2. These data sets mirror the configuration of the PPA, encompassing solely the PPA stations and their immediate neighboring stations. Consequently, the simulated data sets previously discussed in Chapter 7 were not applicable in this chapter. However, these new data sets were processed utilizing the identical Offline sequence.

In the present study, the datasets, both simulated and observed, were sourced from the author of [77]. Consequently, the information presented in Tables 8.1

and 8.2 has been adapted from these [77].

Data sets	A1	A2	A3	A4
Primary	Proton	Helium	Oxygen	Iron
Hadronic Interaction Model	EPOS-LHC			
Energy log [eV]	18.0-20.0			
θ [°]	0-65			
ϕ [rad]	0-2π			
CORSIKA Library	Napoli			
CORSIKA Files	3000 per bin	1500 per bin		
Offline Sequence	Modified SdSimulation Reconstruction Upgrade			
Offline version	Trunk rev 33552			
Detectors	WCD+SSD			
PMTs per WCD	3			
Stations list	PPA Adapted			
Electronics	UB			
ToTd and MoPS?	No			
Generated Events (Within PPA)	32053	16304	16047	16454
Selected Events	17254	8665	8790	8704
Signal Threshold of 5 VEM	Yes			

Table 8.2: Tabular description of the simulated data sets utilized in this chapter. The dataset structure aligns with the formats exemplified in the sections of Chapter 7. The bins are the CORSIKA files refer to the number of files per 0.5 gap in log[E].

8.2 Pre-Production Array

"In Figure 8.1, the PPA stations are depicted in blue, distinguishing them from the surrounding non-PPA stations, which are shown in grey and lack scintillators. The analysis of improvements brought by AugerPrime requires that the core of the analyzed events must be within the PPA, ensuring accurate measurement by the SSDs and allowing the determination of AugerPrime observables. As described in Chapter 7, a threshold of 1 MIP is set for the SSDs. This criterion often results in a higher number of WCDs than SSDs for certain events. Consequently, the surrounding stations, despite not being part of the PPA, contribute additional

data, enhancing the precision of shower reconstruction.

A specific selection criterion, based on a quality cut, was applied to events with cores within the PPA, in alignment with the simulation procedures outlined in section 6.2. This selection process also incorporated a 6T5 trigger condition, which necessitates the functionality of the hexagon of stations surrounding the most active station, irrespective of their trigger status.

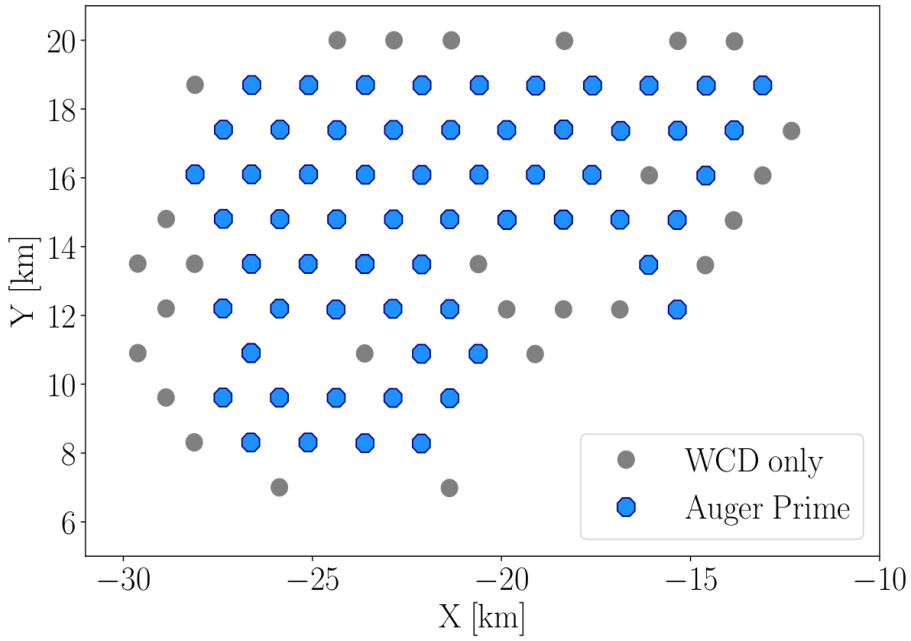


Figure 8.1: The coordinates of the AugerPrime stations in the Pre-Production Array, where each WCD is equipped with a scintillator on top. The adjacent stations, which do not possess SSDs, are also incorporated into the study [77].

Operational since March 2019, the PPA represents the initial deployment of scintillators at the Auger site. The data recorded from March 22 2019 to December 31 2021 by these stations was also examined in a prior doctoral thesis [77, 103]. This chapter focuses on the analysis of these events. As established in Chapters 7, the Pierre Auger Observatory employs the Offline for simulating EASs. Similarly, the Offline package is utilized for processing field data.

The data reconstruction from the PPA was conducted using the `SdSSDDataReconstruction` sequence. This sequence processes events from the Surface Detector, selectively including those events that triggered at least one scintillator-equipped station.

8.3 Event Selection Pre-production Array and Simulation Data

Work in this chapter also applies the same quality controls established for the simulations in Section 6.2 to the PPA events. These cuts are implemented after

Applied Cut	Events	Percentage[%]
Shower Core in PPA	13058	100
Shower reconstruction	13005	99.59
$N_{WCD} > 2$ **and** $N_{SSD} > 2$	9205	70.49
$\theta_{SD} < 55$	8175	62.61
Fitted SSD LDF	8035	61.53
$E_{SD} > 3EeV$	934	7.15
$\theta_{SD} > 20$	786	6.02
6T5	636	4.87

Table 8.3: Selection criteria applied to the data gathered at the PPA from March 2019 to December 2021 [77].

the 5 VEM threshold on WCD data. To further ensure accuracy, a 6T5 trigger was employed to minimize potential errors arising from dysfunctional stations. The sequence and impact of these quality controls are detailed in Table 8.4.

The spatial distribution of the event cores is depicted in Figure 8.25, contrasting all events (represented in blue) with the selected ones (indicated by magenta stars) and their proximity to the SD stations. The analysis only includes events close to AugerPrime stations, allowing for accurate measurements with the scintillators.

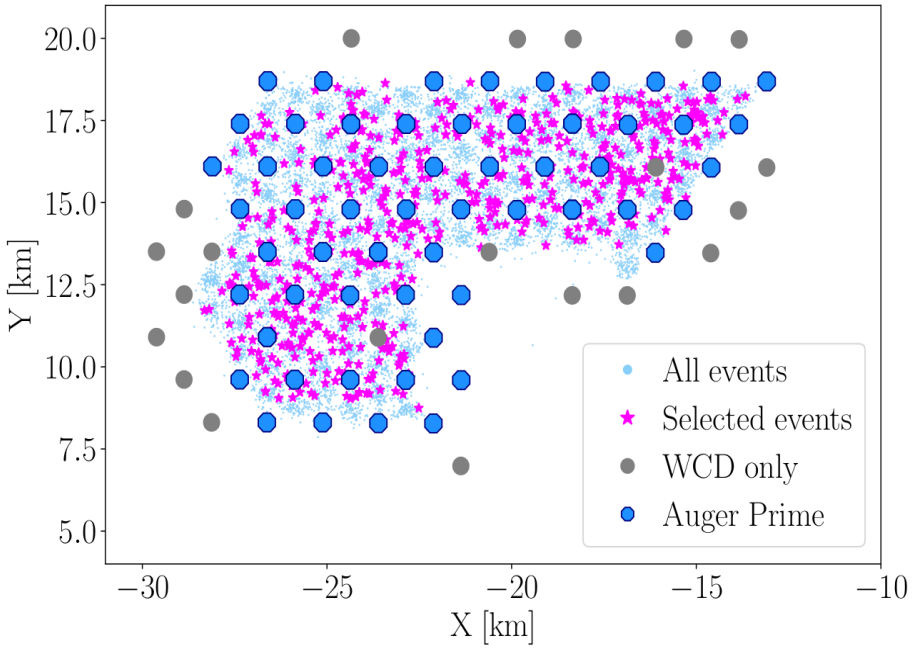


Figure 8.2: The events which took place with in the PPA are presented. Blue and grey circles represent PPA and non-PPA SDs. The core location of events that passed the selection cuts are represented in magenta stars. The location of all shower cores that triggered the SD and are situated within the PPA are marked in light blue [77].

Figure 8.26 presents a monthly tally of events, both pre-and post-quality con-

trol measures. To facilitate comparison, the event counts before the application of quality controls are scaled down by a factor of 10. Notable reductions in event counts are observed in November 2020 and December 2021, attributed to the unavailability of files from the Auger servers during the period of this analysis.

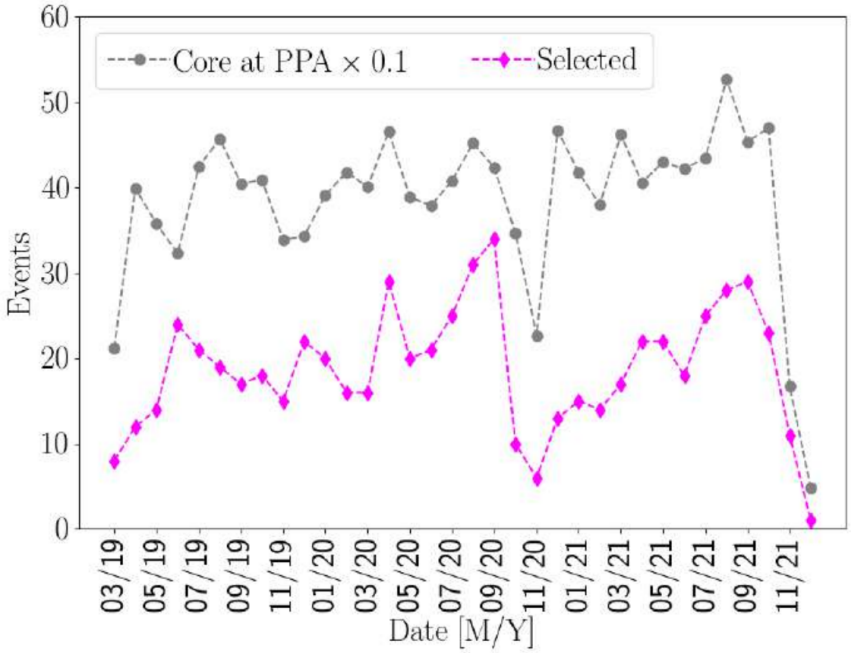


Figure 8.3: Between March 2019 and December 2021, data was collected monthly from the PPA. This dataset includes two distinct categories of event counts. The counts depicted in grey represent the number of events within the PPA which are scaled down by a factor of ten. The magenta counts indicate the number of events that successfully passed the quality checks and were subsequently analyzed in this study. The decline in event numbers was observed in November 2020 and December 2021, which occurred due to the absence of certain files.[77].

8.4 Comparison of observables between observed and simulated showers

The 25 observables set developed in Chapter 7 out of which, Approach-A has been preferred due to its maximum merit factor, allowing the best discriminations among all combinations of observables. Approach-A is combination of TSR, S_{1000} Ratio, Radius of Curvature and S_b^{WCD}. These observables are used for the Random Forest training in this chapter. The distributions of all six observables are shown in Figure 8.4.

The energy distribution in the figure is highly skewed, containing mostly low-energy events between 3-5 EeV. Since we are attempting to discriminate the highest-energy primary particle, these low-energy showers bring a deviation in our analysis.

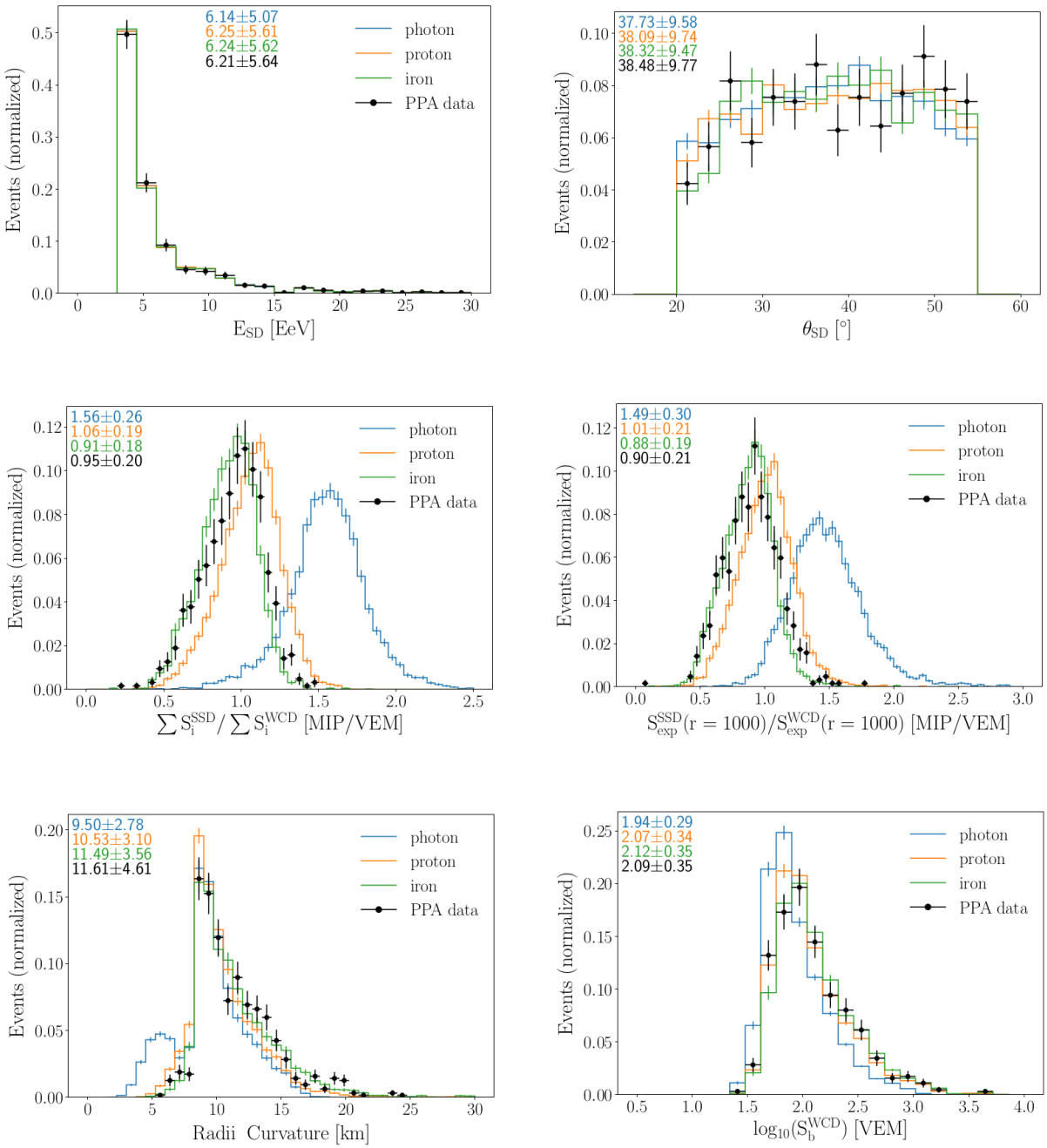


Figure 8.4: The distributions of the reconstructed energy (top-left), zenith angle (top-right),TSR (middle-left), S_{1000} ratio (middle-right), the radius of curvature (bottom left) and s_b^{WCD} (bottom right), determined from the WCDs, are examined after implementing a 5 VEM signal threshold. The data, obtained from the PPA, is also presented with simulated events triggered by photon, proton, and iron particles. The photon showers are also visible as these figures are adopted from [77] To ensure comparability, these simulated events have been weighted to align with the energy distribution observed in the field data. The mean and standard deviation for each distribution are indicated in the top left corner of the respective color representation.

In Figure 8.5, the comparison of four principal observables across various simulations and PPA data is presented. This figure includes density plots for each observable, illustrating their interrelationships. The correlation plots within this figure suggest that the PPA data predominantly aligns with the outcomes of the hadronic simulations.

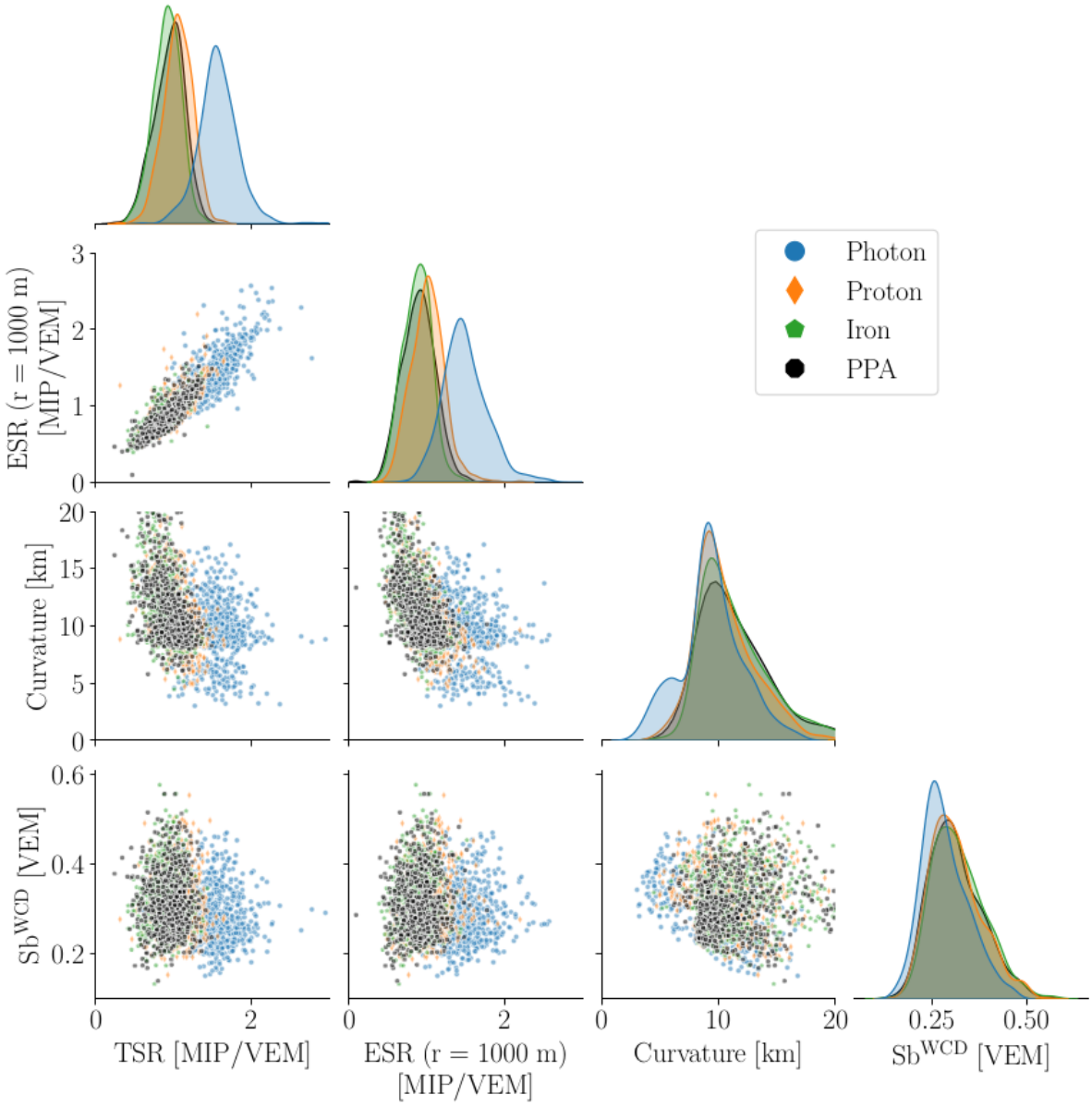


Figure 8.5: The correlation and density plots, also known as pair plots for the four principal observables are presented, examining both PPA data and simulated showers induced by photon, proton, and iron. This visualization aids in understanding the relationships and distributions within the data set, providing insight into the MVA process. The photon data is visible in the plot as it is sourced from [77].

8.5 AugerPrime Pre-Production Array Analysis with Random Forest

In this Section, the PPA events are analyzed with Random Forest to understand their mass composition. As described in the previous chapter, the RF model is trained using the simulated dataset, which is employed to categorize the PPA events by predicting a number for each event between 0.0 to 4.02. These numbers represent each particle type and their value is chosen as the natural log of atomic mass of the primary particle. The values assigned to proton, helium, oxygen, and iron events are 0.0, 1.39, 2.77, and 4.02, respectively.

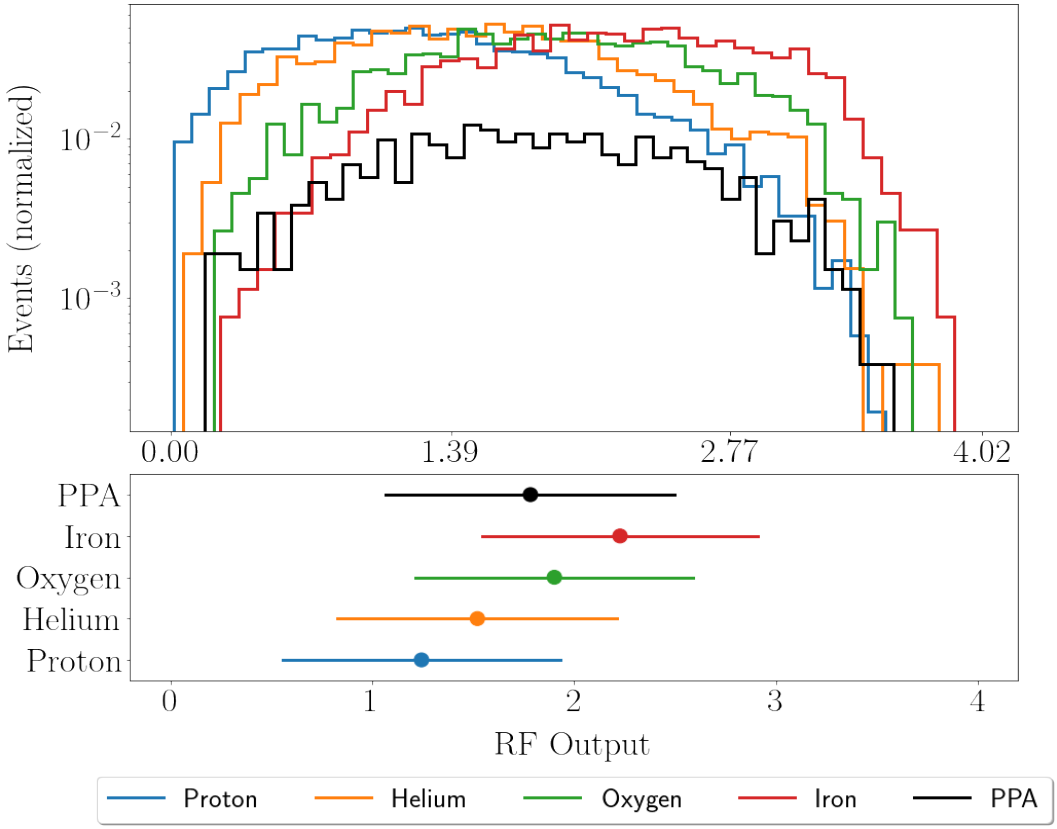


Figure 8.6: The distribution shows predicted outcomes of the testing subsets for all hadronic particle showers along with the prediction of the primary particle observed with PPA (black). In the training phase, protons were labeled as 0, helium as 1.39, oxygen as 2.77, and iron as 4.02. The mean and the standard deviation for the distribution of each particle type are shown in the bottom plot.

In Figure 8.6 the predicted value of the testing subset and PPA dataset is presented. As expected, there is no strong discrimination between different primaries. Also, there is no skewness observed in the prediction of PPA data which can lead to any conclusion. However, few events remain outside the proton or the iron distributions, such events can be categorized as *not-proton* events and *not-iron* events. In total, there are six such events, out of which 5 are Not-Iron events and 1 is Not-Proton event. These events are summarized in Table 8.4. It is noticeable

Auger Event Number	Category	Predicted Value	Reconstructed Energy [EeV]
191961137300	Not-Iron	0.171275	6.123
202770150100	Not-Iron	0.196287	5.057
190867342300	Not-Iron	0.23125	4.757
191557020900	Not-Iron	0.194225	3.751
201333467000	Not-Iron	0.227062	13.325
193490355701	Not-Proton	3.58371	21.275

Table 8.4: Tabular presentation of six events that are categorized as either "Not-Proton" or "Not-Iron".

that all the categorized low-energy events are *Not-Iron* events.

The correlation plots between the RF output and each of the six input variables are shown in Figure 8.7. Since most of the observed events have low energy the value of reconstructed energy and the observable S_b for PPA is shifted towards lower values.

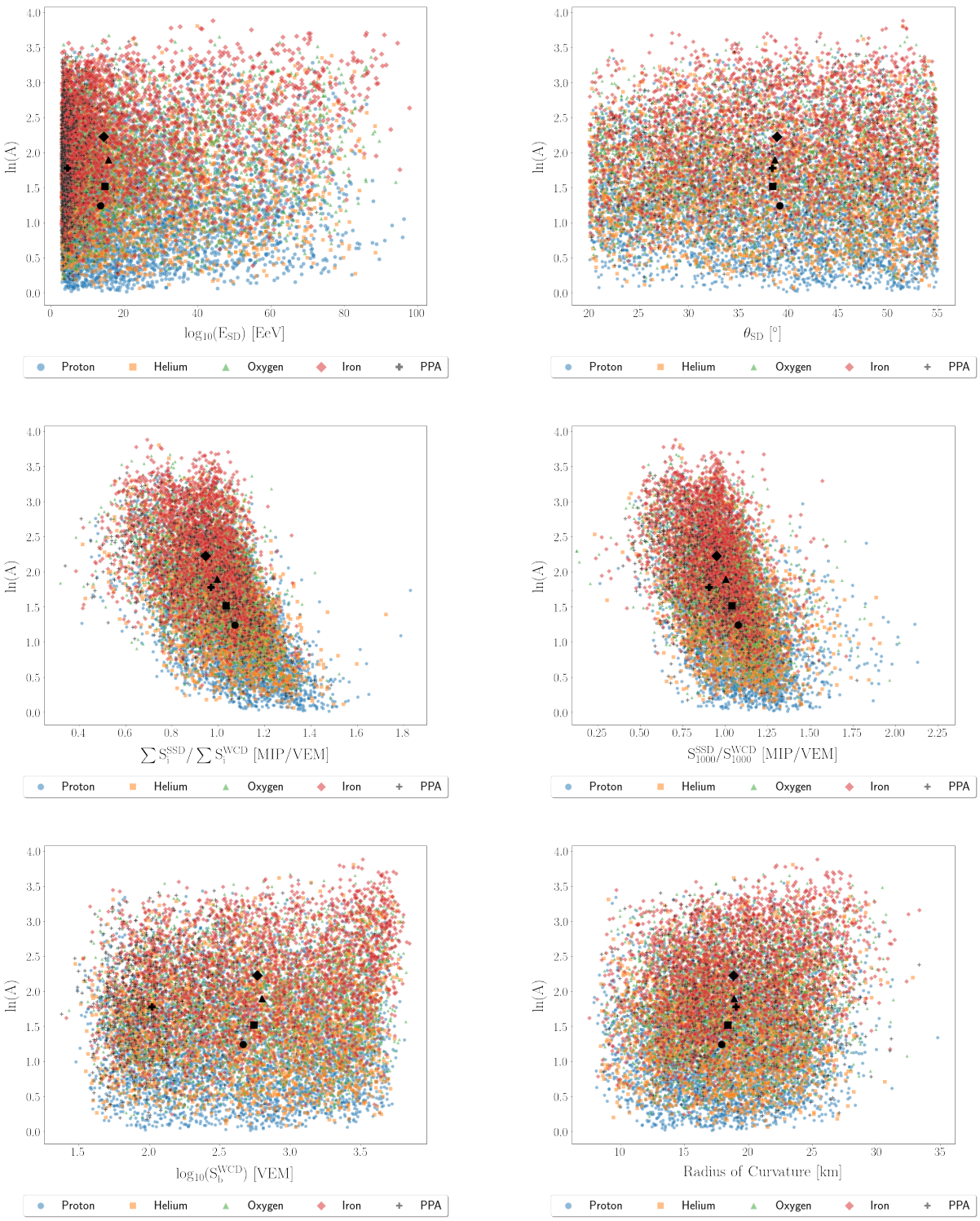


Figure 8.7: Correlation distributions are presented to illustrate the relationship between the predictions of the RF output and the value of each observable for the respective event. The predicted values of all hadronic particles and PPA are presented. The median point for both the x and y axes is indicated by a black marker. The markers used to represent different particles are distinguished by their shapes. In this context, plus, circles, squares, triangles, and rhombuses correspond to PPA, protons, helium, oxygen, and iron, respectively.

8.6 Result Comparison

Telescope Array (TA) in Utah, USA is functionally similar to Pierre Auger Observatory, spanning the area of 700 km^2. The results using the hybrid data TA are presented in [104]. In this analysis, authors have used the MVA with different observables. The observables such as area-over-peak of the SD waveforms at 1200 m, shower front curvature parameter, and signal asymmetry in SD upper and lower layers are analyzed with Boosted Decision Trees (BDT) classifier.

The results from events at TA are presented in Figure 8.8 and the results obtained in this thesis work are shown in Figure 8.9. Both the presented results are in good agreement but no strong mass discrimination is achieved with these works. The results obtained with SDs of both observatories remain uncertain and show no large trend but the hybrid events at TA hint towards lower mass composition of Ultra-High Energy Cosmic Rays. The better results of hybrid events come from the simultaneous observation of lateral and longitudinal development of the Extensive Air Shower.

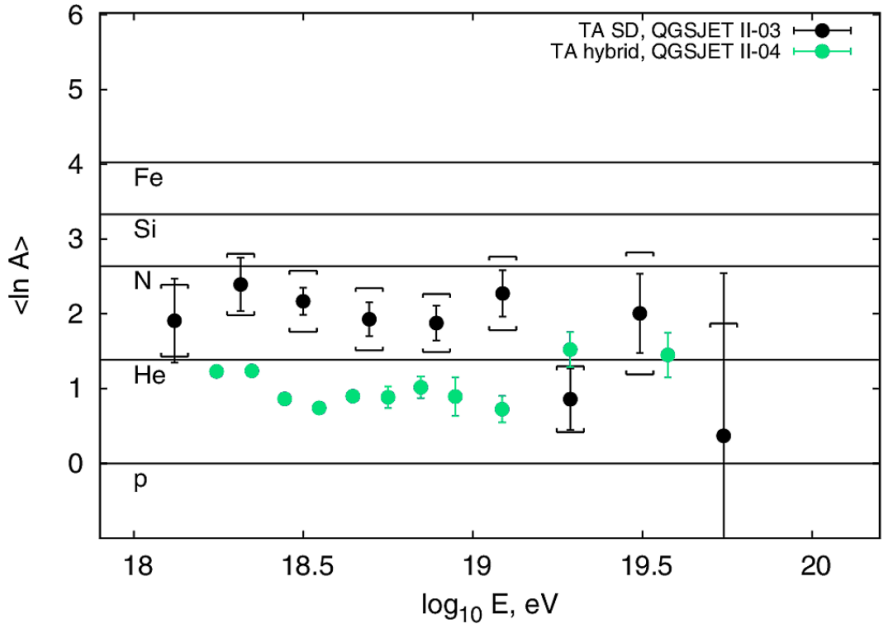


Figure 8.8: The predicted primary mass of events at the Telescope Array, Utah. These events are observed with Surface Detectors are shown in black and events observed with Surface Detectors and Fluorescence Detectors are presented in green. The atomic mass of different hadronic particles is represented with horizontal lines [104].

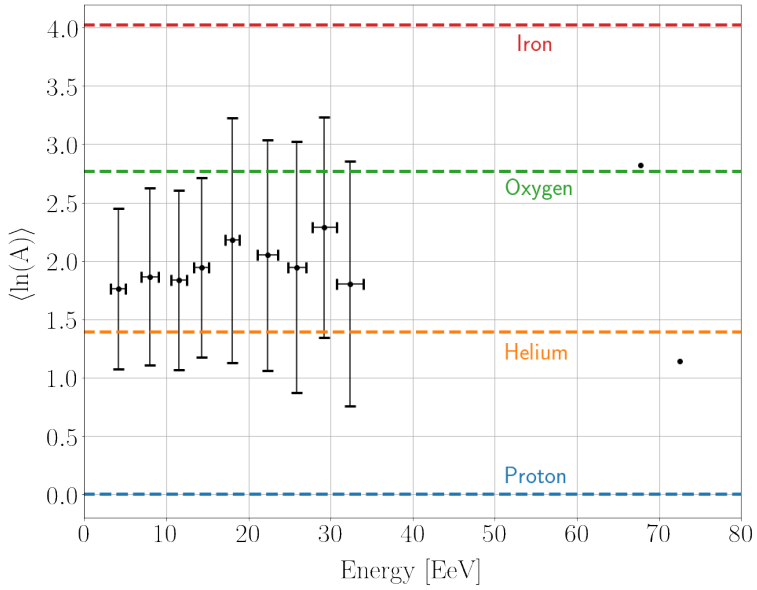


Figure 8.9: The predicted mass composition of events at Pierre Auger Observatory obtained in this thesis work. The predicted value is presented as a function of reconstructed energy. The error bars are the standard deviation of the values in each bin. The defined value of different hardon particles is presented as a horizontal line for comparison. The color code used for these particles is uniform throughout the thesis.

CHAPTER 9

CONCLUSION AND FUTURE PROSPECTS

Due to the complex and similar nature of the hadron-induced events, their clear mass identification is a difficult task. This is an issue addressed at the PAO and its upgrade provides better results. This is realized by comparing the Merit Factors of AugerPrime observables with WCD only observables.

Combining the improved ability of the observatory with the highly reliable Random Forest Regression method provides one of the best possible outcomes, despite the large uncertainty of the results.

The results of the composition sensitivity of the AugerPrime upgrade of the Pierre Auger The observatory obtained in this thesis work is in good agreement with the results obtained at the Telescope Array.

However, the extraction of analysis-ready observable is a complex process as it is sensitive to mistakes. The misinterpretation of data can lead to confusion and mistagging inside the model training. At the same time, better understanding during the selection process can enhance the result. One such improvement that can be made in the future is by recovering the signals of the saturated stations. This could improve the results as almost all the saturated stations are the hottest stations containing the most showers. Efficient reconstruction of signals at saturated stations can improve the LDF reconstruction and better electron to muon ratio. One such method to recover the signal is by fitting the signal to Moyal Distribution [105].

The Pre-Production Array is a preliminary phase of installation of Surface Scintillator Detector. By the time author completed this thesis work more than 1200 SSDs were installed. With this in full effect, more data will be obtained which can further improve the analysis. As it can be seen in our results 8.9, the statistical uncertainties are large leading to difficulty in the conclusion of results. Also, in future analysis large training datasets can be developed for better machine learning models.

The Random Forest Regression method is a highly powerful and robust machine learning method. It is very resistant to overfitting and missing data which

makes Random Forest a reliable method even without having a complete understanding of the data. However, there are other machine learning models available which might provide better discrimination. One such method is the application of a Deep Neural Network. An advantage of DNN is it can handle multidimensional data. This can allow direct use of signals at each detector instead of analyzing TSR, leading to exploration of possible underlying patterns at the detector level.

ACKNOWLEDGEMENT

As I culminate this significant phase of my academic journey, I am filled with immense gratitude towards many who have contributed to the completion of my master's thesis. Their support, guidance, and encouragement have been pivotal in shaping not only this thesis but also my personal and professional growth.

Foremost, I extend my deepest appreciation to my supervisors, Dr. Thomas Hebbeker and Dr. Christopher Wiebusch. Their guidance, support, and invaluable insights have been the cornerstone of my research. Their expertise and dedication to academic excellence have inspired me profoundly. I am also deeply thankful to Dr. Paulo Ferreira, whose mentorship and practical advice have been instrumental in navigating the complexities of my research topic. Their patience and willingness to share their knowledge have been greatly beneficial.

I owe everything to my family – my parents, my brother, and my guru, for their unwavering faith in me and their unconditional love and support. They have been my pillars of strength, always encouraging me to pursue my dreams with determination and resilience. Their sacrifices and belief in my abilities have been the driving force behind my every action in life.

To my relatives, who have always shown an interest in my studies and cheered me on through every milestone, your kindness and encouragement mean the world to me. Your presence in my life is a constant source of joy and comfort.

My heartfelt thanks go out to my friends, who have been with me throughout this academic journey. Your companionship, understanding, and help during the times I needed it most have not only eased the stress of academic life but have also created cherished memories. The laughter, discussions, and unwavering support from you all have made this journey enjoyable and memorable. My connection with each of you is unique and pure, and I am very lucky who feel this way.

In conclusion, I am grateful to everyone who has been a part of my journey, directly or indirectly. Even the slightest contributions to my academic and personal growth are deeply appreciated and will always be remembered.

BIBLIOGRAPHY

[1] V. F. Hess. "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten". In: *Physikalische Zeitschrift* 13 (Nov. 1912), pp. 1084–1091.

[2] Claus Grupen. "The History of Cosmic Ray Studies after Hess". In: *Nuclear Physics B - Proceedings Supplements* 239-240 (2013). Proceedings of the 9th workshop on Science with the New Generation of High Energy Gamma-ray Experiments: From high energy gamma sources to cosmic rays, one century after their discovery, pp. 19–25. ISSN: 0920-5632. DOI: https://doi.org/10.1016/j.nuclphysbps.2013.05.003. URL: https://www.sciencedirect.com/science/article/pii/S0920563213003708.

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

[4] T. Antoni et al. "KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems". In: *Astroparticle Physics* 24.1 (2005), pp. 1–25. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2005.04.001. URL: https://www.sciencedirect.com/science/article/pii/S0927650505000691.

[5] W. D. et al. Apel. "Ankle-like feature in the energy spectrum of light elements of cosmic rays observed with KASCADE-Grande". In: *Phys. Rev. D* 87 (8 Apr. 2013), p. 081101. DOI: 10.1103/PhysRevD.87.081101. URL: https://link.aps.org/doi/10.1103/PhysRevD.87.081101.

[6] R. Aloisio, V. Berezinsky, and A. Gazizov. "Transition from galactic to extragalactic cosmic rays". In: *Astroparticle Physics* 39-40 (2012). Cosmic Rays Topical Issue, pp. 129–143. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2012.09.007. URL: https://www.sciencedirect.com/science/article/pii/S092765051200182X.

[7] Veniamin Berezinsky, Askhat Gazizov, and Svetlana Grigorieva. "On astrophysical solution to ultrahigh energy cosmic rays". In: *Phys. Rev. D* 74 (4 Aug. 2006), p. 043005. DOI: 10.1103/PhysRevD.74.043005. URL: https://link.aps.org/doi/10.1103/PhysRevD.74.043005.

[8] Alessandro De Angelis and Mário João Martins Pimenta. *Introduction to Particle and Astroparticle Physics*. Springer Milan, 2015. DOI: 10.1007/978-88-470-2688-9. URL: https://doi.org/10.1007%2F978-88-470-2688-9.

[9] Richard E. Lingenfelter. "The Origin of Cosmic Rays: How Their Composition Defines Their Sources and Sites and the Processes of Their Mixing, Injection, and Acceleration". In: *The Astrophysical Journal Supplement Series* 245.2 (Dec. 2019), p. 30. DOI: 10.3847/1538-4365/ab4b58. URL: https://dx.doi.org/10.3847/1538-4365/ab4b58.

[10] T. Pierog et al. "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider". In: *Phys. Rev. C* 92 (3 Sept. 2015), p. 034906. DOI: 10.1103/PhysRevC.92.034906. URL: https://link.aps.org/doi/10.1103/PhysRevC.92.034906.

[11] Felix Riehn et al. "A new version of the event generator Sibyll". In: (2015). arXiv: 1510.00568 [hep-ph].

[12] S. Ostapchenko. "Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model". In: *Phys. Rev. D* 83 (1 Jan. 2011), p. 014018. DOI: 10.1103/PhysRevD.83.014018. URL: https://link.aps.org/doi/10.1103/PhysRevD.83.014018.

[13] W.D. Apel et al. "The spectrum of high-energy cosmic rays measured with KASCADE-Grande". In: *Astroparticle Physics* 36.1 (2012), pp. 183–194. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2012.05.023. URL: https://www.sciencedirect.com/science/article/pii/S0927650512001284.

[14] Sven et al. Schoo. "A new analysis of the combined data from both KASCADE and KASCADE-Grande". In: *PoS* ICRC2017 (2017), p. 339. DOI: 10.22323/1.301.0339.

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

[16] A. et al. Aab. "Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications". In: *Phys. Rev. D* 90 (12 Dec. 2014), p. 122006. DOI: 10.1103/PhysRevD.90.122006. URL: https://link.aps.org/doi/10.1103/PhysRevD.90.122006.

[17] D. J. et al. Bird. "The Cosmic-Ray Energy Spectrum Observed by the Fly's Eye". In: 424 (Mar. 1994), p. 491. DOI: 10.1086/173906.

[18] William Hanlon. "Telescope Array 10 Year Composition". In: *PoS* ICRC2019 (2019), p. 280. DOI: 10.22323/1.358.0280.

[19] Jose Bellido. "Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above $10^{17.2}$ eV and Composition Implications". In: *PoS* ICRC2017 (2017), p. 506. DOI: 10.22323/1.301.0506.

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

[21] Anne Zilles. *Introduction to Cosmic Rays and Extensive Air Showers*. Cham: Springer International Publishing, 2017, pp. 1–13. ISBN: 978-3-319-63411-1. DOI: 10.1007/978-3-319-63411-1_1. URL: https://doi.org/10.1007/978-3-319-63411-1_1.

[22] Sonja Mayotte. "Study of the Cosmic Ray Composition Sensitivity of AugerPrime - Probing the Prospects of the Upgrade to the Pierre Auger Observatory with a Deep Learning Approach". PhD thesis. Bergischen Universität Wuppertal, 2021, p. 155. DOI: 10.25926/99h7-fc44. URL: https://elekpub.bib.uni-wuppertal.de/urn/urn:nbn:de:hbz:468-20210812-123827-3.

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

[24] Koichi Kamata and Jun Nishimura. "The Lateral and the Angular Structure Functions of Electron Showers". In: *Progress of Theoretical Physics Supplement* 6 (Feb. 1958), pp. 93–155. ISSN: 0375-9687. DOI: 10.1143/PTPS.6.93. eprint: https://academic.oup.com/ptps/article-pdf/doi/10.1143/PTPS.6.93/5270594/6-93.pdf. URL: https://doi.org/10.1143/PTPS.6.93.

[25] K Greisen. "Cosmic Ray Showers". In: *Annual Review of Nuclear Science* 10.1 (1960), pp. 63–108. DOI: 10.1146/annurev.ns.10.120160.000431. eprint: https://doi.org/10.1146/annurev.ns.10.120160.000431. URL: https://doi.org/10.1146/annurev.ns.10.120160.000431.

[26] Karl-Heinz Kampert and Alan A. Watson. "Extensive air showers and ultra high-energy cosmic rays: a historical review". In: *The European Physical Journal H* 37.3 (July 2012), pp. 359–412. DOI: 10.1140/epjh/e2012-30013-x. URL: https://doi.org/10.1140%2Fepjh%2Fe2012-30013-x.

[27] M. Aguilar et al. "ISOTOPIC COMPOSITION OF LIGHT NUCLEI IN COSMIC RAYS: RESULTS FROM AMS-01". In: *The Astrophysical Journal* 736.2 (July 2011), p. 105. DOI: 10.1088/0004-637x/736/2/105. URL: https://doi.org/10.1088%2F0004-637x%2F736%2F2%2F105.

[28] M. Hof et al. "ISOMAX: a balloon-borne instrument to measure cosmic ray isotopes". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 454.1 (2000). Proc. of the 1st Int Symp. on Applications of Particle Detectors in Medicine, Biology and Astrophysics, pp. 180–185. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(00)00823-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900200008238.

[29] F. Acero et al. "THE FIRST FERMI LAT SUPERNOVA REMNANT CATALOG". In: *The Astrophysical Journal Supplement Series* 224.1 (May 2016), p. 8. DOI: 10.3847/0067-0049/224/1/8. URL: https://dx.doi.org/10.3847/0067-0049/224/1/8.

[30] B. Keilhauer and et al. "Nitrogen fluorescence in air for observing extensive air showers". In: *EPJ Web of Conferences* 53 (2013), p. 01010. DOI: 10.1051/epjconf/20135301010. URL: https://doi.org/10.1051/epjconf/20135301010.

[31] A. Haungs. "Cosmic rays in the knee energy range". In: *Astrophysics and Space Sciences Transactions* 7.3 (2011), pp. 295–301. DOI: 10.5194/astra-7-295-2011. URL: http://www.astrophys-space-sci-trans.net/7/295/2011/.

[32] Douglas Bergman and for the Telescope Array Collaboration. "Telescope Array: Latest Results and Expansion Plans". In: *Journal of Physics: Conference Series* 1468.1 (Feb. 2020), p. 012078. DOI: 10.1088/1742-6596/1468/1/012078. URL: https://dx.doi.org/10.1088/1742-6596/1468/1/012078.

[33] Christian Glaser et al. "An analytic description of the radio emission of air showers based on its emission mechanisms". In: *Astroparticle Physics* 104 (2019), pp. 64–77. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2018.08.004. URL: https://www.sciencedirect.com/science/article/pii/S0927650518301579.

[34] A. et al. Balzer. "The H.E.S.S. central data acquisition system". In: *Astropart. Phys.* 54 (2014), pp. 67–80. DOI: 10.1016/j.astropartphys.2013.11.007. arXiv: 1311.3486 [astro-ph.IM].

[35] T. Huege and W.D. Apel et a. "The LOPES experiment—Recent results, status and perspectives". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 662 (Jan. 2012), S72–S79. DOI: 10.1016/j.nima.2010.11.081. URL: https://doi.org/10.1016%2Fj.nima.2010.11.081.

[36] Tim Huege and. "Radio detection of cosmic rays with the Auger Engineering Radio Array". In: *EPJ Web of Conferences* 210 (2019). Ed. by I. Lhenry-Yvon et al., p. 05011. DOI: 10.1051/epjconf/201921005011. URL: https://doi.org/10.1051%2Fepjconf%2F201921005011.

[37] Yousaf Butt. "Beyond the myth of the supernova-remnant origin of cosmic rays". In: *Nature* 460.7256 (2009), pp. 701–704.

[38] ZHANG Yi-ran and LIU Si-ming. "The Origin of Cosmic Rays from Supernova Remnants". In: *Chinese Astronomy and Astrophysics* 44.1 (2020), pp. 1–31. ISSN: 0275-1062. DOI: https://doi.org/10.1016/j.chinastron.2020.04.001. URL: https://www.sciencedirect.com/science/article/pii/S0275106220300011.

[39] "The young massive stellar cluster Westerlund 1 in gamma rays as seen with H.E.S.S." In: *Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021)*. Sissa Medialab, Aug. 2021. DOI: 10.22323/1.395.0789. URL: https://doi.org/10.22323%2F1.395.0789.

[40] "Acceleration of petaelectronvolt protons in the Galactic Centre". In: *Nature* 531.7595 (2016), pp. 476–479.

[41] Thomas K. Gaisser, Ralph Engel, and Elisa Resconi. *Cosmic Rays and Particle Physics*. 2nd ed. Cambridge University Press, 2016. DOI: 10.1017/CBO9781139192194.

[42] Rafael Alves Batista et al. "Open Questions in Cosmic-Ray Research at Ultrahigh Energies". In: *Frontiers in Astronomy and Space Sciences* 6 (June 2019). DOI: 10.3389/fspas.2019.00023. URL: https://doi.org/10.3389%2Ffspas.2019.00023.

[43] J. Abraham et al. "Correlation of the Highest-Energy Cosmic Rays with Nearby Extragalactic Objects". In: *Science* 318.5852 (Nov. 2007), pp. 938–943. DOI: 10.1126/science.1151124. URL: https://doi.org/10.1126%2Fscience.1151124.

[44] P. Abreu et al. "Update on the correlation of the highest energy cosmic rays with nearby extragalactic matter". In: *Astroparticle Physics* 34.5 (Dec. 2010), pp. 314–326. DOI: 10.1016/j.astropartphys.2010.08.010. URL: https://doi.org/10.1016%2Fj.astropartphys.2010.08.010.

[45] Charles D. Dermer and Soebur Razzaque. "ACCELERATION OF ULTRA-HIGH-ENERGY COSMIC RAYS IN THE COLLIDING SHELLS OF BLAZARS AND GAMMA-RAY BURSTS: CONSTRAINTS FROM THEiFERMI GAMMA-RAY SPACE TELESCOPE/i". In: *The Astrophysical Journal* 724.2 (Nov. 2010), pp. 1366–1372. DOI: 10.1088/0004-637x/724/2/1366. URL: https://doi.org/10.1088%2F0004-637x%2F724%2F2%2F1366.

[46] B. Theodore Zhang et al. "High-energy cosmic ray nuclei from tidal disruption events: Origin, survival, and implications". In: *Physical Review D* 96.6 (Sept. 2017). DOI: 10.1103/physrevd.96.063007. URL: https://doi.org/10.1103%2Fphysrevd.96.063007.

[47] Glennys R. Farrar and Tsvi Piran. *Tidal disruption jets as the source of Ultra-High Energy Cosmic Rays*. 2014. arXiv: 1411.0704 [astro-ph.HE].

[48] M. Del Santo and et al. "The puzzling source IGR J17361–4441 in NGC 6388: a possible planetary tidal disruption event". In: *Monthly Notices of the Royal Astronomical Society* 444.1 (Aug. 2014), pp. 93–101. DOI: 10.1093/mnras/stu1436. URL: https://doi.org/10.1093%2Fmnras%2Fstu1436.

[49] Robert C. Duncan and Christopher Thompson. "Formation of Very Strongly Magnetized Neutron Stars: Implications for Gamma-Ray Bursts". In: 392 (June 1992), p. L9. DOI: 10.1086/186413.

[50] Ke Fang, Kumiko Kotera, and Angela V. Olinto. "NEWLY BORN PULSARS AS SOURCES OF ULTRAHIGH ENERGY COSMIC RAYS". In: *The Astrophysical Journal* 750.2 (Apr. 2012), p. 118. DOI: 10.1088/0004-637x/750/2/118. URL: https://doi.org/10.1088%2F0004-637x%2F750%2F2%2F118.

[51] B. P. Abbott et al. "Multi-messenger Observations of a Binary Neutron Star Merger". In: *The Astrophysical Journal* 848.2 (Oct. 2017), p. L12. DOI: 10.3847/2041-8213/aa91c9. URL: https://doi.org/10.3847%2F2041-8213%2Faa91c9.

[52] Karl-Heinz Kampert, Miguel Alejandro Mostafa, and Enrique Zas and. "Multi-Messenger Physics With the Pierre Auger Observatory". In: *Frontiers in Astronomy and Space Sciences* 6 (Apr. 2019). DOI: 10.3389/fspas.2019.00024. URL: https://doi.org/10.3389%2Ffspas.2019.00024.

[53] The Pierre Auger Collaboration. *The Pierre Auger Observatory Upgrade - Preliminary Design Report*. 2016. arXiv: 1604.03637 [astro-ph.IM].

[54] A.K. Tripathi et al. "A systematic study of large PMTs for the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 497.2 (2003), pp. 331–339. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(02)01792-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900202017928.

[55] HZC Photonics. *XP1805 PMT*. 2023. URL: http://lampes-et-tubes.info/pm/XP1805.pdf (visited on 10/16/2023).

[56] GIORGIO MATTHIAE and. "THE AUGER EXPERIMENT STATUS AND RESULTS". In: *Astroparticle, Particle and Space Physics, Detectors and Medical Physics Applications*. WORLD SCIENTIFIC, June 2008. DOI: 10.1142/9789812819093_0040. URL: https://doi.org/10.1142%2F9789812819093_0040.

[57] Gregory Snow for Pierre Auger Collaboration. "AugerPrime looks to the highest energies". In: *CERN Courier* (June 2016). URL: https://cds.cern.ch/record/2232590/files/vol56-issue5-p029-e.pdf.

[58] I. Allekotte et al. "The surface detector system of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 586.3 (Mar. 2008), pp. 409–420. DOI: 10.1016/j.nima.2007.12.016. URL: https://doi.org/10.1016%2Fj.nima.2007.12.016.

[59] D. Nitz. "The front-end electronics for the Pierre Auger Observatory surface array". In: *IEEE Transactions on Nuclear Science* 51.3 (2004), pp. 413–419. DOI: 10.1109/TNS.2004.828507.

[60] Hugo Rivera. "Measurement of the energy spectrum of cosmic rays above 3×10^{17} eV using the infill array of the Pierre Auger Observatory". In: *Internal Note of the Pierre Auger Collaboration.* GAP-2013-121 (2013).

[61] G. Silli et al. "Study of the efficiency for the SD-433 array". In: *Internal Note of the Pierre Auger Collaboration.* GAP-2021-001 (2021).

[62] "The Pierre Auger Cosmic Ray Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 798 (2015), pp. 172–213. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2015.06.058. URL: https://www.sciencedirect.com/science/article/pii/S0168900215008086.

[63] J. Abraham et al. "The fluorescence detector of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 620.2 (2010), pp. 227–251. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2010.04.023. URL: https://www.sciencedirect.com/science/article/pii/S0168900210008727.

[64] Radomir Smida. "Scintillator detectors of AugerPrime". In: *PoS* ICRC2017 (2017), p. 390. DOI: 10.22323/1.301.0390.

[65] KURARAY Co. Ltd. *Wavelength Shifting Fibers.* 2023. URL: http://kuraraypsf.jp/psf/ws.html (visited on 10/18/2023).

[66] Hammamatsu Photonics K. K. *R9420 PMT.* 2023. URL: https://www.hamamatsu.com/content/dam/hamamatsu-photonics/sites/documents/99_SALES_LIBRARY/etd/R9420_TPMH1296E.pdf (visited on 10/18/2023).

[67] Jarosław Stasielak. "AugerPrime - The upgrade of the Pierre Auger Observatory". In: *International Journal of Modern Physics A* 37.07 (Mar. 2022). DOI: 10.1142/s0217751x22400127. URL: https://doi.org/10.1142%2Fs0217751x22400127.

[68] Julian Kemp. "Development of a silicon photomultiplier based scintillator detector for cosmic air showers". Dissertation. Aachen: RWTH Aachen University, 2020. DOI: 10.18154/RWTH-2020-12243. URL: https://publications.rwth-aachen.de/record/808708.

[69] Tiina Suomijarvi. "New electronics for the surface detectors of the Pierre Auger Observatory". In: *"PoS"* ICRC2017 (2017), p. 450. DOI: 10.22323/1.301.0450.

[70] Gioacchino Alex Anastasi. "AugerPrime: The Pierre Auger Observatory upgrade". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1044 (2022), p. 167497. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2022.167497. URL: https://www.sciencedirect.com/science/article/pii/S0168900222007896.

[71] Bjarni Pont. "A Large Radio Detector at the Pierre Auger Observatory - Measuring the Properties of Cosmic Rays up to the Highest Energies". In: *PoS* ICRC2019 (2019), p. 395. DOI: 10.22323/1.358.0395.

[72] Ana Martina Botti. "The AMIGA underground muon detector of the Pierre Auger Observatory - performance and event reconstruction". In: *PoS* ICRC2019 (2019), p. 202. DOI: 10.22323/1.358.0202.

[73] B Daniel and (for the Pierre Auger Collaboration). "The AMIGA enhancement of the Pierre Auger Observatory". In: *Journal of Physics: Conference Series* 632.1 (July 2015), p. 012088. DOI: 10.1088/1742-6596/632/1/012088. URL: https://dx.doi.org/10.1088/1742-6596/632/1/012088.

[74] X. Bertou et al. "Calibration of the surface array of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 568.2 (2006), pp. 839–846. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2006.07.066. URL: https://www.sciencedirect.com/science/article/pii/S0168900206013593.

[75] Darko Veberič et al. for the Offline team. "The New Trigger Settings". In: *Internal Note of the Pierre Auger Collaboration.* GAP-2005-035 (2005).

[76] A. Coleman. "The New Trigger Settings". In: *Internal Note of the Pierre Auger Collaboration.* GAP-2018-001 (2018).

[77] Paulo Ferreira. "Search for ultra-high energy photons with the AugerPrime upgrade of the surface detector of the Pierre Auger Observatory". Dissertation. Aachen: RWTH Aachen University, 2023. DOI: 10.18154/RWTH-2023-00774. URL: https://publications.rwth-aachen.de/record/888737/files/888737.pdf.

[78] J. Abraham et al. "Trigger and aperture of the surface detector array of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 613.1 (2010), pp. 29–39. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2009.11.018. URL: https://www.sciencedirect.com/science/article/pii/S0168900209021688.

[79] A. Aab et al. "Reconstruction of events recorded with the surface detector of the Pierre Auger Observatory". In: *Journal of Instrumentation* 15.10 (Oct. 2020), P10021. DOI: 10.1088/1748-0221/15/10/P10021. URL: https://dx.doi.org/10.1088/1748-0221/15/10/P10021.

[80] J. Abraham et al. "Upper limit on the cosmic-ray photon flux above 1019eV using the surface detector of the Pierre Auger Observatory". In: *Astroparticle Physics* 29.4 (2008), pp. 243–256. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2008.01.003. URL: https://www.sciencedirect.com/science/article/pii/S0927650508000157.

[81] D. Ravignani and G. Silli. "The optimal distance to measure the shower size with the 433- metre surface detector". In: *Internal Note of the Pierre Auger Collaboration.* GAP-2021-019 (2021).

[82] Tobias Schulz. "Estimating Mass Composition with AugerPrime : Bestimmung der Massenzusammensetzung mit AugerPrime". 51.13.03; LK 01. MA thesis. Karlsruher Institut für Technologie (KIT), 2021. 103 pp. DOI: 10.5445/IR/1000135140.

[83] "The Pierre Auger Cosmic Ray Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 798 (Oct. 2015), pp. 172–213. DOI: 10.1016/j.nima.2015.06.058. URL: https://doi.org/10.1016%2Fj.nima.2015.06.058.

[84] J. Hersil et al. "Observations of Extensive Air Showers near the Maximum of Their Longitudinal Development". In: *Phys. Rev. Lett.* 6 (1 Jan. 1961), pp. 22–23. DOI: 10.1103/PhysRevLett.6.22. URL: https://link.aps.org/doi/10.1103/PhysRevLett.6.22.

[85] Mockler, Daniela. "Measurement of the cosmic ray spectrum with the Pierre Auger Observatory". In: *EPJ Web Conf.* 209 (2019), p. 01029. DOI: 10.1051/epjconf/201920901029. URL: https://doi.org/10.1051/epjconf/201920901029.

[86] E. dos Santos and A. Yushkov. "Extending the Naples CORSIKA shower library for Auger studies." In: *Internal Note of the Pierre Auger Collaboration.* GAP-2018-043 (2018).

[87] V. Novotny et al. "Comparison between the Napoli and Praha CORSIKA productions." In: *Internal Note of the Pierre Auger Collaboration.* GAP-2019-068 (2019).

[88] A. et al. Aab. "Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events". In: *Phys. Rev. D* 91 (3 Feb. 2015), p. 032003. DOI: 10.1103/PhysRevD.91.032003. URL: https://link.aps.org/doi/10.1103/PhysRevD.91.032003.

[89] H.P. et al. Dembinski. "Report on Tests and Measurements of Hadronic Interaction Properties with Air Showers". In: *EPJ Web Conf.* 210 (2019), p. 02004. DOI: 10.1051/epjconf/201921002004. URL: https://doi.org/10.1051/epjconf/201921002004.

[90] Felix Riehn. "Measurement of the fluctuations in the number of muons in inclined air showers with the Pierre Auger Observatory". In: *PoS* ICRC2019 (2019), p. 404. DOI: 10.22323/1.358.0404.

[91] A. et al. Aab. "Measurement of the Fluctuations in the Number of Muons in Extensive Air Showers with the Pierre Auger Observatory". In: *Phys. Rev. Lett.* 126 (15 Apr. 2021), p. 152002. DOI: 10.1103/PhysRevLett.126.152002. URL: https://link.aps.org/doi/10.1103/PhysRevLett.126.152002.

[92] Javier G. Gonzalez. *The Offline Software of the Pierre Auger Observatory: Lessons Learned*. 2012. arXiv: 1208.2154 [astro-ph.IM].

[93] Rene Brun and Fons Rademakers. "ROOT — An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81–86. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(97)00048-X. URL: https://www.sciencedirect.com/science/article/pii/S016890029700048X.

[94] Pedro et al. Abreu. "Update of the Offline Framework for AugerPrime". In: *PoS* ICRC2021 (2021), p. 250. DOI: 10.22323/1.395.0250.

[95] B. Eichmann, M. Kachelrieß, and F. Oikonomou. "Explaining the UHECR spectrum, composition and large-scale anisotropies with radio galaxies". In: *Journal of Cosmology and Astroparticle Physics* 2022.07 (July 2022), p. 006. DOI: 10.1088/1475-7516/2022/07/006. URL: https://doi.org/10.1088%2F1475-7516%2F2022%2F07%2F006.

[96] Antonio Bueno et al. "A New Observable to Infer the Chemical Composition of Cosmic Rays of Ultra High Energy". In: *Internal Note of the Pierre Auger Collaboration.* GAP-2017-068 (2017).

[97] G. Ros et al. "A new composition-sensitive parameter for ultra-high energy cosmic rays". In: *Astroparticle Physics* 35.3 (Oct. 2011), pp. 140–151. DOI: 10.1016/j.astropartphys.2011.06.011. arXiv: 1104.3399 [astro-ph.HE].

[98] Tin Kam Ho. "Random decision forests". In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol. 1. 1995, 278–282 vol.1. DOI: 10.1109/ICDAR.1995.598994.

[99] Fabian Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12.85 (2011), pp. 2825–2830. URL: http://jmlr.org/papers/v12/pedregosa11a.html.

[100] Gilles Louppe et al. *sklearn.ensemble.RandomForestRegressor*. 2023. URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html (visited on 10/18/2023).

[101] Isabelle Guyon et al. "Gene Selection for Cancer Classification Using Support Vector Machines". In: *Machine Learning* 46 (Jan. 2002), pp. 389–422. DOI: 10.1023/A:1012487302797.

[102] Stephanie Glen. *RMSE: Root Mean Square Error*. 2023. URL: https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/ (visited on 10/18/2023).

[103] Álvaro Taboada Núñez. "Analysis of the First Data of the AugerPrime Detector Upgrade". 51.03.03; LK 01. PhD thesis. Karlsruher Institut für Technologie (KIT), 2020. 136 pp. DOI: 10.5445/IR/1000104548.

[104] Hanlon, William. "Measurements of UHECR Mass Composition by Telescope Array". In: *EPJ Web Conf.* 210 (2019), p. 01008. DOI: 10.1051/epjconf/201921001008. URL: https://doi.org/10.1051/epjconf/201921001008.

[105] Ioana C. Mariş et al. "A Phenomenological Method to Recover the Signal from Saturated Stations". In: *Internal Note of the Pierre Auger Collaboration.* GAP-2006-012 (2006).