THE UNIVERSITY
of ADELAIDE

FACULTY OF SCIENCES, ENGINEERING AND TECHNOLOGY

DEPARTMENT OF PHYSICS

Event Reconstruction with the AugerPrime Upgrade

Bradley Manning

Supervisors:
Prof. Bruce DAWSON
Dr. Jose BELLIDO

January, 2024

Contents

Abstract

The origin and mass composition of ultra-high energy cosmic rays (UHECR) remains as one of the biggest unsolved mysteries in astrophysics. The Pierre Auger Observatory (Auger), adopts a hybrid detection technique which utilises fluorescence telescopes and surface detector stations which measure extensive air showers (EAS) produced by cosmic rays interactions with the atmosphere, allowing Auger to reconstruct the energy, arrival direction and mass composition of UHECRs.

Auger has deployed scintillator surface detectors (SSD) on top of their existing surface detector stations, above the existing water-Cherenkov detectors (WCD). The calibration of the new SSD detector has been carefully defined, based on the existing WCD detector. The SSD signal is recorded as the number of minimum ionising particles (MIP), and the calibration has been defined based on simulations until the proposed measurement is performed in the field.

A study into the different types of detector shielding has been performed. Through a custom mock simulation of the Telescope Array surface detector (TASD) incorporated within the Auger SSD simulation framework, the TASD has been simulated alongside the SSD. The results describe in detail how different types of detector shielding, such as stainless steel or aluminium, can significantly impact the quantity of signal measured, as well as change the proportions of signal measured from different electromagnetic particles. Furthermore, using this simulation, an analysis on changing the mass composition assumptions of cosmic rays during reconstruction procedures is performed, testing whether a significant alteration of cosmic ray spectra is seen.

The SSD signal uncertainty model has been recalculated, using real data from SSD station multiplets, improving upon the original model which was defined with Monte Carlo simulations. The signal uncertainty analysis required an SSD Lateral Distribution Function (LDF) to be defined, and so, an SSD LDF parameterisation for the 750 m ground array, and

1500 m ground array has been performed, using the maximum likelihood. The residuals from the SSD LDF parameterised form compared with real data show that the fits perform well for distances from the shower core up to 2000 m, for primary particle energies greater than $10^{18.5}$ eV, at all zenith angles.

The SSD and WCD will simultaneously collect data, allowing researchers to compare signals from the two detectors in a variety of ways. One particular use case is to compare to two signals to understand cosmic ray mass composition. An analysis is performed on the signal ratio (SSD/WCD), where the signal ratio for simulated proton and iron air showers is compared to the signal ratio from real data, outlining the issues that need to be tackled before accurate mass composition analysis can be achieved through study of signal ratios. Similarly, a matrix formalism analysis is performed on different sets of simulated data, to understand and outline some of the best case scenarios that could be achievable, when real data becomes abundant. The results of this thesis will help others understand the difficulties and challenges that need to be overcome when performing detailed SSD analysis.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Bradley Manning
January, 2024

Acknowledgements

I want to start by expressing my gratitude to my supervisors, Dr Jose Bellido and Professor Bruce Dawson for the continued support throughout this journey. Your expertise, patience and commitment to excellence has been invaluable in shaping this research to where it is today.

I am also thankful to the members of the University of Adelaide, High Energy Astrophysics Research group. The weekly meetings, where we shared insightful feedback, constructive criticism, and valuable suggestions, which all greatly contributed to the improvement of this thesis.

I extend this gratitude to all the subject matter experts and friends I made during my travels to Karlsruhe, Germany at the Karlsruhe Institute of Technology. My time there could not have been better. Special thanks to Dr David Schmidt, who has been incredibly supportive and helpful over the many different aspects of this work.

I am indebted to my friends, family and colleagues who have supported me emotionally and intellectually through the years. Your companionship has made the challenging moments much more manageable, and the achievements more joyful.

Thank you all for being a part of this journey.

Bradley Manning
January, 2024

Chapter 1

Cosmic Rays

Cosmic rays are extra-terrestrial, high-energy particles that continuously bombard our planet. They cover a vast range in energy, from the GeV scale to more than 10^{20} eV. Their composition consists mainly of protons and other atomic nuclei. Cosmic rays can be used as a way to gather information about the highest energy astrophysical events and environments, as well as information about the interstellar medium, such as the strength and structure of the magnetic fields through which they propagate. Some of the key challenges currently in astroparticle physics is understanding the origins, acceleration mechanisms, propagation and mass composition of the highest energy cosmic rays.

1.1   History of Cosmic Rays

Cosmic ray history can only be truly appreciated after first investigating the research performed in the late 18th century by French physicist Charles-Augustin de Coulomb. Coulomb used instruments known as electroscopes to conduct experiments centred around static electricity. A typical electroscope from the 18th century is the gold-leaf electroscope, it consists of a vertical metal rod from the end of which hangs two gold leaves. When electric charge is applied to top of the rod, the gold leaves would repel due to the electromagnetic forces [1]. It was expected that an electroscope should maintain its charge if the isolation was perfect, however, Coulomb was surprised when the electroscopes would spontaneously discharge. The possible causes for the spontaneous discharge were not clear at the time.

In the late 1800s, Julius Elster and Hans Geiter were able to improve the technique for careful insulation of electroscopes in a closed vessel; thus improving the sensitivity and ac-

curacy of measurements. In 1899, they isolated the electroscope by placing it inside a heavily shielded metal box. Then, by observing discharge rates they were able to conclude that the source of the radiation must be from outside the box, and due to the required penetrating properties, the source must also be high energy.

In 1912, Austrian physicist Victor Hess conducted an experiment which was considered to be the major breakthrough for cosmic rays. Hess performed a series of high altitude balloon flights while monitoring the radiation of his surroundings with Wulf's electroscopes; an electroscope designed and built by Theodor Wulf in 1910, which offered better precision and accuracy than the existing gold leaf electroscopes available at the time. Hess found that the ionisation decreased as he ascended to heights of around 2 km, before significantly increasing to 3 to 5 times the ionisation level seen at ground. This lead Hess to conclude that the radiation was extraterrestrial in nature, and he also ruled out the possibility of the Sun being responsible for the origin of this radiation by obtaining similar results during night and day. In a series of experiments in 1925, Robert Millikan achieved results that were consistent with the attenuation of radiation originating from the top of the atmosphere; he was the first to use the term that we use today, "cosmic rays".

It is now known that cosmic rays are not electromagnetic radiation, but are energetic charged particles that can cause a cascade of subatomic particles and γ-rays from interactions in the atmosphere. The first piece of evidence describing the secondary particles was discovered by Italian physicist Bruno Rossi in the early 1930s when he performed an experiment in which he placed 3 Geiger counters in a triangular arrangement, and covered them with a lead screen. When the lead screen was in place, Rossi measured a coincidence rate for events per hour; without the screen he measured no coincidence. The measured coincidences could have only been the result of two or more ionizing particles emerging from the lead. This later turned out to be that secondary particles were produced alongside interactions of primary particles with the lead screen [2]. In 1938, French physicist Pierre Victor Auger used particle detectors positioned at high altitudes in the Swiss Alps, and discovered coincidences between particles detected at distances of up to 300m. He came to the conclusion that these cascades of particles were produced by the same source, and uncovered the phenomenon of "extensive air showers" (EAS), which we know of today. It was these cascades that were responsible for the ionising radiation that Hess discovered.

1.2   Cosmic Ray Spectrum

The cosmic ray energy spectrum describes the number of particles arriving at Earth as a function of energy, it can be described by the following inverse power law.

$$\frac{dN}{dE} \propto E^{-\gamma} \qquad \text{m}^{-2}\text{s}^{-1}\text{sr}^{-1}\text{eV}^{-1}, \tag{1.1}$$

where N represents the number of particles arriving at Earth, E as the particle energy and γ is the spectral index. The spectrum falls steeply, and from around 10^{10} eV to 10^{20} eV it can be competently represented by choosing $\gamma \sim 3$, indicating non-thermal acceleration processes (see Figure 1.1) [3]. Scaling the cosmic ray flux by some energy-dependent factor such as $E^{2.6}$ in Figure 1.2 allows various features to be highlighted. These features begin with a steepening at $\sim 3 \times 10^{15}$eV (the *knee*), followed by a further steepening at $\sim 3 \times 10^{17}$ eV (the *second knee*), a change in spectral index at $\sim 3 \times 10^{18}$ eV (the *ankle*), and a suppression of cosmic ray flux beyond $\sim 6 \times 10^{19}$ eV. The physical reasons behind each of these features is one of the key areas of cosmic ray studies, they are generally expressed to be related to the origin, acceleration mechanics and propagation of cosmic rays. More recently, a new feature in the Auger spectrum of cosmic rays above 10^{19} eV has been reported [4]. The new feature is called the "instep", where the steepening is no longer modelled with a simple break, as there is one prior to the high energy fall-off. This means that the Auger spectrum is hence fitted with four power-laws [4].

A popular area of discussion for the cosmic ray spectrum is the suppression, in particular the difference in spectra found at the highest cosmic ray energies between two experiments, the Pierre Auger Observatory and the Telescope Array. Both experiments have results showing the energy spectrum of cosmic rays (see Figure 1.3) which agree within systematic uncertainties until cosmic ray energies exceed 10^{19} eV. There are a range of possible reasons why the two experiments disagree; different analysis methods, different detector design and calibration, and possibly cosmic ray anisotropy since the two experiments view different hemispheres of the sky.
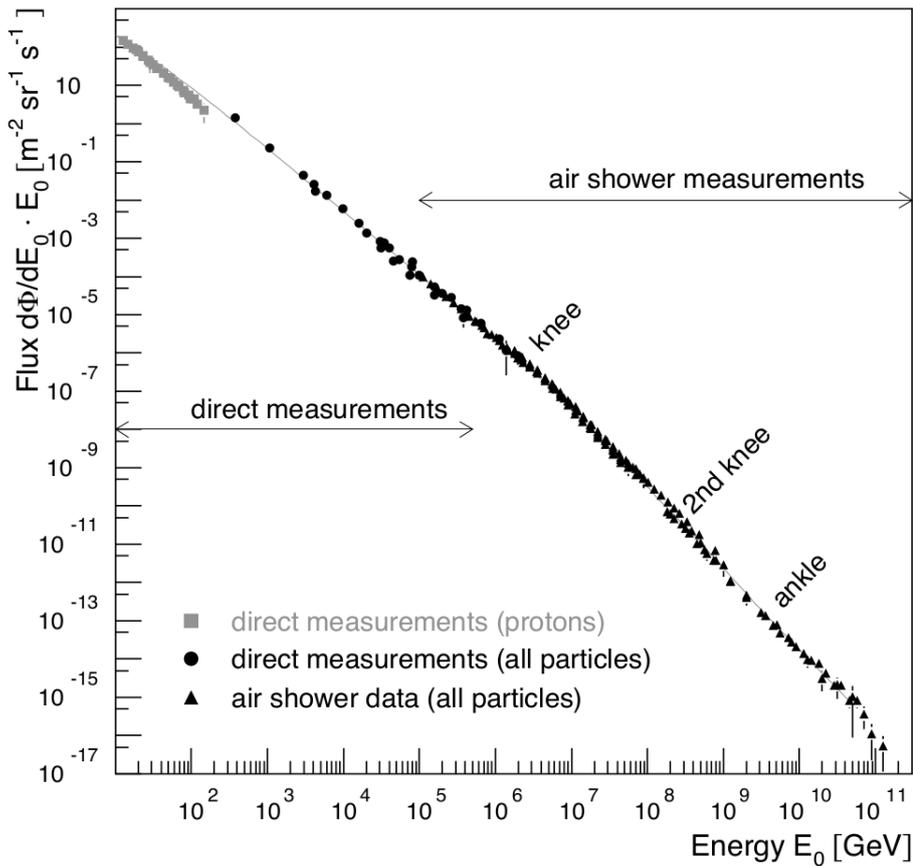
Figure 1.1: The cosmic ray energy spectrum from 1 GeV to 100 EeV. The deviation from a single power law over a large energy range can be seen, with each feature highlighting the energies where more deviation begins. From [3].
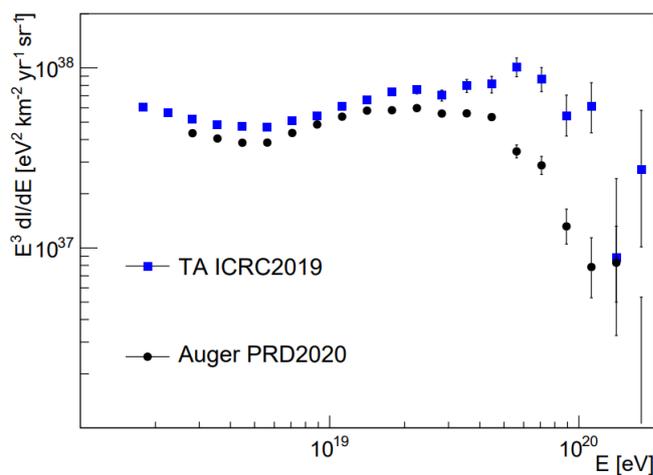
Figure 1.3: Comparison between the Telescope Array (TA) and Auger combined spectra presented at the 37th International Cosmic Ray Conference (ICRC 2021). The TA spectrum is shown in the energy range where Auger data is available. The vertical scale is multipled by E^3. Systematic errors cannot explain the difference seen at cosmic energies $> 10^{19}$ eV. From [6].
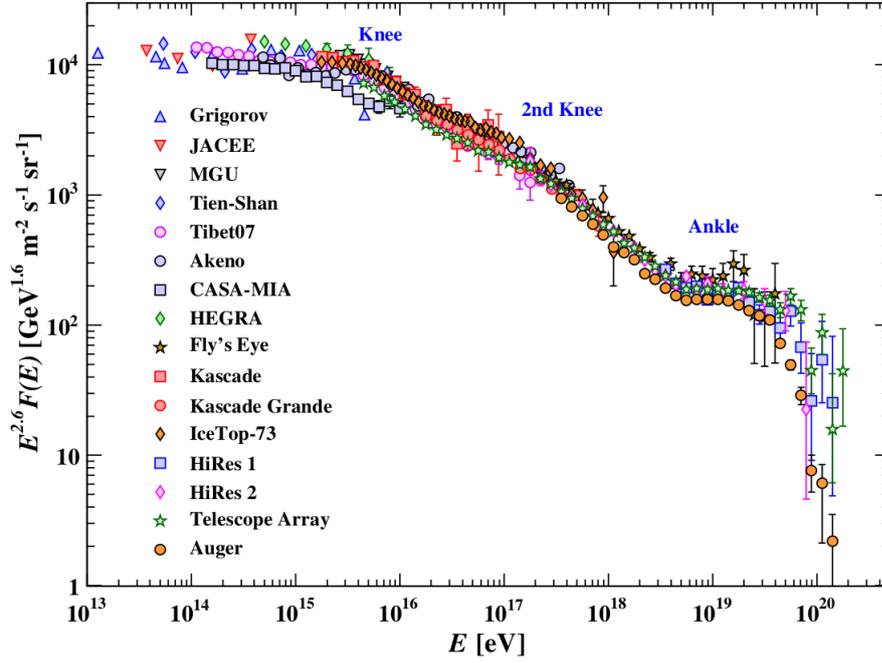
Figure 1.2: The cosmic ray energy spectrum from many different experiments, ranging from 10 TeV to 100 EeV, with the flux multiplied by $E^{2.6}$. By scaling the vertical axis with energy, the structure of each feature in the spectrum is emphasised. From [5]

1.2.1 Greisen-Zatsepin-Kuz'min (GZK) Limit

The cosmic ray flux suppression above $\sim 6 \times 10^{19}$ was first predicted by Greisen [7] and later independently by Zatsepin and Kuz'min [8] shortly after the discovery of the cosmic microwave background radiation (CMBR) by Penzias and Wilson in 1965. The GZK effect predicts the loss of energy of incoming ultra high energy (energies around or above the *ankle*) cosmic rays (UHECR) by photo-pion and pair-production on the CMBR. Ultra high energy protons (p) will interact with CMBR photons (γ_{CMBR}) through pair production

$$p + \gamma_{\text{CMBR}} \longrightarrow p + e^{+} + e^{-} \tag{1.2}$$

in which the proton will lose energy via the production of an electron-positron ($e^{-}e^{+}$) pair. Ultra high energy protons may also interact through photo-pion production

$$p + \gamma_{\text{CMBR}} \longrightarrow n + \pi^{+}$$
$$p + \gamma_{\text{CMBR}} \longrightarrow p + \pi^{0} \tag{1.3}$$
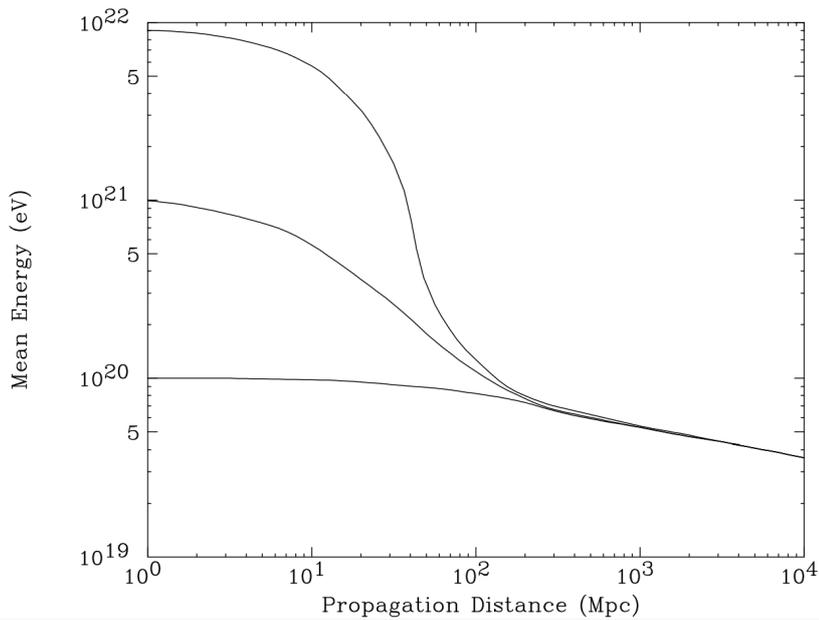
Figure 1.4: A graphical representation of the restriction of cosmic ray sources to within 100 Mpc of our galaxy due to energy loss from the CMBR. Changes in energy as a function of propagation distance for protons of initial energy 10^{20}, 10^{21} and 10^{22} eV are shown. If a 10^{20} eV proton is observed at Earth, its source is likely to be within 100 Mpc, regardless of the initial energy. From [9]. Similar distance limits are applicable to cosmic ray nuclei.

where a neutral or charged pion (π) is created. Each time a proton undergoes pion photoproduction it will lose $\approx 20\%$ of its energy. Nuclei propagating through intergalactic space are also predicted to experience attenuation through interactions with CMBR photons through pair production and photodisintegration [9]. These interactions impose a limit for the distance at which sources of cosmic rays above this energy limit can be located. Protons arriving with energies above 10^{20} eV will not be expected to propagate distances larger than 100 Mpc, independent of their initial energy (see Figure 1.4).

1.3   Cosmic Ray Acceleration Mechanisms

One of the crucial questions in cosmic ray physics is that of their origin. There are two main types of cosmic ray production models, the top-down and bottom-up models. Bottom-up refers to the acceleration of a population of charged particles from low to high energies, while top-down models predict UHECRs to originate from exotic particles. Top-down models presume that UHECRs are decay products of massive particles that may be remnants

from the early stages of the universe. These models predict that the decay of such particles into UHECRs result in the production of many electrons, positrons, neutrinos and photons, however, experimental results on the flux of high-energy photons across a large range of energies disfavour such models [10].

1.3.1   Fermi Acceleration (Top-down Model)

Fermi's original acceleration model, also referred to as 2nd order Fermi acceleration is a stochastic model which considers the interactions between a cosmic ray and interstellar medium (ISM) gas clouds during its propagation through the galaxy. In this model, a cosmic ray enters an ISM gas cloud, which itself is moving at some velocity with respect to the motion of the galaxy, and the cosmic ray scatters off irregularities in the cloud's magnetic field before eventually emerging from the cloud having experienced a change in energy and subsequent acceleration.

Consider the scenario depicted in Figure 1.5, with an ISM cloud travelling at velocity v and a cosmic ray (charged particle) with energy E_1 and momentum p_1 which enters the cloud at an angle θ_1 with respect to the cloud's trajectory. The initial energy of the particle in the ISM cloud reference frame is given by:

$$E_1' = \gamma E_1 (1 - \beta \cos \theta_1) \tag{1.4}$$

where $\beta = v/c$ and $\gamma = 1/\sqrt{1 - \beta^2}$. In the cloud reference frame, the particle scatters elastically and will emerge from the cloud in a random direction with no change in energy:

$$E_2' = E_1' \tag{1.5}$$

where E_1' is given in Equation 1.4. Transforming Equation 1.4 into an inertial reference frame gives the particle's energy after collision as

$$E_2 = \gamma E_2' (1 - \beta \cos \theta_2') \tag{1.6}$$

The fractional energy gain by the particle each time it interacts with the cloud is then given by

$$\frac{\Delta E}{E} = \frac{E_2 - E_1}{E_1} = \frac{1 + \beta \cos \theta_1 + \beta \cos \theta_2' - \beta^2 \cos \theta_1 \cos \theta_2'}{1 - \beta^2} - 1 \tag{1.7}$$
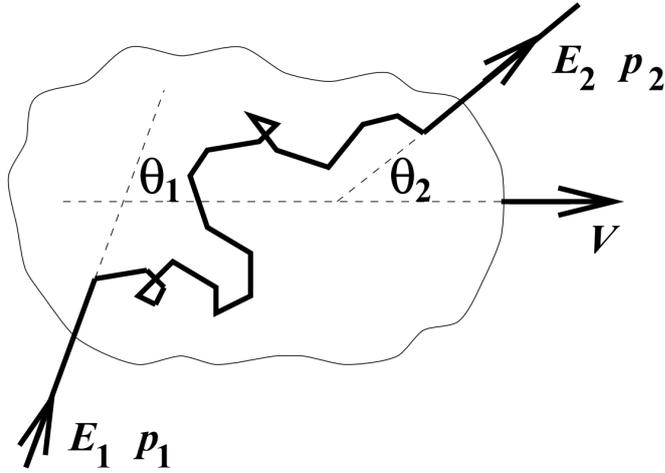
Figure 1.5: A diagram depicting the interaction of a cosmic ray particle of energy E_1 with an ISM cloud moving at velocity v. From [11].

The average fractional energy gain can be determined based on the average angles of collision and scattering, $\cos\theta_1$ and $\cos\theta_2'$. The average value of $\cos\theta_2'$ can be written as

$$\langle\cos\theta_2'\rangle = 0 \tag{1.8}$$

as the particle scatters such that it will emerge from the cloud in a random direction. The average value of $\cos\theta_1$ is dependent on the rate of collisions, and is proportional to their relative velocities. The average value of $\cos\theta_1$ is given as

$$\langle\cos\theta_1\rangle = -\frac{\beta}{3} \tag{1.9}$$

It can be shown by substituting the previous two equations into Equation 1.7, that the energy gain per cloud encounter is second-order in β,

$$\frac{\langle\Delta E\rangle}{E} \propto \frac{4}{3}\beta^2 \tag{1.10}$$

The downside of this acceleration mechanism, is that the energy gain per encounter is small, as $\beta \ll 1$. This is because the probability of a head-on collision (energy gain) is only slightly higher than that of a tail-on collision (energy loss).

1.3.2   Diffusive Shock Acceleration

Galactic cosmic rays are commonly considered to be accelerated through the process of diffusive shock acceleration in sites such as supernova remnants. This example of a bottom-up model requires the confinement of cosmic rays to their acceleration region during the acceleration process. During a supernova explosion, many solar masses of stellar material are ejected at speeds far exceeding the speed of sound in the ISM. An associated shock front propagates radially outwards. The velocity of the shock front v_s depends heavily on the velocity of the ejecta. Similarly to the previously described Fermi acceleration, it can be shown that the average energy gain per shock crossing is linear in β_s

$$\frac{\langle \Delta E \rangle}{E} \propto \frac{4}{3}\beta_s \tag{1.11}$$

This is a much more efficient acceleration mechanism, which can be explained by the fact that the particle sees gas clouds approaching at their own velocity, v_p, which significantly increases the probability of head-on collisions. The Hillas plot shown in Figure 1.6 shows the requirements for an accelerating region capable of containing protons and iron nuclei of energy $> 10^{20}$ eV, based on an maximum possible particle energy attainable in such an acceleration process. The maximum attainable energy (E_{max}) for a given particle of charge Z inside a magnetic field strength B (in μG) is given by:

$$E_{max} \approx ZBR\beta_s \times 10^{18}\text{eV} \tag{1.12}$$

where R is the size of the accelerating region (in kpc), k < 1 [12], and β_s is the shock velocity (in units of c) [3]. Active galactic nuclei (AGN), radio galaxies, neutron stars and gamma ray bursts (GRBs) are some of the only types of sources capable of being UHECR accelerators.
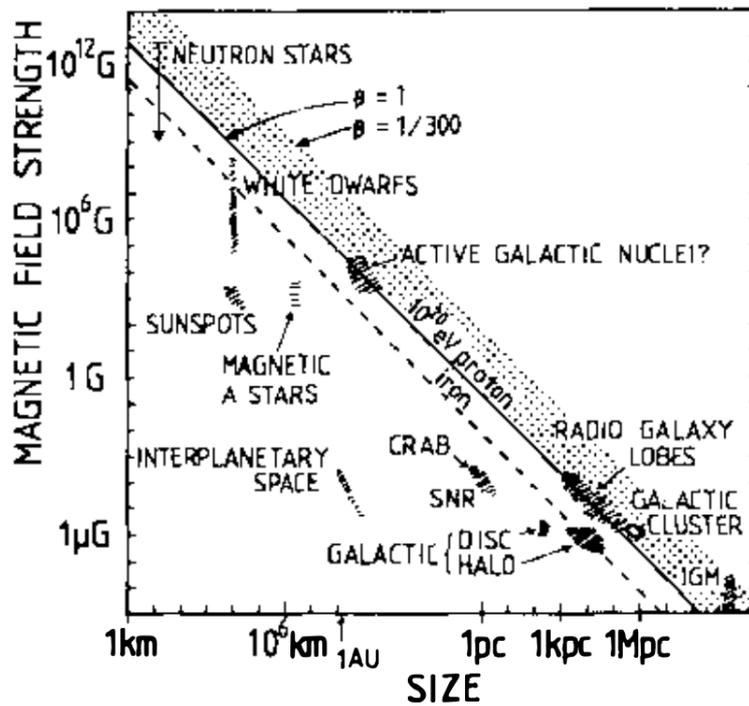
Figure 1.6: The "Hillas plot", which displays the dimensions and magnetic field strength of various astrophysical objects compared to those required for the acceleration of a 10^{20} eV proton or iron nuclei. The dashed line represents the minimum requirements for the acceleration of iron nuclei, whereas the solid line represents the minimum requirements for proton nuclei. Not all of these objects produce shocks, but jets in Active Galactic Nuclei (AGN) do contain large-scale shocks. From [12]

1.4   Cosmic Ray Anisotropy

Given that a cosmic ray is a charged particle, during propagation it will suffer from deflections due to the presence of galactic and extragalactic magnetic fields. Anisotropy studies aim to develop an understanding of the processes through which cosmic rays are created, but complications arise as most cosmic rays lose all information pertaining to their points of origin before arrival at Earth. Neutral particles (such as photons or neutrinos) can be considered as one of the most logical places to look for cosmic ray sources as they are not deflected in magnetic fields, hence their arrival directions point directly to the source. Proton primary cosmic rays at the very highest energy may only suffer deflections of order of a few degrees, hence being able to identify the primary particle of a cosmic ray is an important area of study; an area that the Pierre Auger Observatory upgrade, AugerPrime will expand on. UHECR sources are expected to generate neutrinos through the decay of pions and neutrons during acceleration. Neutrinos and high energy photons are also produced in large quantities in top-down models.

Large-scale anisotropy studies are commonly performed using a classical harmonic search for non-uniformities in the distribution of cosmic ray arrival directions as a function of right ascension. The Pierre Auger Collaboration has reported the existence of a large-scale dipole anisotropy at energies greater than 8 EeV, with a significance of 5σ, the results favour an extragalactic origin for cosmic rays above 8 EeV [13]. A sky map in equatorial coordinates of the cosmic-ray flux above 8 EeV is shown in Figure 1.7. In addition to the dipole analysis, the Pierre Auger Collaboration have also published results indicating a possible anisotropy in the arrival directions of cosmic rays of the highest energies from gamma-ray bright AGN [14].

Additional evidence of extra-galactic origin of UHECRs has been found by Auger with correlations of cosmic ray arrival directions with nearby starburst galaxies. A significance of 4.5 σ is found for the maximum likelihood of starburst galaxies at energies above 38 EeV. There is also an excess observed by Auger at Centaurus A which rejects the isotropic hypothesis with a signifiance of 3.9 σ [15].
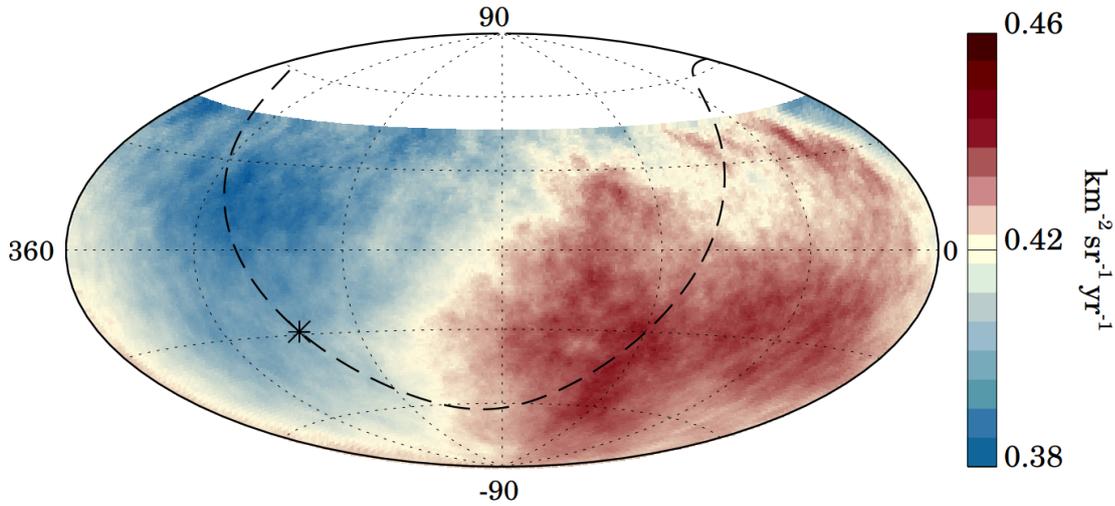
Figure 1.7: A Sky map in equatorial coordinates from the Pierre Auger Obsevatory (Auger), showing the cosmic-ray flux above 8 EeV smoothed with a 45° top-hat function. The Galactic centre is marked with an asterisk, and the Galactic plane is represented by the dashed line. From [13].

1.5 Mass Composition of Cosmic Rays

Cosmic ray mass composition studies aim to determine the type of and abundance of different cosmic ray nuclei as a function of energy. Mass composition knowledge is essential in the interpretation of features in the energy spectrum, and in anisotropy studies and searches for cosmic ray sources. Due to the extremely low flux of UHECR, their mass composition cannot be determined through direct measurement of the primary particles. Instead, the mass composition must be inferred from the properties of the secondary produced in the induced extensive air shower (EAS). X_{max} is one of the most robust mass-sensitive EAS parameters, it is the depth at which a shower reaches maximum size. Measurements of X_{max} from ground-based fluorescence detectors are compared to expectations from various hadronic interaction models to infer the average mass composition of cosmic rays as a function of energy. The latest mass composition results from the Pierre Auger Collaboration are summarised in Figure 1.8, where the measurements of $\langle X_{\mathrm{max}} \rangle$ and $\sigma(X_{\mathrm{max}})$ are shown alongside the predictions for proton and iron nuclei of the hadronic models EPOS-LHC, Sibyll 2.3c and QGSJETII-04 within the energy range $10^{17.2}$ eV to $10^{19.7}$ eV. The X_{max} results are often
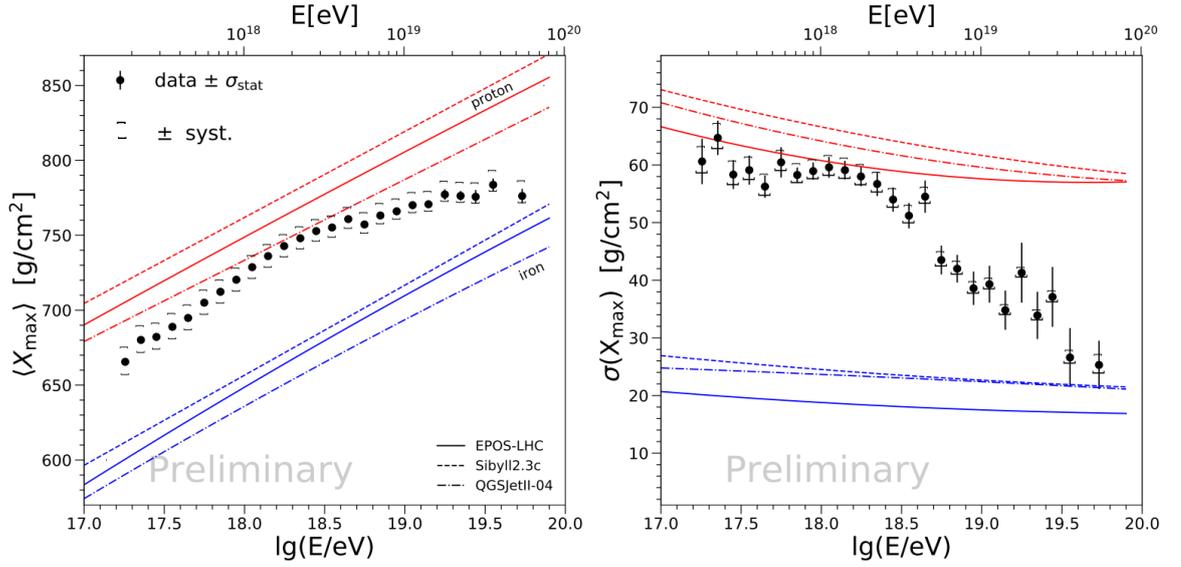
Figure 1.8: (ICRC 2019) Measurements of mean X_{max} (left) and the RMS in X_{max} (right) measured by the Pierre Auger Observatory, compared to the predictions for proton and iron nuclei of the hadronic models EPOS-LHC, Sibyll 2.3c and QGSJETII-04 [16].

interpreted with reference to the elongation rate. The elongation rate being defined as the rate of change of $\langle X_{max} \rangle$ per decade of energy. The elongation rate is (79 ± 1) g/cm²/decade of energy for energies between $10^{17.2}$ eV and $10^{18.33}$ eV, suggesting that the mean primary mass is becoming lighter with increasing energy (when compared to an elongation rate of $\sim$60 g/cm²/decade for constant mass). For energies above $10^{18.33}$ eV, the elongation rate approaches $\sim$26 g/cm²/decade of energy, signifying that the mean primary mass is heading towards heavier nuclei. Similarly, $\sigma(X_{max})$ decreases for energies greater than $10^{18.33}$ eV; a decrease in fluctuations in X_{max} provides further evidence of larger quantities of heavier nuclei as shower-to-shower fluctuations are expected to decrease in magnitude with increasing mass.

Chapter 2

Extensive Air Showers

When a cosmic ray (the "primary" particle) enters the atmosphere it may initiate an EAS by interacting with an air molecule in the Earth's atmosphere. A sequence of particle interactions and decays leads to a cascade of particles travelling along the shower axis, that is, in direction of the primary particle. A thin disk-like structure at the leading edge of the cascade, referred to as the shower front, travels at speeds close to that of light and increases in width following particle interactions and scattering. The size of the shower is characterised by the number of generated particles, and the size increases rapidly following the first interaction until it reaches a maximum after which it then decreases due to losses to the surrounding environment. An air shower consists of three main components, the electromagnetic component, the muonic component and the hadronic component.

- **Electromagnetic component -** This component consists of electrons, positrons and γ-rays produced from a series of electromagnetic particle interactions in the atmosphere. It is by far the largest of the three EAS components.

- **Muonic component -** consists of muons and neutrinos produced in the decay of charged pions and kaons.

- **Hadronic component -** consists of protons, neutrons, charged pions and kaons.

2.1 Heitler's Model

In 1944, Heitler proposed a simple toy model for electromagnetic shower development in his book on the quantum theory of radiation[17]. The model is useful for understanding

the basic concepts and properties of complex particle showers. It assumes that the energy of an incident particle is divided evenly among secondary particles and that interactions in the atmosphere occur only once a fixed atmospheric depth has been traversed. This depth d can be calculated with:

$$d = \lambda_r ln(2) \tag{2.1}$$

where λ_r is the radiation length in the medium. A basic representation of the model is shown in Figure 2.1. A γ-ray of initial energy E_0 interacts with an atmospheric nucleus after travelling the required depth d, producing an electron-positron pair via the process of pair production.

$$\gamma + N \longrightarrow e^+ + e^- \tag{2.2}$$

The electron and positron will both carry half the energy of the photon, $E_0/2$. The electron and positron will then give half of their energy to a γ-ray via the process of bremsstrahlung after travelling a further required depth d.

$$e^\pm + N \longrightarrow N + e^\pm + \gamma \tag{2.3}$$

The γ-rays from this process will then continue on to produce more electron-positron pairs. These processes will continue until the energy of the particles fall below the required energy for further particle production. Below this threshold energy, E_c (E_c = 85 MeV for air [18]), the Heitler model assumes that particles will stop multiplying and dissipate their energy rapidly through ionisation. The depth of the shower maximum is obtained based on the number of interaction lengths. A particle shower developing in this way will contain 2^n particles after traversing an atmospheric depth of $d \times n$, with the energy of each shower particle given by

$$E(n) = \frac{E_0}{2^n} \tag{2.4}$$

The air shower reaches its maximum particle number when the individual particles have energy equal to E_c. that is

$$N_{\max} = \frac{E_0}{E_c} \tag{2.5}$$

which leads to a maximum number of radiation lengths before the shower reaches maximum development, given by

$$n_{\max} = \frac{\ln(E_0/E_c)}{\ln 2} \tag{2.6}$$
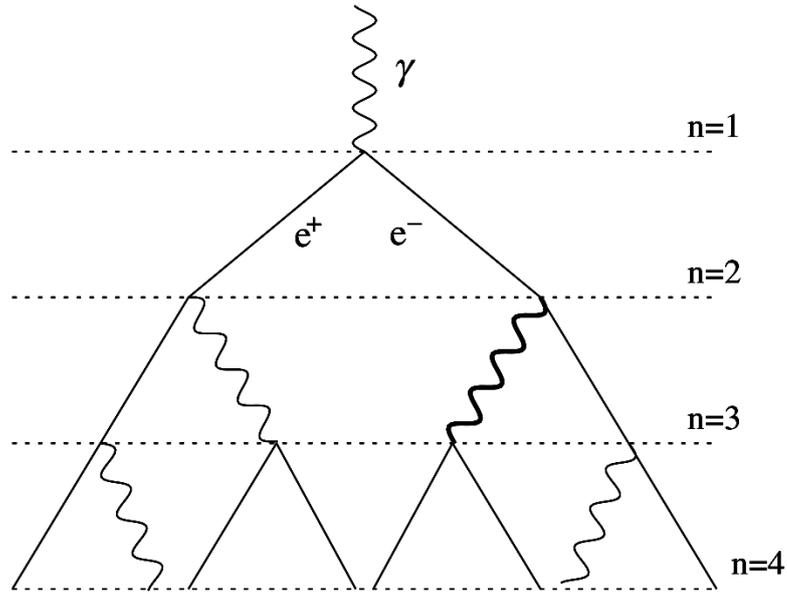
Figure 2.1: A small, simple representation of the first 4 stages of interactions in Heitler's Model. The initial γ-ray will interact via pair production and distribute its energy evenly among the electron-positron pair. The electron-positron pair will then continue to interact via bremsstrahlung, producing additional photons, which will interact further, and so on. From [18].

The depth at which an electromagnetic shower reaches maximum size, $X_{\max}^{\gamma}$, is therefore given by

$$X_{\max}^{\gamma} = n_{\max} \times X \tag{2.7}$$
$$= \lambda_r \ln(E_0/E_c) \tag{2.8}$$

with λ_r as the radiation length of the medium (~ 37 g/cm^2 for air), the elongation rate (change in depth of shower maximum per decade of energy) can be derived to be $\Lambda^{\gamma} = 85$ g/cm^2 per decade. The Heitler Model correctly predicts shower development up to $X_{\max}$. In particular, this model predicts that $X_{\max} \propto \ln(E_0)$ and $N_{\max} \propto E_0$. However, the model does underestimate the ratio of photos to leptons due to the simplifying assumptions.

2.2   Heitler-Matthews Model

The Heitler model can be extended to the case of an air shower initiated by the interaction of a hadron (proton or heavier nucleus) with an atmospheric nucleus to represent what is commonly known as Heitler-Matthew's Model [18]. The primary hadron first interacts with the atmosphere after traversing one hadronic interaction length λ_I where it produces pions and kaons. The first interaction of a primary cosmic ray proton with an atmospheric nucleus, N can be represented as

$$p + n \rightarrow (p, n) + N' + \pi^0 + \pi^+ + \pi^- + K^0 + K^+ + K^- \tag{2.9}$$
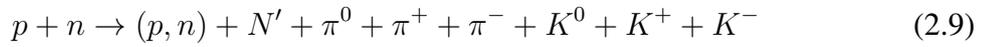
where N' represents the fragmented nucleus. The neutral pions then each decay into two γ-rays, which then continue to cause electromagnetic cascades. Charged pions can either decay or interact, the higher energy charged pions are more likely to interact in the atmosphere, producing more charged and neutral pions. The charged pions that decay will produce muons and neutrinos which form the muonic component of an air shower.

$$\pi^- \longrightarrow \mu^- + \bar{\nu}_\mu \tag{2.10}$$

$$\pi^+ \longrightarrow \mu^+ + \nu_\mu \tag{2.11}$$

After the energy of the charged pions has dropped below the required energy for interaction, almost all of them will exclusively decay into muons and neutrinos. Resulting muons may decay into electrons or positions, which contribute to the electromagnetic component of the shower. Matthews extended the Heitler model to the case of showers initiated by proton or heavier nucleus. The resulting model considers a population of pions and their interactions over an atmosphere modelled as several layers of equal thickness $\lambda_l \times \ln 2$, where λ_l is the interaction length of strongly interacting particles ($\lambda_l \approx 120$ g/cm^2 for pions in air). This model can be used to determine the depth of the shower maximum, $X_{\max}$. $X_{\max}$ for a proton initiated shower, with primary energy E_0 and interaction length λ_l is expressed as

$$X_{\max}^p = X_0 + \lambda_r \ln\left(\frac{E_0}{3N_{ch}E_c}\right) \tag{2.12}$$
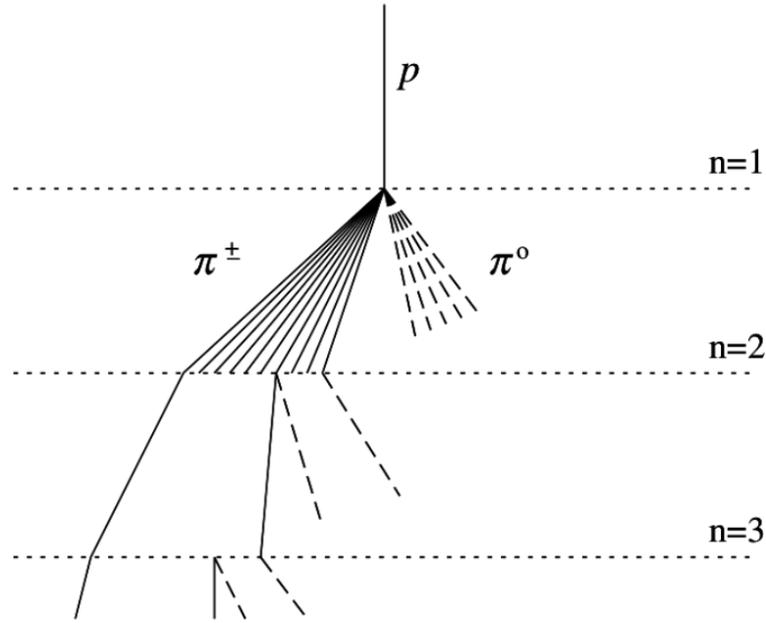
Figure 2.2: Schematic view of an air shower initiated by a hadron. The depiction is exaggerated and is not drawn to scale. From [18].

where X_0 is the atmospheric depth of the first interaction, λ_r is the radiation length of the medium, N_{ch} defines the multiplicity of charged particles produced in hadron interactions ($N_{ch} = 10$, in Matthews' model) and E_c is the critical energy.

It is helpful to express Equation 2.12 in terms of the depth of shower maximum for the electromagnetic cascade from Equation 2.7,

$$X_{\max}^p = X_{\max}^\gamma + X_0 - \lambda_r \ln(3N_{ch}) \tag{2.13}$$

Then, the elongation rate for a pure proton shower can be estimated to be $\Lambda^p \approx 58 g/cm^2$ per decade, which is comparable with results from hadronic air shower simulations.

2.3   Detection Methods

Shower particles that reach ground can be detected by surface detectors such as the Pierre Auger water-Cherenkov tanks, underground muon detectors or scintillators. The amount of signal reported from each detector is proportional to the number of particles detected;
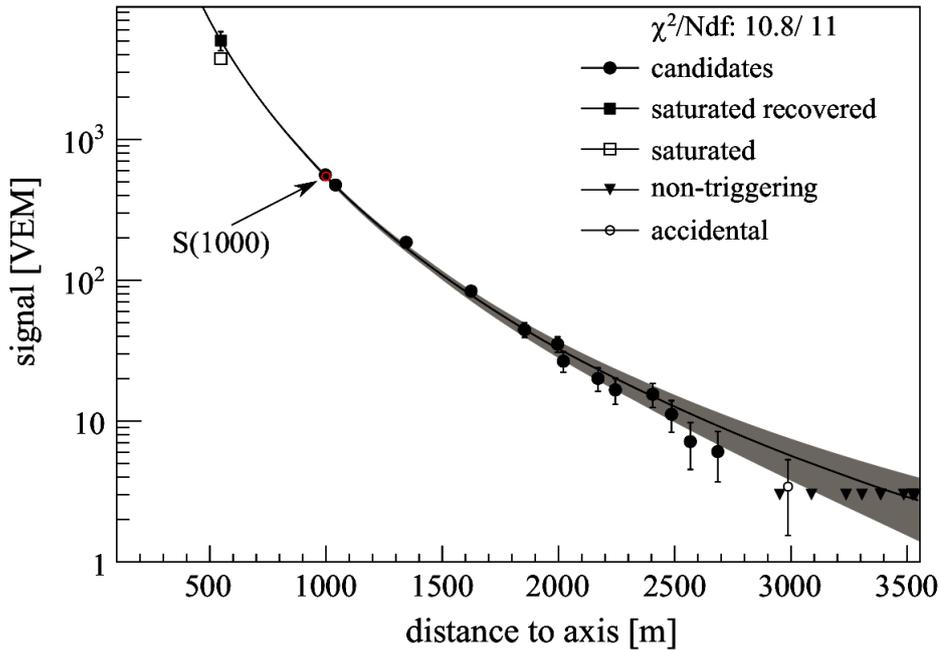
Figure 2.3: A Lateral Distribution Function from the Pierre Auger Observatory water-Cherenkov surface detectors. Each signal is calibrated with respect to the amount of signal a vertical, central through-going muon (VEM) will create. The function highlights the dependence of signal size on distance from the shower core. From [19].

more particles are detected at distances closer to the shower core, which is the position at ground where the shower axis intercepts. Large scale arrays of surface detectors are able to produce a distribution of signal against distance from shower core, known as the lateral distribution function (LDF) (see Figure 2.3). By having different LDFs from different types of detectors, they can be used to describe an air showers particle profile and will be crucial in mass composition analysis.

2.3.1 Surface Detectors

Surface detectors are typically arranged in a ground array network. Ground arrays typically consist of uniformly spaced detectors which may cover significantly several km^2. Larger arrays allow the measurement of larger numbers of particles over reasonable periods of time. extensive air shower (EAS) candidates are detected through timing coincidences between triggering surface detectors and its neighbouring detectors. The density of the measured signal at ground level and the difference in measured triggered times are used to infer
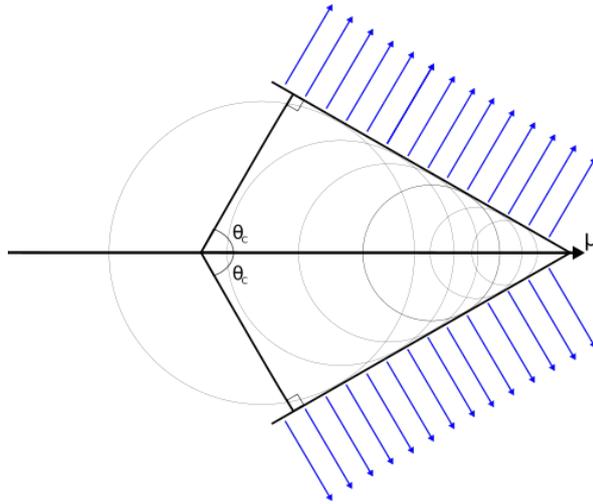
Figure 2.4: Blue lines represent Cherenkov light produced by a relativistic charged particle with a θ_c Cherenkov angle. The charged particle in this case is a muon.

the energy and arrival direction of the primary cosmic ray [20]. Water-Cherenkov detectors and Scintillator detectors are the two types surface detectors used today. Water-Cherenkov detectors consist of tanks which are filled with purified water, which provides a dielectric medium for charged particles. Water-cherenkov detectors typically measure signal through photomultiplier tubes (PMTs) measurements of Cherenkov light photons.

Cherenkov light is a type of electromagnetic radiation emitted when a relativistic charged particle travels through a dielectric medium (such as water). A particle will emit Cherenkov light if it travels through this medium faster than the speed at which light would travel ($v_c = \frac{c}{n}$ where n is the refractive index of the medium) in the medium. Cherenkov light is from the constructive interference of wavelets formed by dipole radiation. The dipole radiation is triggered from polarisation of the medium by the charged particle. If the particle is not travelling faster than light, this dipole radiation interferes destructively and Cherenkov light will not be produced [21]. For the ultra relativistic limit, where $E \gg m_0 c^2$, with m_0 representing the rest mass; the Cherenkov emission angle θ_c (Figure [2.4]) can be found by solving:

$$\cos(\theta_c) = \frac{1}{n} \tag{2.14}$$

where n is the refractive index of the medium that the relativistic charged particle is travelling through. The number of Cherenkov photons emitted per unit length for a particle above the

threshold can be calculated for a wavelength interval λ_1 to λ_2 by using

$$\frac{dN}{dL} = 2\pi\alpha \sin^2(\theta_c)(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}) \tag{2.15}$$

where θ_c is the Cherenkov emission angle and α is the fine structure constant, $\alpha = \frac{1}{137}$.

Scintillator detectors use scintillaing material to convert a fraction of ionisation losses from particles traversing through the material into visible light. The emitted visible light can then be observed by PMTs and converted into electrical pulses. Surface detector arrays are able to be operated with a 100% duty cycle, which is far superior to the duty cycle of fluorescence detectors. Surface detectors, however, are unable to observe the direct development of an air shower and must infer such information from reconstruction parameters and interaction models.

2.3.2 Fluorescence Detectors

The fluorescence technique employs specifically designed detectors which measure nitrogen emission induced by an EAS. Charged particles created within an EAS excite atmospheric nitrogen molecules which upon de-excitation, isotropically emit fluorescence light with wavelengths within the 300-430 nm range [22]. The nitrogen fluorescence spectrum is shown in Figure 2.5. The fluorescence yield produced for a given air shower is proportional to the energy deposited in the atmosphere through ionisation. This allows for a calorimetric energy of an EAS and hence the energy of the primary particle. Fluorescence detectors require cloudless, moonless nights for effective operation, which leaves the duty cycle at around 15%. Fluorescence detectors image the atmospheric nitrogen emission using a camera which is typically constructed from a multitude of pixels, with each pixel viewing a separate region of the sky. The pixels are typically PMTs arranged in a grid, which measure the fluorescence light as it appears like a small spot of light moving across the camera face in conjunction with the shower evolution.
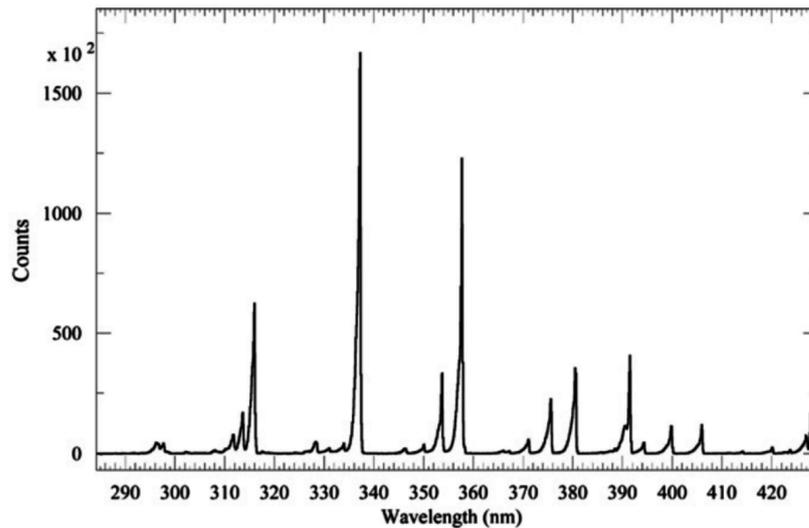
Figure 2.5: The fluorescence spectrum in dry air at 800 hPa [22].

2.4 Past and Current Experiments

The studies in this thesis are primarily based on the the Pierre Auger Observatory (Auger), and to a lesser degree, the Telescope Array Experiment (TA). Auger is discussed in detail in Chapter 3. This section will briefly discuss several other notable cosmic ray experiments which utilised surface detectors, fluorescence detectors or hybrid type detectors.

2.4.1 Volcano Ranch

The first giant shower array detector was constructed at Volcano Ranch, near Albuquerque, New Mexico. It was operated between 1959 and 1963 by a research group led by John Linsley and Bruno Rossi from the Massachusetts Institute of Technology [23]. In the first years, the Volcano Ranch array consisted of 19 3.3m^2 plastic scintillation detectors arranged on a triangular grid with spacing of 442m, spanning a total area of 2km^2. The array spacing was doubled in 1960, to cover an area of 8km^2 [24], as shown in Figure 2.6.

The data recorded by the Volcano Ranch array yielded the first measurements of the energy spectrum of cosmic rays above 10^{18}, giving the earliest evidence of a flattening in the spectrum in that region [25]. Volcano Ranch data was also used for early anisotropy studies, suggesting that ultra high energy cosmic ray (UHECR)s did not appear to arrive from preferred regions of the sky. Volcano Ranch's most energy event recorded was assigned an
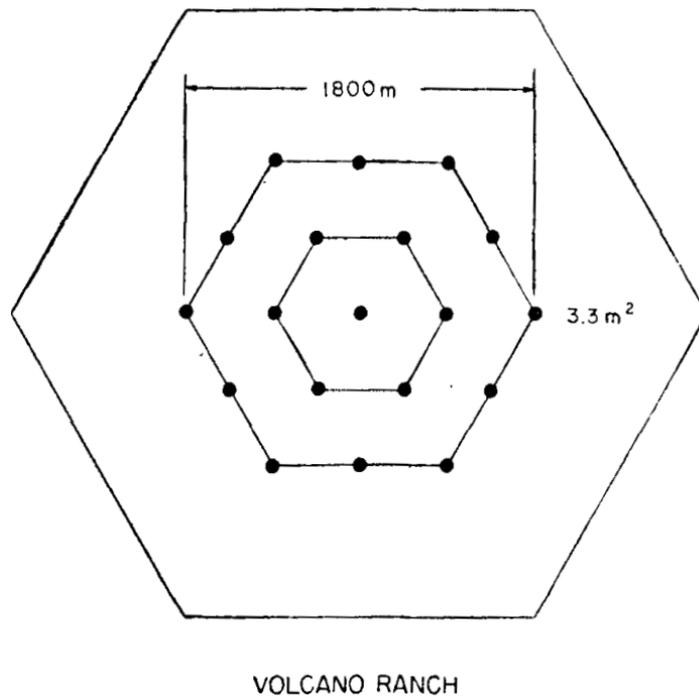
VOLCANO RANCH

Figure 2.6: The layout of the triangular Volcano Ranch array, showing the original position of the 19 scintillation detectors (black dots) and the size of the array (outer hexagon) after expansion in 1960 [24].

energy of 1.4×10^{20} eV, and to this day, it still remains one of the highest energy events ever detected [26].

2.4.2 Haverah Park

Haverah Park is an extensive air shower array that was built as a collaborative experiment between universities of Durham, Leeds, Nottingham and London (Imperial College). Haverah Park operated from 1967 to 1987, the array consisted of water-Cherenkov detectors deployed in different locations over a total area of 12km^2 with irregular spacing due to restricted land access. The final configuration was a central, four detector array with 500m spacing, with an additional 6 sub arrays approximately 2km from the centre [27], as shown in Figure 2.7. The Pierre Auger Observatory later adopted similar technology for its ground array due to the successful long term operation that the Haverah Park array demonstrated with water-Cherenkov detectors.
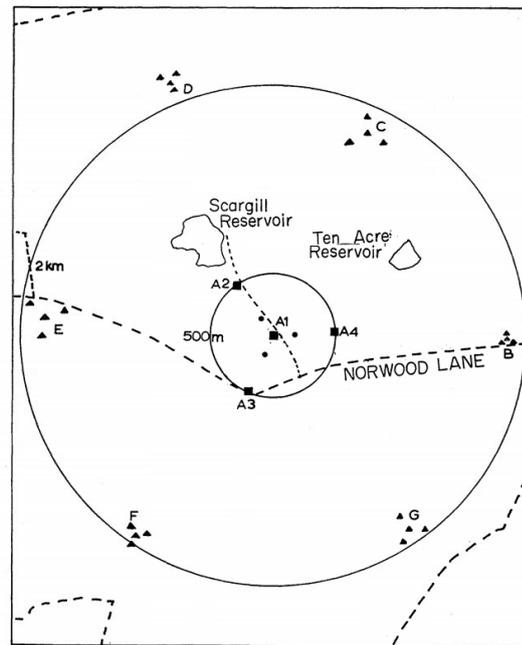
Fig. 1.   The Haverah Park EAS Array.
Detector areas: 150m array   ● 9m², 500m array   ■ 34 m², 2 km array   ▲ 13.5 m².

Figure 2.7: From [28]

2.4.3   Sydney University Giant Air-shower Recorder (SUGAR)

The Sydney University Giant Air-shower Recorder (SUGAR) was the first giant air shower array that was constructed in the Souther Hemisphere. SUGAR was built near the town of Narrabri in New South Wales, Australia where it was was operated by a group from the University of Sydney from 1968 to 1979. The array consisted of 54 pairs of 6m^2 liquid scintillation detectors deployed over an area of 60km ^{2}. Each detector was buried 1.7m underground, to gain a preferential detection to the muonic component of air showers rather than the electromagnetic component. Unfortunately, the distance between each detector proved too great (1.61km), and, even in the largest energy events, the number of stations that measured densities was small [29]. SUGAR's main contributions were in arrival detection studies, due to the data forming a unique set for studies in the Southern Hemisphere.

2.4.4   Yakutsk

The Yakutsk array is located in Oktyomsty near Yakutsk, Russia and has been used to study cosmic rays since 1970. The array has undergone several reconfigurations, with the
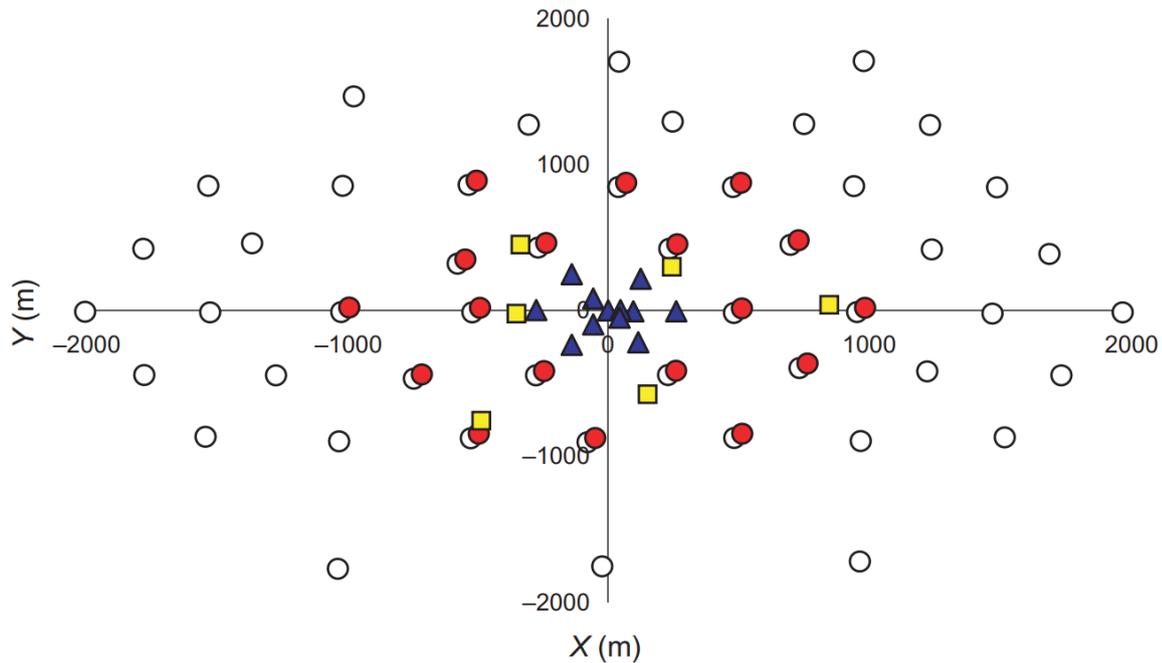
Figure 2.8: The detector configuration of the Yakutsk array. Scintillators are shown as the open black circles, Cherenkov light detectors of C_1 ($\sim$500m spacing) and C_2 ($\sim$50-200m spacing) subsets, are shown as red circles and blue triangles, respectively. Muon detectors are shown as yellow squares [31].

current design consisting of 58 scintillator stations, 48 Cherenkov light detectors, 4 underground muon detectors and 6 radio detectors [30]. At its largest, the array covered $\sim$17km^2 (1990) but has since been reduced to $\sim$10km^2, the detector array configuration can be seen in Figure 2.8. Yakutsk has contributed to the measurement of the energy spectrum of cosmic rays for energies ranging from 10^{15}-10^{20} eV. The array has successfully detected over 10^6 cosmic rays with primary energy above 100 TeV, with the highest energy events exceeding 10^{20} eV [31], and is currently undergoing modernisation to improve measurement accuracy.

2.4.5 Fly's Eye

The Fly's Eye experiment operated between 1981 and 1992 [32], it was located at U.S. Army's Dugway Proving Grounds, in Western Utah, USA. The site offered a clear atmosphere with minimal light pollution, providing an ideal environment for the detection of fluorescence emission. The Fly's Eye detectors were based on the same design as the three prototypes used at Volcano Ranch in 1976. The first detector (Fly's Eye I) began operation in

Figure 2.9: Fly's Eye I detector with mirrors mounted in steel drums with PMT arrays [35]

1981. Fly's Eye I consisted of 67 spherical mirrors, each with a diameter of 1.6m, providing full-sky coverage as shown in Figure 2.9. Sky imaging was performed by a collection of 12-14 PMTs placed at the focus of each mirror, for a total of 880 PMTs across all mirrors. In 1986 a second detector (Fly's Eye II) of similar design was constructed 3.4km from Fly's Eye I. Fly's Eye II contained 36 mirrors, for a total of 464 PMTs; it was capable of operating as a standalone detector, or in combination with Fly's Eye I. Fly's eye II viewed half of the sky in the direction of Fly's Eye I, allowing its data to be used complementary to Fly's Eye I, providing more accurate stereoscopic analyses of air showers [32]. In October 1991, the highest energy cosmic ray to date was measured by Fly's Eye at an observed energy of 3×10^{20} eV [33]. The Fly's Eye experiment was considered to be the first detector to use air fluorescence techniques to successfully measure air showers caused by cosmic rays, demonstrating merits in areas of energy and composition resolution [34]. Notable experiments, such as the Pierre Auger Observatory and the Telescope Array have incorporated the techniques as part of their hybrid detector designs due to the success of Fly's Eye.
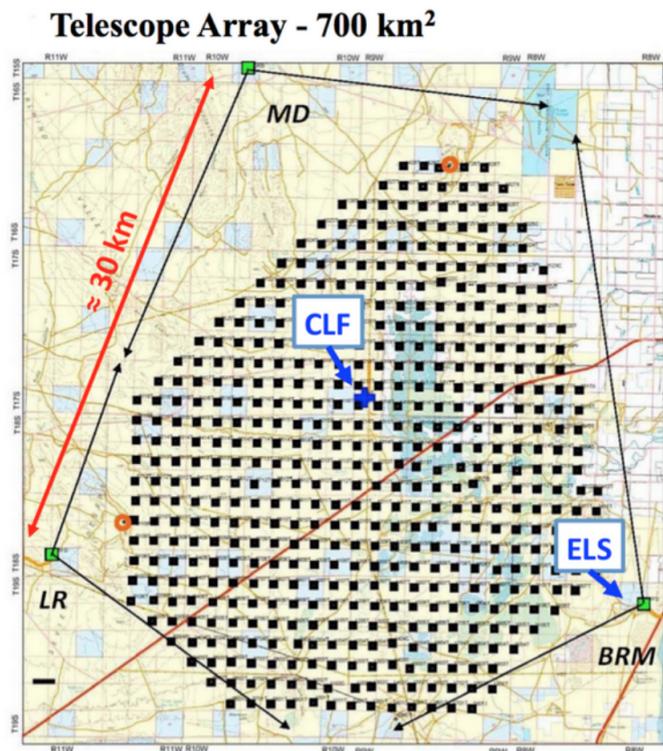
Figure 2.10: Telescope Array Experiment. Surface detectors are represents by dots and the fluorescence detectors are labelled on the perimeter of the site. From [37]

.

2.4.6   The Telescope Array Experiment (TA)

The Telescope Array (TA) project is located in the northern hemisphere in Millard County, Utah (USA), and has been in full operation since May 2008. TA has 507 surface detectors, referred to as "particle counters". The particle counters are placed in a square grid, with a distance of 1.2 km between each particle counter. The total area of the TA surface detector array is around 700 km^2 [36]. Figure. 2.10 shows the layout of the TA surface detector array. TA also has three fluorescence detector stations. Each of the TA particle counters consist of two layers of 3m^2, 1.2 cm thick plastic scintillators with a stainless steel sheet between them, on the top and on the base. Scintillation light is detected when a particle traverses through the scintillators. The amount of energy deposited is proportional to the amount of scintillation light detected. More specific detail on how the TA particle counters work is given in Section 2.4.6.1.
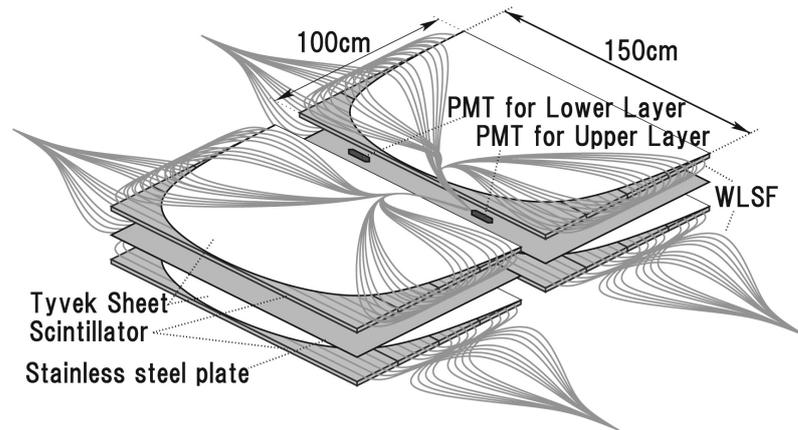
Figure 2.11: Inside of a scintillator box, The WLS fibres are laid along grooves to collect and transmit scintillation light. From [37].

The TA Low-energy Extension (TALE) provides an additional 103 surface detector (SD) stations with variety of spacings, as well as an addiotnal 10 fluorescence detector (FD) telescopes located around the TA MD station. The additional FD telescopes view elevations of $31°$ to $59°$. TALE lowers the energy threshold down to $10^{16.5}$ eV with acceptable resolution and an event rate of around 5000 hybrid events per year.

2.4.6.1   Telescope Array Surface Detector

The TA surface detector or particle counter consists of two segments, 1.5m by 1.0m. Each segment contains a 1.5mm thick stainless steel cover, two 1.2cm thick plastic scintillators with a 1mm thick stainless steel plate and a 1.2mm thick stainless steel base. There are grooves in the scintillator material that contain 104 wavelength shifting fibres. The scintillation light is collected in these fibres and then passed through to a PMT. Each segment is then wrapped with two layers of tyvek, a lightweight and durable material often used for protection.
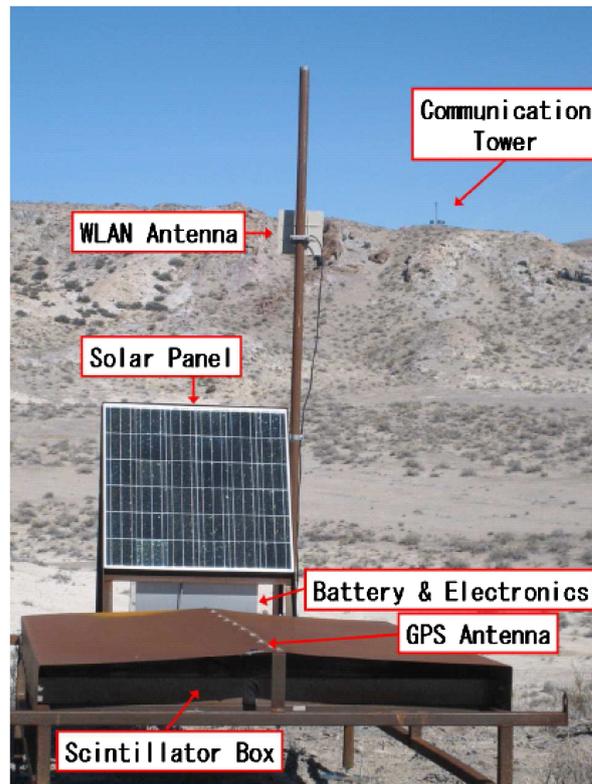
Figure 2.12: One of the TA surface detectors in the field, the scintillator box is underneath the roof. From [37].

Chapter 3

Pierre Auger Observatory (Auger)

The Pierre Auger Observatory is the world's largest observatory dedicated to the measurement of ultra high energy cosmic rays (UHECR), located at an elevation of approximately 1400 m near Malargüe, Mendoza Province, Argentina. The observatory began operation in 2004 while construction was still ongoing, and was later completed in 2008. Auger uses a hybrid detection mechanism, utilising two main types of detectors, the surface detector (SD) and fluorescence detector (FD). The SD consists of an array of water-Cherenkov detectors (WCDs) which measure energy deposited by particles from extensive air showers (EASs) at ground[38]. The FD consists of 27 telescopes deployed at four different locations overlooking the SD. The FD observes the longitudinal development of EAS via their fluorescence emission. Figure 3.1 shows the layout of the surface detector array. A large array allows Auger to collect large amounts of data, however, even with an array of this size, the highest energy cosmic rays are still extremely rare events due to the nature of the cosmic ray spectrum.

31
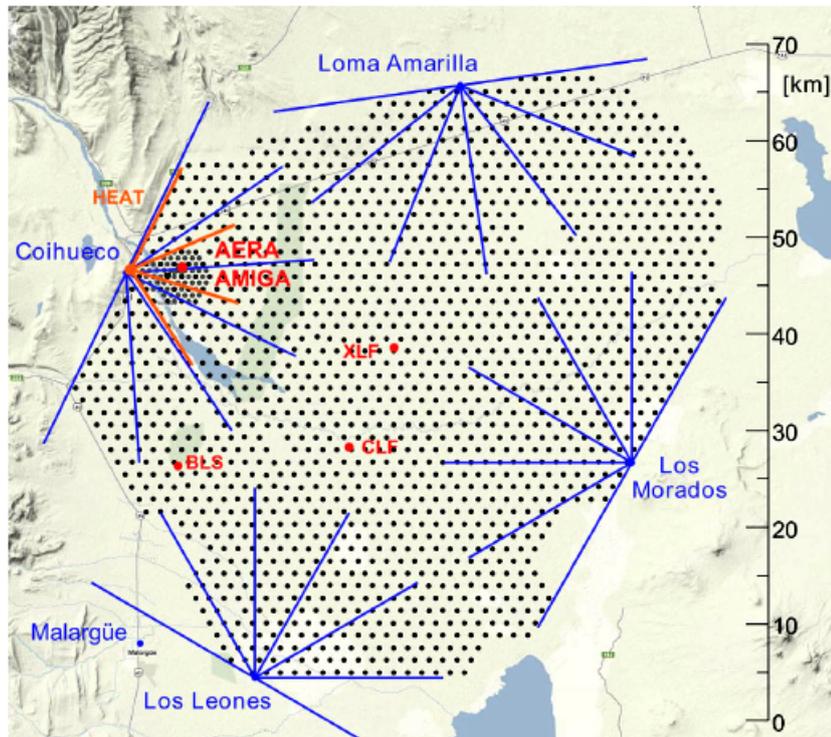
Figure 3.1: Layout of the Pierre Auger Observatory. Each dot represents one of the 1660 SD stations. Lines on the field represent the field of view of each fluorescence telescope. The orange lines represent the high-elevation telescopes for observation of showers induced by lower energy primaries. The denser array with station spacing of 750 m is also depicted. The two laser facilities (XLF and CLF) at the observatory used for calibration of the fluorescence telescopes are shown as well. From [39].

3.1   Water-Cherenkov Detector (WCD)

The Pierre Auger Observatory surface detector array covers a total area of around 3000 km^2. The largest array is known as SD-1500. SD-1500 is comprised of 1660 self-contained, WCDs arranged hexagonally such that each WCD is equidistant from six neighbouring stations with a grid spacing of 1.5 km [40]. The array is fully efficient detecting cosmic rays with zenith angles of up to 60° for energies greater than 3×10^{18} eV [41]. Each WCD and its associated electronics, communication systems and any upgrades is referred to as an SD "station", one of which is depicted in Figure 3.2. In order to observe EAS induced by cosmic rays with lower energy primaries, the SD-1500 array was extended by including a nested array with 750 m spacing. This nested array is known as SD-750, or the "infill" array. SD-750

Figure 3.2: Picture of a WCD. (source?)

lowers the energy threshold to 3×10^{17} eV for $\theta < 55°$. The SD-750 consists of 71 stations, covering a total area of $\sim$27 km^2, for which data acquisition started in September 2007, and deployment finished in September 2012.

Later, in Januarry 2013, an even smaller nested array with 433 m spacing was deployed. This small array is known as SD-433, and uses the SD-750 station Cathy Turner (ID 1764) as the central station. SD-433 is primarily used to study the second knee of the cosmic ray spectrum, by measuring EAS of energies below 10^{17} eV [42].

3.1.1   Design

A single WCD consists of a 3.6 m diameter polyethylene cylindrical tank, with a diffuse, highly-reflective liner inside. A tank is filled with ultra-pure water up to a depth of 1.2 m, resulting in a total detection volume of slightly over 12000 L. Relativistic charged particles from the electromagnetic and muonic components of an EAS produce Cherenkov radiation as they traverse through this water volume, which is collected by three 9-inch photomultiplier tubes (PMTs). The PMTs are Photonis XP1805/D1 photomultiplier tubes, that are symmetrically distributed along the surface of the liner, 1.2 m from the centre of the tank. These PMTs are facing downwards through windows of clear polyethylene. Each PMT has two outputs,
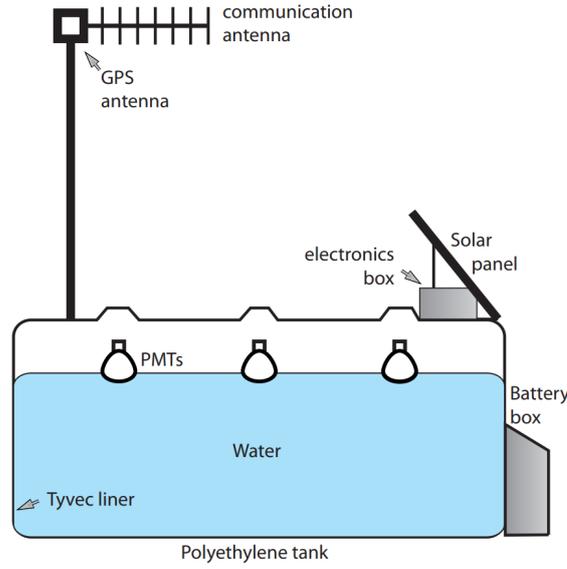
Figure 3.3: A schematic of the Auger WCD. From [43].

one on the anode (referred to as the low gain channel) and one on the dynode (referred to as the high gain channel). The charge gain between the dynode and anode has been chosen such that there is a factor of 32 between the two. This enables a higher dynamic range in order to measure signals close to the shower core. Each station is self-contained with a 10 W battery which is charged by a solar panel attached to the top of the station. A WCD schematic with its components is depicted in Figure 3.3. The electronics (referred to as the unified board (UB)) samples the analog pulses produced at the base of the PMTs with two flash analog to digital converters (FADCs) of 10 bits at a frequency of 40 MHz. If any trigger conditions are fulfilled, the digitised signals which are stored in cyclic memory are transferred to central data acquisition system (CDAS) for subsequent analysis as FADC traces. The size of the time bins in each trace is defined by the sampling rate of the electronics, namely

$$1 \text{ bin} = \frac{1}{40 \text{ MHz}} = 25 \text{ ns}, \tag{3.1}$$

such that the total duration of the trace is 768×25 ns $= 19.2$ μs. The amplitude of the signal is measured in terms of FADC counts, for a maximum of 1023 counts for 10 bit electronics. The data acquisition is governed by a hierarchical system of triggers implemented in hardware and software. The first two levels, T1 and T2 are conformed locally by each station in real-time. There are four different types of station triggers. The threshold (Th) trigger is a simple

trigger which requires at least 1 bin of a trace to be above a specific value relative to the peak amplitude of a vertical muon pulse (I_{VEM}), this is 1.75 I_{VEM} and 3.2 I_{VEM} for T1 and T2 level triggers, respectively. The time-over-threshold (ToT) trigger requires at least 13 bins within a 3 μs sliding window to be above a threshold of 0.2 I_{VEM}. The time-over-threshold de-convoluted (ToTd) trigger which uses a more complicated algorithm which applies a de-convolution to remove the tail of Cherenkov light pulses. After de-convolution, the standard ToT trigger algorithm is applied. Finally, the multiplicity of positive steps trigger (MoPS) which has its own independent algorithm, which accumulates the number of consecutive positive steps in the signal while also recording the amplitude of each step. The trigger passes when there is at least 4 instances where there is sufficient consecutive amplitude increases to distinguish signal from background noise. The third level, T3 is formed by combinations of T2 station triggers at CDAS, while the fourth and fifth levels, T4 and T5 are processed offline to select EAS events, provided that there are at least three station triggers [41].

3.1.2 Reconstruction at the Station Level

Calibration

Station signals are calibrated to the measurement of the peak position of charge histograms produced by vertical through-going atmospheric background muons, called the vertical equivalent muon (VEM). The aim of the calibration procedure is to identify how much integrated charge is produced by a 1 VEM signal. Each station is continuously monitoring the integrated charge and amplitude of background signals through online station electronics. The online process is required as part of maintaining a consistent trigger definition, and hence the constant energy threshold of the array. Stations store integrated charges over a duration of 60 seconds. Due to the high background rate of muons and other particles, there are plenty of particles incident on individual stations such that 60 seconds of data produces well defined distributions of charge. Charge produced by background muons correspond to a distinct peak in the charge distributions. This peak position is then used to convert the electronic analogue-to-digital converter (ADC) signal into VEM. Individual stations cannot distinguish between vertical muons and omnidirectional muons,as such the Auger collaboration performs manual calibration procedures using additional detectors, such as pairs of scintillators or resistive plate chambers (RPCs), to detect vertical coincidence.
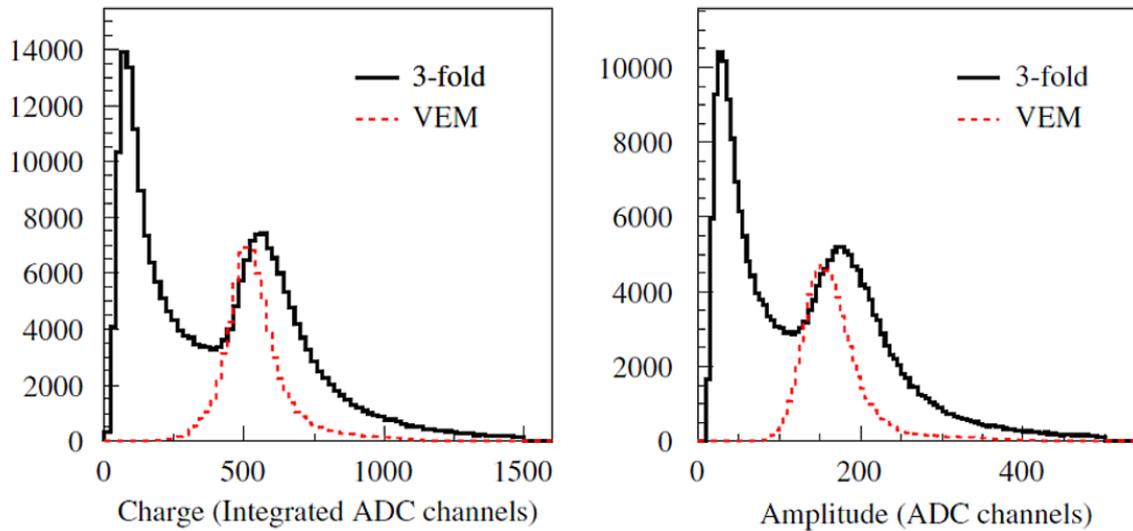
Figure 3.4: 3-fold charge histogram (left) and 3-fold pulse height (right) histograms produced by the online SD station electronics (3-fold distributions represent the sum of the 3 PMT distributions). The black distributions represent the distributions from all omnidirectional particles. The first peak of the black distribution corresponds to signal produced by low energy particles. Whereas, the second peak corresponds to higher energy through-going particles, which are predominately muons. The dashed red distributions correspond to signals produced purely by vertical through-going muons, as determined through external means. From [44]

Since the VEM is defined clearly as vertical, through-going muons, the manual calibration procedure will calculate a value from measuring the ratio between peaks from vertical through-going muon charge histograms, and omnidirectional background particles (not exclusively muons) charge histograms [44]. Figure 3.4 shows an example when summing the signals from each of the WCD PMTs (3-fold). The 3-fold criteria exaggerates the differences between the peak positions of each histogram, and it should be noted that calibration is usually completed by considering each PMT individually. The exaggeration is due to the spacing of the PMTs and their differing responses to omnidirectional background particles. A more descriptive explanation of the calibration procedures is given in Chapter 5.

Uncertainties in the Measurement of Signal

The statistical uncertainties on the integrated signal and on the start time of the signal are obtained from studies using configurations with doublet stations. Doublet stations are

pairs of stations placed only ∼11 m apart and thus sampling the same area of a shower. The uncertainty in the start time strongly influences the angular resolution of the reconstruction of the shower geometry. The uncertainty is primarily comprised of the fluctuations of the arrival times of the first particles, as well as the time resolution of the GPS time-tagging system (∼10 ns), and the 40 MHz FADC sampling (7 ns). These variables depend on a convolution between the thickness of the shower front and the density of particles therein, as well as the cross-section of the detector [45]. The effects can be described with a general time variance model for which the uncertainty of the trace start time t_{start} is estimated as

$$\sigma^2_{t_{start}} = a^2 \left(\frac{2t_{50}}{n} \right)^2 \frac{n-1}{n+1} + b^2, \tag{3.2}$$

where t_{50} is the length of the time interval which contains the first 50% of the total signal. n is the effective number of particles determined from the total signal S. Here, $n = S/l(\theta)$ where $l(\theta)$ is the mean track-length (relative to the vertical) of through-going particles crossing the detector at zenith angle θ. Coefficients a and b are parameterised as functions of the zenith angle of the shower and were determined using doublet stations.

The signal uncertainty of non-saturated stations is based on a Poisson-like parameterisation,

$$\sigma^2_S = f_s(\theta)S \quad \text{where} \quad f_s(\theta) = 0.865(1 + 0.593(\sec\theta - \sec 35°)) \tag{3.3}$$

The procedure for obtaining the coefficients 0.865 and 0.593 is outlined in more detail in Chapter 6, where the methodology which was originally used for obtaining the WCD signal uncertainty model is applied to the scintillator surface detectors (SSDs) to obtain the SSD signal uncertainty model.

Signal Asymmetry

Signal asymmetry refers to an uneven distribution of signals recorded by the detectors in response to a cosmic ray-induced air shower. In practise, various factors can lead to signal asymmetry. The Pierre Auger Observatory's SD array, which consists of a grid of detectors placed on the ground, is one of the largest contributors to signal asymmetry. Depending on the geometry of the detector array and the position of the shower core (the point where the shower axis intercepts ground), signals may be received unevenly across the detector grid. Signal asymmetry is much more noticeable for inclined air showers. Particles approaching
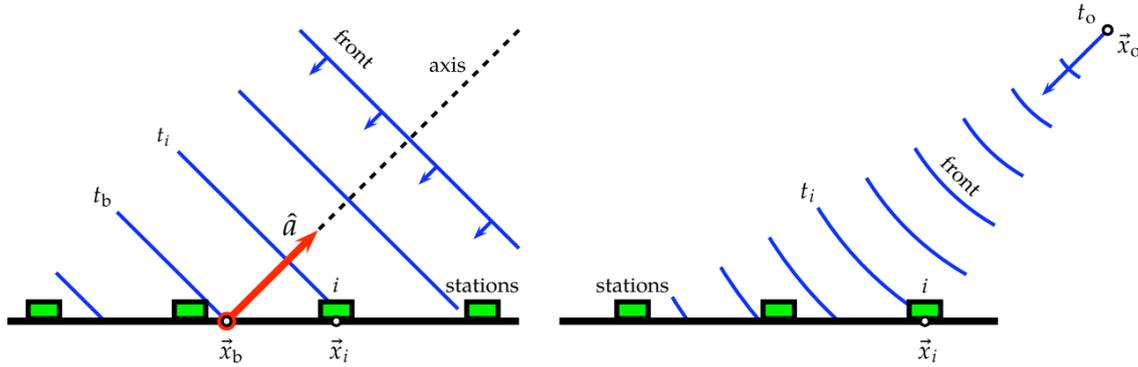
Figure 3.5: Schematic representations for two types of shower front approximations, a plane front approximation (left) and a spherical shower front (right). The shower front reaches station i at position $\vec{x}_i$ at time t_i [46].

ground from underneath the shower axis are considered early particles, and travel through less atmosphere than particles approaching ground from above the shower axis. In most cases, this results in less signal being measured on the late side of an air shower, as more particles are attenuated in the atmosphere. It is important to note that the reconstruction process of the air shower geometry implicitly accounts for this signal asymmetry.

Shower Geometry

The first rough estimate of the axis $\hat{a}$ of the shower and the impact position $\hat{x}_c$ of the core on the ground is obtained from the reconstruction seed triangle consisting of three stations. That is, a station with its two nearest neighbours in a non-aligned configuration, where all stations must have passed a station level trigger. Under the assumption that particles in the shower front move in a plane perpendicular to the shower axis (See Figure 3.5 *left*) with the same speed as the core of the shower (assumed to be the speed of light), the first estimate of the time t_i that the shower front reaches the station i at position $\vec{x}_i$ may be inferred through a projection onto the shower axis as

$$- \hat{a}(\vec{x}_i - \vec{x}_b) = c(t_i - t_b) \tag{3.4}$$

where $\vec{x}_b$ defines the signal-weighted centre of mass of stations in an event, with weighted bary-time t_b. Figure 3.6 shows a graphical representation of the magnitude of signals in the SD-1500 from the $\overline{\text{Off}}\underline{\text{line}}$ event display. The colour gradient from yellow to red describes
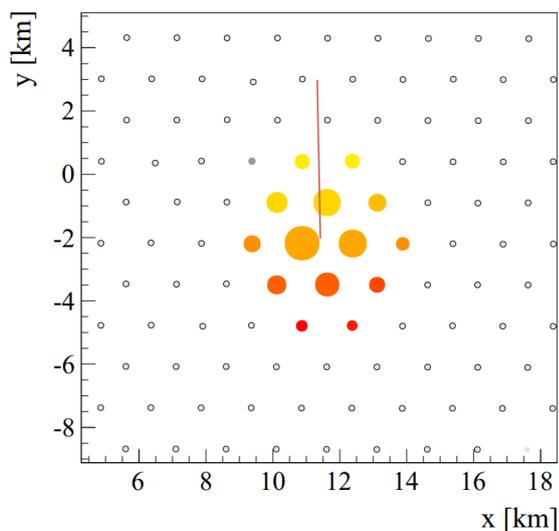
Figure 3.6: A graphical representation of a group of stations in the SD-1500 triggered by a shower. The size of the markers is proportional to the logarithm of the signal. The colours represent the arrival time of the shower front from early (yellow) to late (red) [47].

the trigger timing, representing early to late, alongside the red line showing the fitted shower axis. The first estimate of the shower axis can be determined by minimising the sum of squares of differences in predicted and measured start times,

$$\chi^2 = \sum_i \frac{(c(t_i - t_b) + \hat{a} \cdot (\vec{x}_i - \vec{x}_b))^2}{\sigma_i^2} \tag{3.5}$$

where σ_i is the uncertainty in the start time of the signal in station i. Afterwards, the model of the shower front is extended to a spherical shower front (See Figure 3.5 *right*). The shower development is approximated as starting at time t_0 from a virtual point of origin $\vec{x}_0$ and propagating towards the ground in the shape of a sphere, concentrically inflating with the speed of light [46]. The arrival time of such a shower front can be expressed as

$$c(t_i - t_0) = |\vec{x}_0 - \vec{x}_i|. \tag{3.6}$$

The radius of curvature of the shower front R_c is determined during reconstruction by minimising

$$\chi^2 = \sum_i \frac{(c(t_i - t_0) + |R_c\hat{a} - \vec{x}_i|)^2}{\sigma_i^2} \tag{3.7}$$

The radius of curvature of the shower front can only be fit as a free parameter for events which have a station multiplicity of at least five. Events which have fewer than five stations do not have sufficient information to reconstruct the curvature. Low multiplicity events use a parameterised model based on events with high station multiplicity. The shower axis can then be calculated with the assistance of the shower core position $\vec{x}_c$ determined from the lateral distribution function (LDF) reconstruction,

$$\hat{a} = \frac{\vec{x}_0 - \vec{x}_c}{|\vec{x}_0 - \vec{x}_c|} \tag{3.8}$$

The calculation of the shower impact position $\vec{x}_c$ using the LDF can be combined with the timing information used to calculate the direction of the shower axis $\hat{a}$ to form an iterative or global fit.

Shower size

The shower size is estimated through a maximum likelihood fit of the parameters of the lateral distribution function, $S(r)$, to the signals of triggered stations. The LDF is a function of the perpendicular distance to the shower axis, r. The functional form of the LDF is written as

$$S(r) = S(r_{\text{opt}}) f_{\text{LDF}}(r), \tag{3.9}$$

where $S(r_{\text{opt}})$ is the shower-size estimator. $f_{\text{LDF}}(r)$ could be any specified functional form that describes the data well, such as a power law function, Nishimura-Kamata-Greisen (NKG) function, or the modified NKG function which is most commonly used at the Pierre Auger Observatory (Auger). An example of an LDF fit is shown in Figure 3.7. The optimal value for distance r_{opt} is chosen so that the variability of the shower-size estimator $S(r_{\text{opt}})$ with respect to the shower-to-shower fluctuations is minimised. The shower-size estimator is dependent on mass composition, as showers induced by identical primaries can be sampled at different stages of development due to shower-to-shower fluctuations. The WCD LDF maximum likelihood fitting procedure is additionally constrained by the absence of triggers in stations which would have measured little, or no signal.

The reconstruction procedure uses a slightly modified NKG function [48] of the form

$$f_{\text{NKG}}(r) = \left(\frac{r}{r_{\text{opt}}}\right)^{\beta} \left(\frac{r + r_{\text{s}}}{r_{\text{opt}} + r_{\text{s}}}\right)^{\beta + \gamma}, \tag{3.10}$$
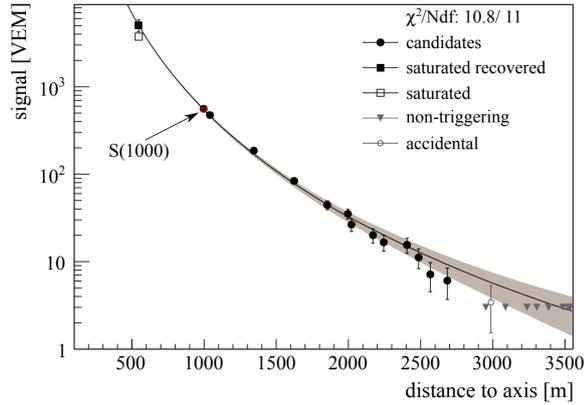
Figure 3.7: An example of a Lateral Distribution Function (LDF) fit to the measured signal (VEM) of water-Cherenkov detectors as a function of distance to the shower axis. (SOURCE?)

with fixed $r_s = 700$ m. The optimum distance r_{opt} is dependent on detector spacing. Through Monte-Carlo (MC) simulations, r_{opt} has been found to be 1000 m for SD-1500 and 450 m for SD-750. β and γ are energy and zenith angle dependent slope parameters. β and γ have been parameterised as a function of zenith angle and $S(r_{opt})$. The parameterisations are used for events with insufficient station multiplicity. To freely fit slope parameter β, events must have a station multiplicity and spatial distribution providing a sufficient lever arm. There must be at least 2, 3, or more stations within a radius of $400 < r/\text{m} < 1600$ and at least two of the stations must be separated by at least 900 m (with station multiplicity of 2), 800 m (station multiplicity of 3) or 700 m (station multiplicity of more than 3). The lever-arm criteria ensures that the LDF slope is sufficiently constrained around the optimum distance r_{opt}. γ must fulfil a similar criteria, but within a radius range of $1000 < r/\text{m} < 2000$. The fitting method of the LDF is described in much more detail in Chapter 7, where the SSD LDF is parameterised based on MC simulations using a method adopted from the WCD procedure.

Energy Reconstruction

After the LDF fitting procedure, an estimate of the shower energy can be obtained by using $S(r_{opt})$. For data within the SD-1500, this corresponds to signals measured at a distance of $r_{opt} = 1000$ m from the shower axis. For a fixed primary energy, $S(1000)$ has a dependence on zenith angle due to the increasing amount of atmospheric attenuation as zenith angle increases. The dependence of $S(r_{opt)}$ on zenith can be removed by using a method known

as the Constant Intensity Cut (CIC) method. The idea behind the CIC method is that, due to the isotropy of the primary cosmic ray flux, showers generated by primary particles of the same energy and composition will arrive at the detector at the same rate. The CIC method eliminates the dependence by converting $S(1000)$ to a quantity known as S_{38}, which is the energy estimator if the shower had arrived at a zenith angle of 38°; 38° is the median of the zenith angle distribution of incident air showers for $\theta < 60°$. Similarly, the SD-750 energy estimator is converted from $S(450)$ to S_{35} (35° is the median angle for $\theta < 55°$) using the CIC method. The attenuation curve $f_{\text{CIC}}(\theta)$ is fitted with a third-order polynomial,

$$f_{\text{CIC}}(\theta) = ax^3 + bx^2 + cx + 1, \tag{3.11}$$

with constants $a = -1.30 \pm 0.45$, $b = -1.68 \pm 0.01$, $c = 0.980 \pm 0.004$, and $x = \cos^2(\theta) - \cos^2(38°)$. S_{38} can be calculated, with

$$S_{38} = \frac{S(1000)}{f_{\text{CIC}}(\theta)}, \tag{3.12}$$

with zenith angle θ, and energy estimator S_{38}. The energy estimator, S_{38}, is converted into energy through a calibration with the E_{FD}. A strong correlation between S_{38} and E_{FD} is shown in Figure 3.8. The correlation is fit with the following function

$$E_{\text{FD}} = A \left(\frac{S_{38}}{\text{VEM}} \right)^B, \tag{3.13}$$

where $A = (1.86 \pm 0.03) \times 10^{17}$ eV and $B = 1.031 \pm 0.004$ are the best fit parameters obtained from high quality hybrid events which independently trigger both the SD and FD [4].
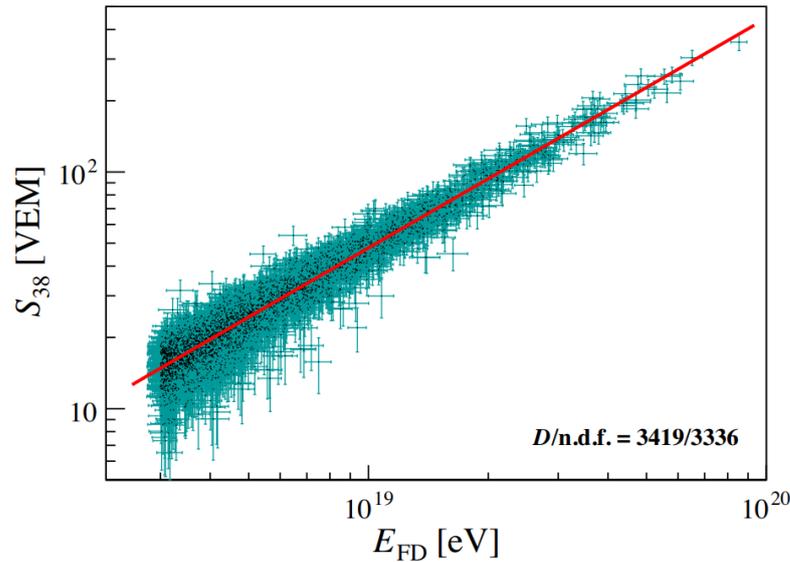
Figure 3.8: Correlation between the energy estimator S_{38} and the FD energy. The red line is the result from fitting Equation 3.13. From [4].

3.2   Fluorescence Detector

Auger has a total of 27 fluorescence telescopes which are located at four sites on the perimeter of the SD array. These sites are known as Los Leones, Los Morados, Loma Amarilla and Coiheuco. The four sites are located along the southern, eastern, northern and western boundaries of the observatory, respectively. A picture of the FD site Los Leones is shown in Figure 3.9 [47]. Each of the Fluorescence detectors are located approximately 1400 m above sea level, with the exception of Coihueco and HEAT, which are located in the Andes foothills at approximately 1700 m above sea level. Each telescope has a field of view of 28.1° in elevation and 30° in azimuth. The arrangement allows for an FD triggering efficiency of 100% for energies above 10^{19} eV over the entire SD array. Photons produced via fluorescence emission by an EAS as it traverses the atmosphere enter telescope bays and are reflected by a 13 m^2 segmented mirror into a telescope camera. Each camera consists of a 22 row by 20 column grid of PMTs. A schematic of one telescope bay is shown in Figure 3.10.

Each of the four FD sites is equipped with a lidar, infrared cloud cameras, and a weather station to monitor the atmosphere. Atmospheric monitoring is important as the FD uses the atmosphere as a calorimeter. There are also two laser facilities, the Central Laser Facility

Figure 3.9: FD building at Los Leones during the day. (SOURCE?)
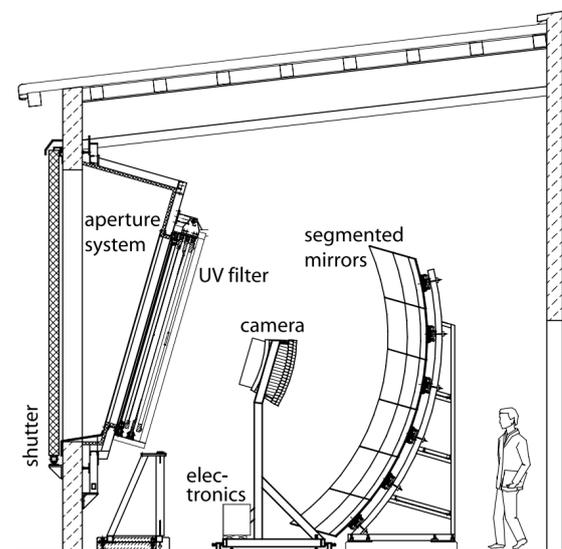
Figure 3.10: Schematic of a telescope bay. From [47].

(CLF) and the Extreme Laser Facility (XLF) located centrally in the SD array. The Global Data Assimilation System (GDAS) is used during event reconstruction to inform atmospheric conditions.

3.3 Pierre Auger Observatory Upgrade (AugerPrime)

The Pierre Auger Upgrade, also known as AugerPrime, is an ongoing upgrade to the current surface detector array [49]. AugerPrime will consist of new 4 m^2 scintillators placed on

top of most existing WCDs. These scintillators are referred to as scintillator surface detectors (SSD). AugerPrime will also consist of updated electronics with an extended dynamic range, an additional 'small' PMT, an array of buried muon detectors and an extended duty cycle for the fluorescence detectors. Scintillator construction is completed, and most of the SSDs are deployed in the field, but are not activated as they await upgraded unified boards (UUBs). During the wait for construction and delivery of UUBs, SSDs have been activated on top of a small subset of WCD stations that have original UB electronics; these stations replace one WCD PMT input with a SSD PMT input, leaving the station with 2 WCD PMTs and the SSD PMT. UUB electronics will have enough ports to operate all WCD, SSD and small PMTs simultaneously. This small subset forms an array of stations called the scintillator surface detector pre-production array (SSD PPA).

The scintillators on top of each WCD will provide a complementary measurement that will allow for the sampling of shower particles with two different detectors that have different responses to muons and electromagnetic particles. All surface detector stations will soon be upgraded with improved electronics. The improved electronics will handle both WCD and SSD signals. The new electronics also aims to improve data quality, trigger and processing capabilities, as well as calibration and monitoring of the stations themselves.

3.3.1   Scintillator Surface Detector (SSD)

The first prototypes of scintillator units in Auger were called the Auger Scintillator for Composition-II, or "ASCII". They were developed by the Auger Bariloche group, starting data acquisition in 2010. First prototypes were around 0.25m^2 in size and operated for approximately one year as a proof of concept. Afterwards, seven 2 m^2 scintillator detectors were deployed at the central hexagon of the SD-750. The results from this phase showed that the 2 m^2 scintillators were roughly half the signal of the WCD (in their respective units). The signal fluctuations are of the order $\sqrt{S}$, implying that an area of approximately 4 m^2 for the scintillator detector would yield similar accuracy in measurements of both detectors. The decision was then made for the design of the Scintillator Surface Detector (SSD), with an area of approximately 4 m^2 that are placed on top of each existing WCD.

In September 2016, twelve prototype stations were deployed in the engineering array (EA), which is situated in the SD-1500. The prototype stations are fully upgraded stations, containing an SSD and new UUB electronics. As of December 2021, approximately 1430 out

of 1660 stations were upgraded with an SSD, but were not upgraded with UUB electronics. The SSDs on these stations will not be activated until UUB electronics are installed.

The SSDs have a total surface area of 4 m^2 including the aluminium casing. Inside the casing there are two modules, each approximately 2 m^2 in size. Each module is filled with 24 plastic scintillator bars which are read out by wavelength-shifting (WLS) fibres that guide the light to a PMT. A rendering of an SSD module can be seen in Figure 3.11, and a rendering of an upgraded station with the SSD on top can be seen in 3.12. The bars, produced by Fermi National Accelerator Laboratory (FNAL) have dimensions of 160 cm long, 1 cm thick and 5 cm width. Each bar is made by extruding liquid polystyrene (Polystryrene Dow Styron 663 W), containing two WLS dopants, PPO (1%) and POPOP (0.3%) [50], shifting the wavelength of the photons produced in the scintillator from UV to the visible range of the electromagnetic spectrum. The spectrum emission of the scintillator bars produced by FNAL is shown in Figure 3.13. Each bar is covered with an outer layer of a 0.25 mm polystyrene mixed with 15% of titanium dioxide (TiO$_2$), which has diffusive reflective properties to reduce light losses and to protect the scintillating material from external sources.
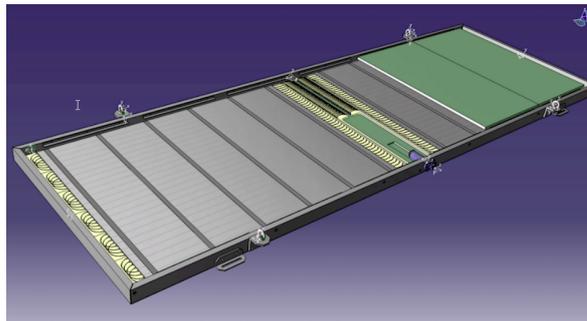
Figure 3.11: Rendering of an SSD module. Taken from the AugerPrime design report. From [51].

The optical fibres chosen for the SSD are the Y-11 (300) MSY from Kuraray with a 1 mm diameter. These fibres have a multi-cladding structure; photons are trapped inside the fibre at multiple surfaces with different refraction indices, resulting in a higher photon yield. As S-type fibres, they are more durable and flexible, resulting in a minimum bending radius of 100 mm as opposed to that of 200 mm for non S-type fibres. These fibres are also doped with wavelength shifting dye K-27, causing photon wavelengths below 490 nm (blue light) to be absorbed and then re-emitted at a peak of around 500 nm (green light) as shown in
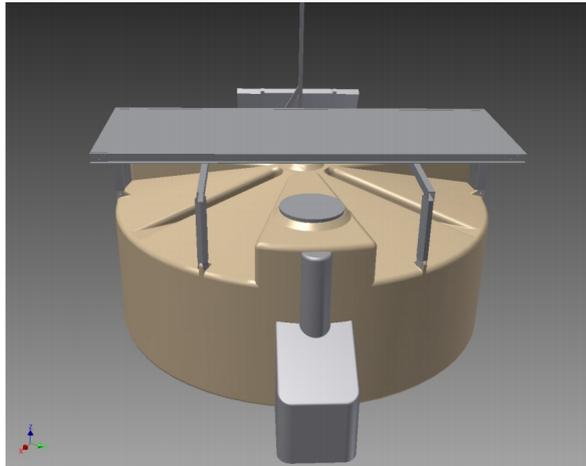
Figure 3.12: Rendering of an Upgrade station with the scintillator on top. Taken from the AugerPrime design report. From [51].
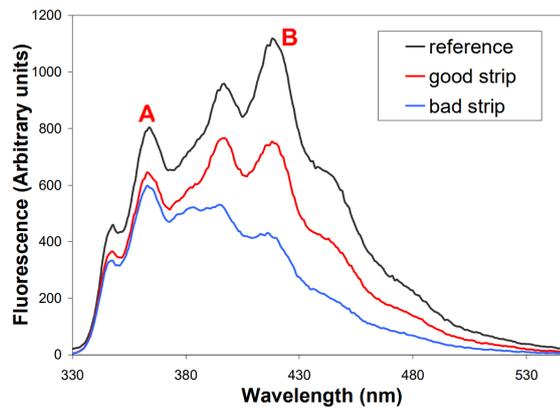
Figure 3.13: Fluorescence spectra of scintillators produced by FNAL and chosen for the SSD. Peaks A (365 nm) and B (420 nm) correspond to the emission peaks of the PPO and POPOP dopants, respectively. From [50].

Figure 3.14. As photons in the SSD may travel through several meters of fibre, shifting the wavelength is beneficial as it increases the attenuation length of the photons to 3.5 m [52].

3.3.2   Other Upgrades

- **Underground Muon Detector (UMD)**

   As a component of the AMIGA enhancement to Auger, the denser SD-750 is being upgraded with 30 m^2 scintillator-based Underground Muon Detectors (UMDs). They
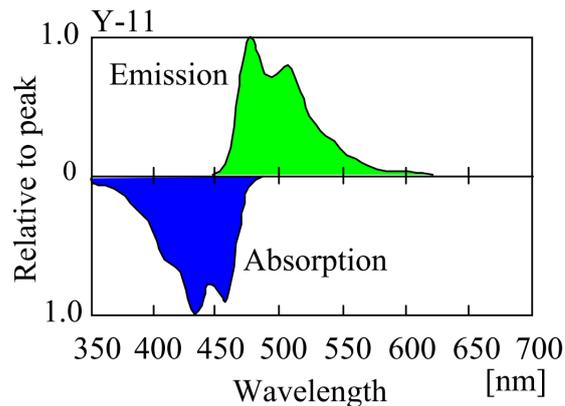
Figure 3.14: Spectrum of absorption and emission for the wavelength-shifting dye K-27[53] in the Y-11 (300) MSY fibres produced by Kuraray [52] and chosen for the SSD.

will be buried at a depth 2.3 m underground to attenuate the electromagnetic component of air showers. They will significantly enhance mass composition reconstruction on an event-by-event basis as well as provide direct calibration of the methods used to reconstruct muon numbers from other SD components.

- **Upgraded Surface Detector Electronics**
  Improvements to the current SD electronics will upgrade the current UB to a newer high performance electronics system for the SD array. The current electronics setup supports 6 channels, a sampling frequency of 40 MHz and 10 bit ADC counting. All 6 channels are being used with the high and low gain signals for the 3 PMTs in each WCD. New electronics support 10 channels, 120 MHz sampling frequency and 12 bit ADC counting. The additional channels will allow the current SD to support more PMTs. The additional channels will be used to support the addition of a small PMT and an SSD to each existing WCD.

- **Small Photo-Multiplier Tube (SPMT)**
  A fourth, "small" PMT, the Hamamatsu-R8619-22 [54] can easily be installed in each SD by exploiting an unused and easily accessible 30 mm window on the Tyvec bag. The smaller light collection will extend the dynamic range of the WCDs to match the upper limits planned for the SSDs. The saturation point is expected to be at around 200 m from the shower core, for showers at energies of around 10^{20} eV.

3.4 $\overline{\text{Off}\underline{\text{line}}}$ Framework

The Auger collaboration has developed a software framework known as $\overline{\text{Off}\underline{\text{line}}}$ for over 20 years, to help analyse data collected by the observatory. Since then, the software has become a general purpose object-oriented framework that is primarily written using the C++ programming language. $\overline{\text{Off}\underline{\text{line}}}$ consists of a sophisticated back-end infrastructure that allows the support over a broad range of data analysis and simulation procedures; it can handle multiple different input formats and output formats [55]. It can handle the data format from the raw data acquisition system, the $\overline{\text{Off}\underline{\text{line}}}$ internal format, and the "ADST" external format [56]. The ADST format is used for end-user analysis with the input/output components of the ROOT framework [57]. Simulations hosted by $\overline{\text{Off}\underline{\text{line}}}$ can integrate with common air shower Monte Carlo packages such as CORSIKA [58], CONEX [59], AIRES [60], CoREAS [61], and ZHAireS [62], while also allowing custom input. Most simulations within $\overline{\text{Off}\underline{\text{line}}}$ are modelled within Geant4 [63], complemented with custom classes which track particles such as Cherenkov and fluorescence photons within simulated fibres. It also contains a simplified front-end system called `EventBrowser`, built with the user-interface components provided by ROOT, which include features which give any collaborator the ability to quickly display reconstructed cosmic ray events, simulated detector responses from simulated air showers, FD reconstruction, and much more. Through a series of carefully designed classes and functions, the $\overline{\text{Off}\underline{\text{line}}}$ framework retains its flexibility with XML configurations and modular design approach; allowing users to insert and extract certain components of a simulation or reconstruction at will. As part of the work completed throughout this thesis, I have extended the `EventBrowser` front-end system to integrate natively with the new SSD data, providing easy access data and visualisations for users within the Pierre Auger Collaboration.

Chapter 4

Combined Pierre Auger Scintillator Surface Detector and Telescope Array Scintillator Detector Simulation

The Pierre Auger Observatory (Auger) and the Telescope Array Experiment (TA) have reported their measurements of UHECR energy spectra observed in the southern and northern hemisphere, respectively. The Auger-TA energy spectrum working group, organised in 2012, has been working to understand the uncertainties in energy scale in both experiments, their systematic differences, and differences in shape of the spectra [64]. Within systematic uncertainties, there is an overall agreement of the energy spectra measured by the two observatories below 10 EeV, however, a significant difference remains at the largest energies (see Figure 1.3). The differences between the two spectra is left unexplained, the differences at high energies persist even in the common declination band, with a constant energy shift of ±4.5%. An additional energy-dependent shift of ±10%/decade after 10^{19} eV is required to get agreement between the two spectra [64]. Figure 4.1 shows the Auger and TA energy spectra with the energy shifts applied. Previous works acknowledge the uncertainties arising from differences between methods used by the two observatories to reconstruct energy, that is, the Auger Constant Intensity Cut (CIC) method, and the TA Monte-Carlo energy lookup table (See Figure 4.2). The aim is to test whether the TA assumption of the primary particle mass being proton when creating energy lookup tables can have an impact on the cosmic ray energy spectra at energies above 10^{19} eV. However, recreation of a TA energy lookup
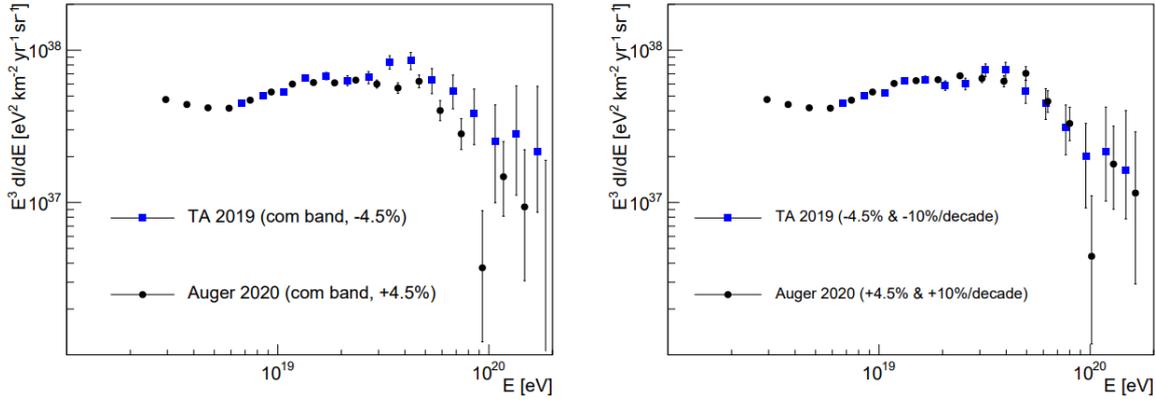
Figure 4.1: Left: Auger and TA spectra in the common declination band ($-15.7° < \delta <$ $24.8°$) with a constant energy shift $\pm 4.5\%$, Right: with an energy-dependent shift $\pm 10\% \times$ $\log_{10}(E/10^{19}\text{eV})$ for $E > 10^{19}$ eV. From [64].

table will require a simulation of the TA scintillator detector. Firstly, this study introduces a custom built Monte-Carlo simulation within the Auger $\overline{\text{Off}\underline{\text{line}}}$ framework which models the TA scintillator and Auger scintillator surface detector (SSD) simultaneously. This simulation demonstrates the complexity behind detector shielding and how it can affect the amount of signal that is measured by the detectors, and verifies the validity of using one detector to approximate the other. Afterwards, the custom simulation is used to produce the TA energy lookup tables, which are then employed to reconstruct the TA energy spectra under a different primary particle mass assumption.
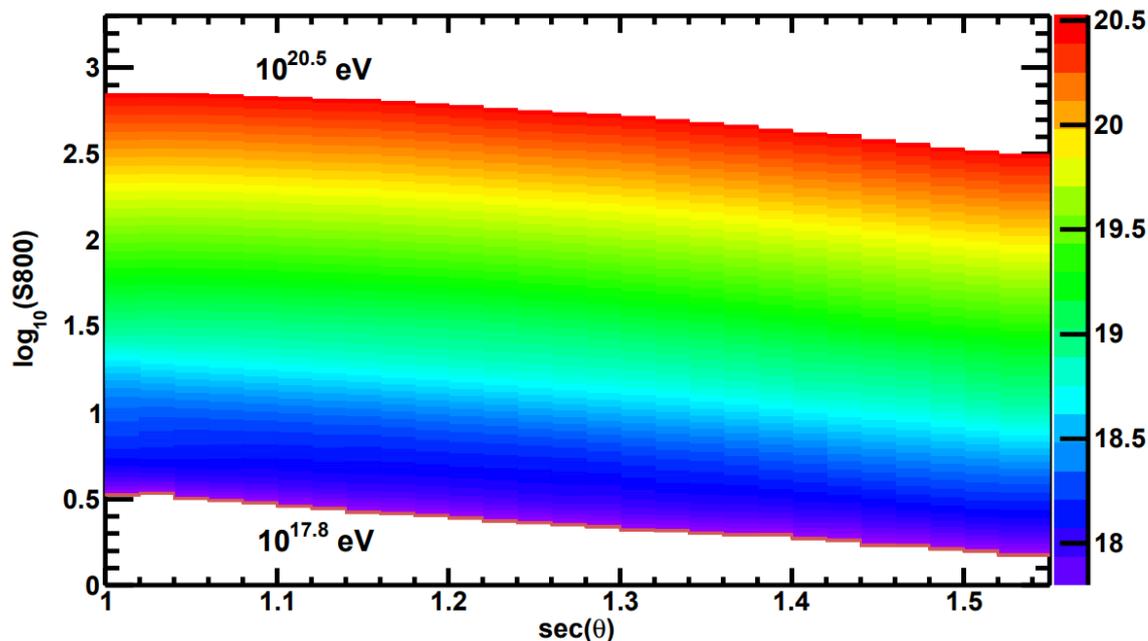
Figure 4.2: TA SD Energy Estimation Table derived from a CORSIKA surface detector Monte-Carlo for proton showers. The x-axis corresponds reconstructed zenith angle as $\sec(\theta)$, the y-axis corresponds to $\log_{10}$ of the signal (MIP) of the TA scintillator detector at 800 m from the shower axis, and the z-axis corresponds to the $\log_{10}$ of *true* energy. It was created from the TA SD simulation library which consisted of 16,800 CORSIKA proton showers with $\epsilon = 10^{-6}$ thinning. The hadronic model QGSJET-II.03 is used the high-energy interactions. From [65]

4.1   Simulation and Detector Design

One may assume that if the scintillator materials and calibration procedures between two scintillator detectors are similar, then the detectors will return a similar signal in minimum ionising particle (MIP) per m^2. However, this is only true for detectors which have similar designs, including their detector shielding. Both the Auger and TA collaborations calibrate their detectors to a MIP. The Auger calibration process is covered in Chapter 5; in general, calibration is to the most probable signal produced by uniformly distributed, vertical muons. Below, we describe the simulation framework which was used to verify the validity of approximating the TA scintillator detector with the SSD in Section 4.3, and to recreate the TA energy lookup tables for cosmic ray spectra analysis in Section 4.4.

Simulations have been completed using a customised, self-modified version of the $\overline{\text{Off}}\underline{\text{line}}$ standard application `SdSimulationReconstructionUpgrade` which contain an embedded model of the TA scintillator detector near the Auger SSD. The modifications allow the user to simultaneously simulate a TA scintillator detector and an Auger SSD. As both detectors are simulated simultaneously in the same simulation world, shower particle sampling is consistent between both detectors, each individual air shower can provide signals from both detectors, and the two detectors are always at similar distances from the shower axis. Alongside a consistent calibration to MIP, these features ensure that the only significant difference between the two detectors is the shielding, as described below.

4.1.1 Pierre Auger Scintillator Surface Detector Simulation Design

The Auger SSD aluminium housing (outlined in Section 3.3.1), which shields approximately 4 m^2 of 1 cm thick scintillator, has a top layer thickness of 1 mm. The scintillator material is placed within 10 mm thick extruded polystyrene. A schematic view of the SSD detector can be seen in Figure 4.3. The scintillator material is "sandwiched" between two layers of extruded polystyrene foam. There is also a layer of expanded polystyrene between the aluminium casing and the extruded polystyrene layer. The extruded polystyrene and expanded polystyrene layers are simulated, however, they are not expected to have a large impact on signal as their densities are small (0.032 g/cm^3 and 0.01414 g/cm^3, respectively). This leaves the Auger SSD with approximately 1 mm of aluminium shielding above the scintillator material. The existing model of the Auger SSD within the $\overline{\text{Off}}\underline{\text{line}}$ framework is used, but, for consistency with the TA scintillator detector implementation, the signal responses have been modified to be based purely on energy deposited in the scintillator material.

4.1.2 Telescope Array Scintillator Detector Simulation Design

The TA scintillator detector consists of 3 m^2 of 1.2 cm thick scintillator material housed within a stainless steel scintillator box, with an iron roof, as seen in Figure 4.4. The iron roof, designed to protect the scintillator from large temperature variations, is 1.2 mm thick. A scintillator box is placed underneath the iron roof; within this box are two layers of scintillator material, each 1.2 cm thick. The scintillator layers are separated by a 1 mm thick stainless steel plate. The top and bottom covers of the scintillator box are composed of stainless steel, with thicknesses of 1.5 mm and 1.2 mm, respectively [37].
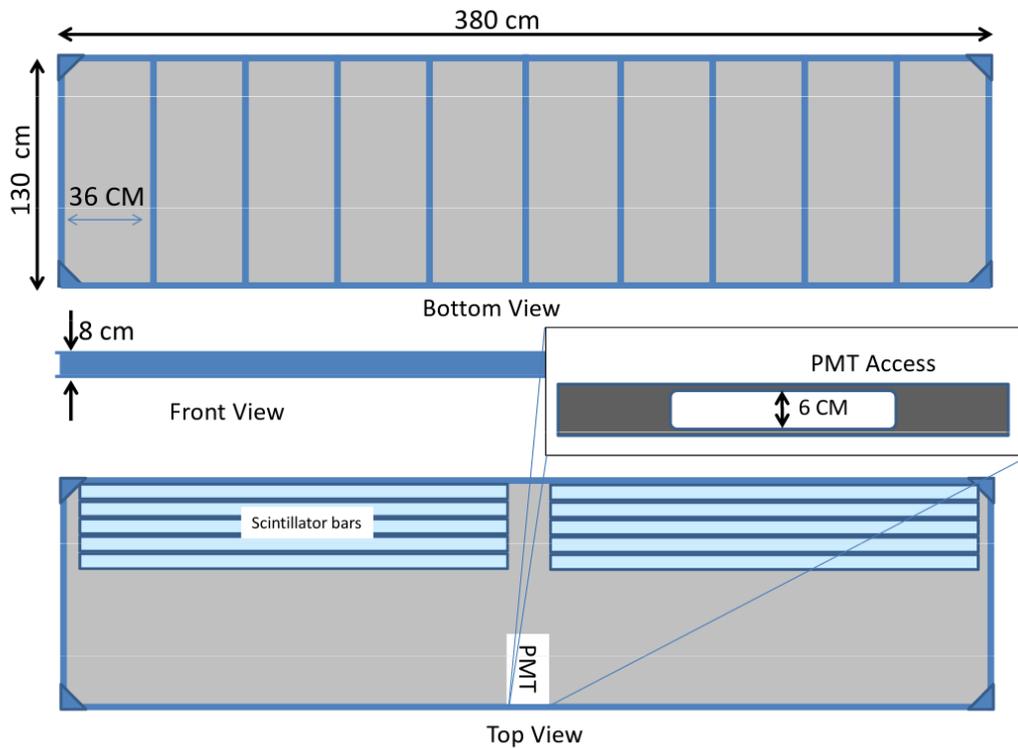
Figure 4.3: A schematic view of the Auger Observatory SSD aluminium external box. From [51].

The custom simulation models the top-most scintillator layer, the stainless steel cover, and the iron roof. This means that there is an iron roof approximated as a flat 1.2 mm thick iron sheet, then the 1.5 mm stainless steel top cover, and finally, a 1 cm thick scintillator layer. Note that the simulation keeps the thickness of the scintillator material constant for both detectors at 1 cm (TA scintillator is actually 1.2 cm in the field). This was done so that the difference between the two detectors would purely be the housing, or the "shielding" of the scintillator material. This change does not impact results as both detectors are individually calibrated to their own MIP. The TA detector has been implemented without the introduction of complex electronics; its signal responses are purely based on the amount of energy deposited in the scintillator material, that is, it is assumed to have 100% light collection efficiency.
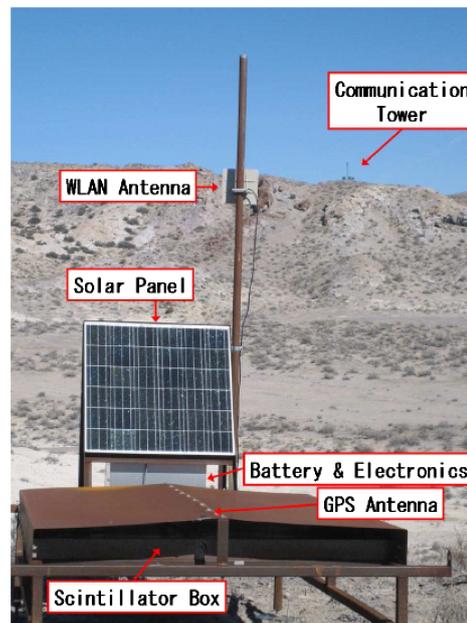
Figure 4.4: A deployed TA SD in the field. The electronics box and scintillator box are on an iron frame. An electronics unit is installed under the solar panel, and the scintillator box is mounted on the platform under the roof. From [37].
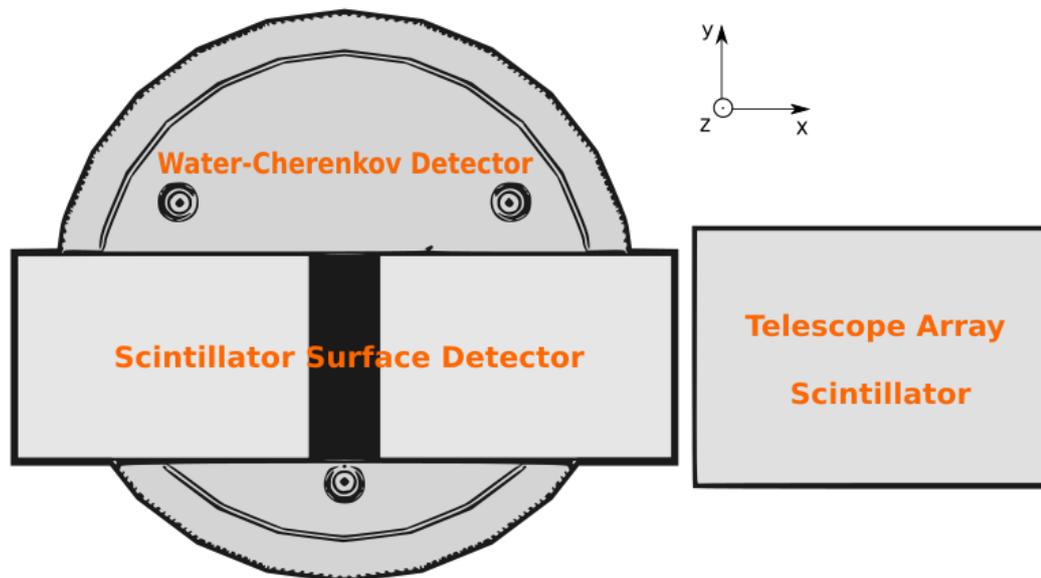
Figure 4.5: A bird's-eye visualisation of the detector volumes used within the GEANT4 `G4StationSimulator` simulation module. The TA scintillator detector volume shares the same y and z origin co-ordinates as the SSD. However, it is displaced +3 m along the x axis.

4.2   Simulation Method

The SSD and TA scintillator are simulated simultaneously, in the same injection cylinder[1]. A bird's-eye visualisation of the GEANT4 detector volumes can be seen in Figure 4.5. The TA scintillator detector volume is depicted as the rectangular shape to the right of the water-Cherenkov detector (WCD) and SSD. In the GEANT4 $[x, y, z]$ right-hand co-ordinate system, the TA and SSD detector volumes share y and z co-ordinates for their origin, but the TA detector is displaced +3 m along the x axis. Through direct modification of the $\overline{\text{Off}}\underline{\text{line}}$ GEANT4 `G4StationSimulator` modules, the simulation can be executed alongside other well known $\overline{\text{Off}}\underline{\text{line}}$ simulation modules, such as the COsmic Ray SImulations for KAscade (CORSIKA) sampling module, `CachedShowerRegenerator`. A more detailed description of the sampling process is described in Section 4.2.1 below. All CORSIKA air showers are obtained from the Auger Napoli database, which contains thousands of unique simulations with a range of energy, zenith angle and mass composition. This study uses one library from the database, the parameters of the library can be seen in Table 4.1. In total, there are 12,000 unique proton and iron CORSIKA showers. The detection of each individual shower has been simulated once, using a custom simulation, based on the $\overline{\text{Off}}\underline{\text{line}}$ standard application `SdSimulationReconstructionUpgrade`. The application has been modified to include a model of the TA scintillator detector as described in Section 4.1.2. In order to reproduce the TA energy lookup table, this study employs the QGSJET-II.03 hadronic interaction model. This model was chosen due to its utilisation by the TA collaboration during the reconstruction of their energy table.

The air showers in the database contain low energy cuts; 0.1 GeV for hadrons and muons, 0.25 MeV for electrons and photons. These cuts help reduce the amount of computational power and time required. CORSIKA stops tracking and recording individual particles once they reach energies below the cut-off. Energy thresholds should not be too high, as low energy particles (around 5 MeV) will still produce substantial amounts of signal in thin scintillators. CORSIKA particle files may be thinned particle files; unthinned files contain information on every particle at ground (and other elevations, if desired). Unthinned files are extremely large in size and are computationally intensive to use, and so, common practise is to use CORSIKA simulations which have undergone "particle thinning". Particle thinning

---

[1]The injection cylinder size has been increased to a radius of 5 m to compensate for the additional detector volumes.

Primaries	Proton, Iron
lg(E/eV)	18.5 - 20.0, energy spectrum index -1
$\theta/°$	0-65, flat in $\cos^2\theta$
$\phi/°$	0 - 360, uniformly distributed
Had Int. Model	QGSJet-II.03

Table 4.1: Properties of the CORSIKA files used from the Auger Napoli database. This study used a total of 12,000 unique proton showers and 12,000 unique iron showers from this set.

dramatically reduces computational time by utilising 'thin sampling' or 'variance reduction' as specified in [58]. When thinning is active, all particles below a specified fraction of the primary particle energy are exposed to the thinning algorithm. The air showers in the database used for this study contain zenith angles, ranging from 0-65°, flat in $cos^2(\theta)$, with thinning fraction of $\epsilon_{th} = E/E_0 = 10^{-6}$.

4.2.1   Particle Sampling

The $\overline{\text{Off}\underline{\text{line}}}$ module `CachedShowerRegenerator` reads the input CORSIKA particle file, and calculates the number of particles that are to be thrown into the injection cylinder within the GEANT4 detector simulation. CORSIKA particle files which have undergone particle thinning undergo a simulation procedure known as 'dethinning' prior to detector simulations. The dethinning process is required when processing thinned air shower files; dethinning returns particle samples in a thinned file to realistic values. The particle dethinning procedure is performed while determining the number of incident particles on a particular detector, at some distance. In order to determine this number, a subset of the CORSIKA footprint (particles at ground) around the detector in the shower plane is used to create a sampling area. The sampling area is given a radial width as a fraction of detector core distance δ, and an azimuthal angle α in the shower plane which surrounding the detector at position r in the shower plane; this is expressed mathematically as

$$A_{sampling} = \frac{\delta\alpha}{2\pi}((r + \delta r)^2 - (r - \delta r)^2).$$ (4.1)

In this work, α is chosen to be $\pm15°$, and δ is 0.10 (which is the $\overline{\text{Off}\underline{\text{line}}}$ default). The sampling area is depicted graphically in Figure 4.6. Since this area is defined in the shower plane, and particles are contained only at ground level in the CORSIKA particle file, each
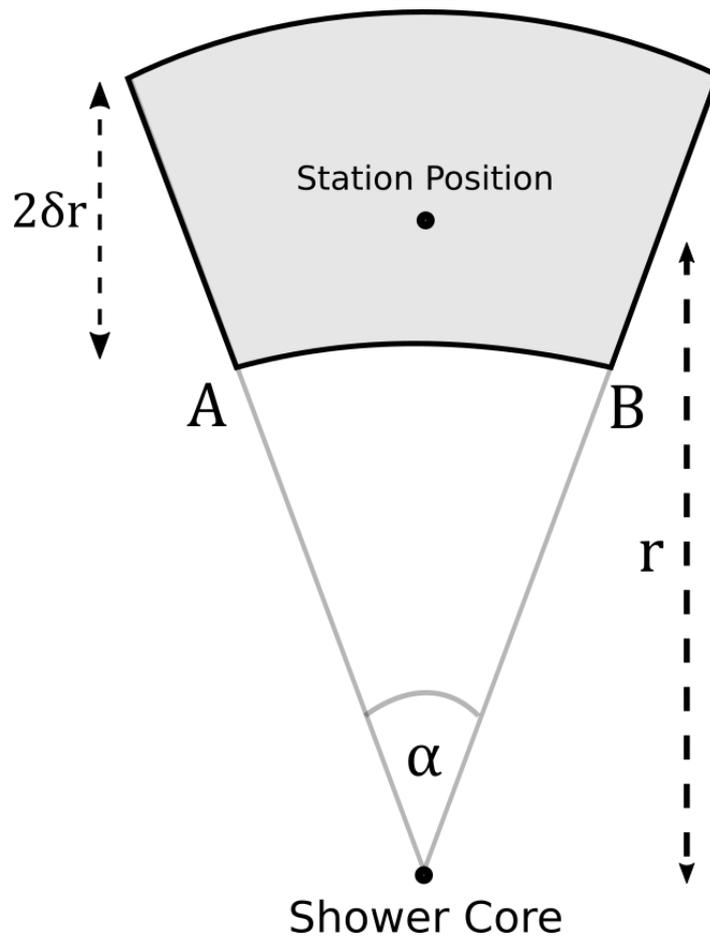
Figure 4.6: The shaded area depicts the area which contains the subset of particles eligible for dethinning around a particular station at distance r from the shower axis. In this study, $\delta = 0.1$, $\alpha = 15°$. This area is defined in the shower plane; its shape will stretch when projected to the ground plane, depending on primary particle zenith angle.

particle must be projected into the shower plane. Here, it is assumed that it is approximately correct that if each particle is projected along the direction of the shower axis, then the result will be similar to projecting each particle along their own direction. Figure 4.7, shows the definition of the sampling area in the shower plane, and its original area at ground, before projection [66].

After a subset of particles has been defined for a particular detector, the number of particles that are to be simulated can be calculated. This number is calculated simultaneously with the dethinning procedure. For thinned CORSIKA showers, each unique particle from
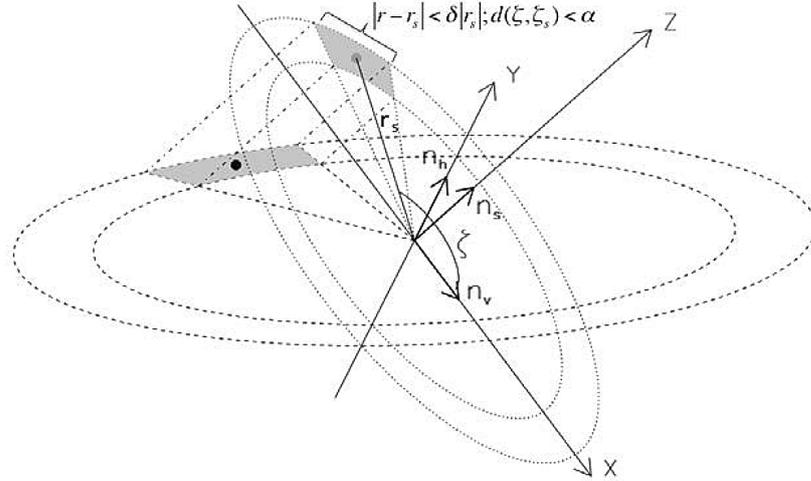
Figure 4.7: Depiction of the sampling area [67]. The sampling area is an area in the shower plane with the detector at core distance r_s and azimuth ζ_s. The sample area has width $2\delta r_s$ and azimuth angle α. From [66].

the sampling area will have a predetermined weight value which can be used to enhance the probability of the particle being injected into the GEANT4 simulation. The number of injections for one unique particle is determined by sampling a Poisson distribution which has a mean value (μ) equal to the particle weight (w), multiplied by the proportion of projected area which the injection cylinder encompasses inside the sampling area. It is written as:

$$\mu_{Poisson} = w\frac{A_{\text{injection}}(\theta)}{A_{\text{sampling}}} \tag{4.2}$$

where θ is the zenith angle of the unique particle. The Poisson distribution is sampled once, returning an integer value N. That unique particle will be thrown N times into the GEANT4 injection cylinder. Each throw has a random probability of being thrown through the side of the injection cylinder as opposed to the top. The probability increases as the zenith angle of the unique particle increases; a larger zenith angle results in the projected area of the side of the injection cylinder also being larger. The value of weight will change depending on the type and intensity of thinning used during the CORSIKA simulation. Heavily thinned showers may not accurately represent the original shower after de-thinning.

4.2.2   Procedure for Obtaining Data

- **Downloading the Napoli Database simulated CORSIKA air showers:**

  All CORSIKA air shower files used in this analysis were downloaded from the Napoli
  database (a database created by the Pierre Auger Collaboration). Each CORSIKA
  simulation writes its associated files to a user defined format. In this case, the format
  is of the type *"DATXXXXXX"*, with a generated series of numbers which have some
  representation of that simulation. For example, each *XXXXXX* mod 5 may represent
  air showers with iron as the primary particle. Each simulation returns a *.part* file which
  contains the particle information for each particle at ground level, a *.long* file which
  contains the longitudinal profile of the shower, and a *.inp* file which contains the input
  parameters. It is possible to read and generate a particle list at ground from only the
  .part file.

- **Reading the CORSIKA particle file and generating a particle list:**

  The CORSIKA particle file contains a wide range of particle information for each
  particle at ground with energies above the energy cuts. The most important parameters
  include the name, position, energy, zenith angle, azimuth angle and momentum of
  each particle. Particles are written with zero azimuth angle corresponding to magnetic
  north.

- **Simulating the CORSIKA air shower in GEANT4:**

  Each individual particle undergoes the dethinning and sampling (as described earlier)
  to determine whether or not the particle should be simulated. Particles deemed eligible
  for detector simulation are injected onto the edge of a large non-physical cylinder
  known as the "injection cylinder". It is crucial that this cylinder is large enough to
  completely surround any detectors within it; this ensures that detector edge effects and
  other geometrical effects are correctly simulated. The probability of the particle being
  injected through the side, or the top of the injection cylinder is dependent on the zenith
  angle of each unique particle. The probability is equal to the ratio of the projected area
  of the side and the projected area of top of the cylinder. The simulation modules place
  a ring of twelve stations, each station 800 m from the true shower axis (this shower
  axis corresponds to the Monte Carlo core position). Post-processing is performed to
  combine the signals from the twelve stations into an average. This approach is taken
  to minimise the effect of signal asymmetry.

- **Writing the output to data files:**

  The detector simulation will output signals, PMT traces, and other useful data into
  an $\overline{\text{Off}}\underline{\text{line}}$ file known as an 'ADST'. However, this work does not use these ADST
  files directly because the SSD detector simulation in these files have undergone an
  electronics simulation. As stated earlier, the implementation of the TA detector does
  not simulate electronics, so, the results here use results from the SSD simulation prior
  to simulating electronics. To keep assumptions between the TA and SSD simulations
  as consistent as possible, outputs for both detector simulations is the energy deposit
  (MeV) per particle, for each simulated detector. All required information for post-
  processing of data is simply output to a custom file during simulation.

- **Analysing the data:**

  The data is output to a file which contains the following information for each injected
  particle:

  - Particle ID (gamma = 22, mu- = 13, mu+ = -13, e+ = -11, e- = 11),

  - Particle Name (gamma, mu-, mu+, e+, e-),

  - Detector ID (ID corresponding to a particular detector at a certain position in the
    array. This ID can be used to calculate distance to the shower core, or to identify
    stations which were part of the twelve station ring),

  - Particle injection position (x, y, z) in metres, with respect to the centre of the
    injection cylinder,

  - Particle momentum (x, y, z) in GeV/c,

  - Energy deposit from each track step inside the relevant detector.

  The fields needed to create an energy lookup table (See Figure 4.2) only includes the
  Detector ID, and the Energy Deposit within a specific detector. While the other fields
  are not explicitly used in this study, the others are used in understanding the impacts
  of detector shielding in Section 4.3.

4.2.3 Assumptions

The custom simulation does not model electronics for the TA scintillator detector. In-
stead, it is assumed that the TA detector has 100% light collection efficiency. It simplifies

comparisons if the simulation of the Auger SSD followed similar assumptions. As such, the
Auger SSD simulation has been modified to record results prior to simulating electronics.
The output from the custom simulation is the simulated energy deposit (in MeV) for each
incident particle, in each detector. This energy deposit assumes a 100% collection efficiency;
it does not consider any losses due to attenuation in fibres, or through photomultiplier tube
(PMT) collection efficiencies. For analyses here, comparisons are performed by directly
comparing the energy deposits in the respective detectors, after calibrating them to MIP.

4.3   The Effect of Shielding on Signal from Scintillator Detectors

Here, an investigation into the differing responses from the Auger SSD and the TA scintillator is completed using the scintillator detector simulation described in Section 4.1. Results show that even though both detectors are scintillator based, they have differing responses to different types of particles due to unique detector design.

4.3.1   Results

Figure 4.8 shows the energy deposit in a 1 cm thick scintillator from electrons that have traversed through 1 mm thick aluminium shielding (like Auger SSD), or 1.2 mm thick iron roof and 1.5 mm thick stainless steel shielding (simiar to TA detectors). Every electron that deposited energy within the scintillator material is included in the histograms, and the same library of CORSIKA simulated air showers is used in both cases.

It is known that the 1 mm thick aluminium shielding, with a density of 2.7 g/cm^3, is far less dense than the combination of 1.2 mm thick iron roof (7.8 g/cm^3) and a 1.5 mm stainless steel cover (8 g/cm^3). It can be roughly calculated for vertically incident particles, based on electrons depositing 2 MeV/g/cm^2, that the aluminium shielding will impose an energy threshold of at least 0.5 MeV. On the contrary, the iron roof and stainless steel shielding will impose an energy threshold of around 4 MeV for vertical particles; this threshold increases as the particle zenith angle increases. This behaviour is reflected within the colour scale of the 2D histogram, and the overlaying averages (black lines). It can be seen that the iron roof and stainless steel shielding average energy deposit is small and flat for electrons with energies less than 5 MeV. The energy deposits corresponding to this flat region are from electrons that do not fully traverse through the scintillator material. For energies greater than 5 MeV, there is an upwards incline to a steady average as the electron energy increases to sufficient energy to fully traverse through the shielding and scintillator material from any incident angle. Similarly, the aluminium shielding shows very similar trends, but with a much larger quantity and different energy threshold. The energy spectrum of electrons from the ensemble of simulations is shown in Figure 4.10a, the majority of electrons are just below 5 MeV, and so, it should be expected that a significant number of a electrons will not reach the scintillator material after traversing through an iron roof and stainless steel shielding to produce any signal.
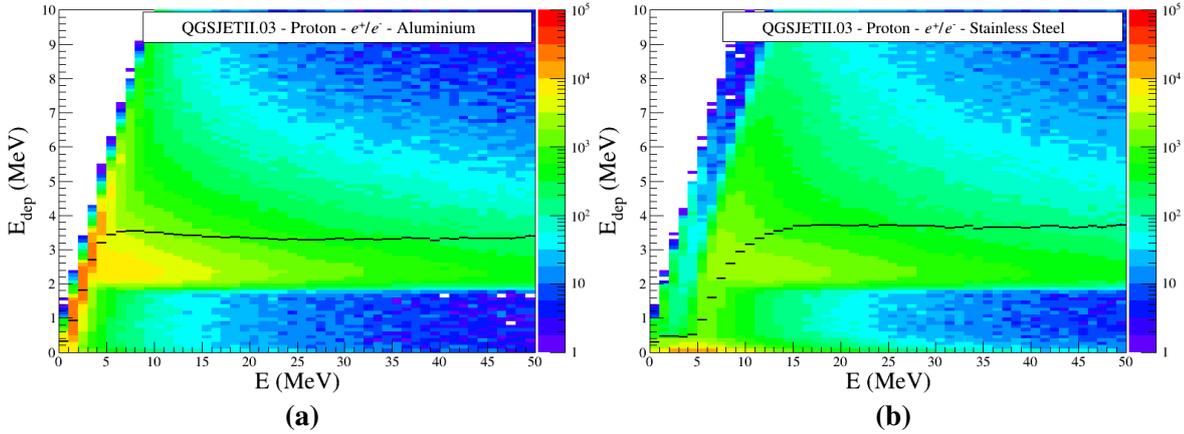
Figure 4.8: (a) A graphical representation of the energy deposit in an aluminium shielded scintillator material, for a given electron energy. (b) A graphical representation of the energy deposit in an iron plus stainless steel shielded scintillator material, for a given electron energy. In both cases, the z axis colour scale represents the number of particles, and the distribution of electrons is from an ensemble of CORSIKA simulated air showers obtained from within the Napoli database. The proton showers are simulated with energies ranging from $10^{18.5}$ eV to 10^{20} eV, and zenith angles ranging from $0°$ to $65°$. The scintillator detector material is positioned 800 m from the shower axis. The black line represents the average energy deposit for each particle energy bin.

Figure 4.9 depicts the energy deposited in the scintillator material from muons after traversing through aluminium shielding (Auger SSD) or an iron roof and stainless steel shielding (TA SD). Muons typically have a much larger energy than electrons, and so it is expected that the majority of muons will traverse entirely through either type of shielding, as well as the scintillator material. The majority of muons have energies well above the aforementioned threshold of 4 MeV for the iron roof and stainless steel shielding, as shown in Figure 4.10b, which depicts the energy spectrum of muons for the simulation dataset. Based on the fact that muons will likely always traverse entirely through, it is expected that there will be little to no increase in signal, or a very small increase in signal from having more shielding material. Additional material may result in larger signals as it will give high energy muons a larger probability to produce secondary particles, such as delta rays. At first glance in Figure 4.9, there is no noticeable difference in signal produced from muons due to changing types of shielding.
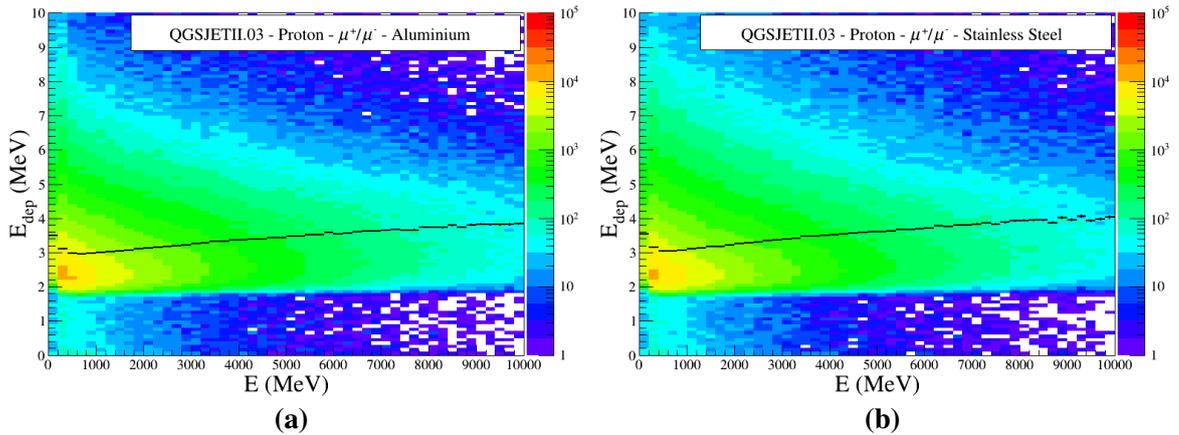
Figure 4.9: (a) A graphical representation of the energy deposit in an aluminium shielded scintillator material, for a given muon energy. (b) A graphical representation of the energy deposit in an iron plus stainless steel shielded scintillator material, for a given muon energy. In both cases, the z axis colour scale represents the number of particles, and the distribution of electrons is from an ensemble of CORSIKA simulated air showers obtained from within the Napoli database. The proton showers are simulated with energies ranging from $10^{18.5}$ eV to 10^{20} eV, and zenith angles ranging from $0°$ to $65°$. The scintillator detector material is positioned 800 m from the shower axis. The black line represents the average energy deposit for each particle energy bin.

Figure 4.11 shows the energy deposit in the scintillator material from gamma rays that have traversed through aluminium shielding, or an iron roof and stainless steel shielding. Gamma rays do not deposit energy within the scintillator material themselves, they must undergo a particle interaction that produces secondary charged particles that can deposit energy. One example is pair production, where a photon produces an electron and positron pair, which can deposit energy within the scintillator material. The energy spectrum of the gamma rays can be seen in Figure 4.10c. It is expected that a significant increase in shielding material will result in a larger proportion of gamma rays interacting within the shielding and producing secondary particles. This can be seen in Figure 4.11, where there is a much higher average energy deposit from gamma rays with iron roof and stainless steel shielding as opposed to aluminium shielding.

In summary, it has been found that electrons will deposit less energy with the iron roof and stainless steel shielding because, as shown in the energy spectrum of electrons, many electrons will not have sufficient energies to pass through the shielding. Muons will deposit

(a)

(b)

(c)
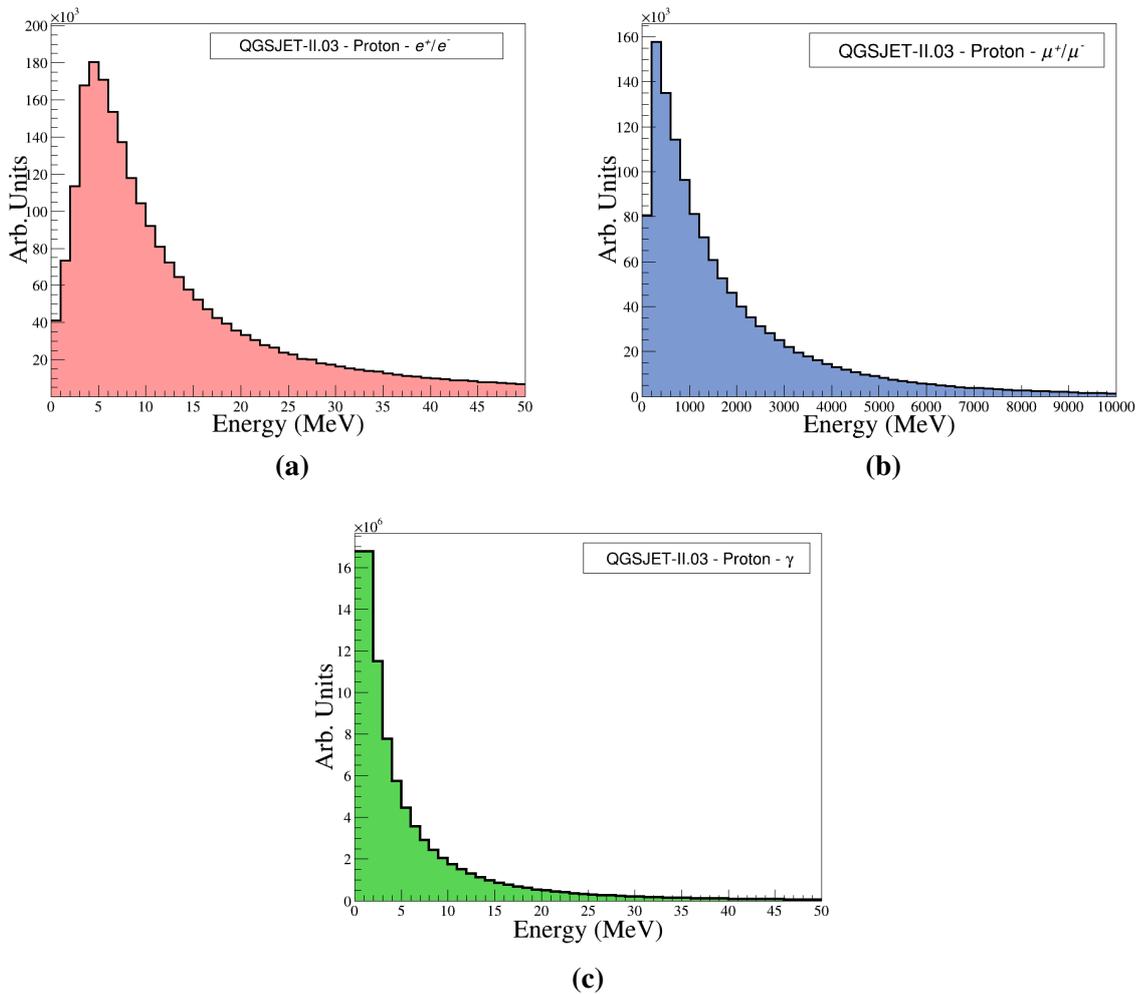
Figure 4.10: From the same, randomly chosen simulated air shower: (a) Energy distribution
of electrons, (b) energy distribution of muons, and, (c) energy distribution of gamma rays.
The distributions are conformed from their respective simulated particles that are located
800 m from the shower axis. The particles are from an ensemble of CORSIKA simulated
air showers obtained within the Napoli database. The proton showers are simulated with
energies ranging from $10^{18.5}$ eV to 10^{20} eV, and zenith angles ranging from $0°$ to $65°$.

similar amounts of energy, because they have high energies and will not be stopped by the
shielding. Gamma rays will produce more signal via secondary particles with the iron roof
and stainless steel shielding, as there is a larger probability of interacting before reaching the
scintillator material.
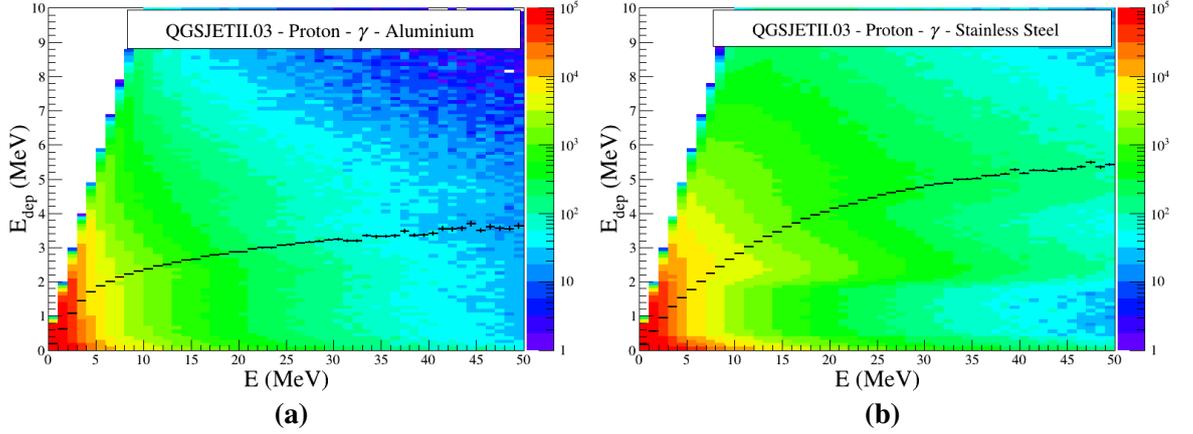
Figure 4.11: (a) A graphical representation of the energy deposit in an aluminium shielded scintillator material, for a given gamma ray energy. (b) A graphical representation of the energy deposit in an iron plus stainless steel shielded scintillator material, for a given gamma ray energy. In both cases, the z axis colour scale represents the number of particles, and the distribution of electrons is from an ensemble of CORSIKA simulated air showers obtained from within the Napoli database. The proton showers are simulated with energies ranging from $10^{18.5}$ eV to 10^{20} eV, and zenith angles ranging from $0°$ to $65°$. The scintillator detector material is positioned 800 m from the shower axis. The black line represents the average energy deposit for each particle energy bin.

The relative differences in individual component signals due to different shielding is simply calculated from the ratio of component signals from each respective type of shielding. That is; let δR_i ($i = e, \gamma, \mu$) represent the relative differences in signal S, between stainless steel shielding and aluminium shielding. Then the relative difference in signal is written as

$$\delta R_i = \frac{S_i^{\text{SS}} - S_i^{\text{Al}}}{S_i^{\text{Al}}} \qquad (4.3)$$

where SS symbolises the TA stainless steel and iron roof scintillator shielding, and Al represents the Auger aluminium SSD shielding. The differences due to changing from aluminium shielding to stainless steel and iron roofing for each of the three major air shower components of interest (electrons, gamma rays and muons) is written explicitly as

$$\delta R_e = \frac{S_e^{\text{SS}} - S_e^{\text{Al}}}{S_e^{\text{Al}}}, \delta R_\gamma = \frac{S_\gamma^{\text{SS}} - S_\gamma^{\text{Al}}}{S_\gamma^{\text{Al}}}, \delta R_\mu = \frac{S_\mu^{\text{SS}} - S_\mu^{\text{Al}}}{S_\mu^{\text{Al}}} \qquad (4.4)$$

Figure 4.12 shows the relative differences in signals at 800 m from the shower axis due to having an iron roof and stainless steel shielding, instead of aluminium shielding. For an ensemble of proton CORSIKA showers, it was found that electrons produce approximately 40% less signal in the heavier, iron roof and stainless steel shielding. Muon signal is shown to be slightly larger, at around 7%, and the signal produced from gamma rays initiated secondary particles is shown to be upwards of 90% larger in the heavier shielding. It is also found that these contributions are independent of air shower primary particle energy, as they are largely determined by the properties of the individual secondary particles. Figure 4.13 shows the same relative differences as earlier, but, for an ensemble of iron CORSIKA showers. The relative differences in component signals due to shielding are independent of air shower mass, however, the total measured signal is expected to be dependent on mass as the relative proportions of individual shower components is mass dependent (shower primaries of heavier mass typically have higher proportions of muon content in total signal).

The total signal, which is calculated as the sum of the three individual signal components from electrons, gamma rays, and muons

$$S_{total} = S_e + S_\gamma + S_\mu \tag{4.5}$$

can be substituted into Equation 4.3 (where $i = total$) to give

$$\delta R_{total} = \frac{[S_e^{SS} + S_\gamma^{SS} + S_\mu^{SS}] - [S_e^{Al} + S_\gamma^{Al} + S_\mu^{Al}]}{S_{total}^{Al}} \tag{4.6}$$

Then, using $\delta R_i S_i^{Al} = S_i^{SS} - S_i^{Al}$, the equation can be rearranged into a form that does not require signals from the other shielding type (apart from δR_i).

$$\delta R_{total} = \frac{\delta R_e S_e^{Al} + \delta R_\gamma S_\gamma^{Al} + \delta R_\mu S_\mu^{Al}}{S_{total}^{Al}} \tag{4.7}$$

If it is true that δR_i is independent of energy, then one could simulate a different type of shielding for a small ensemble of showers (of the same mass), and then use the calculated δR_i values to estimate simulated signal from detectors with the different shielding for any energy using Equation 4.7. In this particular case, it is possible to calculate how much more (or less) signal an individual component contributes to the total signal due to having stainless steel shielding instead of aluminium shielding. Figure 4.14 and Figure 4.15 show how much
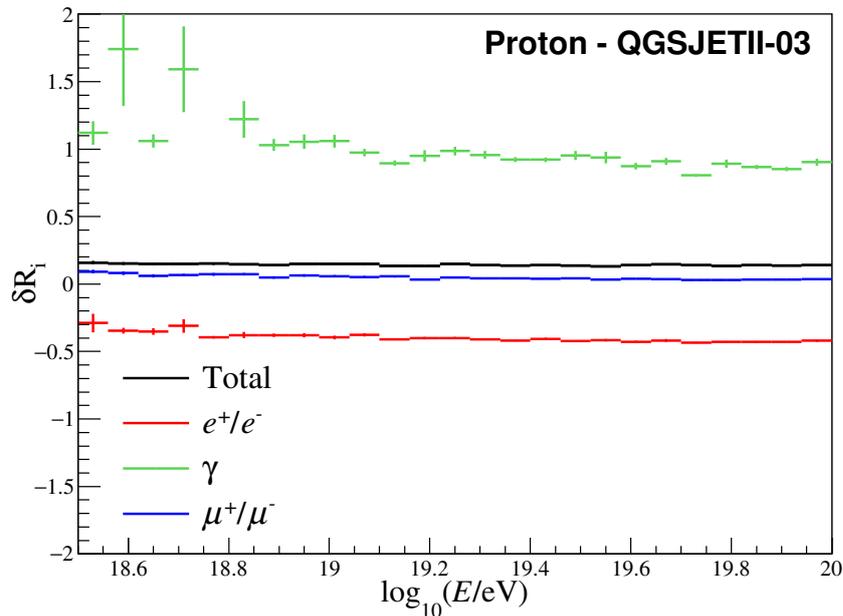
Figure 4.12: The relative differences in signals at 800 m from the shower axis due to having stainless steel shielding and an iron roof, instead of aluminium shielding, for different proton air shower components. Electrons (e^+/e^-) are depicted in red, gamma rays (γ) in green, and muons (μ^+/μ^-) in blue. The total relative difference is shown in black. The electron signal is simulated to be approximately 40% less in the heavier, stainless steel and iron roof shielding. The muon signal is shown to be slightly larger (around 7%), and the signal from gamma rays is shown to be upwards of 90% larger in the heavier shielding. The total signal is measured to be approximately 16% larger with the stainless steel shielding and iron roofing. All contributions are independent of the air shower primary particle energy.

more each individual component contributes to the total signal due to changing shielding. Here, both proton and iron primary particles show a very similar set of results, with the iron results have a slightly less difference in total signal due to a higher proportion of total signal belonging to muons, which are minimally affected by shielding.
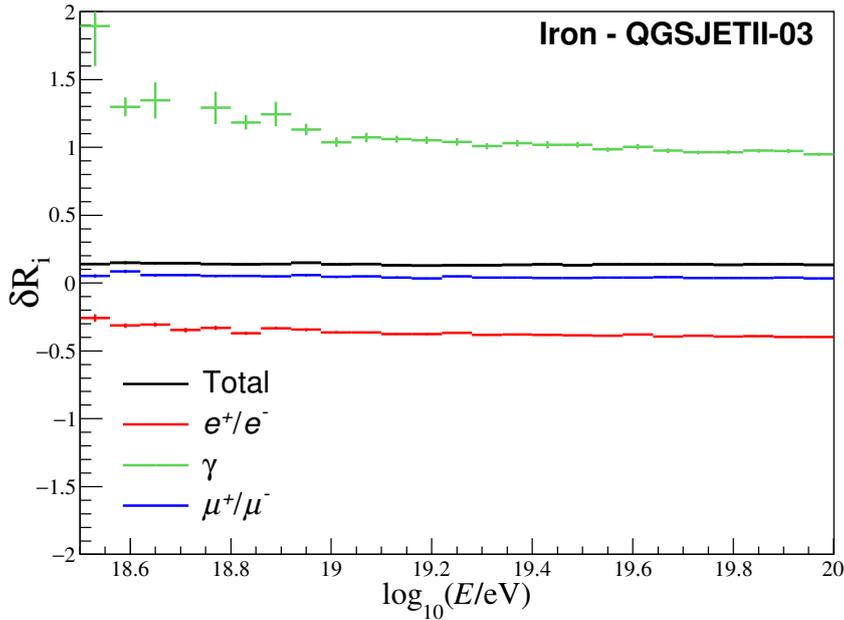
Figure 4.13: The relative differences in signals due to having stainless steel shielding and an iron roof, instead of aluminium shielding, for different iron air shower components. Electrons (e^+/e^-) are depicted in red, gamma rays (γ) in green, and muons (μ^+/μ^-) in blue. The total relative difference is shown in black. The electron signal is simulated to be approximately 40% less in the heavier, stainless steel and iron roof shielding. The muon signal is shown to be slightly larger (around 7%), and the signal from gamma rays is shown to be upwards of 90% larger in the heavier shielding. The total signal is measured to be approximately 14% larger with the stainless steel shielding and iron roofing. All contributions are independent of the air shower primary particle energy. Large fluctuations towards the lower energies are more representative of a lack of sufficient statistics.
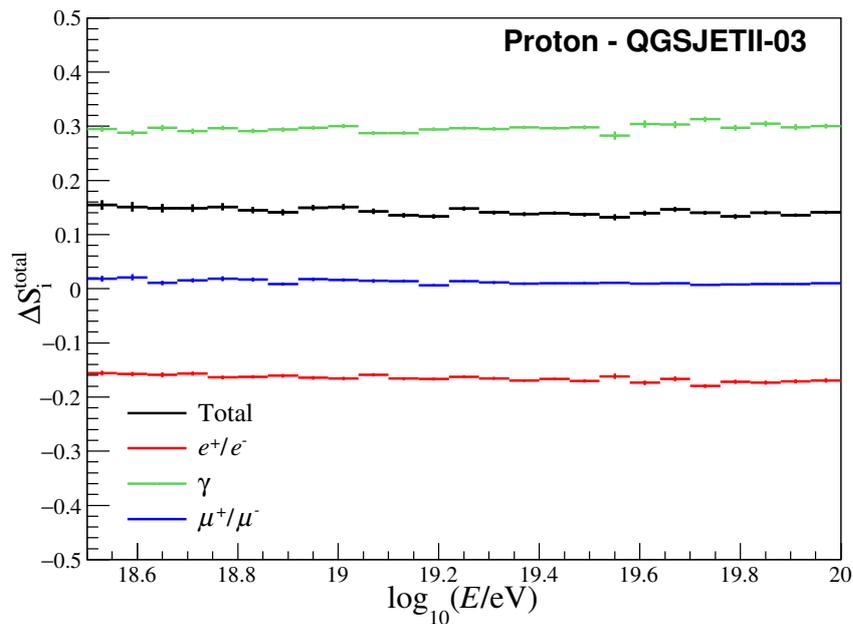
Figure 4.14: The relative increases, or decreases in total signal at 800 m from the shower axis for different proton air shower components, due to changing the shielding from aluminium to stainless steel and iron roofing. The reduced amount of electron signal due to the change in shielding results in $\approx 15\%$ less total signal. Muons produce a small amount of additional signal, which results in an increase of $\approx 2\%$ in total signal. Gamma rays have a significant increase in produced signal, which results in an additional $\approx 30\%$ total signal. All relative increases or decreases in total signal are independent of energy.
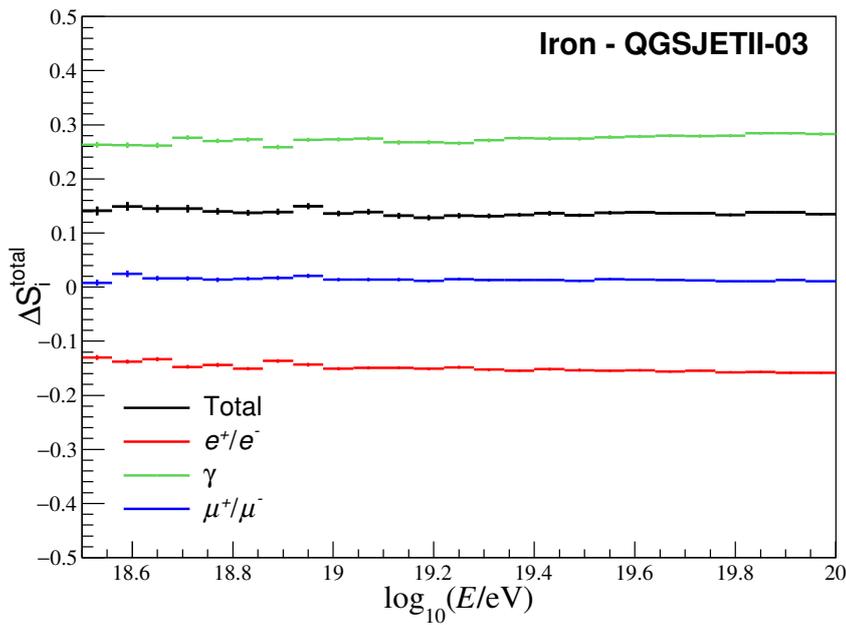
Figure 4.15: The relative increases, or decreases in total signal at 800 m from the shower axis for different iron air shower components, due to changing the shielding from aluminium to stainless steel and iron roofing. The reduced amount of electron signal due to the change in shielding results in $\approx 15\%$ less total signal. Muons produce a small amount of additional signal, which results in an increase of $\approx 2\%$ in total signal. Gamma rays have a significant increase in produced signal, which results in an additional $\approx 28\%$ total signal. All relative increases or decreases in total signal are independent of energy.

4.4   Pierre Auger and Telescope Array Energy Spectra

4.4.1   Energy Lookup Tables

Energy lookup tables are used by the Telescope Array Experiment (TA) to estimate the energy of cosmic ray primary particles with the help of a measured detector signal at 800 m and a reconstructed zenith angle of the cosmic ray primary particle. The energy lookup table that TA uses as part of their reconstruction procedure is shown in Figure 4.2. We aim to reproduce this energy lookup table as accurately as possible, within the Auger $\overline{\text{Off}\underline{\text{line}}}$ framework. Energy lookup tables can be created with simulations as both the zenith angle and energy of the primary particle are known from Monte Carlo truth. The only additional component that needs to be simulated is the detector signal response. After all simulations described earlier were completed, there was a large amount of data available that could be used to create a simulated energy lookup table using the simulated detector response at 800 m. First, a 3D graph is filled; primary particle zenith angle in $\sec(\theta)$ is placed on the x-axis, simulated and calibrated signal (calibrated to a MIP defined as the mode of energy deposited per muon from a distribution of ten-thousand 2 GeV muons) per detector (800 m from the shower core) on the y-axis and shower primary energy in $\log_{10}(E/\text{eV})$ on the z-axis. For each simulated CORSIKA shower, the zenith angle and energy of that shower is known, and the detector response at 800 m corresponding to that zenith angle and energy is simulated.

Shower to shower fluctuations were taken into account by sampling a large enough data set such that data could be separated into bins along the x and y axes. The number of bins along these axes is dynamic, and has been programmed such that the number of entries per bin is at least (on average) equal to T, where T is a user-defined threshold. In this work, T has been set to 10 entries per bin in the results and figures. With a sufficient number of entries per bin, errors can be calculated for each bin; all errors have been calculated as standard errors of the mean. The raw data (3D Graphs) from the simulations are shown for proton, iron and a 50/50 proton/iron mix in Figures 4.16 and 4.17. This data is binned, and did not contain enough data within it to have comparable bin sizes to the original TA surface detector (SD) energy estimation table. Therefore, each set was fit with multiple 2-D functions using the binned data with error bars. It was found that the best fit that converged was a function linear in $\log_{10}(S_{800})$ and quadratic in $\sec\theta$ out of all possible combinations from linear to quartic. This compatible fit was then evaluated over a larger number of bins
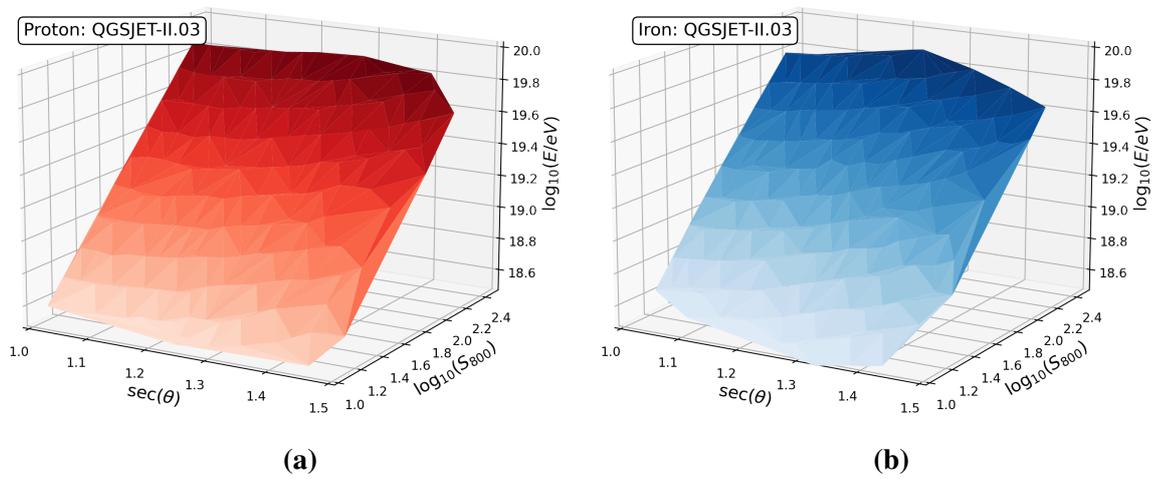
Figure 4.16: (a) Simulated Energy Estimation Table created with QGSJET-II.03 proton air showers. (b). Simulated Energy Estimation Table created with QGSJET-II.03 iron air showers.
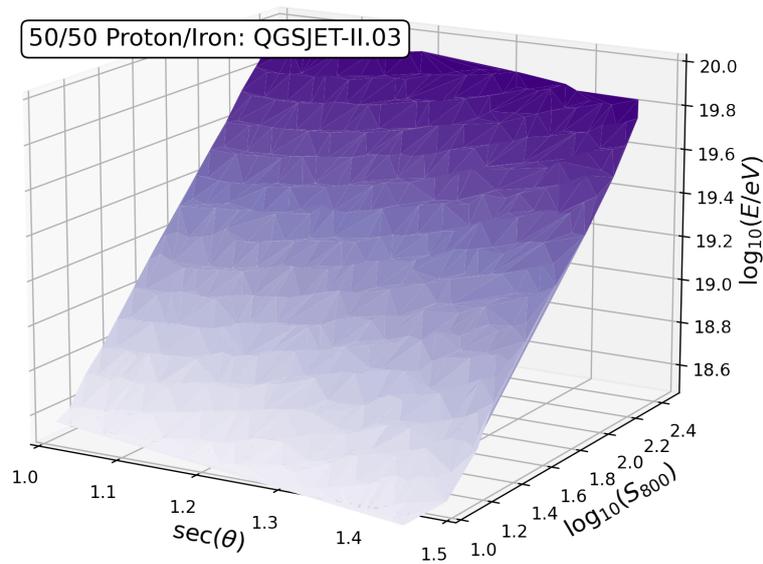
Figure 4.17: Simulated Energy Estimation Table created with a 50/50 ratio of QGSJET-II.03 proton air showers and QGSJET-II.03 iron air showers.

to create a smooth looking energy table, as seen in Figures 4.18 and 4.19. Figure 4.18a, specifically, is the re-creation of the TA SD energy estimation table. It is similar in both
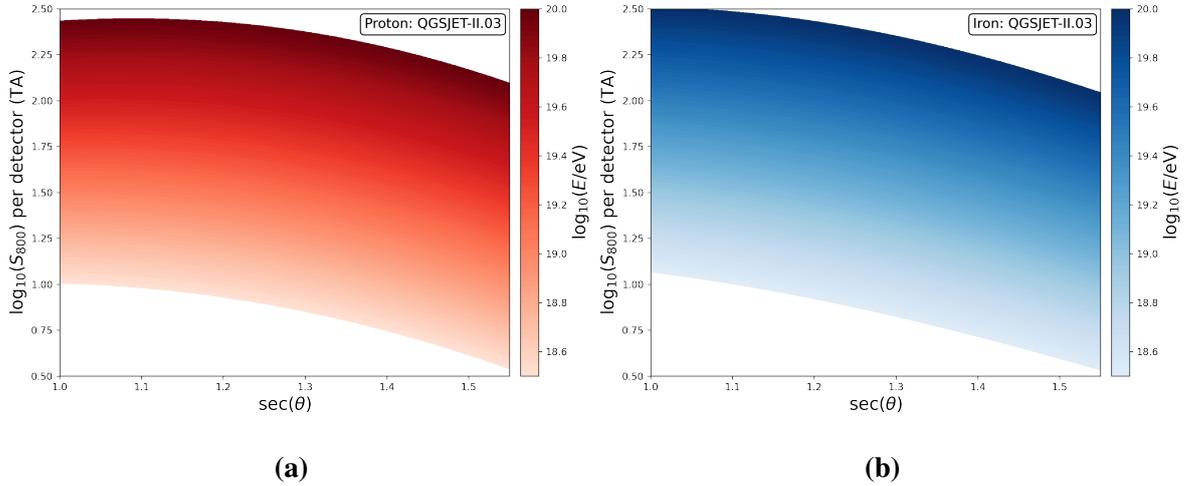
Figure 4.18: (a) Simulated Energy Estimation Table created with QGSJET-II.03 proton air showers. (b) Simulated Energy Estimation Table created with QGSJET-II.03 iron air showers. The table shown is an evaluation of a fitted function, which is quadratic in $\sec(\theta)$ and linear in $\log_{10}(S_{800})$.

shape and magnitude, however, the energy range is not as large. This was acceptable as the simulations used here could never exactly replicate the TA detector properties, and no electronics were simulated; it is assumed that there is 100% light collection efficiency. As the calibration procedure also assumes 100% light collection efficiency, this assumption does not impact the outcomes of this study.

4.4.2 Altering the Cosmic Ray Energy Spectrum with different Energy Lookup Tables

As mentioned earlier, it is known that the the Pierre Auger Observatory (Auger) and TA energy spectra (as shown in Figure 1.3) show a systematic difference that grows with energy. Here, the TA assumption of a using a proton primary particle during energy reconstruction is tested, alongside whether or not changing this assumption significantly affects the reconstructed energy of the cosmic rays, hence altering the spectrum. The energy lookup tables that have been created for the proton, iron and 50/50 proton/iron primary particles can be used to alter the existing energy spectrum which TA has defined. The TA cosmic ray energy spectrum has two important features; the ankle starting around $10^{18.7}$ eV, and the suppression
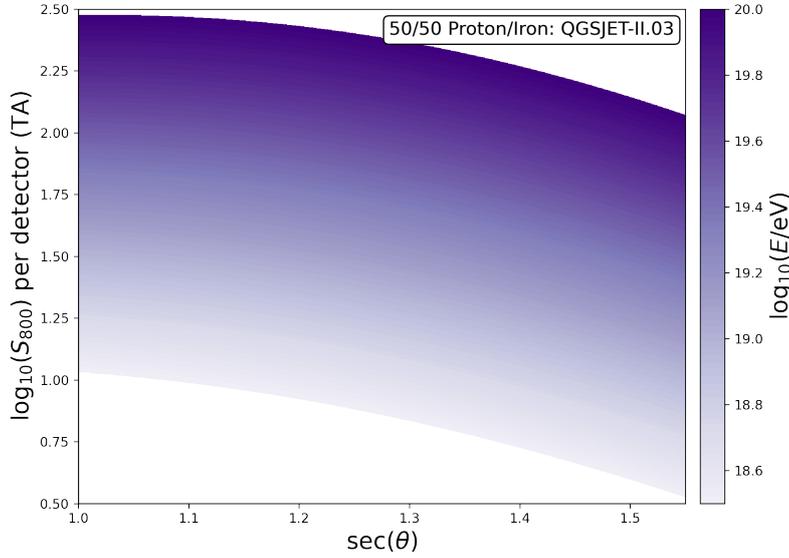
Figure 4.19: Simulated Energy Estimation Table created with a 50/50 ratio of QGSJET-II.03
proton air showers and QGSJET-II.03 iron air showers. This table is an evaluation of a fitted
function, which is quadratic in $\sec(\theta)$ and linear in $\log_{10}(S_{800})$

starting around $10^{19.8}$ eV. Due to these features, the flux follows the shape of a broken power
law function:

$$J_{\text{BPL}}(E) = K\left[h(E_1 - E)E^k + h(E - E_1)h(E_2 - E)E_1^{k-l}E^l \right.$$
$$\left. + h(E - E_2)E_1^{k-l}E_2^{l-m}E^m\right] \tag{4.8}$$

where h is the step function, defined as:

$$h(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \tag{4.9}$$

In Equation 4.8, K is the overall normalisation constant, E_1 is the energy where the first break
(ankle) occurs, E_2 is the energy where second break (suppression) begins, and k, l, m are the
spectral indices. The values used in these tests are from the TA 2019 Energy spectrum results
in [68], and are as follows: $K = 2.24 \pm 0.06 \times 10^{30}$ eV^{-1}m^{-2}s^{-1}sr^{-1}, $E_1 = 10^{18.69 \pm 0.01}$ eV,
$E_2 = 10^{19.81 \pm 0.03}$ eV, $k = -3.2 \pm 0.02$, $l = -2.68 \pm 0.02$, and $m = -4.84 \pm 0.48$.

Method

The effect on the spectrum due to different primary mass is tested by first converting the TA energy spectrum function to expected scintillator detector signals, using the proton energy tables created through this work. The expected scintillator detector signals can then be looked up in the other energy tables, and a different cosmic ray energy will be determined, dependent on the cosmic ray mass assumption. The formulated method is as follows:

- A random energy is sampled within between $10^{18} < E/eV < 10^{20.5}$, which is a slightly larger range than the actual energy range of $10^{18.5} < E/eV < 10^{20}$. This ensures that any potential edge effects are not missed.

- Each sampled energy is converted to an expected detector signal using the proton energy table (Figure 4.18a), when combined with a randomly sampled zenith angle from a $\cos(\theta)\sin(\theta)$ distribution for $0° \leq \theta \leq 45°$. This range of zenith angles reflects the original TA energy table.

- The expected detector signals and zenith angles are then looked up in a different energy table, such as the iron energy table, to return a new reconstructed cosmic ray energy. This cosmic ray energy will be representative of the mass used to create the energy lookup table.

- The new energy is then plot into a ROOT profile graph on top of the original TA energy spectrum for comparison. This process is iterated millions of times to construct a new spectrum.

Results

Figure 4.20 shows the outcome after altering the spectrum with the two different energy tables created during this work, that is, the iron and 50/50 proton/iron mix tables. It has been found that changing the mass assumption during reconstruction from proton to iron does not provide a large enough shift in the spectrum to account for the discrepancies between the Auger and TA energy spectra. Understandably, the 50/50 proton/iron mix lies between the proton and iron data. Figure 4.21 shows the relative change in primary energy as a function of zenith angle. If only zenith angles with a value of 0 degrees were being considered, the shift from proton to iron would be almost 10 times more significant, a shift that significant
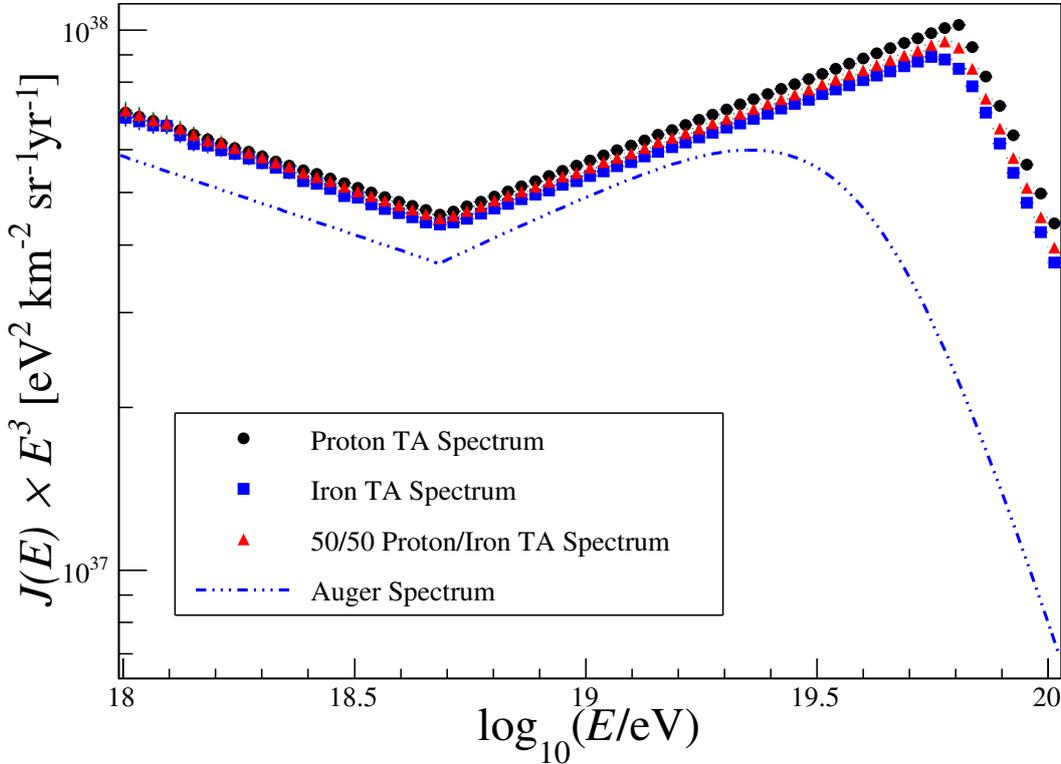
Figure 4.20: The effect of changing the primary particle mass assumption when reconstructing the cosmic ray energy via a Monte-Carlo energy lookup table. The black circles represent the original TA energy spectrum as defined from published results. The blue squares show the energy spectrum after deconstructing through a proton energy table, and reconstructing with an iron energy table. The red triangles show the energy spectrum after deconstructing with the proton energy table and then reconstructing with a 50/50 proton/iron energy table. The largest shift, from proton to iron, is not significant enough to move the points outside of range of uncertainty from the TA results. For comparison, the Auger spectrum has been plotted as the blue dashed line.

would result in the TA spectrum approaching the Auger spectrum. Zenith angles at the other side of the distribution will result in the TA spectrum moving away from the Auger spectrum. The mean value represented by the red point corresponds to the relative change that is to be expected if all zenith angles between 0 and 45 degrees are considered. This result reflects the differences in shape between the simulated energy estimation tables. Negative values of $\Delta E/E$ can be interpreted as proton showers producing less detector signal at 800 m than iron
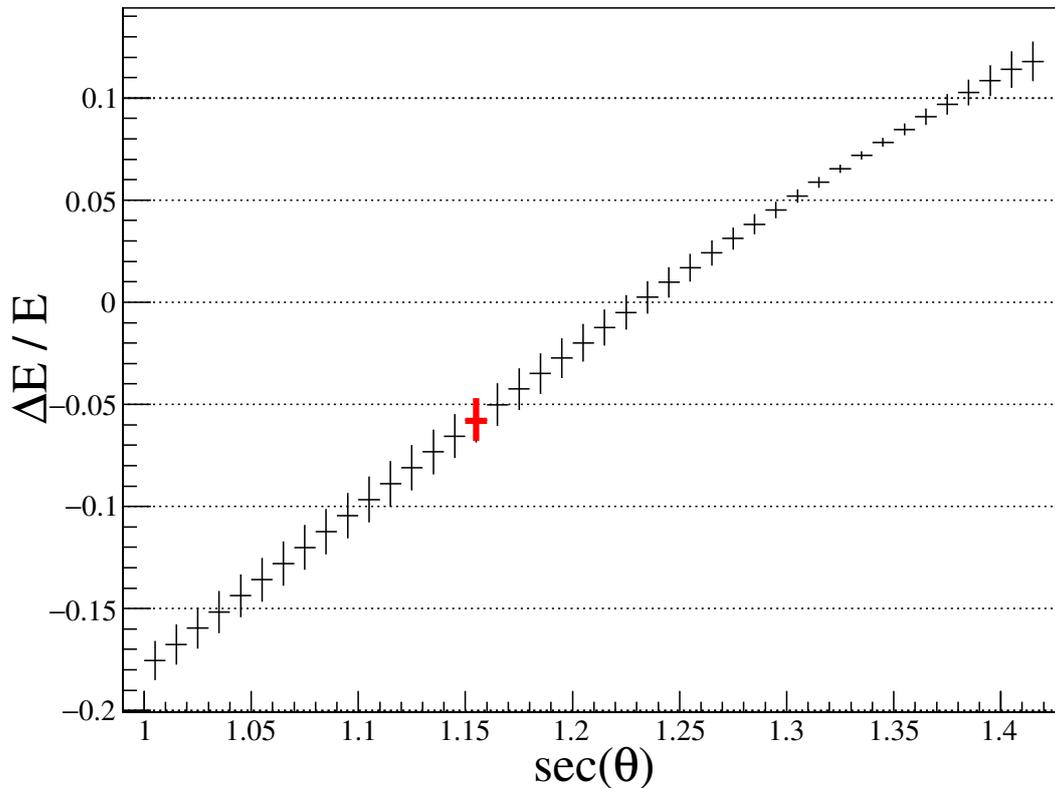
Figure 4.21: The relative change in reconstructed cosmic ray primary particle energy after changing the primary mass from proton to iron as a function of zenith angle. The red point corresponds to the mean zenith angle from a $\sin\theta\cos\theta$ distribution ranging from 0 to 45 degrees. Negative relative changes shift the TA spectrum closer to the Auger spectrum.

showers, for a particular shower energy and zenith angle. The same signal will be produced by an iron shower of lower energy, hence the negative shift. The upwards trend is formed by the differences in the lateral distributions of particles between iron and proton showers, and depicts iron showers requiring a higher energy shower to produce the same amount of signal at 800 m. Since proton and iron primary cosmic rays have one of the largest differences in terms of primary particle mass, it is safe to say that assumptions in primary particle mass have a minor, second-order effect on the energy spectrum, and are not wholly responsible for the differences between the Auger and TA energy spectra, at least with the specific parameters that have been used in this study.

4.4.3   Conclusion

Simulations were created to analyse the effect of shielding on a scintillator detector signal. Detectors with the same size of scintillator material, but different types of shielding were bombarded with the same distribution of particles. It was found that a scintillator detector with shielding similar to the TA SD, would produce approximately 15% more signal than the same detector with shielding similar to the Auger scintillator surface detector (SSD). Thicker shielding enhances signal from gamma rays, but reduces signal from electrons, and muons were largely unaffected. This study was completed for detectors placed 800 m from the shower axis. The proportions and factors used here will have a dependence on distance to the shower axis, because the distribution of particles changes with this distance. It would be expected that as the distance to the shower axis decreases, the difference in signal due to different shielding would be increasing, as there will be an increased amount of electrons and gamma rays with respect to muons. As a result, careful consideration needs to be taken when designing detectors and detector simulations, as shielding which is often designed to protect the scintillator material will affect the amount of signal shower components may produce.

An extensive simulation procedure was used to recreate expected signals from the TA SD. As the electronics setup of the TA SD was not well understood, the custom simulations output raw energy deposits in MeV. The results were analysed and simulated energy estimation tables were created. In an attempt to understand the differences between the Auger and TA cosmic ray energy spectra, these energy estimations were used to reverse engineer the TA energy spectrum under the proton mass assumption. Through use of the iron and 50/50 proton/iron mix energy estimation tables, the TA energy spectrum was recreated under different assumptions of mass. It was found that assuming a different type of mass composition does not yield a significant enough change in the TA energy spectrum. The affect of changing mass does not appear to be wholly responsible for the differences between TA and Auger.

Chapter 5

Defining the Auger Minimum Ionising Particle (MIP)

The development of AugerPrime will allow collaborators to capitalise on the distinct responses of the WCD and the SSD. Each detector reacts differently to similar types of particles. Any analysis which relies on the relative sizes of each detector signal will require a consistent and precise calibration of both detectors. Additionally, as some proposed algorithms rely on simulated signals, the definition of the corresponding calibration units ("VEM" or "MIP") must be identical for measured and simulated signals. Here, the WCD calibration procedure governing how WCD signals are calibrated to units of VEM for both real measurements and simulations is reviewed. Then, an accurate, concise definition of a "minimum ionising particle" with respect to the Pierre Auger Observatory (Auger)'s SSDs is suggested, combined with a discussion of effects that systematic biases in the calibration may have on reconstructed muonic signals; which is discussed further in Chapter 8. To conclude, a calibration measurement using resistive plate chamber (RPC) equipment is proposed, with some additional comments on the ratio between the measured peak in the calibration charge histogram and the peak from vertical, through-going, and uniformly distributed charged particles in the scintillator surface detector (SSD).

5.1   Calibration

The calibration of signals is an important aspect of event reconstruction at Auger. Before AugerPrime, the Auger SD consisted of only water-Cherenkov detector (WCD) detectors.

WCDs measure signals that are calibrated to the vertical equivalent muon (VEM), as described in Section 3.1.2. The definition of a VEM is defined by carrying out measurements of the peak position of background particle charge histograms, often referred to as the muon hump, because this position is primarily comprised of muon traces. The measurement of background particles allows for a consistent and continuous calibration to be performed on each station every minute. The concise definition of the VEM corresponds to muons which are vertically through-going (they traverse entirely through top to bottom of the water) and central (they are incident at the top, central position of the tank). The three WCD photomultiplier tubes (PMTs) will measure similar signals for central muons. The peak position for vertical, central muons is at a different position than the peak position for background muons because of the differences in angular distribution. Since the peak positions differ, data must be corrected by the ratio between the two peak positions such that the definition of the VEM is defined by vertical, central through-going muons for all data used in event reconstruction. The addition of SSDs to Auger have created an additional calibration unit, the *Auger Minimum Ionising Particle (Auger MIP)*. The definition of the Auger minimum ionising particle (MIP) is defined by uniformly distributed vertical, through-going muons. That is, muons which are incident on any part of the SSD, always traverse through the entire SSD, and traverse vertically through the SSD. To summarise the differences, 1 VEM corresponds to the average charge from one vertical, central muon in the WCD and the Auger MIP will correspond to the average charge from one vertical muon incident at a uniformly distributed random location on the SSD.

Similarly to the WCD, the SSD muon hump positions are at different locations for uniformly distributed, vertical, through-going muons and omnidirectional background muons. In future studies such as mass composition analysis, WCD and SSD signal correlations or WCD and SSD lateral distribution function studies will involve combinations of the VEM and MIP. The addition of the MIP means that the VEM is no longer relative to itself; it is now crucial that the definitions of the VEM and MIP are well-defined, clear and precise with respect to each other. For example, as explained in Chapter 8, the matrix formalism method for extracting the muon content of SSD and WCD signals requires precise definition and calibration of the MIP and VEM.

5.1.1   Charge Histograms

The SSD and WCD charge histograms used for calibration are created by accumulating the integral of pulses from many incident particles traversing through each detector. For example, one vertical incident muon will traverse through the SSD and produce a signal pulse as in Figure 5.1. The peak of the pulse is found from the 25 ns time bin which contains the largest flash analog to digital converter (FADC) count, or equivalently, the largest number of photoelectrons when using simulations. The charge of a trace is calculated by integrating the pulse. As the peak and charge from many different pulses are accumulated, the peak and charge histograms are generated. An example of a charge histogram can be seen in Figure 5.2. Histograms of the peak are not explicitly used for calibration, however, they have other uses such as setting detector thresholds.

5.1.2   The Vertical Equivalent Muon Calibration Procedure

The signal of the WCD is calibrated to the average signal produced by a vertical and central through-going muon, known as a VEM. Obtaining a measured sample of muons satisfying this definition is not realistic for detectors in the field, so a histogram of the charges produced by all triggering omnidirectional atmospheric particle pulses is instead created. Then, the peak position of the charge histogram is corrected by a calculated factor to convert measured signals into the correct units of VEM. The correction factor may be obtained experimentally through the use of the Muon Auger RPC for the Tank Array (MARTA). The MARTA RPCs are composed of millimetre-thick gaseous volumes housing tetrafluorethane (R-134a), that are contained by highly resistive parallel plates. High-voltage electrodes are applied to these plates, creating an intense and uniform electric field. As ionising particles pass through the detector, electrons induce signals in specific readout electrodes. Processing of data provided from the readout electrodes for two or more RPCs, displaced vertically, can be used to identify vertically incident particles [69]. Two RPCs could be arranged such that there is one below the WCD and one above the WCD. Vertical, central muons (identified by vertically coincident measurements from the relevant readout electrodes) can then be selected from the background flux of particles. The RPC setup allows for the measurement of two charge histograms from the same background flux to be created, albeit with one histogram containing data only for vertically coincident measurements. These two histograms
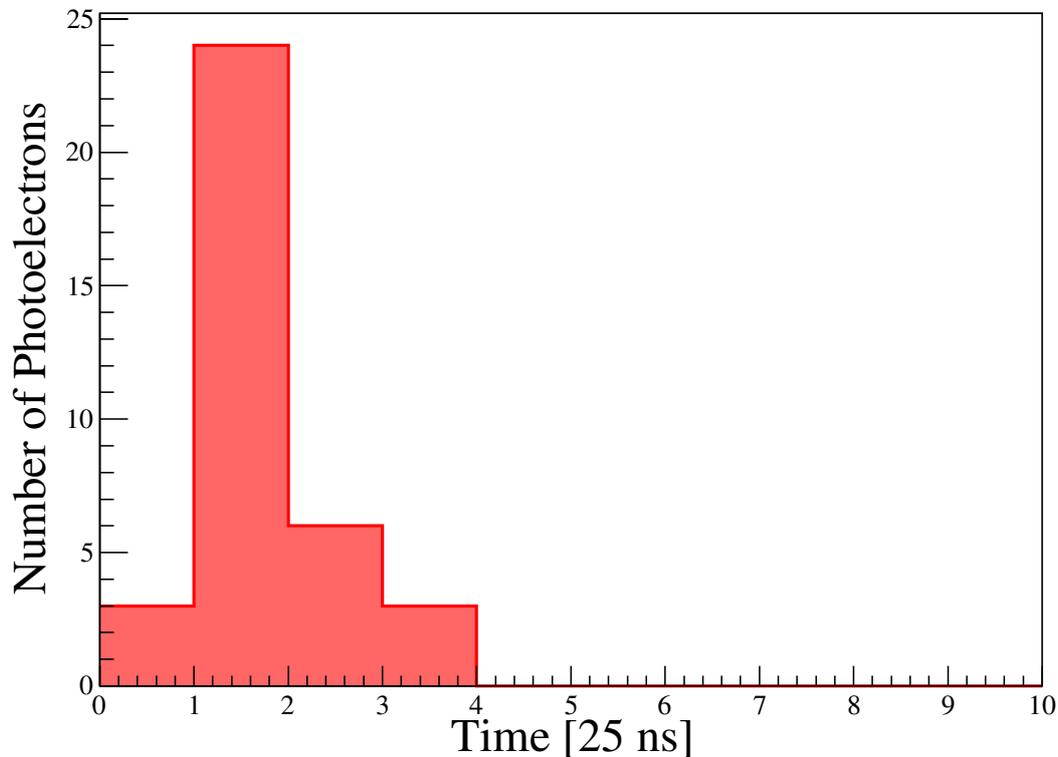
Figure 5.1: A simulated signal trace from one vertically incident muon on the SSD. The photoelectron count in the time bin with the largest signal is the "peak" value, and over many incident particles, accumulates to form the peak histogram. In this particular case, the peak is 24 photoelectrons, as seen in the bin corresponding to time between 25 ns and 50 ns. The charge is calculated as the integral of a pulse; which equates to 36 photoelectrons for this pulse. The charge histogram is conformed by accumulating the integral of pulses from many incident particles.

are used to calculate the correction factor needed to calibrate signals to the VEM unit. Previous experiments have been performed in 2003 and 2006, using a similar procedure with a triggering muon telescope which is composed of two small scintillator detectors to identify vertically incident particles. The results obtained in 2003 and 2006 can be seen in Table 5.1. There are currently no measured values for the SSD, but a proposal for how a measurement could be taken is proposed later.

As WCDs cannot measure a VEM, each individual WCD measures an omnidirectional background charge histogram every 60 seconds from background particle pulses. The back-
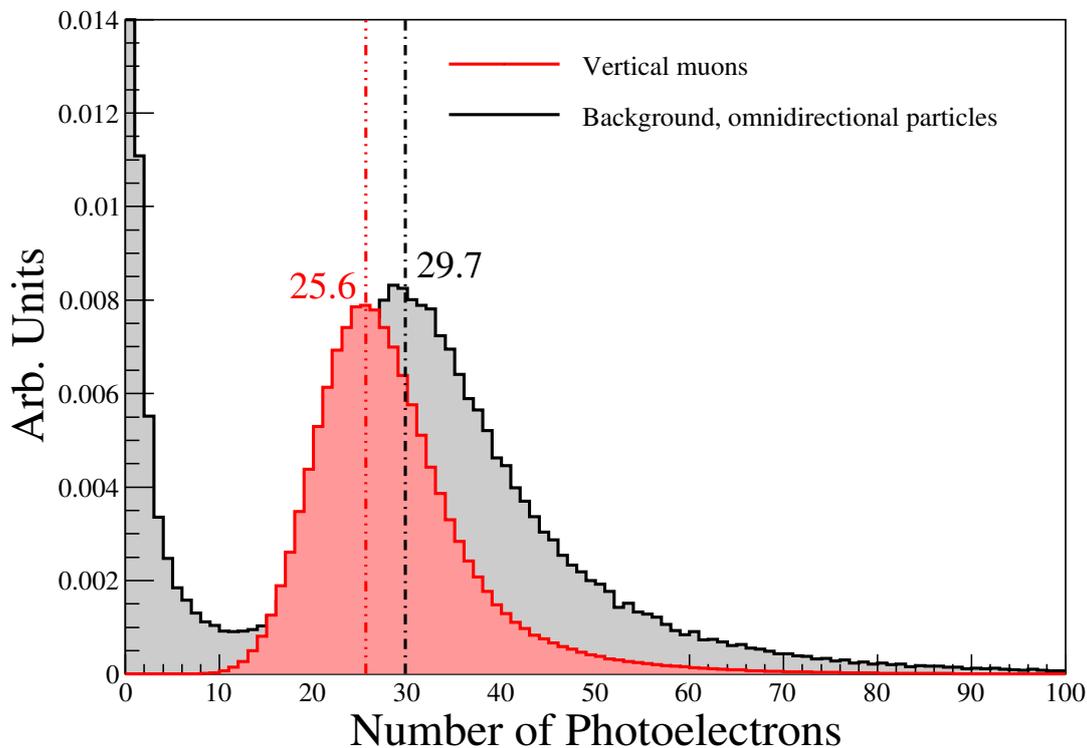
Figure 5.2: Calibration charge histograms from an SSD simulation. The peak positions are fit with a quadratic around the peak. The fitted peak position for the vertical, uniformly distributed and omnidirectional particles are 29.7 and 25.6 photoelectrons, respectively. This leaves the ratio between omnidirectional and vertical particle peaks at around 1.16 ± 0.02.

ground charge histograms are included within the data stream when the WCD triggers. During event reconstruction, the correction factor is applied to the charge histogram from station data, to calibrate to VEM units.

Figure 5.3 shows an example of the VEM calibration procedure completed using simulations, for a single WCD PMTs. The fitted peak position for the VEM and omnidirectional particles are 77.5 and 80.0 photoelectrons, respectively. This gives a correction factor of around 1.03; the simulation analysis for the WCD has been performed to verify that the procedure returns accurate results, as they can be directly compared to measurements from 2003 and 2006. Figure 5.4 shows the ratio between the peak positions of the omnidirectional charge histogram, when considering the total signal measurements from all three

		Ratio
Single WCD PMT	**2003**[70]	0.99
	2006[44]	1.03 +/- 0.02
	2019 (Simulation, Figure 5.3)	1.03 +/- 0.02
Threefold WCD	**2003**[70]	1.13
	2006[44]	1.09
	2019 (Simulation, Figure 5.4)	1.07 +/- 0.01
	2020[71]	1.08 +/- 0.01
SSD	**2003**	-
	2006	-
	2019 (Simulation, Figure 5.2)	1.16 +/- 0.02

Table 5.1: Table containing previous calibration factors obtained from measuring the differences between the peak positions of the vertical, central muon WCD charge histogram and the background, omnidirectional muon WCD charge histogram. There are currently no measured values for the SSD. Calibration procedures in $\overline{\text{Off}}\underline{\text{line}}$ are performed at the single PMT level. All results from 2019 were obtained from my own simulations.

PMTs (threefold). The WCD calibration in $\overline{\text{Off}}\underline{\text{line}}$ calibrates PMTs individually, where the ratio between the two peak positions is simulated to be 1.03, which is shown in Figure 5.3.

5.1.3 The Auger Minimum Ionising Particle Calibration Procedure

The definition of the MIP should be defined in a similar way to the VEM. Here, it is proposed that the Auger MIP be defined as uniform, vertical and through-going muons incident on the SSD. The peak positions from SSD charge histograms should be used to calibrate data to the Auger MIP using the same procedure as the WCD.

A measurement has yet to be taken in order to calculate this ratio for both unified board (UB) SSDs and upgraded unified board (UUB) SSDs. Note that in the future, all existing UB electronics will be upgraded to the new AugerPrime UUB electronics. A calibration procedure similar to the procedure used for the WCD is proposed in Section 5.3. Until this measurement is taken, a first calculation from my simulations is being used (by the Pierre Auger Collaboration) to correct incoming data.
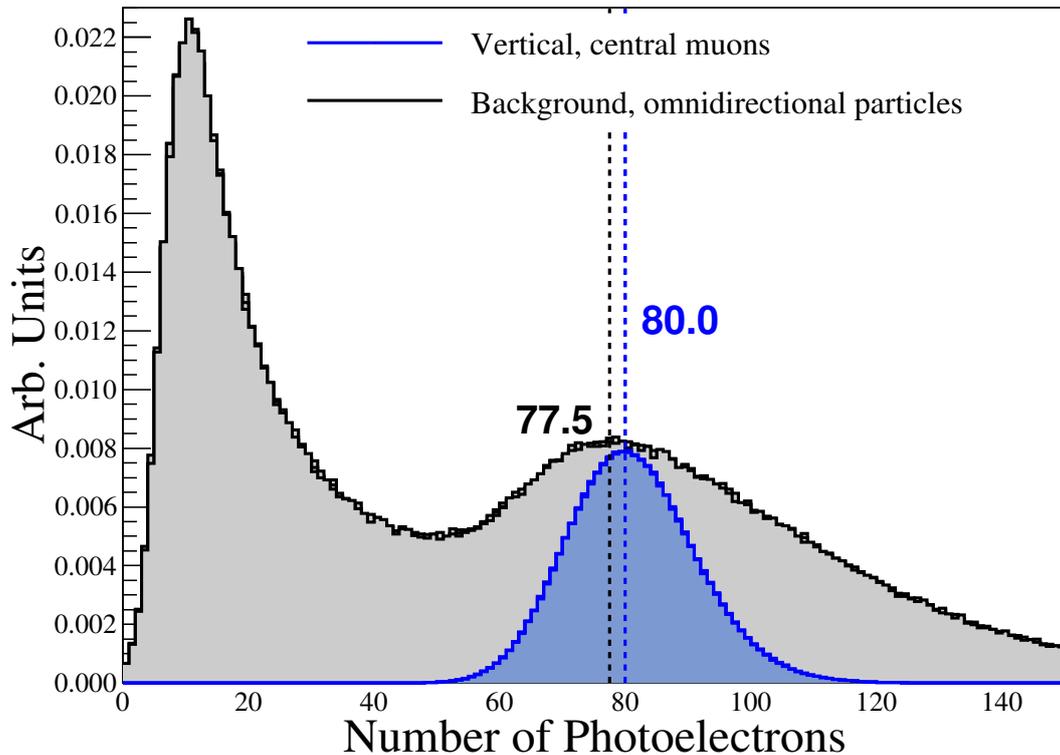
Figure 5.3: Calibration charge histograms for a single WCD PMT from my O̅ff̲l̲i̲n̲e̲ WCD simulation. Black line: charge histogram created from a simulated flux of omnidirectional background particles. Blue line: charge histogram created from the same simulated flux of omnidirectional background particles as the black line, however, only vertical through-going muons are selected. Peak positions are fit with second degree polynomials around the maximum. The fitted peak position for the VEM and omnidirectional particles are 77.5 and 80.0 ADC counts respectively. The ratio between the two peak positions is then calculated to be around 1.03. This simulated analysis can be performed using ADC counts, energy deposits or number of photoelectrons interchangeably as the ratio between the peak positions remains consistent.
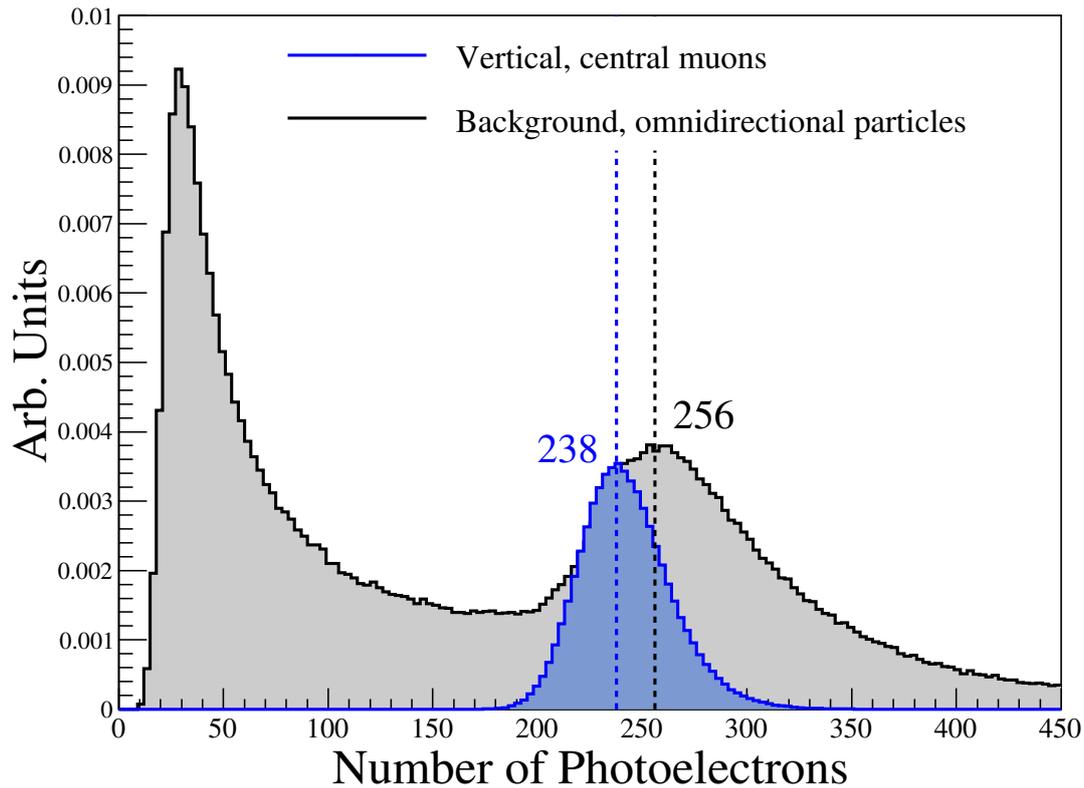
Figure 5.4: Charge histograms for the sum of the three WCD PMTs from my $\overline{\text{Off}}\underline{\text{line}}$ WCD simulation. Through the use of my simualtions, the ratio between the omnidirectional and vertical VEM peak is simulated to be 1.07. The threefold coincidence charge histogram may be useful for understanding and analysis, but it is not explicitly used during event reconstruction as calibration is performed for each PMT.
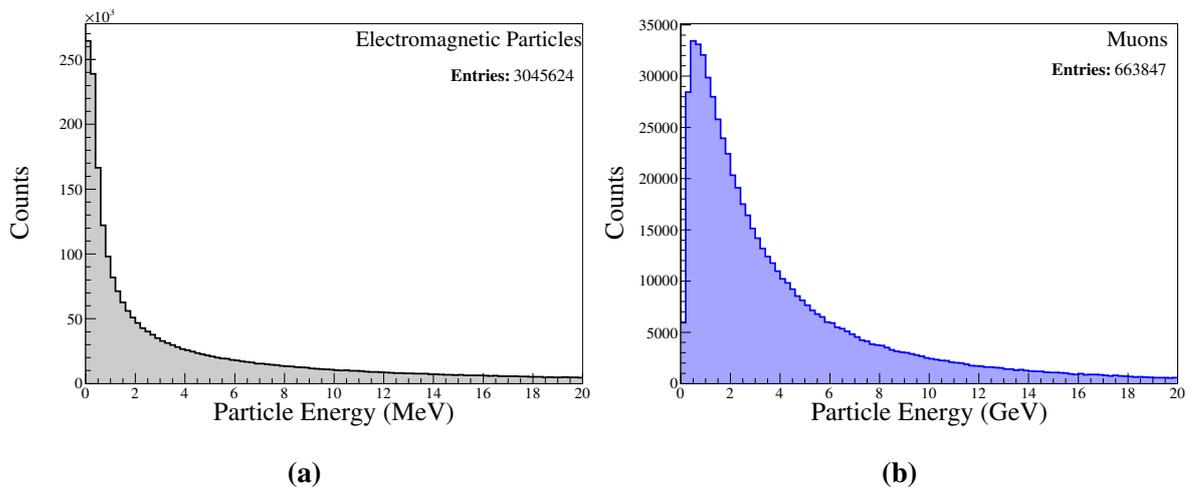
Figure 5.5: (a) The energy spectrum of one hour of background electrons and photons at Malargue. (b) The energy spectrum of one hour of background muons at Malargue. The simulations in this study use an extended version of these background particles which spans 24 hours.

5.2   Minimum Ionising Particle Calibration Simulation

$\overline{\text{Off}}\underline{\text{line}}$ contains packages and applications which simulate calibration histograms from injecting background particles, or vertical muons. The background particle energy spectrum for 1 hour of background flux at Malargue is shown in Figure 5.5.

I have updated and improved one of the $\overline{\text{Off}}\underline{\text{line}}$ applications to function with the new $\overline{\text{Off}}\underline{\text{line}}$ framework for SSDs. It takes an input background particle file, which is essentially a large list of particles and their energies, and then the application outputs omnidirectional charge histograms, similar to the histograms from real data, and is also able to produce a charge histogram for uniformly distributed, vertical through-going muons. The charge histogram for the vertical MIP is defined differently to the vertical, central muon WCD charge histogram. In the scintillator, it is not required to have central muons, and so vertical muons are injected uniformly across the scintillator casing. The peak of the omnidirectional and vertical MIP charge histograms were measured and compared (See Figure 5.2), and the ratio between the two was found to be 1.16 ± 0.02. This is much larger than the 2003 and 2006 measurements (shown in Table 5.1) for each PMT in the WCD. This factor needs to be taken into account when calibrating to the Auger MIP, which is done inside the $\overline{\text{Off}}\underline{\text{line}}$ reconstruction module `SdCalibrator`.

5.2.1 Effect of the Implied Water-Cherenkov Detector Trigger

In reality, the SSD operates in slave mode to the WCD because they do not have their own trigger. As the SSDs do not trigger by themselves, and will only record data if the WCD triggers. To reflect this in simulations, SSD charge values were only recorded if the WCD PMT pulses passed a threefold threshold trigger at $0.1 \times \mathrm{VEM}_{peak}$. Which is to say, the resulting trace from the summation of the three WCD PMTs must have a given 25 ns time bin where the number of photoelectrons exceeds $0.1 \times \mathrm{VEM}_{peak}$. For data, VEM_{peak} is with respect to the sum of the three PMTs, and is dynamically updated every 4 hours using online station histograms. For simulations, this value has been hardcoded to the VEM_{peak} calculated from 1000 vertical, central muons. The value was 91 photoelectrons, leaving the trigger at 9 photoelectrons. Regardless, Figure 5.6 shows that no difference in the SSD charge histogram peak position was found due to having an implied WCD trigger.

5.2.2 Simulating Photoelectrons instead of FADC Counts

The histograms from the $\overline{\mathrm{Off}}\underline{\mathrm{line}}$ simulation are calculated from the number of photoelectrons detected by the PMTs in each detector. In both situations, the number of photoelectrons recorded for the histograms is the number of photoelectrons that would consequently be input to an electronics simulation; this ensures effects such as photon loss in fibres are taken into account. The ratio of the two peak positions will be the same regardless of the simulating number of photoelectrons or FADC counts, as there will be a constant factor describing how many FADC per photoelectron. This factor is approximately 1.55 FADC per photoelectron, but the factor is not used in this study. As proof of concept, a similar calibration procedure was quickly run using an improved version of the $\overline{\mathrm{Off}}\underline{\mathrm{line}}$ simulation application which included the electronics simulation; the improvements were developed after the completion of this study. The improved application can output histograms in FADC, instead of number of photoelectrons. The two SSD calibration histograms, vertical and omnidirectional from the improved simulation can be seen in Figure 5.7, where the ratio between the two peaks is calculated to be approximately 1.15, and agrees with the original result of this work (Figure 5.2).
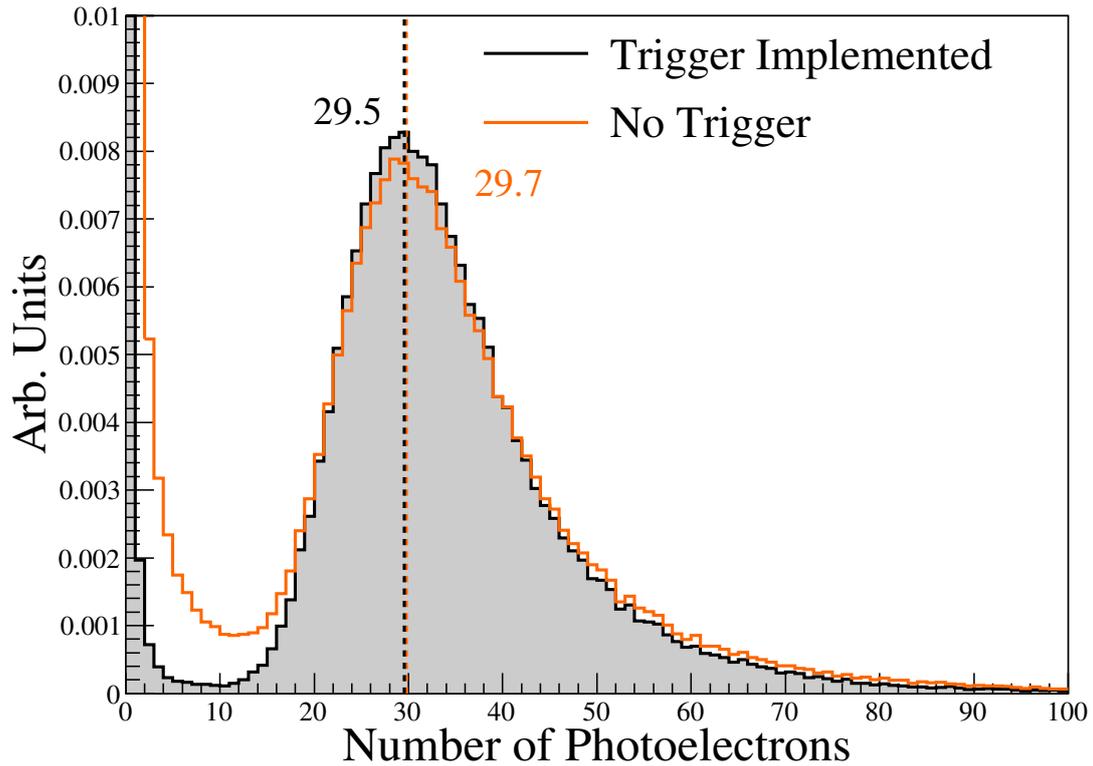
Figure 5.6: Charge histogram for the SSD PMT with an implied trigger of 10% of the VEM$_{peak}$, calculated from 1000 vertical, central muons (9 photoelectrons). It was found that the WCD trigger criteria does not affect the peak position of the SSD charge histograms. However, as expected, there is a noticeable effect of the trigger filtering out low numbers of photoelectrons.
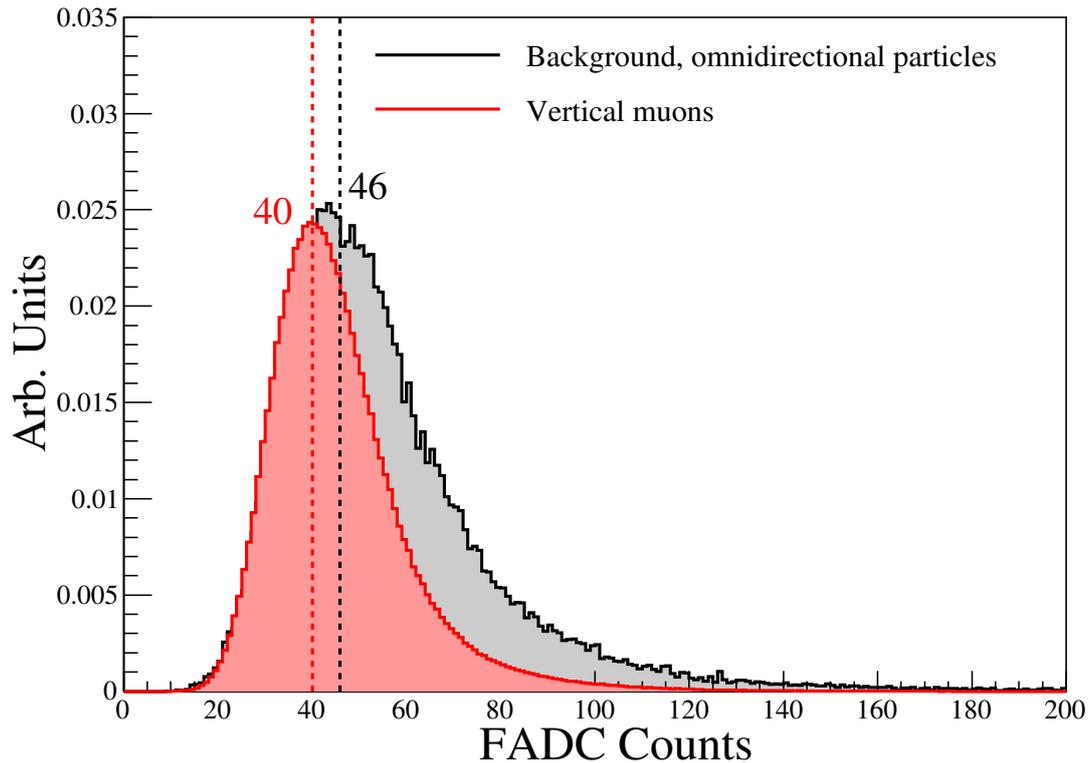
Figure 5.7: Calibration charge histograms from the more recently developed $\overline{\text{Off}}\underline{\text{line}}$ simulation which can output charge as FADC counts. The peaks of the histograms are calculated by taking the maximum value of a quadratic fit around the maximum value of each histogram. The ratio between the fitted peaks of the omnidirectional and vertical charge histograms is 1.15, consistent with what is seen in Figure 5.2.

5.3   Proposed Measurement

An omnidirectional to vertical correction factor has been derived using simulation procedures, where the WCD simulated correction factor is within range of the WCD measured correction factor, however, it is still preferable to obtain a measured factor for the SSD to use in event reconstruction. There is a clear variance in the measured values obtained in the WCD. The variance may be attributed to ageing of the detectors, such as the impact due to reduced liner reflectivity, or reduced water quality. The SSD may suffer similarly due to ageing effects which affect the scintillation light yield. The best approach would be to perform calibration measurements on a regular basis.

Through discussion with the the Laboratory for Instrumentation and Experimental Particle Physics (LIP) group in Portugal, a possibility of two measurement options were agreed on, with a possible third measurement discussed for future application. Extra care needs to be taken when considering the two measurement options, A and B, expressed in Figure 5.8 because they affect the definition of the *Auger MIP*. It has been suggested that MARTA[69] is to be used for the measurement, as it is based on RPCs.

RPCs should cover the entirety of one of the two scintillator modules because the scintillators are designed such that the same distribution of particles incident on either module will produce the same signal. This allows the use of RPCs which have less surface area than the SSD. There is another viable alternative if it is difficult to obtain large enough RPCs, and that is to measure vertical coincidence at a particular position on the SSD that corresponds to the mean signal of the scintillator module. The position which corresponds to the mean signal of the entire scintillator module is shown in Figure 5.9. This would require that the RPCs need only to be larger than the width of a scintillator module (1.2 m), rather than the length as well. Performing the measurement over the specific distance which corresponds to the mean signal, will produce a charge histogram with the same peak position as a measurement over the entirety of one of the two scintillator modules. In both cases, the measurement should, at the very least, contain enough statistics to mimic the online calibration of the WCD. If measuring over the entirety of one of the scintillator modules, this should be a background flux of particles of over 60 seconds or more.

Whichever size of RPC setup is used, it is important to ensure that a *uniform vertical MIP* is being measured by covering both scintillator modules with RPCs, or, covering one scintillator module with RPCs, or, covering one scintillator module at the key location with
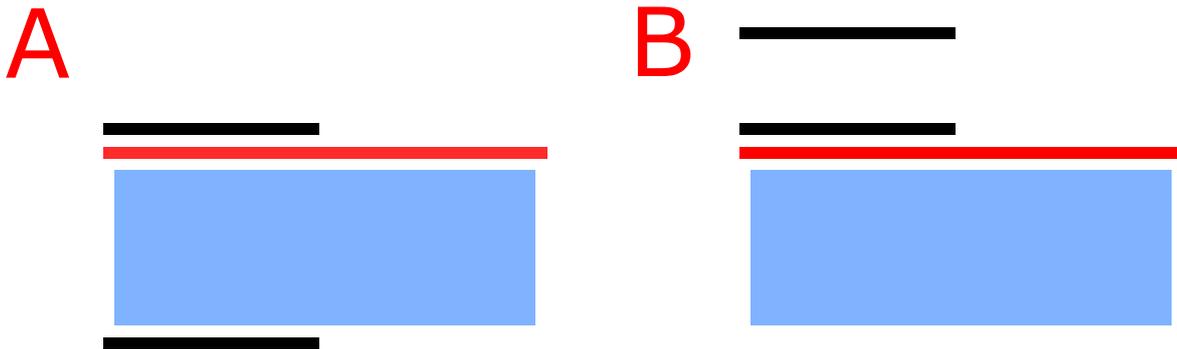
Figure 5.8: Two different proposals provided to the LIP group for the calibration measurement. Option A: This proposal involves the use of two MARTA RPCs; one is placed above the SSD, covering the entirety of one scintillator module, and the other is placed underneath the WCD. Option B: The two MARTA RPCs are placed above the SSD and WCD, covering the entirety of one scintillator module. The resolution of vertical coincidence will improve with increased distance between the two RPCs.

RPCs. Unfortunately, depending on how precise the Auger MIP is desired to be, the proposed measurement options A and B will change how it is defined. The first proposal, option A, involves the use of two RPCs. One RPC is placed above the entirely of one scintillator module, the other RPC is placed underneath the WCD. Here, the RPCs will measure vertical coincidence of particles which travel through the full height of the SSD and the WCD. The vertical coincident histogram here implies an energy cut on the measured particles. That is, the particles must have sufficient energy to traverse through the scintillator and 1.2 m of water. The definition of the Auger MIP through option A will imply an energy threshold of at least ≈ 240 MeV. It may be questionable whether or not this is truly a MIP.

The second proposal, option B, consists of both RPCs being placed above the scintillator. This method will not imply such a large additional energy threshold on each particle like in option A. Measuring the vertical coincident histogram through option B leaves us with a definition of the Auger MIP which corresponds to *uniformly distributed vertical minimum ionising particles*. To be clear, either option will work fine, but the definition of the MIP will change depending on which option is chosen.

LIP collaborators have declared option A as the most simple because it is a translation of the traditional WCD calibration procedure. Option B has been stated to be more mechanically difficult but achievable with a 50 cm separation between the RPCs. It is noted that an increased distance between the two RPCs improves the resolution of vertical coincidence.

Figure 5.9: Number of expected photoelectrons as a function of position along the SSD length. This plot has been created using muon tower data from Karlsruhe Institute of Technology (KIT), as well as the $\overline{\text{Off}}\underline{\text{line}}$ simulation. The blue line corresponds to the position where the measured number of photoelectrons at that distance is equal to the mean number of photoelectrons measured across all distances. From [72].

Unfortunately, due to the worldwide pandemic from 2019 to present, the LIP team has been unable to travel to the Pierre Auger Observatory to perform the necessary preparations and measurements for any calibration procedure. The measurement will be performed sometime in the near future.

5.4   Conclusion

For a reliable interpretation of the ratio between the two sub-detector measurements, it is essential they are expressed in units (VEM and MIP) with exactly the same definitions in simulations as in data. Analysis such as the matrix formalism method will require the use of accurately defined simulation-derived parameters. Any analysis which uses SSD signals will also require a precise and consistent definition of the Auger MIP to be compared. As such, simulations have been performed using a modified version of the $\overline{\text{Off}}\underline{\text{line}}$ simulation packages to accurately estimate the calibration factor required to convert the peak position of the online station background, omnidirectional charge histograms to a corresponding unit *Auger MIP*, defined as *uniformly distributed, vertical, through-going muons*. The correction factor was found to be 1.16 ± 0.02, and has been implemented in the current $\overline{\text{Off}}\underline{\text{line}}$ calibration packages.

Chapter 6

SSD Signal Uncertainty Model

At the the Pierre Auger Observatory (Auger) Observatory, the integrated signal of an SD station is composed of the energy deposit by muonic and electromagnetic components of an air shower. Previous studies have shown that the uncertainty of the signal measurement can be described as a Poisson-like term with a normalisation constant that depends on the zenith angle of the primary cosmic ray particle. This dependence reflects the increasing contribution to the signal of the muonic component of the shower. Both the muon/electromagnetic particle ratio, and the muon track length (and therefore signal) increase with zenith angle [73]. The previously suggested Poisson-like term for a given detector signal S, is written as:

$$\sigma = f(\theta) \times \sqrt{S} \qquad (6.1)$$

with the *spectral factor*, $f(\theta)$ describing a linear relation of the secant of the zenith as

$$f(\theta) = a(1 + b(\sec\theta - \sec 35°)) \qquad (6.2)$$

The signal S is given in units of VEM for the water-Cherenkov detector, and in units of MIP for the scintillator surface detectors. Completion of the AugerPrime upgrade will attach an scintillator surface detector (SSD) to all water-Cherenkov detector (WCD) stations; each air shower event which contains WCD triggered stations will have an accompanying SSD signal for the same station. These signals can be used to fit unique lateral distribution functions (LDFs), which describe the expected signal of the respective detector as a function of its

distance to the shower axis. The LDF is used to reconstruct important properties of cosmic rays, such as their energy and shower core position.

The uncertainty of the signal measurement is needed for use as a weight in the LDF fit, for both the WCD and SSD. Doublets and triplets were originally used to derive the WCD signal uncertainty model as in [73]. Doublets and triplets are SD stations that have other stations within a distance of $\approx$15 m, providing more than one independent measurement of signal in the same location. In the present work, the array with 750m spacing (SD-750) is used, it is a subset of the larger array with 1500m spacing (SD-1500), and has a sensitive energy range of $10^{17.5} - 10^{18.5}$ eV. Figure 6.1 shows the location of SSD multiplets in November 2019.
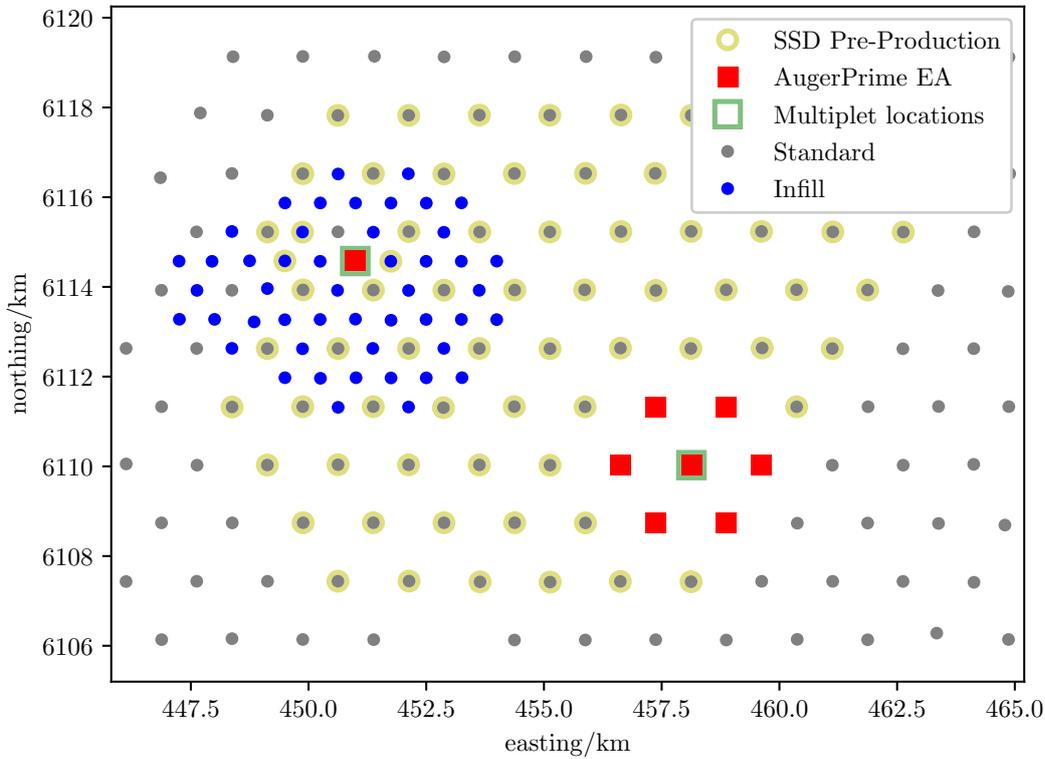
Figure 6.1: A map containing a portion of the Pierre Auger Observatory detector array. The SD-750 grid is represented by the blue stations inside of the standard 1500m grid. The two multiplet locations can be seen in more detail in Figure 6.2, where the left multiplet corresponds to a set of 5 nearby stations, and the right multiplet corresponds to a set of 3 nearby stations.

6.1   Signal Uncertainty Model

The SSD signal uncertainty model has been studied previously, and was found to be of the form described earlier, in Equation 6.1. The signal uncertainty model for the SSD was firstly derived in 2018 [74] by means of simulating doublet stations at 1000 m from the shower core and at azimuthal angles of 90° and 270° in the shower-plane. This was done to reduce bias introduced by attenuation effects, as particles at these angles would traverse through similar amounts of atmosphere. These doublet stations were labelled *pseudo doublets* as they do not accurately resemble the ≈15 m spacing between doublet stations at Auger. The SSD signal uncertainty model was revisited in 2019 [75] and refined using a slightly different approach. The main changes were in the simulation framework, which included updated material properties of the SSD. A probability density function represented as a positive-side normal distribution (or "truncated" normal distribution) was used to describe the probability of observing a signal *S*, and a different likelihood function was used.

The SSD signal uncertainty model has been revisited during the present work with some significant changes. This work analyses a series of doublet stations from *real* data to derive the signal uncertainty model, as opposed to using *pseudo doublets* from simulations in previous works. The uncertainty of the signal measurement is determined with doublet station pairs from the quadruplet set of SSD stations inside the SD-750. An LDF approximation derived from simulations is used to apply a correction factor to bias on the uncertainty. This bias is introduced by the steepness of the LDF, which is imposed by a non-zero distance between stations in a multiplet. Correcting for the bias is necessary to help determine the parameters of a data-driven uncertainty model.

6.2   Data Set

All SSD stations operate in slave mode to the WCD. SSDs cannot trigger themselves, and will transmit all their data when a WCD triggers, even if no particles were detected. Operation in slave mode incurs an implicit trigger threshold on the SSD which is directly related to the WCD trigger. An unbiased SSD data set will require threshold cuts on expected values of signal (vertical equivalent muon (VEM)) that are above the WCD trigger threshold, otherwise, results will be biased with upward signal fluctuations.

The following cuts were applied on the SSD SD-750 data set:

- Data from Apr 2019 - May 2021 are used. Data ranging from 17th Feb 2020 to 20th May 2020 is excluded, as the SSD doublet pairs were out of commission.

- Smaller, unknown periods where SSDs have failed to transmit data, or have been out of commission are avoided by only considering stations in the multiplet which have non-zero SSD signal (minimum ionising particle (MIP)). This can include integrated baseline.

- WCD "bad periods" are excluded, as they are known to have issues with event reconstruction. "Bad periods" are period identified by the collaboration where there are clear issues with data.

- All events must have a successful WCD event reconstruction and must have also passed the 6T5 WCD trigger, to exclude low quality events.

- To ensure full efficiency during event reconstruction, the zenith angle range (from WCD event reconstruction) has been limited to $[0°, 55°]$. Below $10^{17.8}$ eV, the range has been limited to even lower angles according to the equation below from [76]

$$\sec(\theta_{cut}) = 1.29 \cdot \log_{10}(E/\text{eV}) - 21.17. \tag{6.3}$$

- Stations must not have low-gain saturated signals, or have low-gain recovered saturated signals. Saturated SSD stations should be avoided until a method to recover saturated SSD signals has been implemented in event reconstruction.

- The stations potentially used for the LDF fit must be candidates, to exclude random and lonely stations. Random and lonely stations are stations which are included in the data, but may not be part of the event. They are usually large distances from the reconstructed shower core position.

- The expected signal (from the SSD LDF) must be above 10 MIP to avoid WCD trigger bias, and to guarantee that the signal distribution is approximately Gaussian. See below for more detail.

- The distance to the shower core from the multiplet stations must be greater than 200 m to minimise the effect of bias in signal due to the steepness of the LDF.
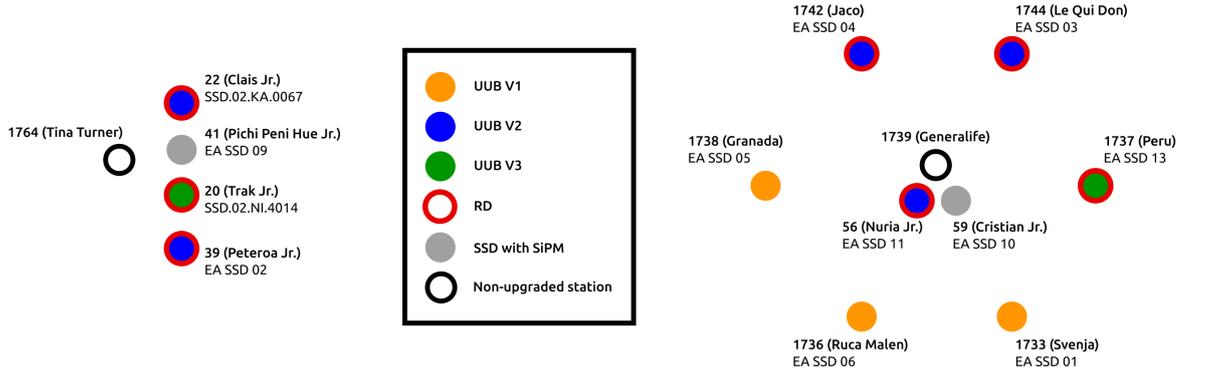
Figure 6.2: Map of the quadruplet and doublet series of SSD stations inside the SD-750 and Engineering Array. From [77]

.

6.3   Analysis Method

The analysis method used is based on a signal uncertainty study used for the WCD [76]. The map in Figure 6.2 shows a quadruplet series of SSD stations close together in the SD-750, as well as a doublet series in the Engineering Array, each of these series contains a mixture of upgraded and non-upgraded stations. Nuria Jr. and Cristian Jr. could be used as a doublet, however Nuria Jr. has a standard SSD while Cristian Jr. has an SSD with a silicon photomultiplier. The vast majority of SSDs being installed for AugerPrime will be standard, and so, this doublet pair is not considered. In the quadruplet series, stations Clais Jr., Pichi Peni Hue Jr., Trak Jr. and Peteroa Jr. have SSDs, while Pichi Peni Hue Jr. contains an SSD with a silicon photomultiplier. In this work, two doublet pairs consisting of Clais Jr., Trak Jr. (separated by 18.0 m) and Trak Jr., Peteroa Jr. (separated by 11.6 m) are considered.

When analysing doublet stations in an event where both WCDs have triggered, there are two independent measurements of SSD signal at nearly the same distance from the shower axis. With the assumption that the SSD signal measurements are Gaussian distributed and that the distributions of both signals are identical, the signal uncertainty can be extracted by subtracting one signal from the other. That is, the probability distribution function for the two signals S_1 and S_2, is a normal function with variance σ^2, that can be represented as

$$\rho(S_1) = \rho(S_2) = N(\langle S \rangle \, ; \sigma^2) \qquad (6.4)$$
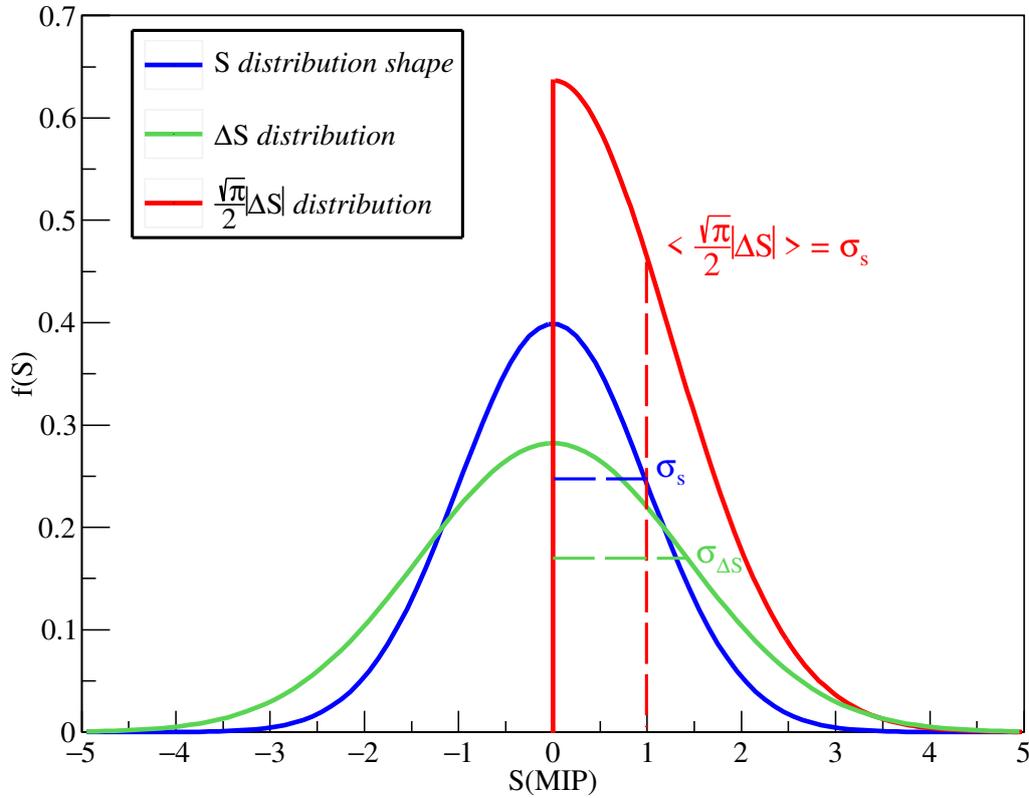
Figure 6.3: Distributions of ΔS (green) and $\frac{\sqrt{\pi}}{2}\Delta S$ (red) for assumed Gaussian distributed S (blue).

Note that, S_1 is the SSD signal from one station of the doublet and S_2 is the SSD signal from the other. The difference of the two signals, labelled ΔS, is simply

$$\Delta S = S_1 - S_2 \tag{6.5}$$

Then, we form the distribution of ΔS by taking the difference of the two Gaussians, to create a new distribution with variance $2\sigma^2$. The PDF of ΔS is written as

$$\rho(\Delta S) = N(0; 2\sigma^2). \tag{6.6}$$

If the absolute value of ΔS is used instead, the distribution is a half-normal distribution which can be used to estimate the uncertainty of S[1]. The uncertainty of S can then be obtained from the expected value of $|\Delta S|$:

$$\begin{aligned}
\langle |\Delta S| \rangle &= \int_0^\infty x \frac{\sqrt{2} e^{\frac{-x^2}{4\sigma^2}}}{\sqrt{2}\sigma\sqrt{\pi}} dx \\
&= \frac{1}{\sigma\sqrt{\pi}} 2\sigma^2
\end{aligned} \tag{6.7}$$

$$\langle |\Delta S| \rangle = \frac{2}{\sqrt{\pi}}\sigma, \text{ or } \sigma = \frac{\sqrt{\pi}}{2} \langle |\Delta S| \rangle \tag{6.8}$$

All distributions are depicted in Figure 6.3.

6.4   Systematic Difference in Signal of Multiplets

When analysing doublet stations, the fact that the stations of a multiplet do not have exactly the same distance to the shower axis must be considered; they are actually separated by approximately 15 m, depending on the pair being considered. This small distance causes the doublet stations to have different expected signals contrary to the assumptions above, which in turn, introduces a bias on the estimation of uncertainty.

The magnitude of this bias increases with the steepness of the LDF. There are multiple ways of minimising the effect of this bias. One approach is to remove all stations that are closer than 200 m to the shower axis, thus avoiding the steep inner part of the LDF. An alternative approach is to correct for the slope of the LDF using parameterisations of the LDF shape.

The data set used for this SSD signal uncertainty study is not as robust or large as the WCD data set. Due to smaller data set, both of the proposed solutions are implemented. All stations that are closer (to the shower axis) than 200 m are removed. A correction for the slope of the LDF is then applied using a parameterised LDF functional form derived from simulations.

In order to estimate the magnitude of the bias, the LDF shape for the SSD must be known. Prior to this work, there was no parameterised LDF functional form for SSD SD-750 data,

---

[1]The distribution which was used to create the half-normal distribution has variance $2\sigma^2$

and, even if there were, it would be challenging to fit an LDF. The challenge arises from the majority of events being low energy. Low energy events containing doublet pairs usually only contain the SSDs from the doublet pairs, with no other stations containing SSDs. This problem arises because all SD-750 stations do not yet have SSDs (as of 2022). In situations where there is only the doublet pair of SSDs in the event, one cannot easily fit an LDF.

In Chapter 7, the functional form of an SSD LDF for SD-750 has been parameterised using simulations and a previous signal uncertainty model. The parameterisation of the SD-750 SSD LDF shape requires the SSD shower size at 450 m and the zenith angle (θ) of the shower. In standard reconstruction procedures, the shower size would be fit. However, without sufficient multiplicity of SSDs, the only way to estimate the SSD shower size is from the WCD shower size. To help with this, the SSD shower size has been parameterised as a function that is quadratic in shower zenith angle and quadratic in WCD shower size; the study of this parameterisation is outlined in Chapter 7. The functional form can be expressed as

$$
\frac{S_{\text{SSD}}}{S_{\text{WCD}}} = (a + S_{\text{WCD}}(b + cS_{\text{WCD}})) + (d + S_{\text{WCD}}(e + fS_{\text{WCD}}))\sec(\theta)
$$
$$
+ (g + S_{\text{WCD}}(h + iS_{\text{WCD}}))\sec^2(\theta) \quad (6.9)
$$

where S represents $\log_{10}$ of signal at 450 m for either the SSD or WCD. The parameterisation for SSD shower size and SSD LDF shape allows an estimate of the LDF to be calculated for events that have only one SSD station (or doublet). Some examples of approximated LDFs, can be seen in Figure 6.4 and more detail can be seen in Chapter 7.

An LDF can help correct for the systematic difference in signals between doublet stations at distances r and $r + \Delta r$ to the shower core. To consider signals without applying a correction, $|\Delta S|$ from Equation 6.8 is evaluated as

$$
|\Delta S| = |S_1(r_1) - S_2(r_2)| \quad (6.10)
$$

where r is the distance to the shower axis, in the shower plane. Consequently, the distance between the doublet stations Δr is calculated simply by taking the difference of the two distances.

$$
\Delta r = r_1 - r_2 \quad (6.11)
$$
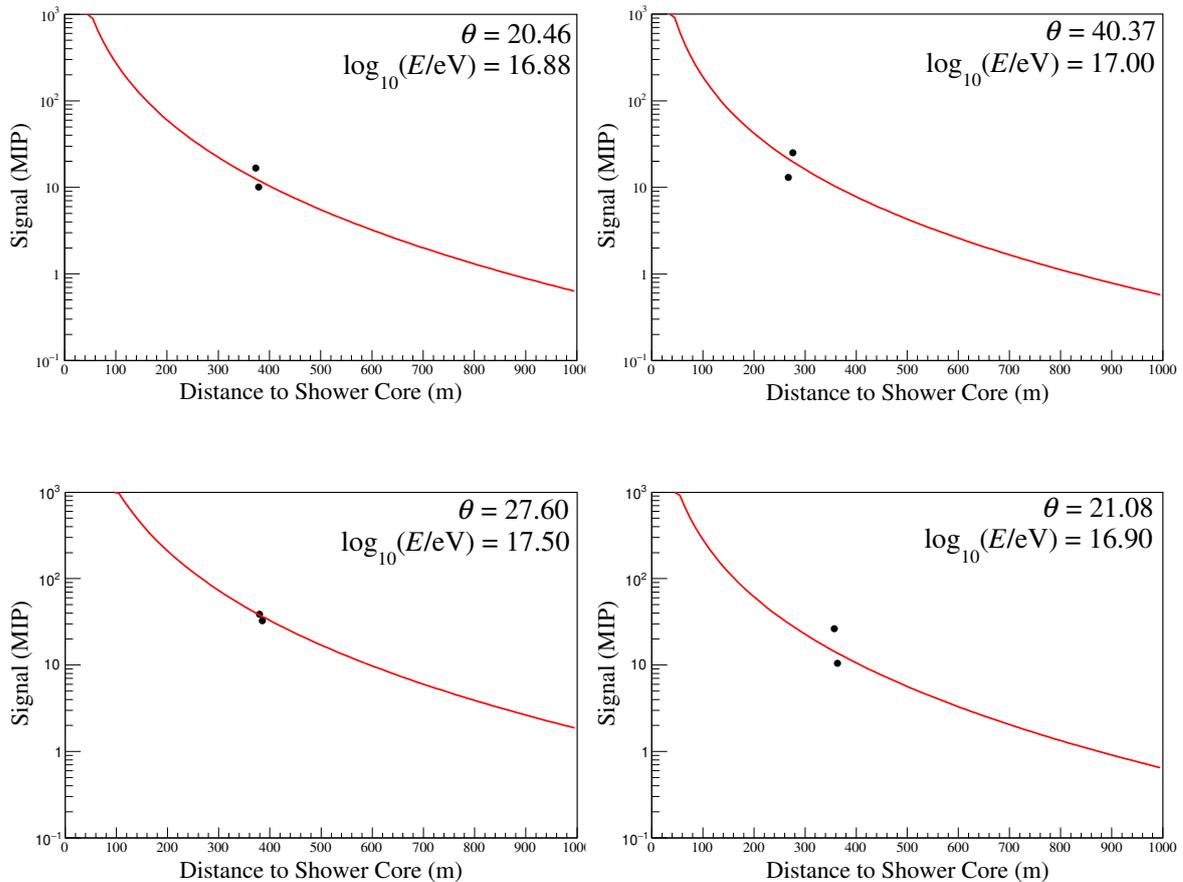
Figure 6.4: Examples of events which contain only one set of doublet SSD stations and no other SSD stations. The shape of the LDF is fixed, calculated from a parameterisation derived in Chapter 7.

Signals which are not corrected have a dependence on distance to the shower core, shower zenith angle, and the magnitude of the signal itself. Stations which have uncorrected signals introduce a bias into the uncertainty analysis because they break the assumption that doublet stations have the same distribution of signal. Strictly speaking, the slope of the LDF is significant enough to cause two stations, separated by distances as small as 11 m to measure unique distributions of signal. To minimise the effect of the spacing between stations, a correction using the slope of the LDF is applied using Equation 6.12.
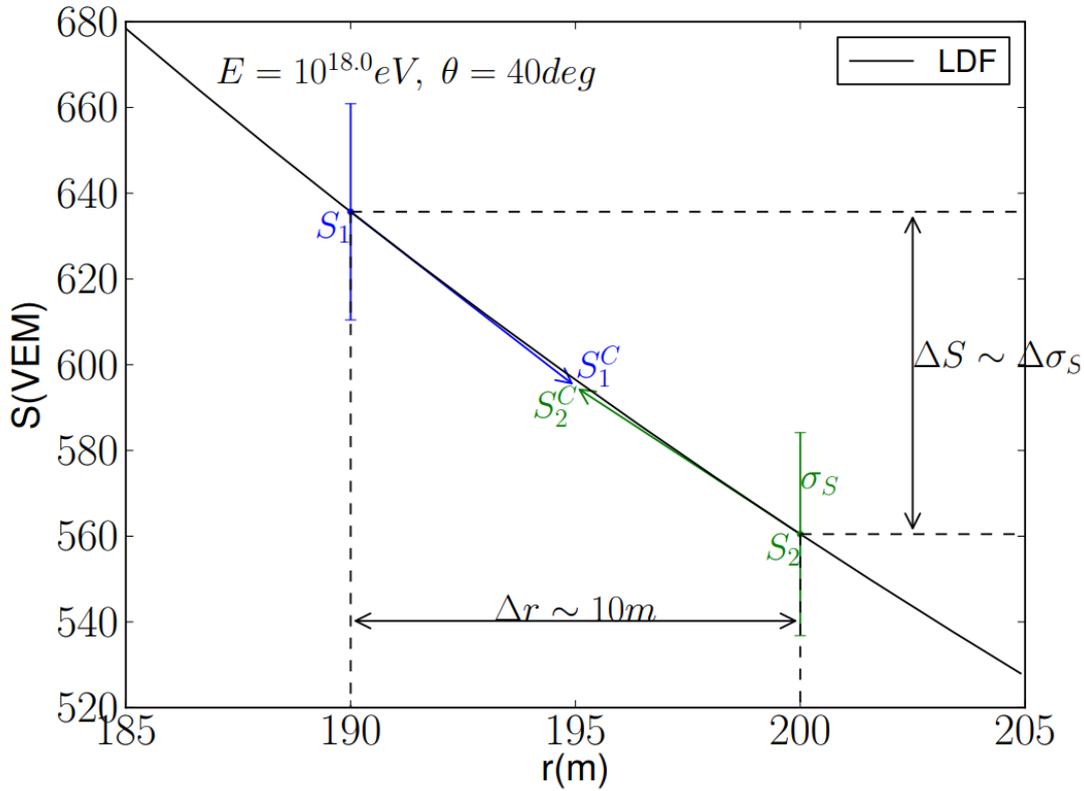
Figure 6.5: The signal correction based on the LDF for the WCD. The correction for the SSD signal is equivalent, but instead uses the SSD LDF. From [76].

Each signal can be corrected by calculating the signal at the mean distance to the shower core for both stations in the doublet using

$$(S_1^C, S_2^C) = (S_1, S_2) \mp \frac{dS_{LDF}}{dr} \cdot \frac{1}{2}\Delta r \tag{6.12}$$

See Figure 6.5 for an illustration of the procedure for correcting WCD signals. The procedure is equivalent for the SSD, but the SSD LDF is used instead. The correction significantly reduces bias in the uncertainty. Section 6.5 contains more detail and figures before and after applying the LDF correction alongside each other, demonstrating the magnitude of the bias. The uncertainty bias could be further reduced with a more accurate understanding of the SD-750 SSD LDF, fitted through multiple stations. The reduction of signal bias for the doublet stations used in this analysis as a function of Δr can be seen in Figures 6.6a

(a) Signal Bias without LDF correction

(b) Signal Bias with LDF correction
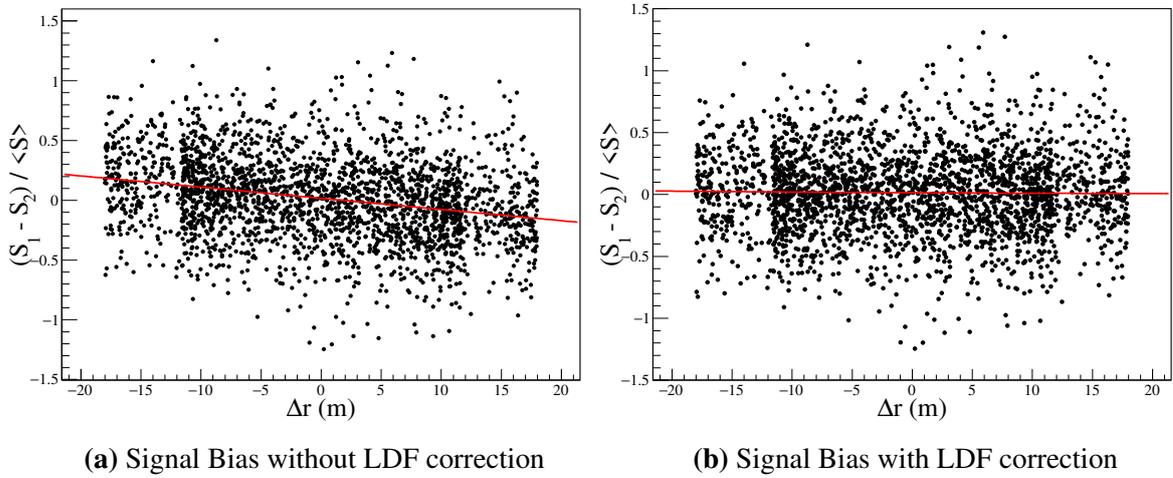
Figure 6.6: (a) Magnitude of the signal bias as a function of Δr (distance between the doublet pair in the shower plane). This bias increases significantly as the distance between the two doublet stations increases. (b) Magnitude of the bias after correcting for the distance between the two stations. The doublet pair stations are separated by 11.6 m and 18.0 m.

and 6.6b. The bias increases significantly as the distance in the shower plane between the two doublet stations increases. There is a clear slope in Figure 6.6a that has been corrected significantly (as seen in Figure 6.6b) by applying the LDF correction.

6.5   Results

The data outlined in Section 6.2 was used to create all of the following results in this chapter. Figure 6.7 shows the development of the spectral factor $f(\theta)$ (Equation 6.2 as a function of $\sec(\theta)$; it contains the final result from this study, as well as the most recent results from prior studies. The entirety of the procedure used for acquiring the result is as follows:

- For each reconstructed event present in the data described in Section 6.2. The signals from doublet pairs corresponding are recorded and stored as values corresponding to S_1 and S_2, alongside their distance to the shower axis, r_1 and r_2.

- The bias in signal due to the LDF slope is corrected by using a parameterised form of the SD-750 SSD LDF, through equation 6.12, using $\delta r = r_1 - r_2$. When performing
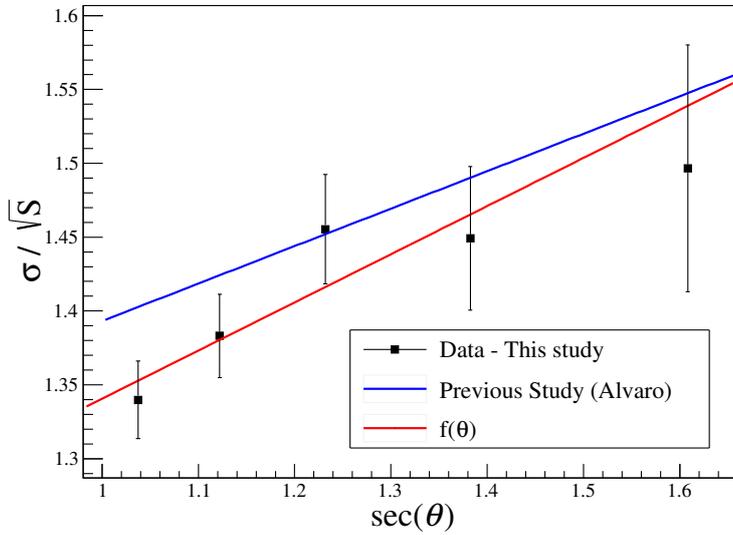
Figure 6.7: Spectral factor $f(\theta)$ vs $\sec\theta$ after correcting for the LDF slope. The blue line represents results obtained from a previous study using Dense Rings and simulations. The red line represents results of fitting the data (black points) in this study.

the correction, extra care was taken to ensure that the correction was moving in the right direction. That is, if r_1 were closer to the shower axis, then the corrected signal should be less than S_1. Figure 6.5 depicts this correction for signals S_1 and S_2.

- After correcting the signals, the average and the difference, as well as the uncertainty can then be calculated as outlined in the analysis method above, in Section 6.3. For the formulae to be accurately represented, multiple events must be considered as the results require that the data is binned. The data must be binned as the signal uncertainty is defined from the expected value of the difference of signals.

One of the assumptions made is that the uncertainty is roughly Poisson. Figure 6.8 shows the signal uncertainty against expected signal for different zenith angle bins. The zenith angle dependence can be seen, as well as the proportional increase by approximately $\sqrt{S}$, (S is the average signal of the doublet) supporting the assumptions made. Results are shown as the signal uncertainty divided by the $\sqrt{S}$, which allows signals from across different ranges to be combined. Similarly, Figure 6.9 shows the uncertainty as a function of signal, with different bins in r. Considering the data available, the radial bins were made quite small, and
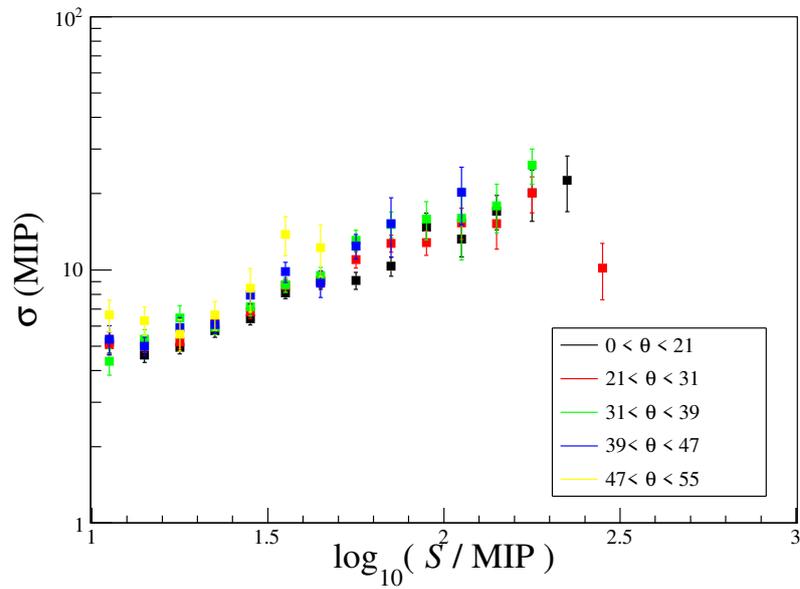
Figure 6.8: Signal uncertainty vs expected signal for different zenith angles bins.
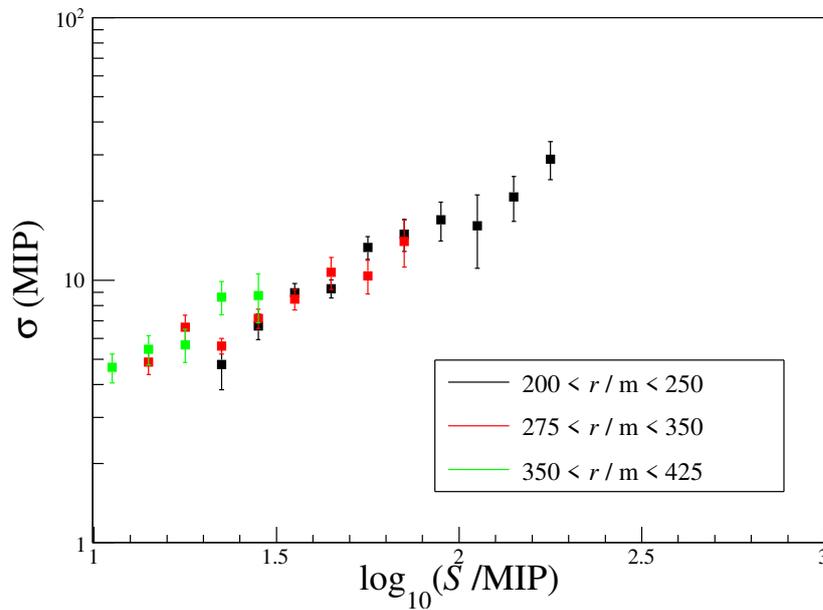
Figure 6.9: Signal uncertainty vs expected signal for different bins in r (distance to the WCD reconstructed shower core, in the shower plane).
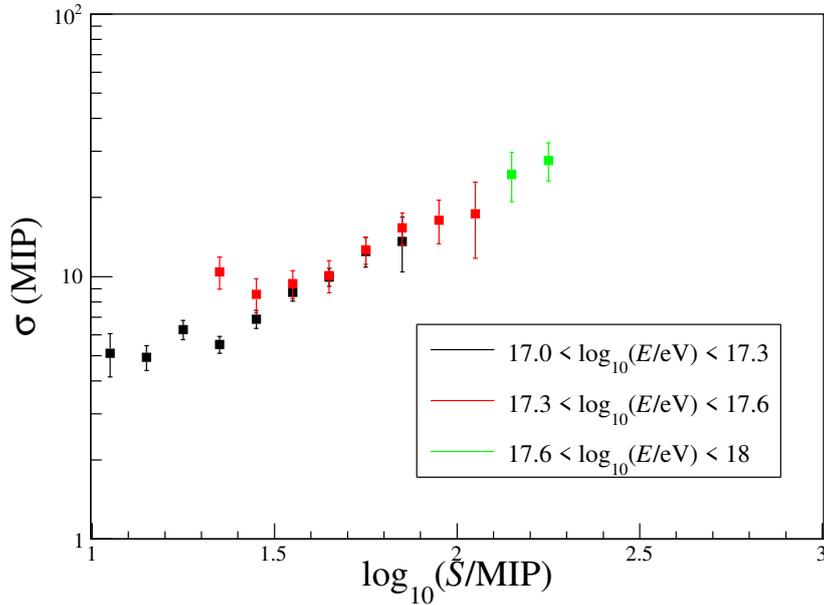
Figure 6.10: Signal uncertainty vs expected signal for different bins in reconstructed primary energy of the air shower.

there is no noticeable trends other than the known Poisson-like factor. Figure 6.10 shows the uncertainty as a function of signal, with different bins in reconstructed shower energy. There is also no noticeable trends other than the known Poisson-like factor. The derivation of the signal uncertainty model is actually the fitting of the spectral factor. Historically, the spectral factor has been fit as a function of zenith angle, as there have been no significant dependencies on signal nor distance to the shower plane. As confirmation that this is still the case, the spectral factor as a function of signal is shown in Figure 6.12. The expectation is that there will be no significant dependency on signal. When the LDF correction has been applied, there is no noticeable dependency. Likewise, for Figures 6.13 and 6.14, when the LDF correction has been applied, there is no noticeable dependency on distance to the shower axis and reconstructed energy.

The correction improves the accuracy of the signal uncertainty model, Figure 6.11 shows the model for data that has not been corrected with the LDF slope. When compared to Figure 6.7, it can be seen that the correction reduces the uncertainty at all zenith angles. Without the correction, the fitted function would not be close to previous results.

Parameter	a	b
$f(\theta)$	1.412 ± 0.017	0.231 ± 0.075

Table 6.1: Coefficients from fitting real data (LDF correction applied) with function $f(\theta)$ from Equation 6.2.

.
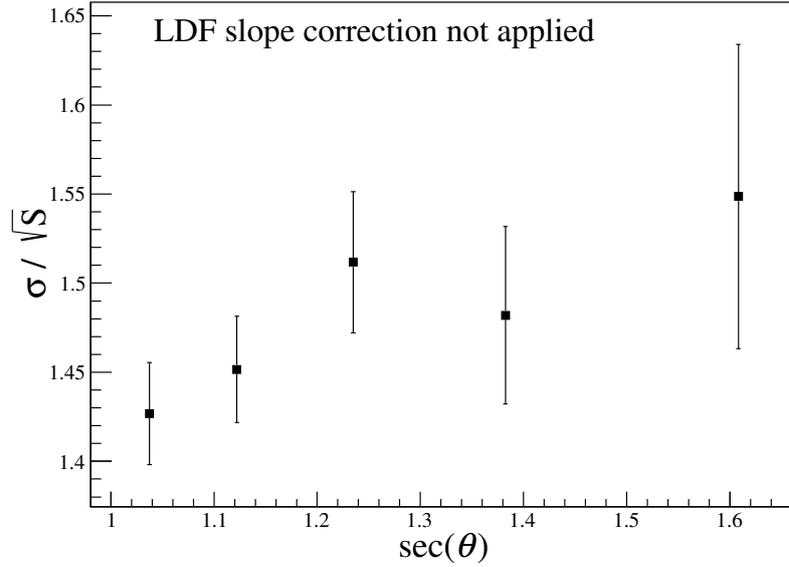
Figure 6.11: Spectral factor $f(\theta)$ vs $\sec\theta$ before applying the LDF correction. The uncertainty is significantly larger when the correction has not been applied.

Data is binned equidistant in $\cos^2\theta$ to provide better statistics for higher zenith angle bins. Error bars are drawn on each point, representing the standard error of the mean for each bin. The parameters a and b were obtained by performing χ^2 minimisation. The fit converges with χ^2/n_{dof} of 0.58 and parameters

$$f(\theta) = (1.41 \pm 0.017) \times (1 + (0.231 \pm 0.075) \times (\sec\theta - \sec 35°)). \tag{6.13}$$

Irrespective of the χ^2, a sloped line is fit based on knowledge acquired from previous results. Signals which are almost equal to 0 MIP have a larger spread compared to larger signals.
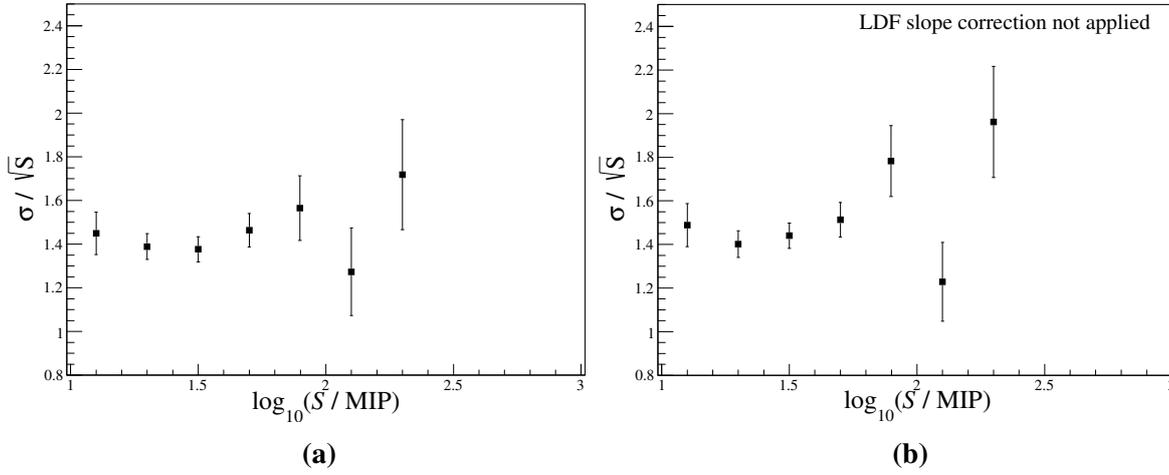
Figure 6.12: Zenith angle range considered here is $30°$ to $40°$. (a) Spectral factor $f(\theta)$ vs expected signal after applying the LDF slope correction. (b) Spectral factor $f(\theta)$ vs expected signal without applying the LDF slope correction. If the assumption that the signal uncertainty is proportional to $\sqrt{S}$ for a given zenith angle range is correct, then the expectation is that $\sigma/\sqrt{S}$ vs $\log_{10}(S/\text{MIP})$ will be flat. The correction applied certainly moves towards the right direction, but could be improved by further understanding of the SSD LDF.

The previously suggested $\sqrt{S}$ Poisson-like behaviour is confirmed, indicating that the uncertainty of the signal measurement arises primarily from the fluctuations in the number of particles hitting the SSD. Previous studies suggest that there could also be a possible dependency on the uncertainty due to energy and distance to the shower core. A dependency could be explained by the differing response from proportionate muonic and electromagnetic components of an air shower [75]. The uncertainty on energy and distance to the shower core has been investigated by binning data within a zenith bin of $30°$ - $40°$. Figures 6.9 and 6.13b show the dependence of the uncertainty on distance to the shower core. The limitations of the current data set prevent data being obtained for distances greater than $\approx$425 m. No significant dependence on distance to the shower core has been found for distances between 200 m and 425 m. Figures 6.10 and 6.14b show the dependence of the uncertainty on primary energy. The current data set lacks sufficient statistics for primary energy above $10^{17.8}$ eV. No significant dependence was found for primary energies ranging from 10^{17} - 10^{18} eV.

(a)   (b)
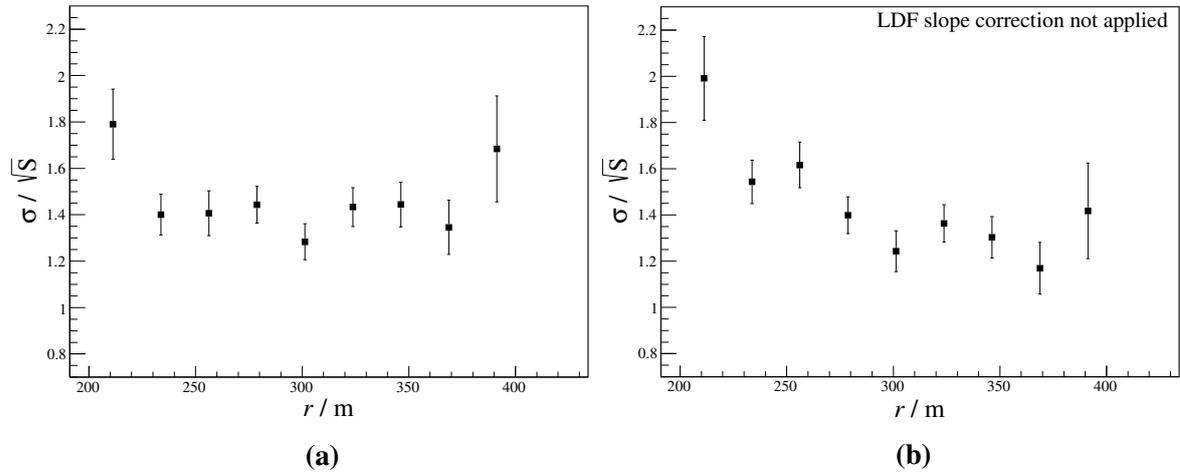
Figure 6.13: Zenith angle range considered here is 30° to 40°. (a) Spectral factor $f(\theta)$ vs r (distance to the WCD reconstructed shower core, in the shower plane). (b). Spectral factor $f(\theta)$ vs r without applying the LDF slope correction. There is a clear dependence on r before applying the LDF correction. Applying the correction has effectively minimised this dependence.

(a)   (b)
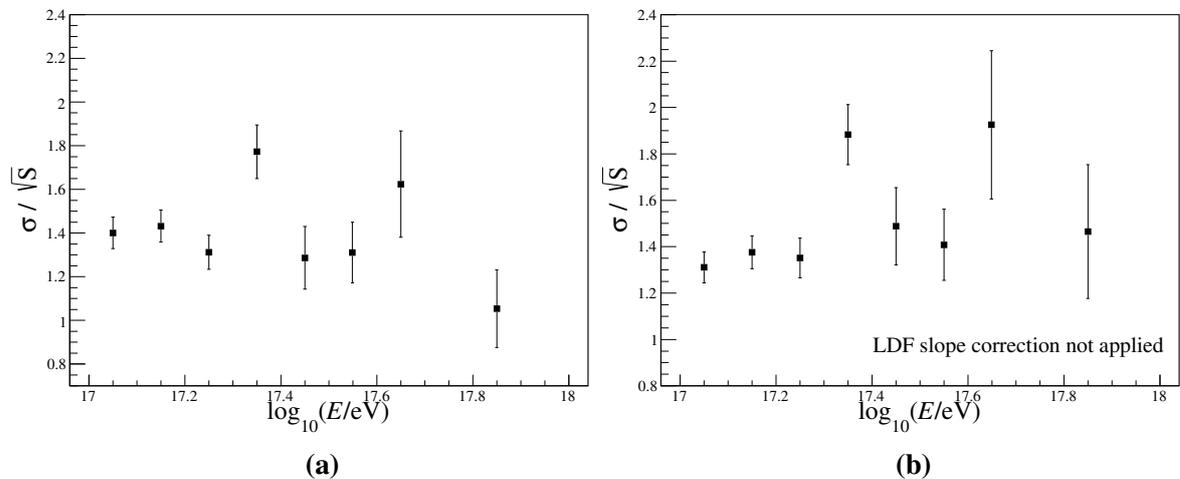
Figure 6.14: Zenith angle range considered here is 30° to 40°. (a) Spectral factor $f(\theta)$ vs reconstructed primary energy of the air shower after applying the LDF correction. (b) Spectral factor $f(\theta)$ vs reconstructed primary energy of the air shower before applying the LDF slope correction.
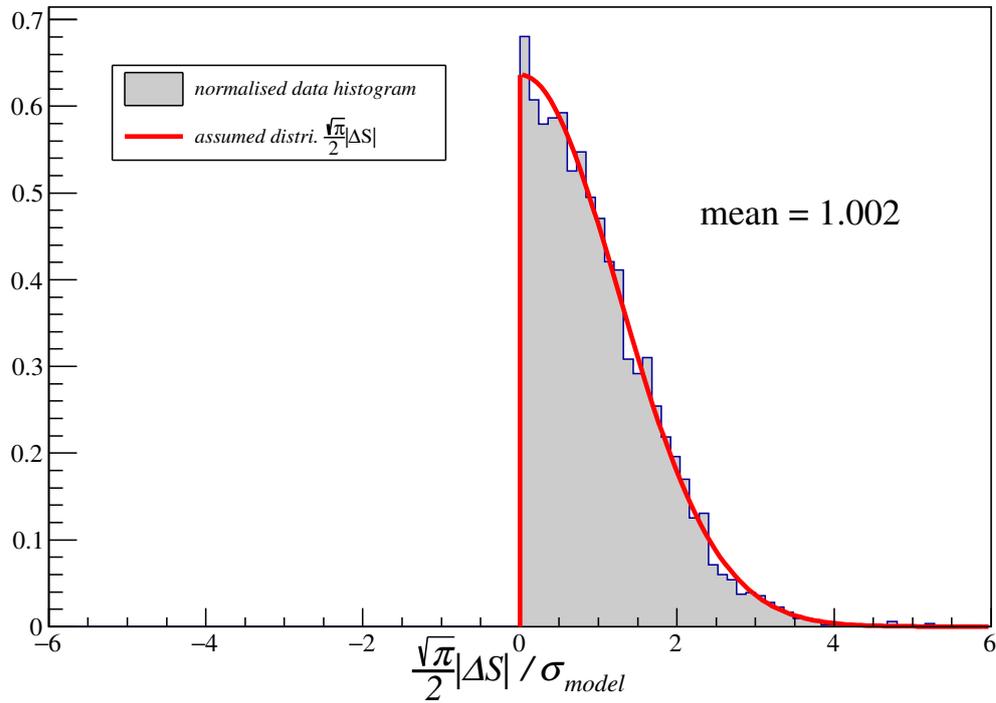
Figure 6.15: Measured distribution (grey) obtained from available SSD data outlined in Section 6.2, and assumed distribution of $\frac{\sqrt{\pi}}{2}|\Delta S|$ (red).

The parameterisation given in Table 6.1 can be used to normalise the values of $\frac{\sqrt{\pi}}{2}|\Delta S|$ from all measurements to crosscheck assumptions about the distributions. Figure 6.15 shows that the measured $\frac{\sqrt{\pi}}{2}|\Delta S|$ distribution agrees with the assumed distribution.

6.6   Conclusion

The SSD signal uncertainty model had previously been derived using simulations. This is the first study that has attempted to use the limited SSD data that is currently available to derive a data-driven uncertainty model. This study follows proven methods with doublet stations that have been used to successfully describe the WCD signal uncertainty. In particular, the uncertainty model has been derived with two doublet pairs consisting of Clais Jr., Trak Jr. (separated by 18.0 m) and Trak Jr., Peteroa Jr. (separated by 11.6 m). The systematic difference in signal between the doublet stations was corrected using an SSD LDF; an LDF that had to be approximated using WCD reconstructed parameters due to the lack of available SSDs within 750 m of the shower core. The only dependence on shower parameters that was deemed significant, given statistical uncertainties and limitations from the data, is with respect to zenith angle.

Chapter 7

SSD Lateral Distribution Function

The lateral distribution function (LDF) is a function that describes the expected signal as a function of distance to the shower axis. This function is commonly used to quantify the energy estimator, S(1000), as well as to reconstruct extensive air showers (EAS). Accurately defining the LDF is crucial in determining the primary energy of a cosmic ray. Unfortunately, because there are currently no analytical solutions for the hadronic-cascade equations, functional forms for the LDF have typically been chosen through empirical means. Currently, the *Observer*[1] water-Cherenkov detector (WCD) reconstruction procedure uses a slightly modified Nishimura-Kamata-Greisen (NKG) function of the form

$$f_{\mathrm{LDF}}(r) = \left(\frac{r}{r_{\mathrm{opt}}} \right)^{\beta} \left(\frac{r + r_s}{r_{\mathrm{opt}} + r_{\mathrm{s}}} \right)^{\beta + \gamma} \tag{7.1}$$

with fixed r_s = 700 m and r_{opt} = 1000 m. These values have been chosen specifically for the Auger array with 1500m spacing (SD-1500). Here, r represents the distance to the shower axis, while β and γ are shape parameters that determine the slope of the LDF. The functional form of the LDF and its associated shape parameters vary with the zenith angle (θ), energy, and mass of the primary particle. Accurately fitting the LDF requires a good estimation of the shower core position, which in turn requires a reliable reconstruction of the shower direction. Fortunately, reconstruction procedures for the Auger WCDs have been extensively studied and are well developed. This allows the WCD reconstructed shower parameters, such as shower direction, core position, and zenith angle to be used until a standalone scintillator

---

[1] *Observer* is a reconstruction task force in the Pierre Auger Observatory (Auger)

surface detector (SSD) shower reconstruction procedure is developed. Here, the shape of the SSD LDF is investigated using Monte Carlo simulations to verify the result from previously completed studies in [75], while also testing whether the modified NKG function (Equation 7.1) remains a suitable functional form to describe the SSD LDF shape.

7.1 SSD LDF Form

Originally, investigation begun with the original NKG function from Equation 7.2.

$$\rho_{\text{NKG}}(r, s, N_e) = \frac{N_e}{(r_{\text{M}})^2} \frac{\Gamma(4.5 - s)}{2\pi\Gamma(s)\Gamma(4.5 - 2s)} \times \left(\frac{r}{r_{\text{M}}}\right)^{s-2} \left(1 + \frac{r}{r_{\text{M}}}\right)^{s-4.5} \quad (7.2)$$

The original NKG function is well known as a good descriptor of the lateral distribution of electrons. It was believed that the lateral distribution of particles from the SSD would fit well with the original NKG function as the signal would be primarily electromagnetic. It was quickly found that the function did not fit well at large distances from the shower axis, due to the limitations of having one slope parameter. The modified NKG function used for the WCD LDF has two shape parameters, which have been previously described as β and α. The original NKG function has a fitted normalisation parameter, N_e (representative of the number of electrons at ground level), and the fitted slope parameter s (also known as shower age). r is the distance to the shower axis, and r_{M} is the Molière radius, which was set to $r_{\text{M}} = 100$ m. One benefit of the original NKG function, because it only has one slope parameter, is that it is useful in studies which investigate correlations of the slope parameter with shower primary particle mass. However, the flexibility of the modified NKG function comes from the second slope parameter α, it allows for finer tuning of the LDF slope at larger distances to the shower axis.

The modified NKG function has provided better fits than the original NKG function, leading to it being chosen as the functional form for the SSD, even though it is the same form used for the WCD. It is difficult to study correlations with shower primary particle mass when there are two slope parameters in the LDF, however, it is possible to set the value of the second slope parameter, γ to 0. If γ is set to 0, then this modified NKG function behaves similarly to the original NKG function, with one slope parameter. Fixing γ this way will allow mass composition studies to investigate correlations of the shape parameter β with primary particle mass, provided that β is fit freely in high station multiplicity events.

7.2   Parameterising the SSD LDF

Chapter 6 has defined a signal uncertainty model which is critical in characterising the uncertainty of station signals. This chapter uses that signal uncertainty model to parameterise the LDF for signals recorded by the SSD. The values for r_{opt} and r_s have been chosen as 1000 m and 700 m, respectively, since the SSDs are co-located with the WCDs, and these values have been optimised for the Auger SD-1500 [78]. In order to obtain the LDF for the modified NKG functional form, shape parameters β and γ must be parameterised with respect to the shower size and geometry. For simplicity, the $\beta + \gamma$ power above the second term in Equation 7.1 is denoted as α from now on, that is,

$$f_{\text{LDF}}(r) = \left(\frac{r}{r_{\text{opt}}} \right)^{\beta} \left(\frac{r + r_s}{r_{\text{opt}} + r_{\text{s}}} \right)^{\alpha} \tag{7.3}$$

The γ parameterisation is calculated as $\gamma = \alpha - \beta$. The benefit of a parameterised LDF shape is that it allows low station multiplicity events to be fit with an LDF of fixed shape. β and α are modelled with an ansatz functional form that is linear in S_{1000} and quadratic in zenith angle (θ). Specifically, the ansatz forms are expressed as

$$\beta(S_{1000}, \theta) = a_\beta + b_\beta s + (c_\beta + d_\beta s) \sec\theta + (e_\beta + f_\beta s) \sec^2\theta \quad \text{and}$$
$$\alpha(S_{1000}, \theta) = a_\alpha + b_\alpha s + (c_\alpha + d_\alpha s) \sec\theta + (e_\alpha + f_\alpha s) \sec^2\theta \tag{7.4}$$

where $s = \log_{10}(S_{1000})$. The zenith angle and shower size dependencies arise from the changes in the lateral distribution of particles with shower age.

7.2.1   Method

The parameterisation of the slope parameters β and α are obtained by simulating COR-SIKA showers from the Napoli database, which includes air showers with continuously distributed primary energies and zenith angles. Detailed properties of the database are described in Table 7.1. Here, different parameterisations for proton and iron primary particles, as well as a parameterisation for a mixed distribution of 50% proton and 50% iron primary particles, are presented. The purpose of the Monte Carlo study is to establish a baseline LDF model to aid in the fitting process of real data.

When determining the optimum energy estimator, S_{1000}, each unique CORSIKA shower is simulated twelve times, with each simulated event featuring a dense ring of twelve stations

Primaries	Proton, Iron
lg(*E*/eV)	18.5 - 20.0, energy spectrum index -1
$\theta/°$	0-65, flat in $\cos^2 \theta$
$\phi/°$	0 - 360, uniformly distributed
Had Int. Models	QGSJet-II.04

Table 7.1: Properties of the COsmic Ray SImulations for KAscade (CORSIKA) files used from the Napoli database. This study used a total of 12000 unique proton showers and 12000 unique iron showers from this set.

at exactly 1000 m from the shower axis. Dense rings are available through a feature available in the $\overline{\text{Off}}\underline{\text{line}}$ collaboration software, it adds extra stations in a ring at any specified distance from the shower axis. The twelve events per unique air shower are then combined, resulting in a large statistical sample for each unique shower. S_{1000} is calculated by averaging the signal values from the 144 stations located in the dense rings. Simulating a large number of events per unique shower provides a significant advantage to the fitting procedure as it increases the number of stations at other core distances. Originally, dense rings were located every 100 m from the shower axis, and the signals were average based on the 144 stations in the dense rings at each distance, but this proved problematic. The LDF parameterisations resulting from using 144 fixed stations located at every 100 m were found to poorly fit real data. This was due to the atypical distribution of station distances from the shower axis, which did not match that of real data. Figure 7.1 illustrates an example LDF fit, where the 144 dense stations create a line of stations located exactly 1000 m from the shower axis. The benefit of combining twelve events into one becomes obvious as there is a good coverage of stations over all distances to the shower axis, while still maintaining a realistic weighting of stations at different distances.

A maximum-likelihood fit of the LDF is performed for each unique shower, with the likelihood function being expressed as a product of the probabilities of observing a signal S given its expectation $\hat{S}$,

$$\mathcal{L} = \prod_{i=1}^{N} p(S_i | \hat{S}_i, \sigma_i) \tag{7.5}$$

The probability distribution function (PDF) describing S_i is represented as $p(S_i | \hat{S}_i, \sigma_i)$, where σ_i is the uncertainty on the expected signal from Chapter 6 (Equation 6.1). A previous study [75] has described the PDF as a truncated normal distribution; a distribution that is similar
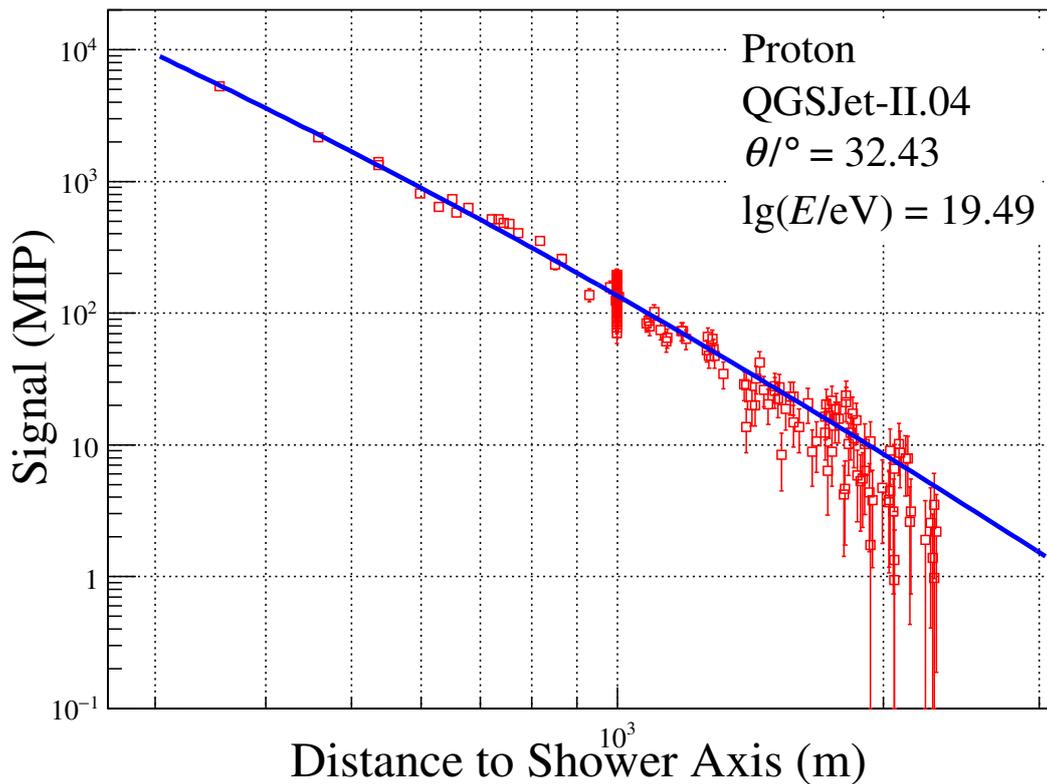
Figure 7.1: Example fit of the LDF to one unique CORSIKA simulated air proton air shower. The same air shower has been simulated twelve times, and the stations from each event have been combined. Each red point is the signal from one SSD station, with its corresponding signal uncertainty. The line of red stations where the distance to the shower axis is 1000 m is created by the dense ring stations. The blue line is the modified NKG function fit.

to a Poisson distribution for smaller signals, but matches closely with a Gaussian distribution for larger signals. This study opts for a different approach, where the PDF utilised is a combination inspired by those used for WCDs. It is a combination of a Gaussian PDF (for $n_i \geq 30$) and a Poisson PDF (for $n_i < 30$). Effective particle number, n_i, is calculated for each signal (S_i) via a conversion given by

$$n_i = p\frac{S_i}{\sec\theta} \tag{7.6}$$

where θ is the primary particle zenith angle, and p is the "Poisson Factor" defined by

$$p = f(\theta)^{-2}. \tag{7.7}$$

$f(\theta)$ is the spectral factor from the signal uncertainty model in Chapter 6 (Equation 6.2). The expected track-length of particles is explicitly taken into account with the $\sec(\theta)$ term. The Poisson factor derived with the explicit track-length dependency is shown in Figure 7.2. This study considers the track-length dependency, enabling the Poisson Factor to be interpreted as the amount the SSD deviates from the assumption that incident particles produce, on average, a signal of $\sec\theta$ / MIP (where a vertical particle with $\theta = 0$ produces 1 MIP). Notably, the minimum constraint of one vertical equivalent muon (VEM) utilised in the WCD model cannot be applied here.

The Poisson factor for the SSD is consistently below one across the range of zenith angles, which suggests that each incident particle typically produces more signal than expected. In contrast, the WCD Poisson factor is notably above one. The differences in the Poisson factors between the two detectors is attributed to differing responses to different types of particles. In the SSD, distributions of signal typically form a Landau distribution. Signals are calibrated to the mode of this Landau distribution, which, due to the asymmetric shape of the Landau distribution, is considerably less than the mean. This results in incident particles, on average, producing more signal than expected.

Once all signals have been converted to particle numbers, the likelihood function is expressed by adding each contribution

$$\mathcal{L} = \prod_i f_{\mathrm{P}}(n_i, \mu_i) \prod_i f_{\mathrm{G}}(n_i, \mu_i) \tag{7.8}$$
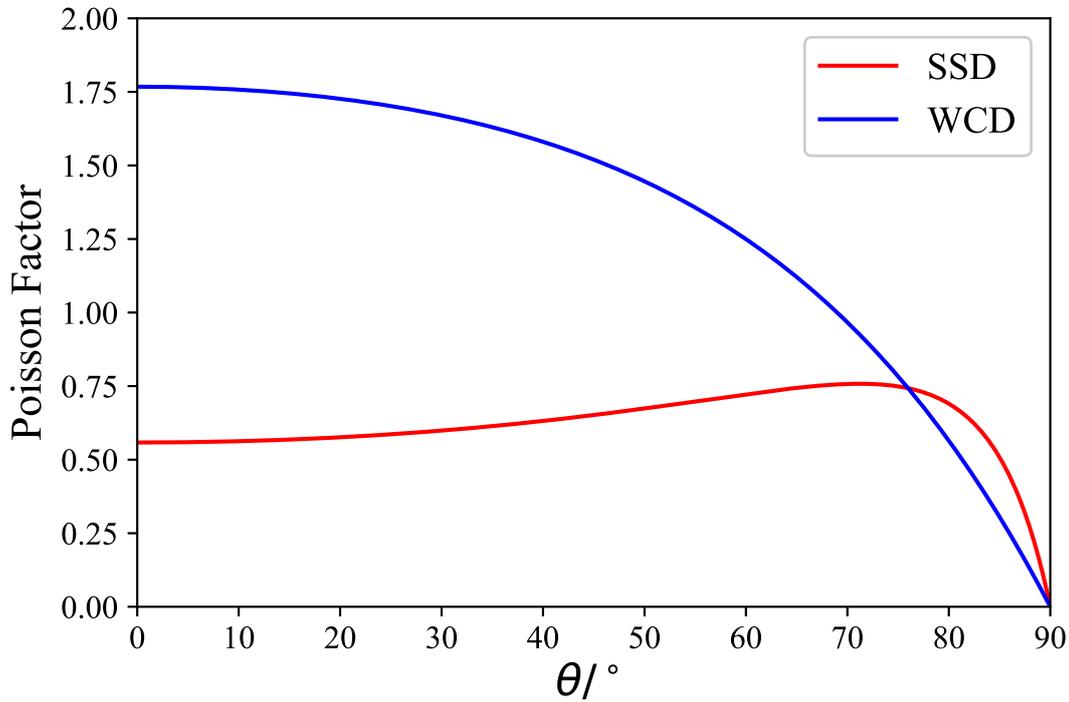
Figure 7.2: Evaluation of the Poisson Factor used to convert signals into effective particle numbers based on the SSD or WCD signal uncertainty model. A value less than one means that, on average, an effective particle produces more signal than expected.

Note that, unlike the WCD likelihood function (Section 3.1.2), the SSD likelihood function does not consist of any terms related to zero signals, or saturated signals. This is because zero signals cannot be well understood without an SSD trigger, and saturated signals do not yet contain the necessary framework to be recovered, and hence should always be excluded. Signals with an effective particle number less than 30 are described by the following Poisson probability distribution

$$f_{\mathrm{P}}(n_i, \mu_i) = \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!} \tag{7.9}$$

where n_i is the effective particle number for a given station, and μ_i is its signal expectation given from the LDF. Signals with an effective particle number larger than 30 are described

by a Gaussian distribution

$$f_{\mathrm{G}}(n_i, \mu_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{\left(-\frac{(n_i - \mu_i)^2}{2\sigma_i^2}\right)\right\} \tag{7.10}$$

where σ_i is given by the signal uncertainty model described in Chapter 6. Contributions to the total log-likelihood for Poisson distributed signals ($n_i < 30$ effective particle number) are expressed as

$$\Delta l_i = -\mu_i - \ln(n_i!) + n_i \ln(\mu_i). \tag{7.11}$$

Gaussian distributed signals ($n_i > 30$ effective particle number), are expressed as

$$\Delta l_i = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_i^2) - \frac{(n_i - \mu_i)^2}{2\sigma_i^2}. \tag{7.12}$$

When fitting functions, it is often easier to use computational packages which minimise. So, when using packages such as ROOT MINUIT, it is more effective to minimise the negative log-likelihood function rather than maximise the likelihood function. This is possible because maximising the likelihood function is equivalent to minimising the negative log-likelihood function. The total negative log-likelihood function is calculated as the sum of the individual components from each station, which allows for a simultaneous fit of the LDF parameters to data from all stations in each event.

$$-l = \sum_i -\Delta l_i \tag{7.13}$$

As stated earlier, the procedure for obtaining the LDF parameterisation involves combining the signals from 12 events of the same unique CORSIKA shower and fitting them with a maximum-likelihood function. However, the current signal uncertainty is not well-suited for describing small signals. For instance, if only a fraction of a minimum ionising particle (MIP) is recorded, such as 0.1 MIP, the uncertainty calculated from Equation 6.1 would yield $\sigma \approx 0.45$ MIP. This uncertainty value is too small, as one or two particles in the detector could have been measured, producing substantially more than 0.45 MIP. Signals that are less than one MIP are rare, because, more often than not, the WCD would not trigger in the situations where these signals occur. The SSD only records data when the WCD triggers, meaning that almost all of these situations are not recorded in data. Regardless, the very few

low signals that are present have been found to have minimal impact on the likelihood fitting procedure, they contribute less than 0.5% to the total likelihood (partly due to having 12 events), resulting in no noticeable changes in the results. Fitting has been performed using the MINUIT package available in ROOT5. The exact procedure is as follows:

1. The initial guesses of β and α are calculated based on previous parameterisations in [75][2], for the reconstructed WCD zenith angle, and S_{1000} calculated from dense stations in each unique shower.

2. All 12 events are combined into one set, and are then fit, with β free and α fixed to the initial guess.

3. Afterwards, β is fixed to the result from the previous step, and then the fit is repeated with β fixed and α free.

4. This procedure is repeated for all unique CORSIKA showers. Fits that did not converge for β and/or α are excluded from results. In the rare circumstances where this occurred, it was typically when the zenith angle was high (approaching $60°$) and the shower energy was low (approaching $10^{18.5}$ eV).

The dependency of β on zenith and shower energy for proton air showers can be seen in Figures 7.3a and 7.4a, respectively. The model has limitations at higher zenith angles where β approaches positive values. Positive values indicate that signal would be increasing with distance, which is non-physical. As such, these parameterisations should not be used for zenith angles which are above $60°$, as larger zenith angles may result in positive β values. Similarly, Figures 7.3b and 7.4b show how α changes with zenith and primary energy for proton air showers. The values for β and α, derived for proton showers are shown in Table 7.2. When considering the slope parameters individually, larger values of β and α represent less steep LDFs. However, the slope parameters β and α are highly degenerate, and may compensate for each other in different ways depending on the fitting procedure.

Figure 7.5a and Figure 7.5b show the slope parameters β and α as a function of zenith angle and shower energy for iron air showers. The slope parameters from iron showers show similar properties to the proton showers, in both magnitude and shape. Iron showers have

---

[2]Note that α is written as γ in some previous works. Here, it is written as α, where α is the exponent of the second term in the LDF function, and $\alpha = \beta + \gamma$.
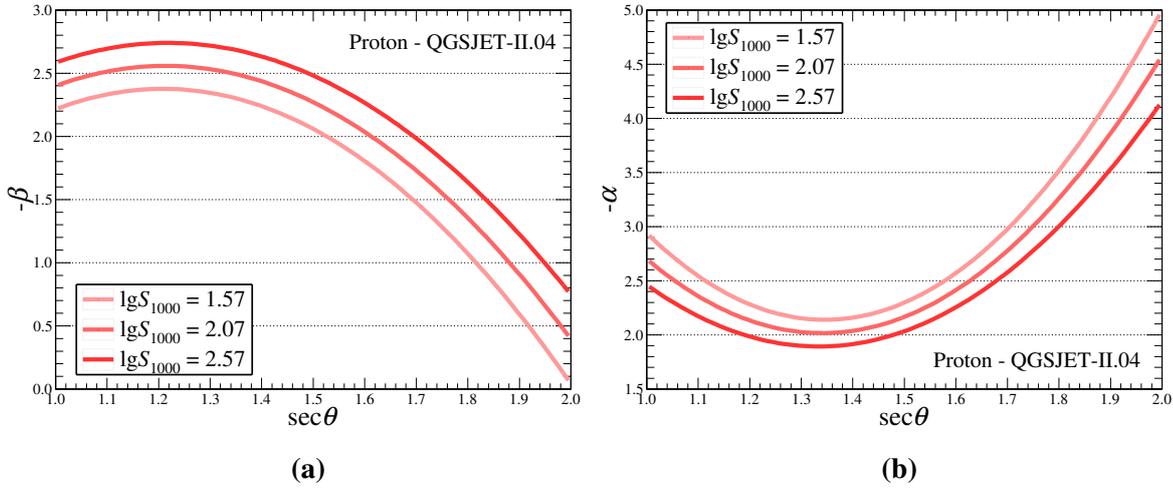
Figure 7.3: (a) The slope parameters β and (b) α as a function of $\sec(\theta)$ for proton air showers, at three different shower sizes.
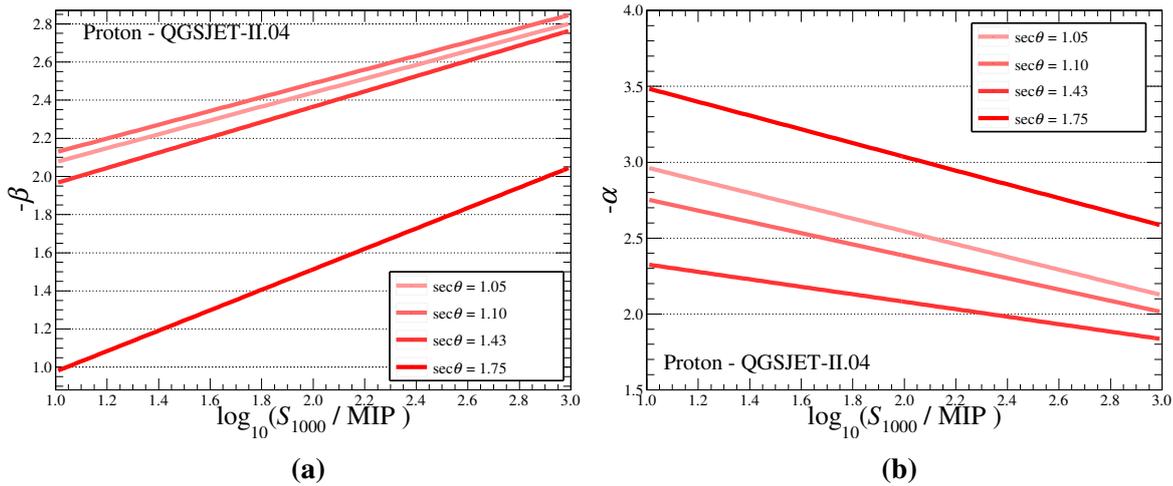
Figure 7.4: (a) The slope parameters β and (b) α as a function of on $\log_{10}(S_{1000})$ for proton air showers, at four different zenith angles.

a slightly smaller value for the magnitude of β which is only discernible when considering averages from large ensembles of data. The differences are more easily understood when looking at the β and α parameterisation values in Table 7.2 (proton) and Table 7.3 (iron).

To apply parameterisations to real data, a mass composition of the instigating shower must be assumed as it is unknown during reconstruction. It would be unreasonable to assume

Proton - QGSJET-II.04						
Parameter	a	b	c	d	e	f
β	4.57	-0.96	-10.65	1.05	4.45	-0.46
α	-19.41	3.28	24.89	-4.38	-9.17	1.58

Table 7.2: Coefficients of the β and α parameterisations for the SSD LDF from the Napoli database proton air showers.
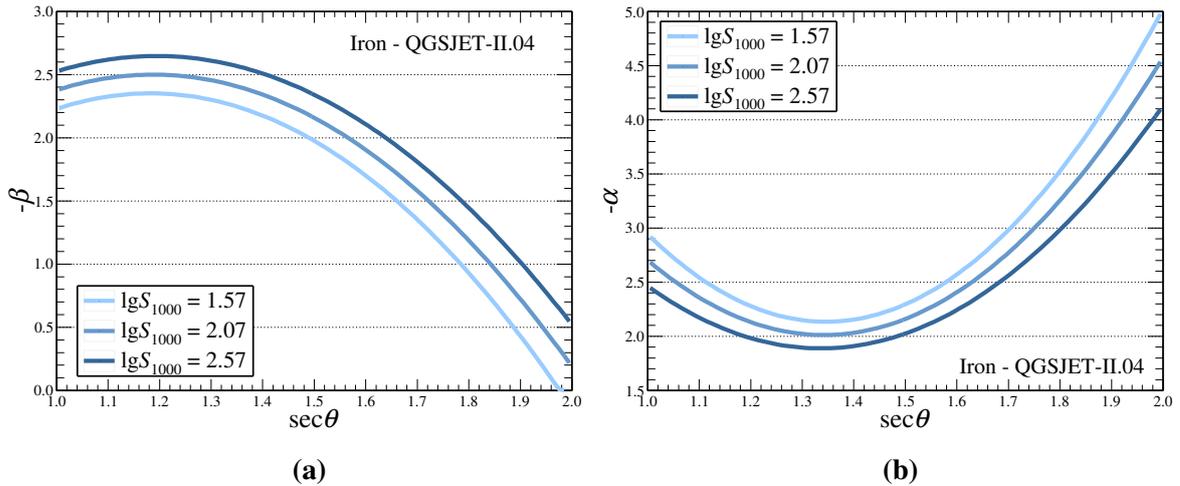
Figure 7.5: (a) The slope parameters β and (b) α as a function of $\sec(\theta)$ for iron air showers, at three different shower sizes.

a pure proton composition, and use the proton parameterisations, likewise, it would also be unreasonable to assume an iron composition. The best approach for low station multiplicity events, is to assume a 50/50 mix of proton/iron; this is based on knowledge that the majority of air showers at energies greater than $10^{18.5}$ eV are either proton or iron. Alternatively, with results from different research like shower X_{max} studies, different proportions of mass compositions may be used at different energies. The parameterisations for a 50/50 mix are shown in Table 7.4, and, as expected, the values are halfway between the proton and iron parameterisations. A parameterisation for a functional form of the SSD LDF which only had one shape parameter, β, was considered. However, it was found that the fits simply did not accurately describe the data as well as having two shape parameters. It should be noted that the best approach for high station multiplicity events is to freely fit β and α while using the parameterisations as an initial guess for the fitting procedure.
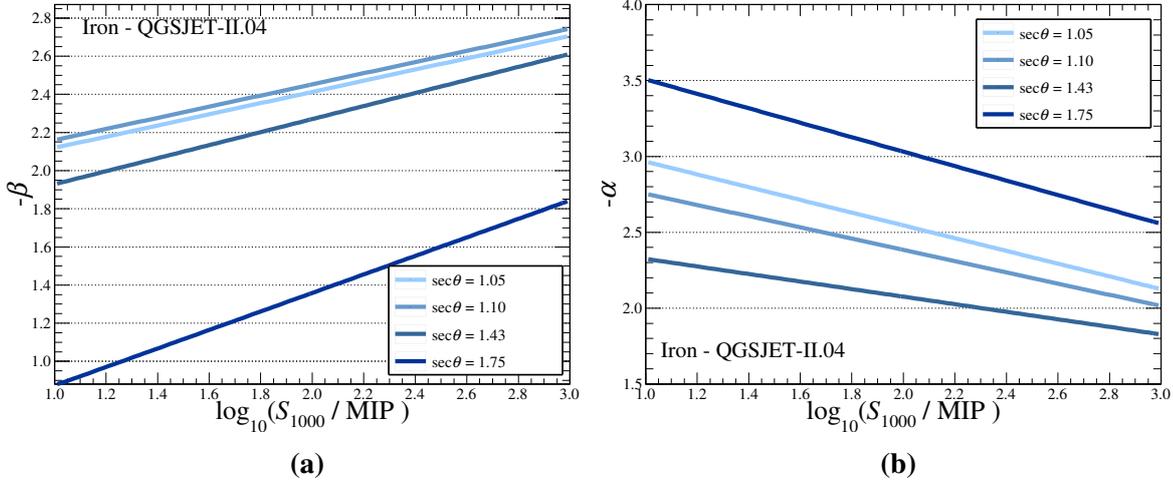
Figure 7.6: (a) The slope parameters β and (b) α as a function of on $\log_{10}(S_{1000})$ for iron air showers, at four different zenith angles.

<div align="center">

Iron - QGSJET-II.04

Parameter	a	b	c	d	e	f
β	4.21	-0.85	-10.42	1.02	4.46	-0.46
α	-19.67	3.39	25.33	-4.57	-9.35	1.66

</div>

Table 7.3: Coefficients of the β and α parameterisations for the SSD LDF from the Napoli database iron air showers.

<div align="center">

50/50 Mix - QGSJET-II.04

Parameter	a	b	c	d	e	f
β	4.39	-0.90	-10.53	1.02	4.45	-0.46
α	-19.59	3.37	25.20	-4.54	-9.30	1.65

</div>

Table 7.4: Coefficients of the β and α parameterisations for the SSD LDF from the 50/50 combination of the Napoli iron and proton air showers.

7.3   Parameterisation of the LDF for SSDs within the 750 m Array

In Chapter 6, it was necessary to have a reconstructed SSD LDF for the array with 750m spacing (SD-750) in order to apply a correction factor for the doublet pairs located close to the shower axis. At the time of this study, the amount of data available with a high station

Primaries	Proton, Iron
lg(E/eV)	17.0 - 18.0, energy spectrum index -1
θ/°	0-65, flat in $\cos^2\theta$
ϕ/°	0 - 360, uniformly distributed
Had Int. Model	QGSJet-II.04

Table 7.5: Properties of the low energy CORSIKA files used that were simulated for the SD-750 SSD analysis. This study used a total of 8000 unique proton showers and 8000 unique iron showers from this set.

Parameter	a	b	c	d	e	f
β	6.263	-2.822	-11.958	3.544	4.552	-1.187
α	-20.647	4.447	27.003	-6.368	-9.973	2.342

Table 7.6: Coefficients of the β and α parameterisations for the SD-750 SSD LDF from the simulation library described in Table 7.5.

multiplicity was limited. The majority of events that were useful for signal uncertainty analysis consisted only of the doublet pairs, and no other SSD stations. An SSD LDF had to be fit based on two data points that were near each other. It is impossible to fit events with such low SSD station multiplicity without using information from the WCD. Therefore, in order to maximise the amount of data that could be used, and to make the signal uncertainty analysis possible, a parameterisation for the SD-750 SSD LDF was created. The parameterisation was produced by following the same procedure and method that was used for the SD-1500 SSD LDF parameterisation, and was made possible due to an existing signal uncertainty model completed in [75]. The library of simulated data used in the SD-750 LDF parameterisation is expressed in Table 7.5. The resulting parameterisation can be seen in Table 7.6.

A parameterisation of the shape of the SD-750 SSD LDF alone, was insufficient for events with such low multiplicity, as there is still not enough information to accurately fit the normalisation. It is not sufficient to take the average of the two stations as the energy estimator S(450) (or normalisation); this would ignore signal asymmetry, the existing signal uncertainty model, and it would assume that the stations are equidistant and at exactly 450 m from the shower axis. The best course of action available was to predict the SSD energy estimator using the WCD energy estimator from a successful WCD reconstruction. The SSD energy estimator, $S(450)_{\text{SSD}}$ has been parameterised as a function of the WCD

Parameter	a	b	c	d	e	f	g	h	i
$\frac{S_{\text{SSD}}}{S_{\text{WCD}}}$	1.18	0.086	-0.06	0.32	0.08	-0.145	0.077	-0.8	0.43

Table 7.7: Parameterisation for the ratio of SSD signal as a function of WCD signal. S represents the $\log_{10}$ of signal for the SSD or WCD.

energy estimator, $S(450)_{\text{WCD}}$. The parameterisation is quadratic in both zenith angle and $\log_{10}(S(450)_{\text{WCD}})$, it follows the functional form described below

$$\frac{S_{\text{SSD}}}{S_{\text{WCD}}} = (a + S_{\text{WCD}}(b + cS_{\text{WCD}})) + (d + S_{\text{WCD}}(e + fS_{\text{WCD}}))\sec(\theta)$$
$$+ (g + S_{\text{WCD}}(h + iS_{\text{WCD}}))\sec^2(\theta) \quad (7.14)$$

where $S = \log_{10}(S(450))$ of either detector, and θ is the zenith angle. The resulting parameterisation can be seen in Table 7.7. Figure 7.7 provides an insight into how well the parameterisation for $S(450)_{\text{SSD}}/S(450)_{\text{WCD}}$ ratio performs. In general, the parameterisation performed satisfactorily for the purpose of allowing events with as little as two stations to be equipped with a fixed shape LDF. As seen in Chapter 6, the parameterised shape of the LDF was also satisfactory, as it helped significantly reduce bias introduced from the small distance between doublet stations without obvious over or under compensation.
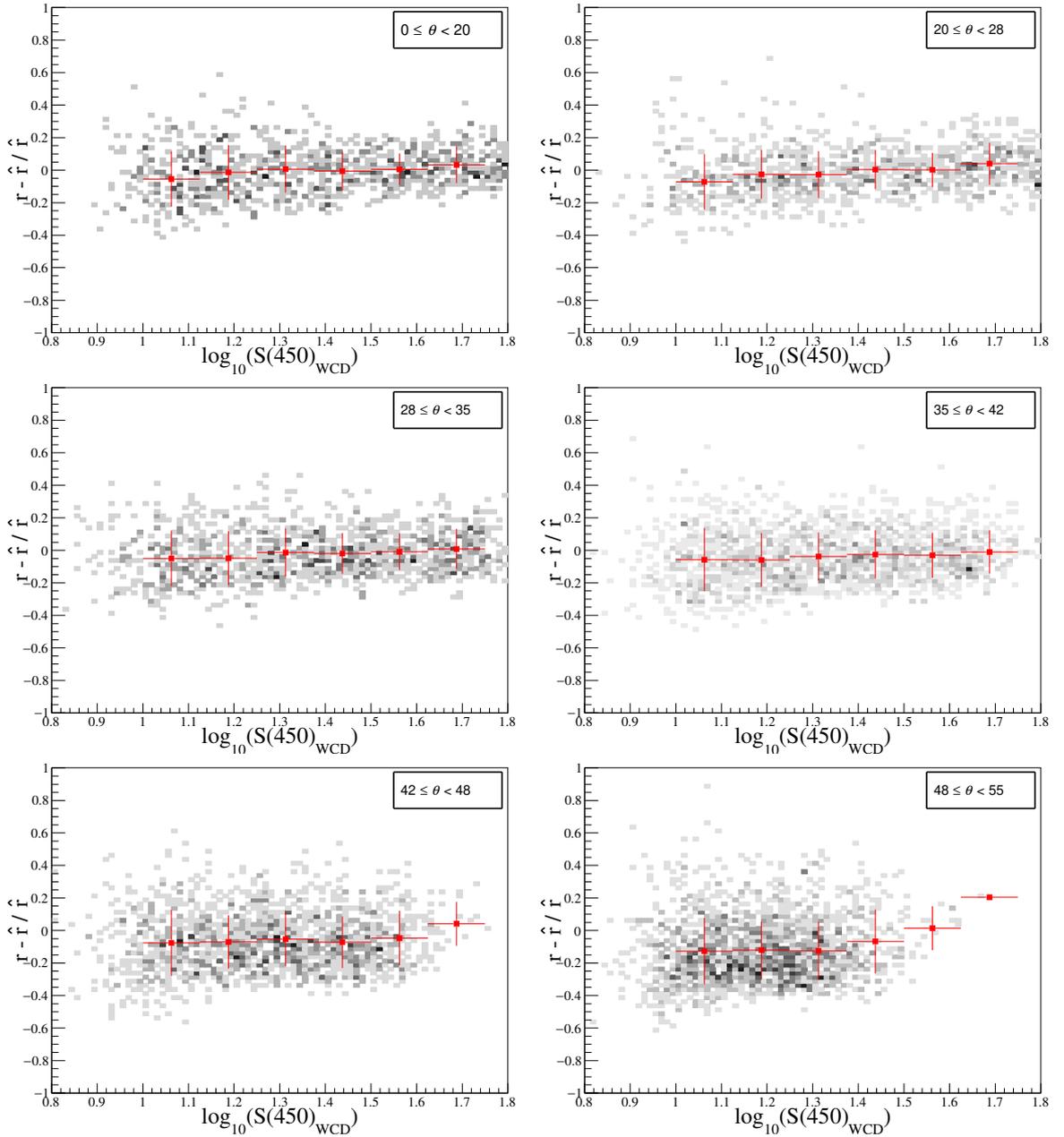
Figure 7.7: The relative difference in the ratio of $S(450)_{\text{SSD}}$ and $S(450)_{\text{WCD}}$ as a function of $S(450)_{\text{WCD}}$ across various zenith angle bins. The ratio is represented as r, and the expected ratio as $\hat{r}$. These plots provide insight into how well the parameterisation in Table 7.7 performs. r is from Monte-Carlo simulated signals, and $\hat{r}$ is from the parameterisation.

7.4   Residuals

The quality of the parameterised fits are judged through two different metrics, the residuals,

$$\text{Res}(S_i) = \frac{S_i - \hat{S}_i}{\sigma[\hat{S}_i]}, \tag{7.15}$$

and the relative difference between the data and the predicted values from the model,

$$\text{Rel}(S_i) = \frac{S_i - \hat{S}_i}{\hat{S}_i}. \tag{7.16}$$

The residuals as a function of distance to the shower axis for simulated 50/50 mix of proton and iron data are shown in Figure 7.8, and the relative differences are shown in Figure 7.9. Figure 7.10 and Figure 7.11 show residuals and relative differences as a function of SSD signal. It is noticeable for the relative differences at large core distances, that there is a downwards trend showing that the expected signal is larger than the measured signal. This feature arises from the large variation in small SSD signals due to large contributions of electromagnetic signal.

Stations that report no SSD signal and are still present in the simulation data because they have valid WCD signal, are completely removed from this average and uncertainty calculations in each graph. These stations result in the band of stations with a relative difference close to -1 that is seen within Figure 7.9. The signal uncertainty model provides an overall good description across distance, even though the model does not describe small, or single particle signals very well. The increased density of particles at large distances to the shower axis is due to the nature of the array spacing, covering more stations as distance from the shower axis increases. This is visible in Figure 7.8.

Figure 7.12 shows the bias in the reconstructed energy estimator, S(1000). There is a clear trend where the energy estimator for lower energy showers is overestimated, and underestimated for higher energy showers. Fortunately, the average bias is within $\pm 5\%$ for all energies and zenith angles. It is noticeable that the variance in the bias is higher for larger zenith angles and lower energies, this agrees with the signal uncertainty model which has a Poisson like nature. After combining the knowledge from all of the figures mentioned above, it appears that while the LDF parameterisation provides a more than satisfactory representation of SSD signals, a perfectionist may take note that low signals may be consistently

overestimated with the parameterisations in their current form. This study gave lower signals more power in the log-likelihood function through a Poisson probability distribution, however, the exclusion of zero signals may have attributed to this small bias. Regardless, further improvement on this particular parameterisation method may prove difficult without a carefully refined uncertainty model for small signals, or, an SSD trigger. Figures showing the residuals, relative differences and S(1000) bias for purely proton and iron data are available in Appendix A. It has been concluded that this parameterisation of the SSD LDF is satisfactory, and has been used in the other chapters of this thesis.
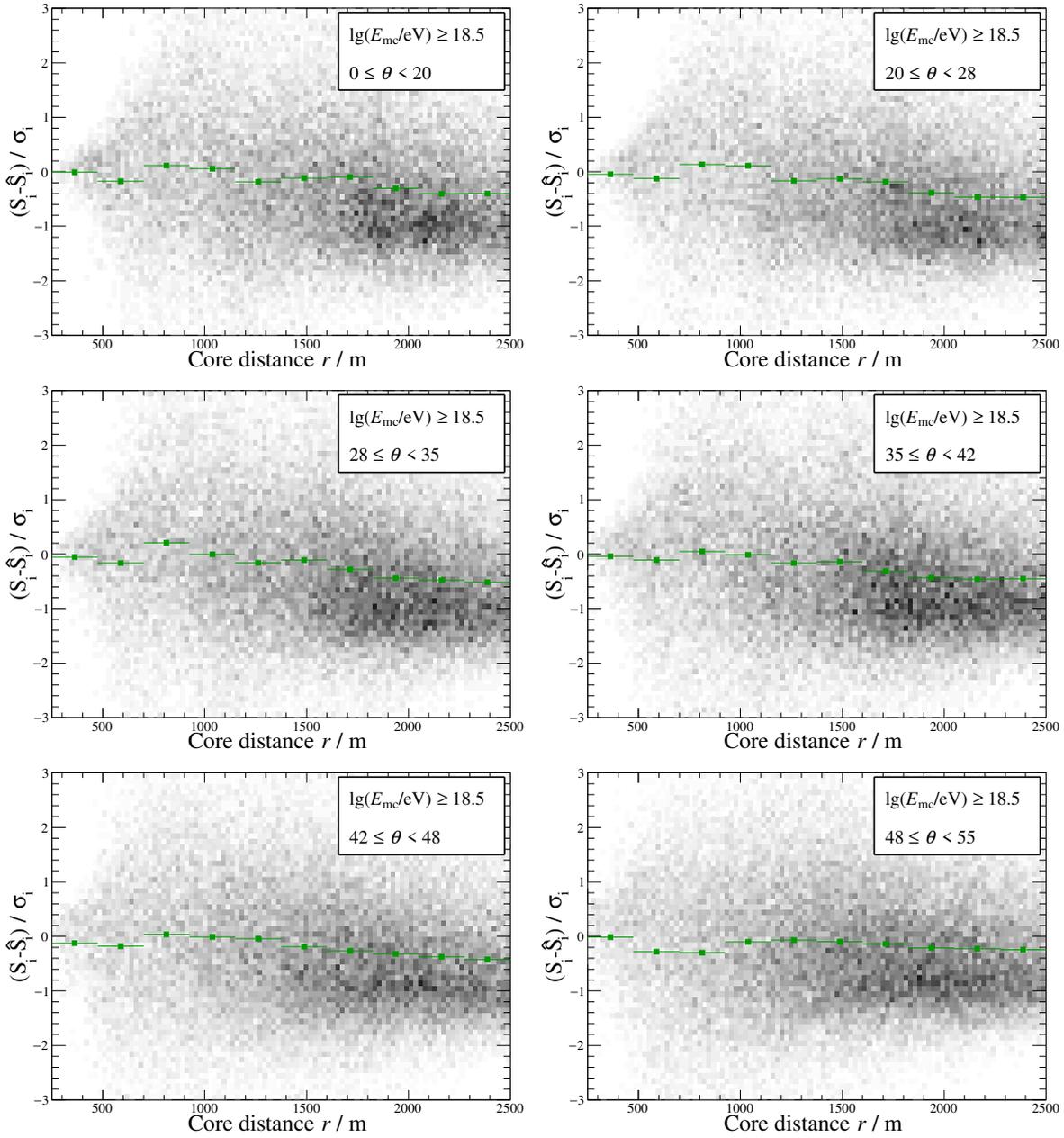
Figure 7.8: Residuals for all simulated and the 50/50 proton/iron LDF parameterisation as a function of distance to the shower axis. The ensemble of data is described in Table 7.1. There are 24000 air showers total, 12000 of which are iron, and the remainder are proton. Shower energies are distributed uniformly, and zenith angles are distributed uniformly in $\sin^2(\theta)$. σ is defined from the SSD signal uncertainty model described in Chapter 6.
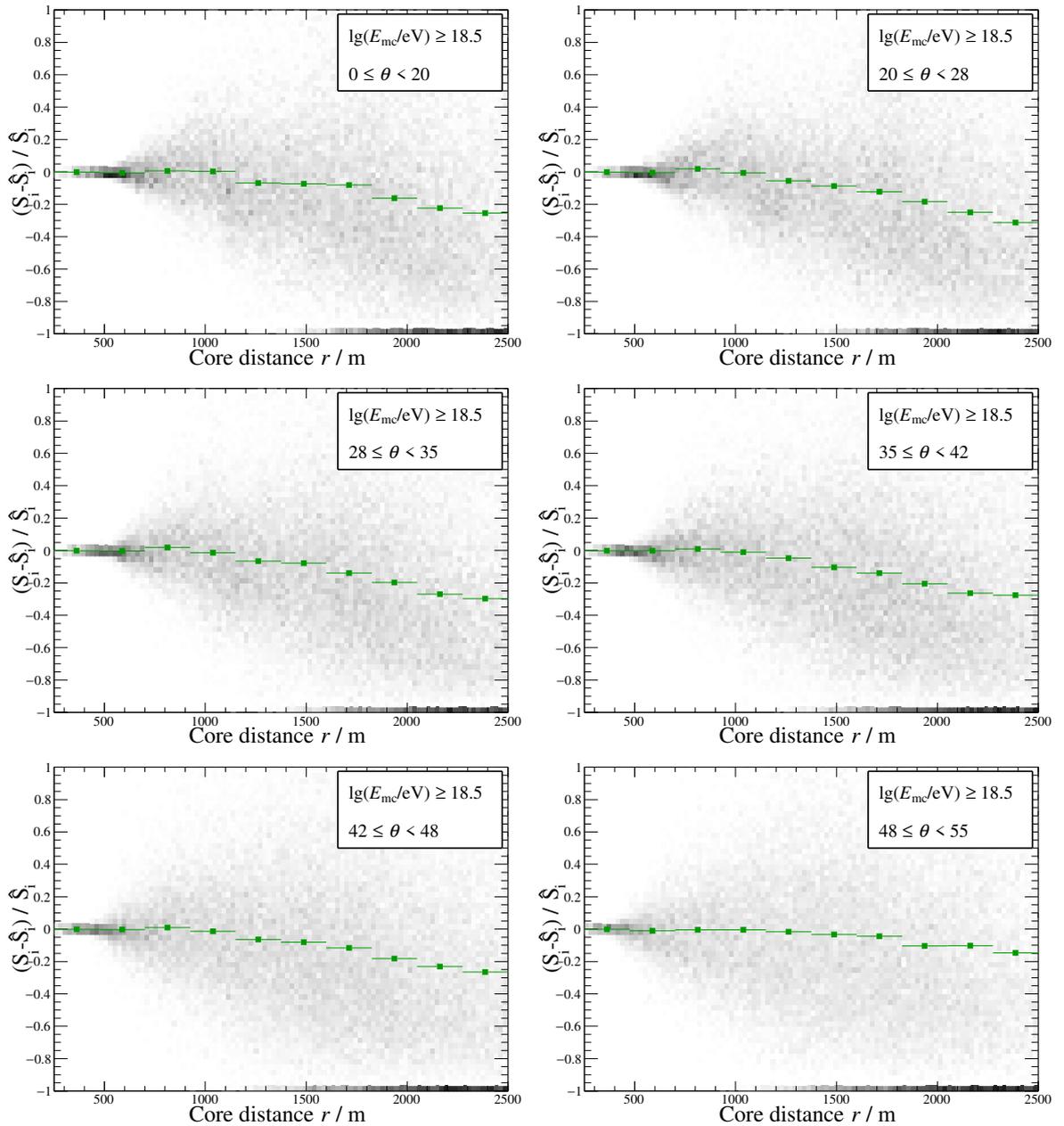
Figure 7.9: Relative differences between measured and expected signal for all simulated data and the 50/50 proton/iron LDF parameterisation as a function of distance to the shower axis. The ensemble of data is described in Table 7.1. There are 24000 air showers total, 12000 of which are iron, and the remainder are proton. Shower energies are distributed uniformly and zenith angles are distributed uniformly in $\sin^2(\theta)$.
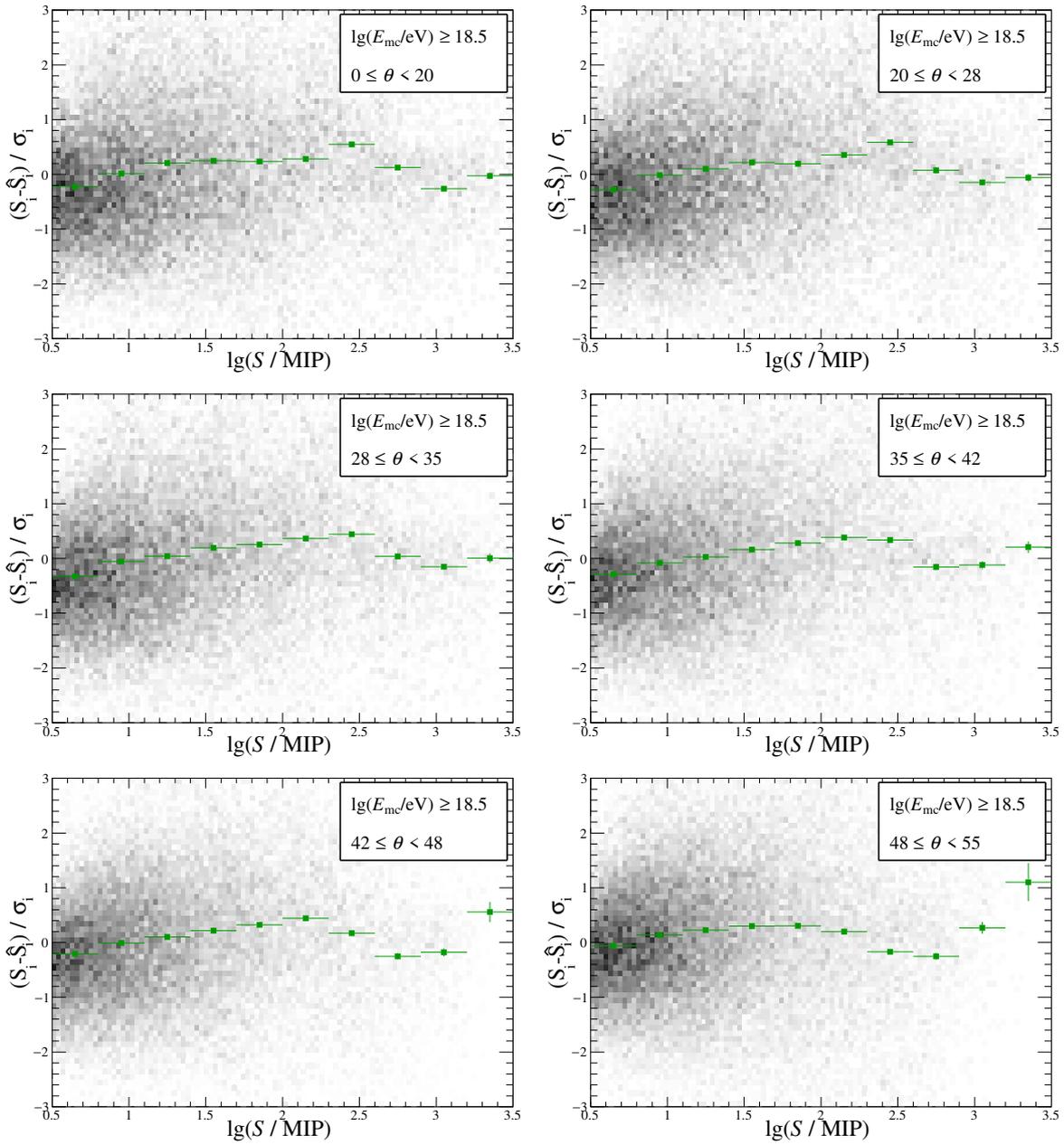
Figure 7.10: Residuals for all simulated data and the 50/50 proton/iron LDF parameterisation as a function of SSD signal. The ensemble of data is described in Table 7.1. There are 24000 air showers total, 12000 of which are iron, and the remainder are proton. Shower energies are distributed uniformly and zenith angles are distributed uniformly in $\sin^2(\theta)$. σ is defined from the SSD signal uncertainty model described in Chapter 6.
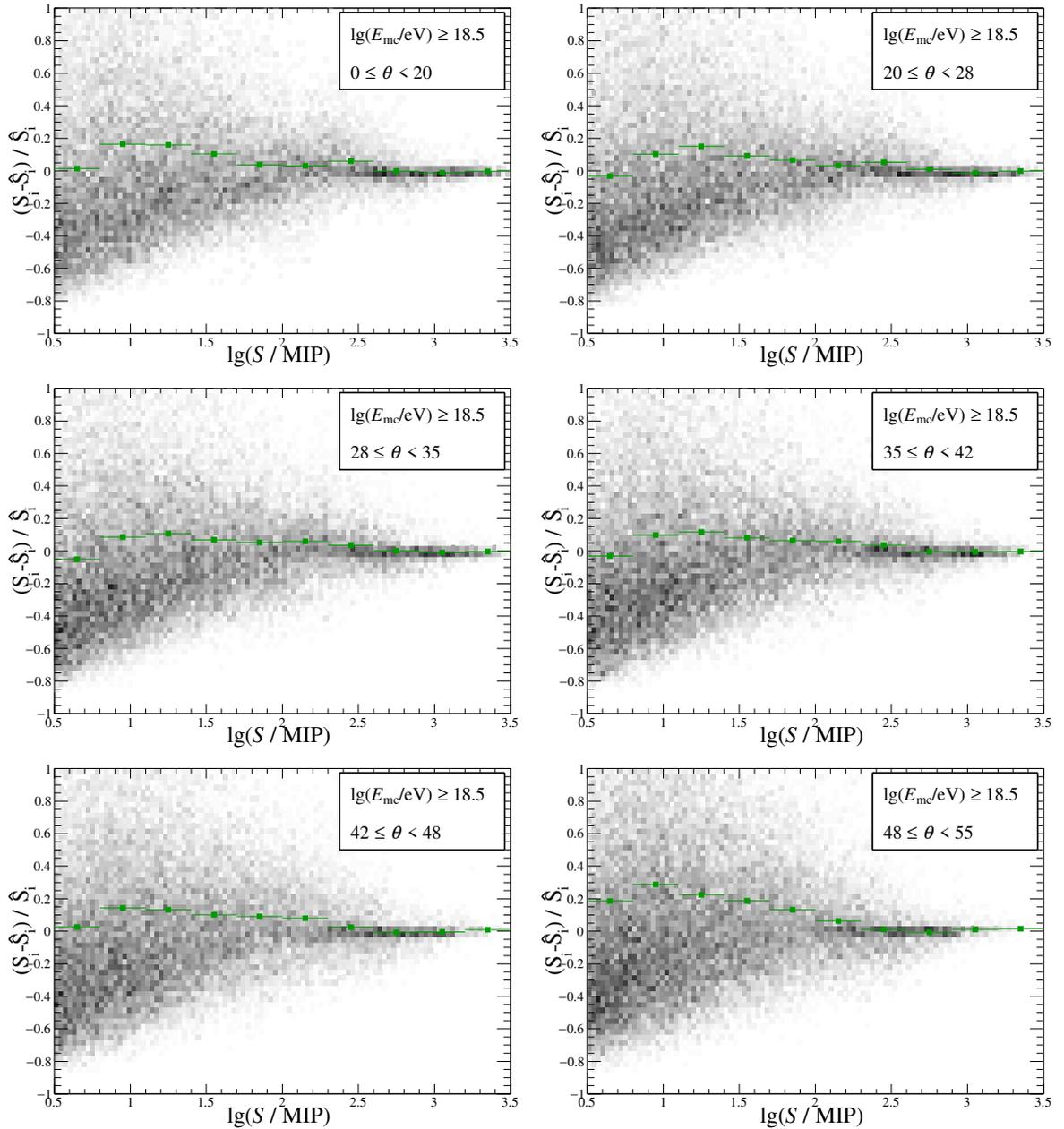
Figure 7.11: Relative differences between measured and expected signal for all simulated data and the 50/50 proton/iron LDF parameterisation as a function of SSD signal. The ensemble of data is described in Table 7.1. There are 24000 air showers total, 12000 of which are iron, and the remainder are proton. Shower energies are distributed uniformly and zenith angles are distributed uniformly in $\sin^2(\theta)$.
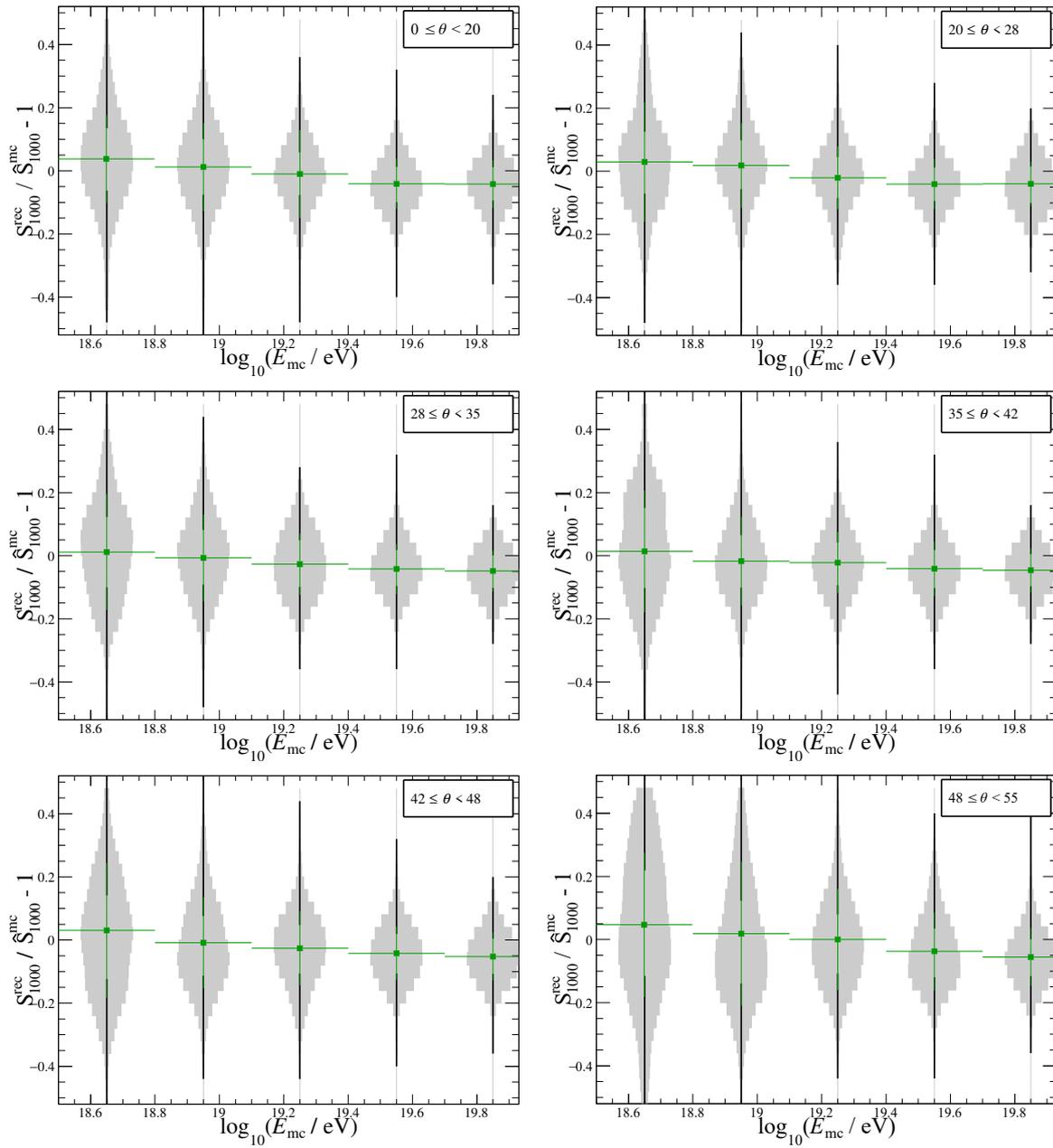
Figure 7.12: Bias in the reconstructed energy estimator S(1000) for all showers, in different bins of shower primary particle energy. The ensemble of data is described in Table 7.1. There are 24000 air showers total, 12000 of which are iron, and the remainder are proton. Shower energies are distributed uniformly and zenith angles distributed uniformly in $\sin^2(\theta)$.

7.5   Validation with Real Data

Real data from the SSD Pre-Production array (SSD PPA) has been reconstructed using the new LDF parameterisation for the 50/50 proton/iron mix. The data consists of events recorded from March 2019 to July 2021. Events have been reconstructed with fixed β and α, so that the quality of fit from the shape of the parameterisation can be interpreted directly from the residuals. The residuals from fits to real data are shown in Figure 7.13. The residuals are written as the difference between expected signal (LDF signal) and measured signal, divided by the expected signal. This is the same as Equation 7.16. It is noticeable that for distances to the shower axis ranging from less than 500 m, and up to 1500 m, that there is, on average, minimal bias. The LDF parameterisations perform well when applied to ensembles of data, but, individually there may be large discrepancies between the measured and expected signal. This difference may be attributed to a variety of factors. Some examples are the signal uncertainty, and the inability to know the correct mass composition of the cosmic ray primary particle. As the LDF is fit with a parameterisation derived from 50/50 proton and iron mass, but any single cosmic ray cannot be a combination of masses, it is expected that no fit will be perfect if the shape parameters of the LDF are fixed. When the distance to the shower axis increases, the fit to measured data worsens. This is largely due to the fact that the SSD signal uncertainty reflects larger proportions of the SSD signal as the SSD signal decreases in magnitude. As mentioned previously in Chapter 6, a SSD signal of 1 MIP, should have an uncertainty of almost 1 MIP.

7.5.1   Selection Cuts

The raw dataset, starting from March 2019 to July 2021 consists of more than 30000 showers with energy above $10^{17.5}$ eV, and zenith angle below $60°$. However, the majority of these events were filtered out with the following selection cuts,

- Each event was ensured to have the 6T5 trigger selection. The 6T5 trigger condition requires that the station with the largest signal in the event (also known as the hottest station), is surrounded by at least six fully functioning stations, these six surrounding stations are not required to be triggered. The 6T5 trigger is usually employed when performing studies related to reconstruction because it filters out events which land
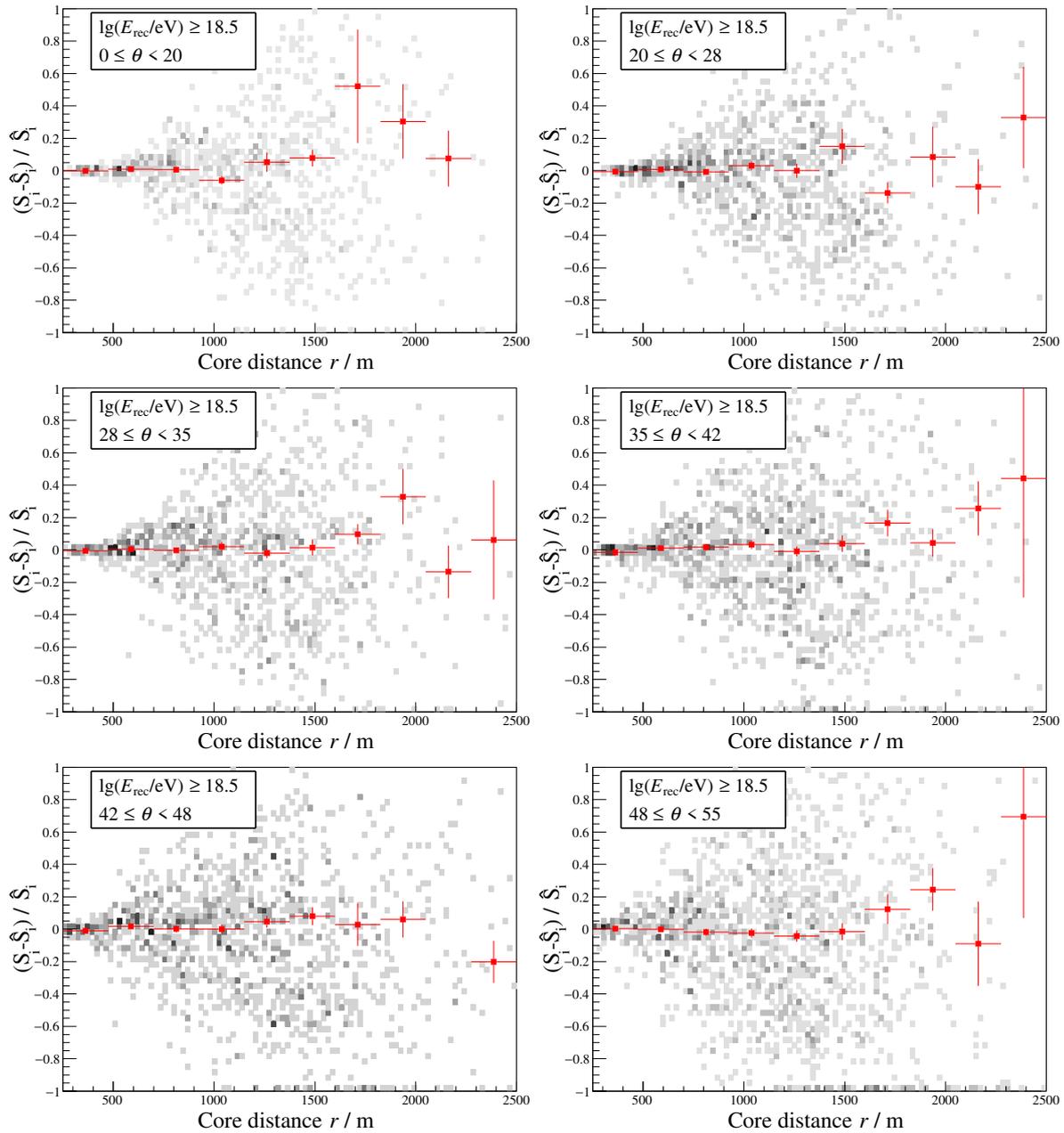
Figure 7.13: Relative differences between measured signal and expected signal as a function of distance to the shower axis for real data obtained from the SSD PPA. The events have been reconstructed using the Auger $\overline{\text{Off}}\underline{\text{line}}$ framework with the new LDF parameterisation configured. There are roughly 2800 high quality events spanning over the different zenith angle bins.

on the edges of the array, preventing them from potentially creating bias in the reconstruction procedure. It also ensured that the majority of stations were operational during early stages of the SSD PPA,

- Each event also required that the WCD LDF was reconstructed, allowing the use of the well-developed WCD reconstruction to filter events based on the WCD reconstructed shower parameters such as energy and zenith angle,

- It was also required that each event had the hottest station equipped with a fully functioning SSD. This ensured that the majority of events would be contained within the SSD PPA,

- Finally, each event had to successfully apply the parameterised SSD LDF function for the residuals to be analysed.

During the two year period (2019-2021), the SSD PPA had undergone expansion, repairs, and testing; it was not 100% operable at all times. After applying all of the selection cuts, more than 90% of events were thrown out and around 2800 high quality events remained. When considering the future capabilities of the Pierre Auger Observatory, the dataset used in this study is not significant. Auger's future consists of no SSD PPA, but instead every station in the field will be equipped with the SSD. This will significantly improve the amount of data that will be available for studies such as this. For now, roughly 30000 reconstructed air showers are available, and roughly 2800 high quality SSD events. The relative differences between measured and expected signals in Figure 7.13, show promise for the LDF parameterisations, and provide high hopes for SSD reconstruction and other analyses as more data becomes readily available in the future.

7.6   Conclusion

A new parameterised functional form of the SSD LDF 1500 m array has been created using the maximum log likelihood method. It uses the same functional form as the WCD LDF, and it uses the same values, $r_s = 700$ m, and $r_{opt} = 1000$ m, because the SSDs are placed above the existing WCDs. The methodology follows a careful fitting procedure which is similar to the WCD procedure, and adds in a Poisson probability density function for small signals.

This work also incorporates the new signal uncertainty model which uses real data, defined in Chapter 6. The final parameterisation has been provided in Table 7.4, which corresponds to a simulated data set of 50% proton air showers and 50% iron air showers. The residuals show that the fits perform very well for large and average signals, but may be consistently overestimating small signals. This overestimation can be attributed to the fact that SSD stations do not have their own trigger, making it difficult to include silent SSD stations into the likelihood function. If there were a SSD trigger, there may be situations where the SSD would report signal, but the WCD may not. However, as there is no SSD trigger, the SSD can only record data if signal is sufficiently large to trigger the WCD. This may reduce the amount of small SSD signals available in our data, resulting in an overestimation in the magnitude of small signals from the LDF parameterisations. The overestimation can also be attributed to the large amounts of uncertainty for small SSD signals which are not reflected in the signal uncertainty model.

An existing signal uncertainty model defined in [75] was used to perform an LDF parameterisation for the 750 m array. It followed the same procedures as the 1500 m parameterisation. This parameterisation enabled opportunities for the new signal uncertainty model to have biases corrected. Due to the successes within the signal uncertainty model; this 750 m LDF parameteristion was deemed satisfactory for its purpose. The work here also enabled a parameterisation of the signal ratio (SSD/WCD) to be performed, allowing the the signal uncertainty model analysis use events which only contained two SSD stations. Overall, all the fits obtained in this study were satisfactory and have been used within all other chapters in this thesis.

Chapter 8

Comparing Signal Ratio and Matrix Formalism Methods for Determining the Mass of Cosmic Rays

Determination of the mass of cosmic rays is one of the many challenging tasks in astrophysics. In order to infer the mass of cosmic rays from experimental data, various methods have been developed. This study focuses on two approaches, the signal ratio method and the matrix formalism method, both of which use data from the Pierre Auger Observatory (Auger) detectors, including the water-Cherenkov detector (WCD) and scintillator surface detector (SSD). The signal ratio method involves measuring the amount of signal deposited by extensive air showers (EASs) in multiple detectors with different sensitivities to different types of particles. By comparing the signal ratio to a simulated expected ratio for a specific mass, the cosmic ray mass can be inferred. The matrix formalism method is a more mathematical approach, which uses a matrix that describes the relationship between the WCD and SSD signals in terms of muonic and electromagnetic components. This allows the muonic content of either detector to be estimated based on the total amount of measured signal. The matrix formalism method requires the derivation of two parameters, a, and b, from simulations. These parameters represent the amount of electromagnetic signal in the SSD with respect to the total measured electromagnetic signal in both detectors, and the amount of muonic signal in the SSD with respect to the total measured muonic signal in both detectors, respectively. The effectiveness of each approach is tested with a metric known as the fig-

145

ure of merit for separating proton and iron induced air showers. The figure of merit can be calculated for a given proton and iron distribution with some observable S using

$$\text{FOM} = \frac{|\langle S_{\text{p}} \rangle - \langle S_{\text{Fe}} \rangle|}{\sqrt{\sigma_{S_{\text{p}}}^2 + \sigma_{S_{\text{Fe}}}^2}} \qquad (8.1)$$

where S_{p} and S_{Fe} refer to the observable produced by proton and iron showers, respectively. Sufficiently large values for the figure of merit (greater than 1.5), can only be achieved if the underlying distributions of data are sufficiently separated with minimal overlap. Values greater than 3 represent minimal overlap and would be remarkable for discriminating mass information. Hence, an in-depth study of the major contributing factors to the uncertainty of the signal ratio or calculated muon signal is presented here through analysis of their sampled distributions.

8.1   Simulation Library

The purpose of this work is to study a very specific, and relatively common real scenario with simulations. This scenario considers proton and iron air showers which have a zenith angle of 38 deg, a primary energy of 10^{19} eV, and varying X_{max}. The COsmic Ray SImulations for KAscade (CORSIKA) library used here consists of 1000 unique proton showers, and 1000 unique iron showers simulated with these parameters and the QGSJET-II.04 hadronic model. Careful handling of the simulation dataset allows the resolutions of the resulting distributions to be broken down into separate components to be further understood. A deep understanding of the resolutions and the components they are comprised of will help in further studies which aim to improve the ability to screen primary mass sensitive information from detector signals. All detector simulations contain dense rings at every 100 m from the shower axis, starting at 400 m to the shower axis, and finishing at 2500 m from the shower axis. This simulation library has been used for both the signal ratio method, and matrix formalism method.

8.2   SSD/WCD Signal Ratio

WCDs and SSDs are both used by the Pierre Auger Observatory to detect cosmic ray air showers. These detectors have different sensitivities to electromagnetic and muonic components of air showers. While SSDs are equally sensitive to both components, WCDs are more

sensitive to the muonic component. Electrons typically do not have enough energy to fully traverse through the water in the WCD, causing them to produce fractions of VEM signal. On the contrary, most muons in air showers have sufficient energy to pass through the WCD and will most likely produce one full VEM. Therefore, the ratio of signals from the SSD and WCD ($S_{\text{SSD}}/S_{\text{WCD}}$), can be used to infer the relative contributions of the electromagnetic and muonic components of signal. This ratio is often expressed as MIP/VEM, where MIP corresponds to the SSD signal, and VEM corresponds to the WCD signal. The measured ratio can be compared to a simulated ratio from a specific mass to provide insights, or possibly even predict the mass of the cosmic ray.

The following study is related to results obtained purely from simulations, where several methods for calculating the signal ratio, each with its own advantages and disadvantages are considered. Each approach is described below.

Best Case: Calculating the signal ratio per station; the ratio is obtained by dividing the SSD signal and the WCD signal of that station. This method does not ignore signal correlations between the SSD and WCD. Signal correlations arise because the detectors are co-located, with the SSD above the WCD. An example can be seen when incident particles on the SSD are more likely to also be incident on the WCD, this is depicted later in Figure 8.3. This best case describes Monte Carlo truth for signal ratios when dense rings are used to calculate the ratio. So here, the shower asymmetry is averaged over.

Method 1: Calculating the signal ratio per event; the ratio is obtained by evaluating the SSD and WCD lateral distribution functions (LDFs) at a particular distance. This will ignore signal correlations because lateral distribution functions result from the combination of all stations within an event. This may result in larger uncertainties, negatively impacting the ability to infer mass.

Method 2: Calculating the signal ratio per event; the ratio is obtained by fitting a function to the lateral distribution of signal ratios, and then evaluating this function at a particular distance. This will not ignore signal correlations as long as the uncertainty for each ratio is calculated with the correlation term included. The uncertainty of the ratio has been calculated using Equation 8.4.
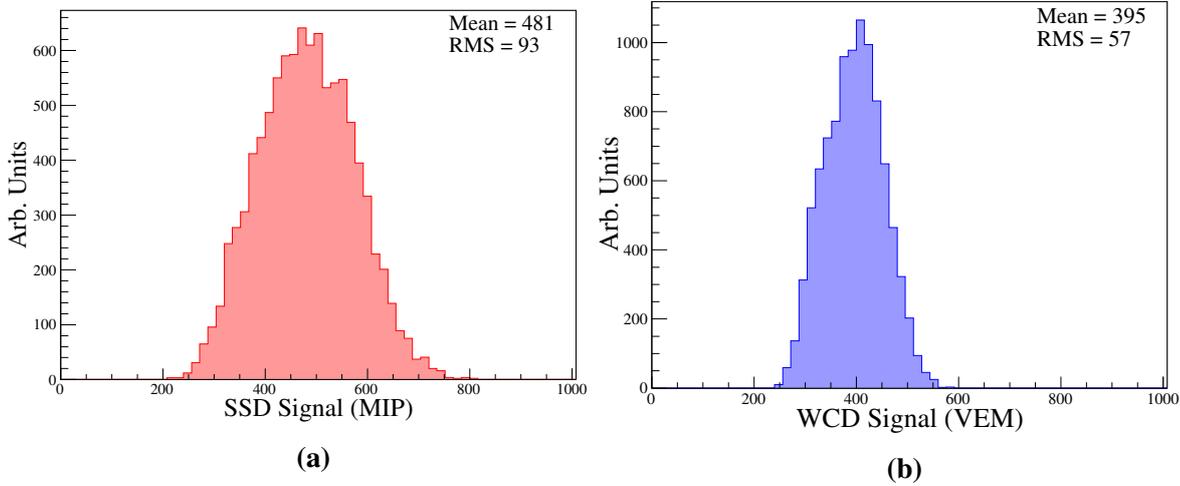
Figure 8.1: (a) SSD signal and (b) WCD signal at 500 m for the 1000 proton showers with a primary energy of 10^{19} eV, zenith angle of $38°$, and random X_{max}.

It is important to note that the ratio of signal depends heavily on the calibration of VEM and MIP. It is also reliant on a well-constrained LDF fit in each reconstructed event (if opting to use LDFs to construct the ratio). For now, as stated in Chapter 5, the calibration of the MIP is currently based on results derived using simulations. For purposes of understanding differences between each method, they have been tested using the same simulated data.

The distributions of SSD and WCD signal at 500 m and 1000 m from the shower axis can be seen in Figure 8.1 and Figure 8.2. Note that 500 m is chosen instead of 400 m, because the signals at 400 m are often saturated. It is these signals which form the foundation of the signal ratio method. The signals from the SSD are designed to have a similar magnitude to the signals from the WCD; it is caused by careful consideration when deciding on the total surface area of 3.8 m^2 for the SSD.

8.2.1 Correlation between SSD and WCD signals

Figure 8.3 presents the signal correlation between the SSD and WCD signals, obtained at a distance of 500 m from the shower axis, using an ensemble of simulated proton showers with an energy of 10^{19} eV and a zenith angle of $38°$. The observed signal correlation arises due to the positioning of the SSD above the WCD. Consequently, there exists a correlation in the signals between the two detectors, as particles traversing through the SSD are more
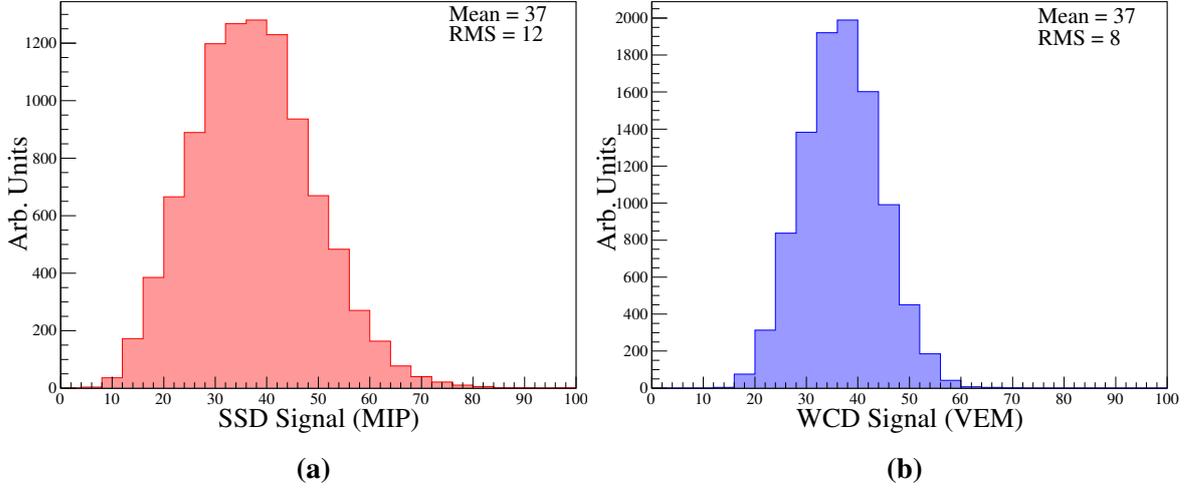
Figure 8.2: (a) SSD signal and (b) WCD signal at 1000 m for the 1000 proton showers with a primary energy of 10^{19} eV, zenith angle of 38°, and random X_{max}.

likely to also traverse through the WCD. The covariance was individually calculated every 100 m for each dense ring of stations in the simulations. The covariance was obtained with

$$\text{COV} = \frac{1}{n}\sum(S_{\text{SSD}} - \mu_{\text{SSD}})(S_{\text{WCD}} - \mu_{\text{WCD}}) \qquad (8.2)$$

where S is the average signal of a given dense ring, and μ is the mean signal of each respective detector across the ensemble of simulations used. The uncertainty on the ratio of signal S, defined as $R = S_{\text{SSD}}/S_{\text{WCD}}$, is then calculated using

$$\delta R = \sqrt{\left(\frac{\partial R}{\partial S_{\text{SSD}}}\sigma_{\text{SSD}}\right)^2 + \left(\frac{\partial R}{\partial S_{\text{WCD}}}\sigma_{\text{WCD}}\right)^2 + 2\frac{\partial R}{\partial S_{\text{SSD}}}\frac{\partial R}{\partial S_{\text{WCD}}}\text{COV}}. \qquad (8.3)$$

The uncertainty of the ratio S equates to

$$\delta R = \sqrt{\left(\frac{1}{S_{\text{WCD}}}\sigma_{\text{SSD}}\right)^2 + \left(\frac{-S_{\text{SSD}}}{(S_{\text{WCD}})^2}\sigma_{\text{WCD}}\right)^2 + 2\frac{1}{S_{\text{WCD}}}\frac{-S_{\text{SSD}}}{(S_{\text{WCD}})^2}\text{COV}}. \qquad (8.4)$$

It is noticeable that for positive correlation between SSD and WCD signals as seen in Figure 8.3, the resulting correlation term will be negative, and negative correlation terms will reduce the uncertainty of the signal ratio (only if one calculates the ratio station by station).
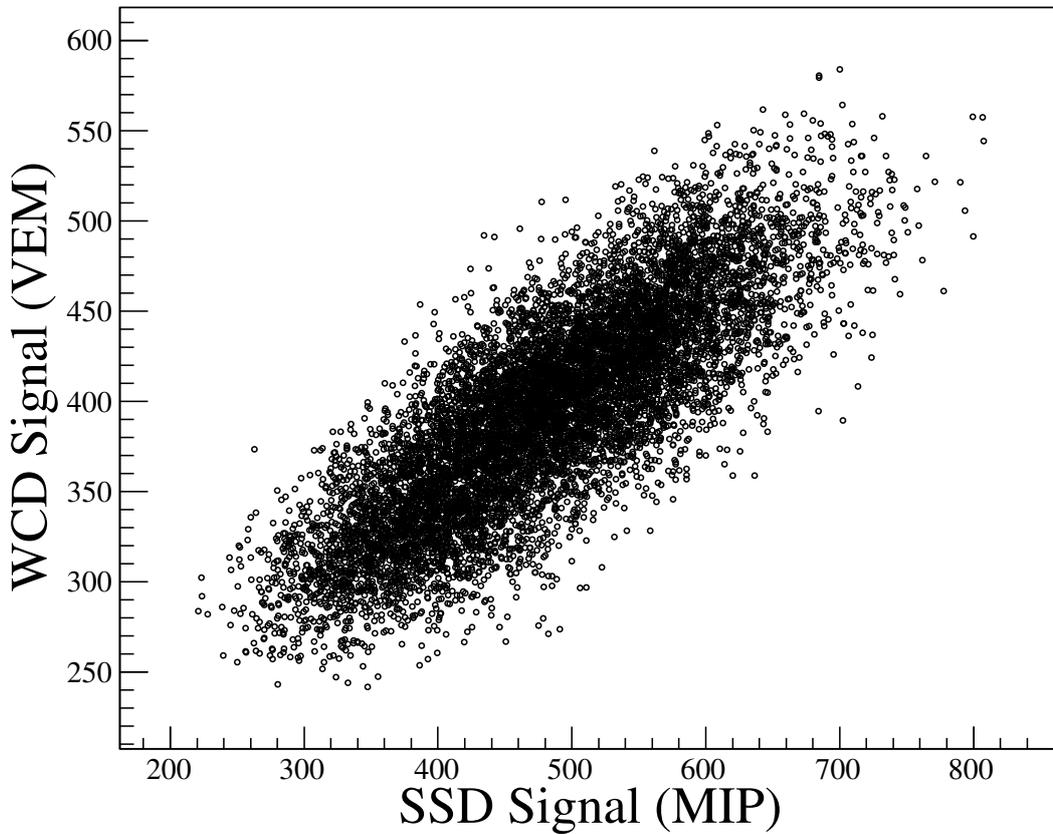
Figure 8.3: Visual representation of the correlation between SSD and WCD by plotting one against the other. This figure shows the significance of the correlation for stations located 500 m from the shower axis. The simulated signals are averages of a dense ring from an ensemble of proton showers with energy 10^{19} eV and zenith angle 38 deg.

8.2.2   Best Case Scenario - Calculating Ratio of Signals Once Per Station with Simulated Dense Rings

This method aims to calculate the ratio per station by directly obtaining the SSD signal and WCD signal from the measured signals of each individual station. It has been observed that the variance of the signal ratio distribution increases with distance from the shower axis. This can be attributed to the widening distributions of contributing signals from the SSD and WCD. Interestingly, the variance of the distribution of signal ratios is smaller than initially expected due to positive correlations between the SSD and WCD signals. These positive

correlations result in a negative correlation term, reducing the variance of the resulting distribution. The *true signal ratio*, which is often referred to later, is the best case scenario where signal ratios have been calculated per station, but averaged over twelve stations within a simulated dense ring.

Figure 8.4 illustrates the distributions of the SSD and WCD signal ratios. The ratio is calculated by averaging the signals from twelve dense stations, which are located at the same distance to the shower axis, and then calculating the ratio. This approach mitigates the effect of signal asymmetry and reduces the variance of the corresponding distribution of signal ratio. This analysis resulted in a mean signal ratio at 500 m of 1.215 MIP/VEM with an RMS of 0.141 MIP/VEM. In Figure 8.4b, the signal ratio is calculated from the dense ring positioned 1000 m away from the shower axis. It is clear that the variance of the distributions are significantly larger at 1000 m from the shower axis, compared to 500 m from the shower axis. Figure 8.5 presents similar graphs for air showers with iron primary particles. Similar trends are seen between proton and iron showers. Notably, as the distance from the axis increases, the variance of the SSD/WCD distribution also increases, reflecting the broader signal distributions observed in SSD and WCD measurements. The mean and RMS has been calculated for each dense ring in both the proton and iron simulation datasets, and are later used to calculate the figure of merit. As stated previously, an increasing uncertainty in the signal ratios will make it more difficult to separate differences due to cosmic ray mass. As such, it would be expected that for this method, the figure of merit will decrease as the distance to the shower axis increases. Figure 8.6 shows the distributions from both proton and iron drawn on top of each other, the amount of overlap between the two distributions expresses clearly how difficult it will be to determine if a random point of data belonged to a specific distribution.

Real data from the SSD pre-production array beginning in April 2019 and ending in May 2021 has been used to calculate the signal ratio per station at 500 m and 1000 m. As real data do not have dense rings, the signal ratios for real data have been calculated from taking the station signals from stations that are within $\pm$ 10% of the respective distance. Figure 8.7 and Figure 8.8 show the trend of signal ratios as a function of shower zenith angle. These results also have QGSJETII-04 proton and iron signal ratios drawn on top, with each point showing error bars which represent the standard error of the mean. For simulations, the signals from the dense rings are averaged, and then the ratio is taken. It is noticeable that real data signal ratios are below both proton and iron signal ratios. The expectation is that the
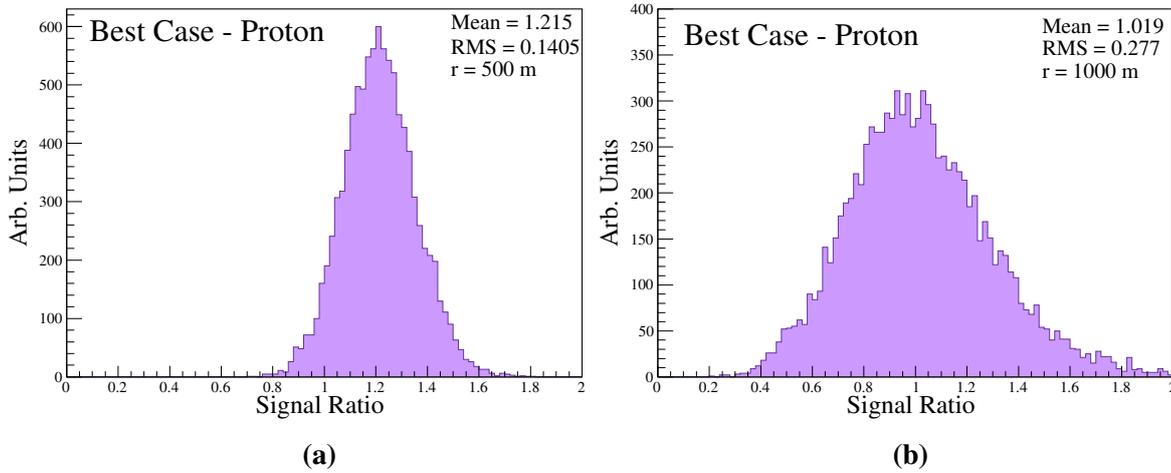
Figure 8.4: Ratio of signal distributions from calculating the signal ratio by directly dividing the SSD signal by the WCD signal. Ratio of signal at (a) 500 m and (b) 1000 m for the 1000 proton showers with primary energy of 10^{19} eV, zenith angle of $38°$ and random X_{max}. In this case, the best case represents Monte Carlo truth as dense stations are used to calculate the signal ratio.
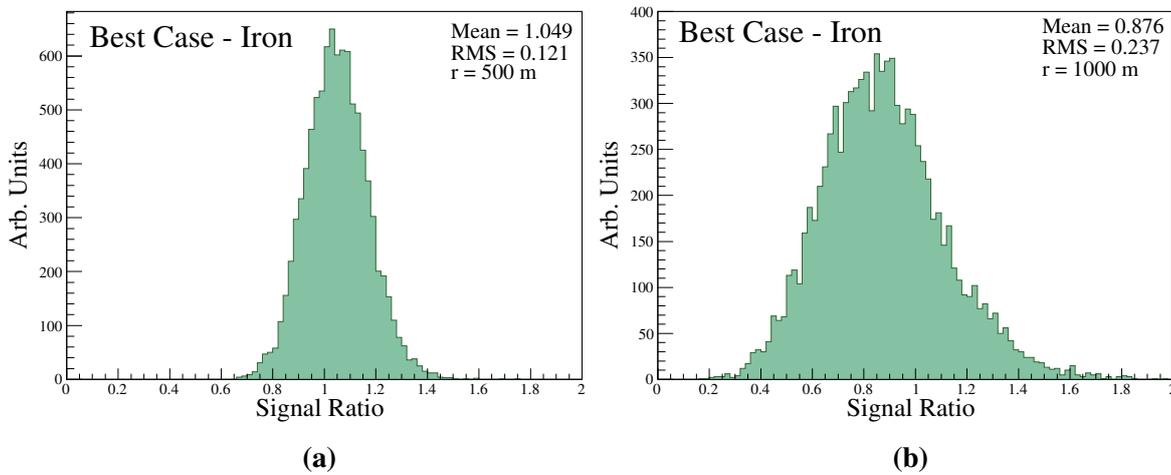
Figure 8.5: Ratio of signal distributions from calculating the signal ratio by directly dividing the SSD signal by the WCD signal. Ratio of signal at (a) 500 m and (b) 1000 m for the 1000 iron showers with primary energy of 10^{19} eV, zenith angle of $38°$ and random X_{max}. In this case, the best case represents Monte Carlo truth as dense stations are used to calculate the signal ratio.
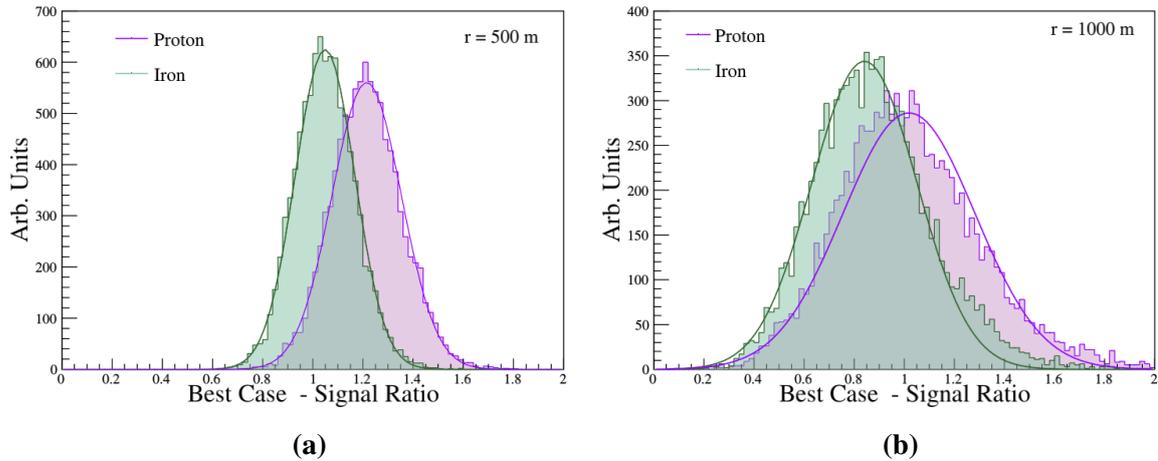
Figure 8.6: Ratio of signal distributions from calculating the signal ratio by directly dividing the SSD signal by the WCD signal. Ratio of signal at (a) 500 m and (b) 1000 m for the 1000 proton showers and 1000 iron showers, both with primary energy of 10^{19} eV, zenith angle of $38°$ and random X_{max}.

signal ratios calculated from data would lie somewhere between the proton and iron results because proton and iron represent the likely extremes. Regardless, the results here disagree with that, and may highlight other more pressing issues with high energy hadronic interaction models, such as the muon deficit, where simulations produce less muons than expected. If muon content were to increase in simulations, we would expect to see the ratios decrease, as WCD are more susceptible to muon signal, are on the denominator of the ratio; this would shift the simulated ratios towards data. A less obvious observation, because muons treat both detectors as volume detectors (they will traverse entirely through the detector), the signal ratios will approach the projected areas of the two detectors as the electromagnetic components contributing to signal diminish.

8.2.3 Method 1 - Calculating Ratio of Signal from Evaluating Lateral Distribution Functions

To be clear, the following components of this chapter reference simulated data, not real data. Method 1 calculates the signal ratio by taking the signal ratio of event-fitted SSD and WCD LDFs. To calculate the signal ratio using this method, an LDF must be fitted for each event. As events with high station multiplicity typically originate from cosmic rays with
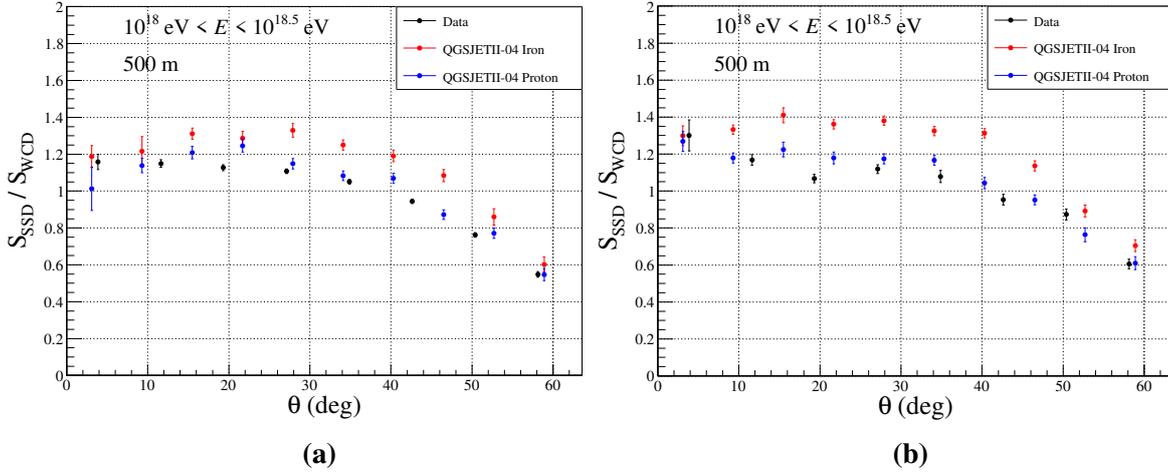
Figure 8.7: (a) The signal ratio as a function of shower zenith angle for shower energies between 10^{18} eV and $10^{18.5}$ eV, and (b), for shower energies between $10^{18.5}$ eV and $10^{19.0}$ eV. All signal ratios are calculated from stations within $\pm$ 10% of 500 m from the shower axis. The error bars represent the standard error of the mean.
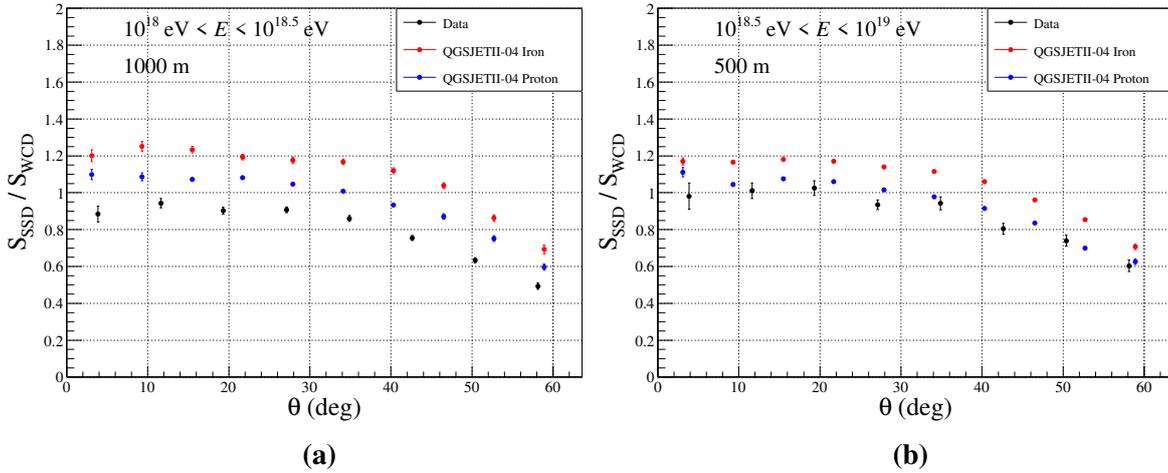
Figure 8.8: (a) The signal ratio as a function of shower zenith angle for shower energies between 10^{18} eV and $10^{18.5}$ eV, and (b), for shower energies between $10^{18.5}$ eV and $10^{19.0}$ eV. All signal ratios are calculated from stations within $\pm$ 10% of 1000 m from the shower axis. The error bars represent the standard error of the mean.

energies greater than 10^{19} eV, and because the SSD detectors are still relatively new to Auger, most events in the dataset lack sufficient station multiplicity to allow a fully unparameterised

fit of LDF for both the SSD and WCD. Therefore, pre-defined LDF parameterisations derived in Chapter 7 for the SSD and WCD are used, and only the energy estimator, S(1000), is fit in each event using stations that are not within dense rings. This imposes a fixed shape of the LDF corresponding to a zenith angle of 38 deg (all simulations here are 38 deg), and may introduce inaccuracies and biases into the determination of the signal ratio. The resulting distributions of signal ratio determined through this method for proton showers are illustrated in Figure 8.9, and the distributions for iron showers are shown in Figure 8.10.

As distances increase from the shower axis, the advantages of using an LDF fit becomes apparent, where the variances from the distributions of signal ratio are much smaller than the previous method. The use of signals from multiple stations in one event allows larger distances, such as 1000 m to use signals, evaluated from the parameterised LDF. These signals typically have smaller variance than the raw signals themselves. The trade-off is that the distribution of signal ratio will have a variance that is similar across all reasonable distances, where "reasonable distances" refer to distances to the shower axis where the LDF is well-defined (500 m to 1500 m). It is expected that this method will behave similarly to the previous method, but with a larger figure of merit for larger distances to the shower axis. Figure 8.11 shows the proton and iron distributions of signal ratio drawn on top of each other. It becomes easy to see that there is still quite a significant amount of overlap between the two distributions, which will make determining mass composition difficult. This method introduces a bias that shifts the mean signal ratio from the correct value obtained by calculating signal ratios per station. Calculating signals per station is considered the correct value because it obtains the signal ratio in the purest way. This method is considered much more realistic than the best case scenario, because it does not utilise the simulated dense rings.

8.2.4 Method 2 - Calculating Ratio of Signal per Event by Evaluating a Lateral Distribution of Signal Ratios

Lastly, method 2 calculates the ratio by evaluating a fit to the lateral distribution of individually calculated station ratios. As explained previously when describing the limitations of method 1, most events do not have high enough station multiplicity to freely fit LDFs. The lateral distribution fit used in this method is fixed to the shape of the SSD LDF and WCD LDF (the shape is fixed according to the values of β and γ in the SSD and WCD
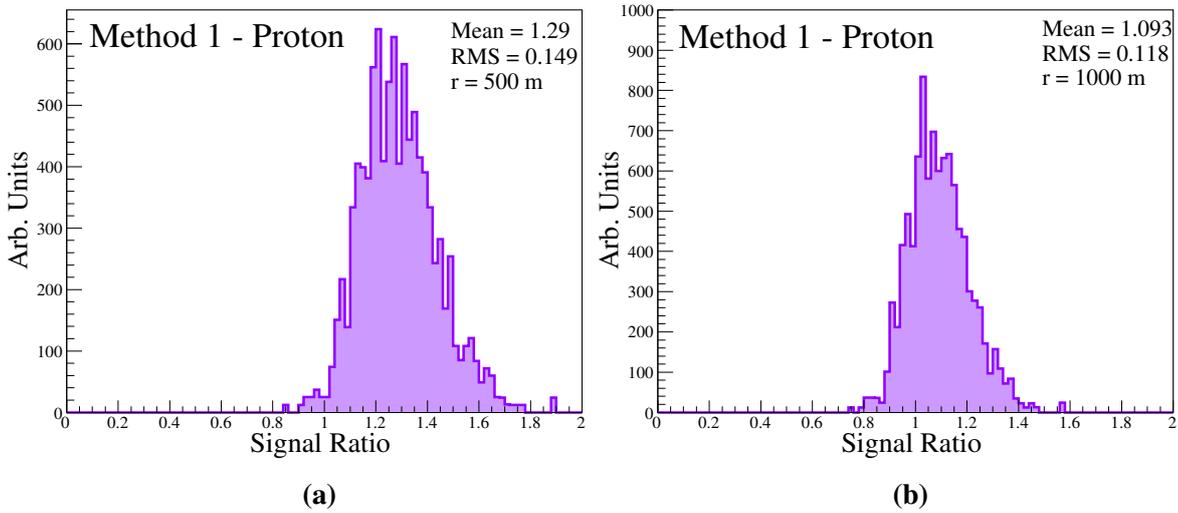
Figure 8.9: Ratio of signal distributions from calculating the signal ratio once per event, using an event fitted LDF for the SSD and WCD signals. Ratio of signal at (a) 500 m and (b) 1000 m for 1000 proton showers with primary energy of 10^{19} eV, zenith angle of 38° and random X_{max}.
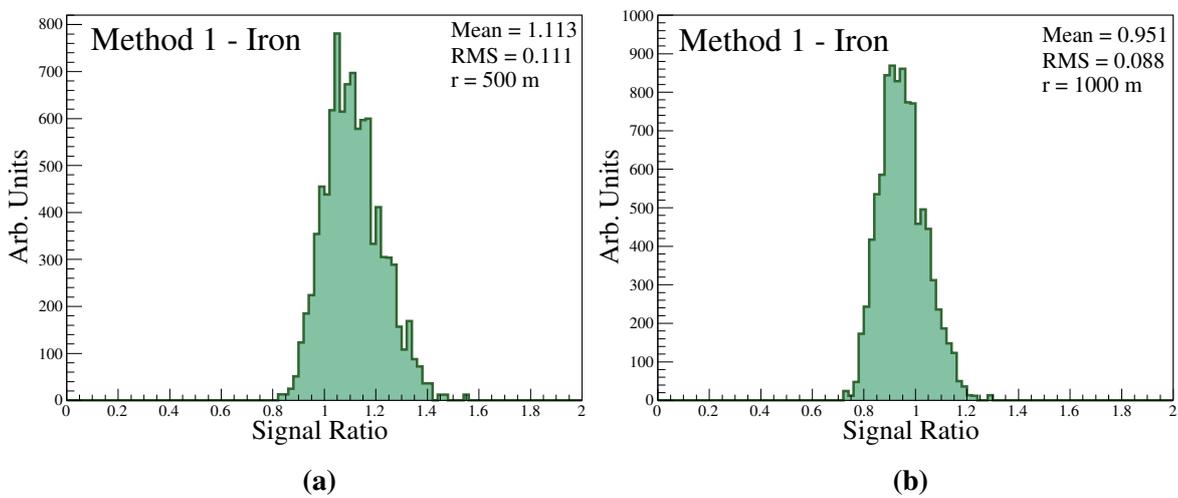
Figure 8.10: Ratio of signal distributions from calculating the signal ratio once per event, using an event fitted LDF for the SSD and WCD signals. Ratio of signal at (a) 500 m and (b) 1000 m for 1000 iron showers with primary energy of 10^{19} eV, zenith angle of 38° and random X_{max}.
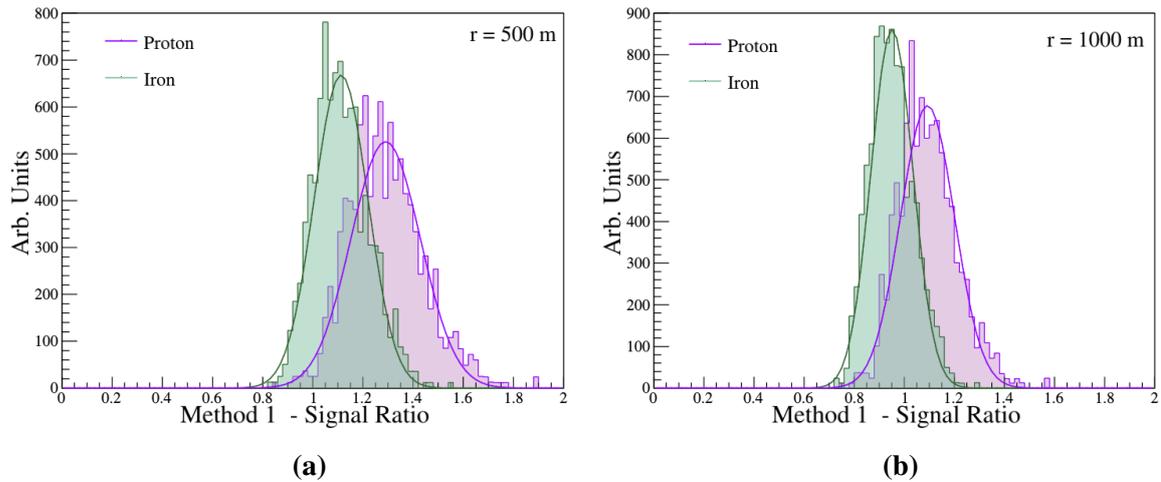
Figure 8.11: Ratio of signal distributions from calculating the signal ratio once per event, using an event fitted LDF for the SSD and WCD signals. Ratio of signal at (a) 500 m and (b) 1000 m for the 1000 proton showers and 1000 iron showers, both with primary energy of 10^{19} eV, zenith angle of $38°$ and random X_{max}.

LDF parameterisations, derived in Chapter 7), but the normalisation of the ratio is freely fit using all stations that are not within a dense ring. The distributions for the proton and iron datasets are shown in Figure 8.12 and Figure 8.13. The combination of the two is shown in Figure 8.14. This method has demonstrated results similar to method 1, with slightly higher accuracy and less biased, due to freely fitting the normalisation of the resulting signal ratio LDF. We say Method 2 is less biased because the mean signals are similar to the true signal ratios. Method 1 produces a noticeable bias in the mean value of the ratio distributions that may be attributed to an incorrect functional form of the LDF. Method 2 suffers from similar issues, however, there is no noticeable bias in the mean value of the ratio with this method because the normalisation of the ratio is freely fitted. The shape of the LDF is not fit because it would require events with a station multiplicity of at least seven, corresponding to the six degrees of freedom in the ratio LDF. The occurrence of real events with such high multiplicity is highly unlikely, and applying an energy threshold of sufficient magnitude to achieve this level of multiplicity would severely limit amount of available data. Overall, this method performs more accurately than method 1, due to freely fitting the normalisation after taking the ratio, however the shape of the ratio LDF is still fixed. This method may be the most
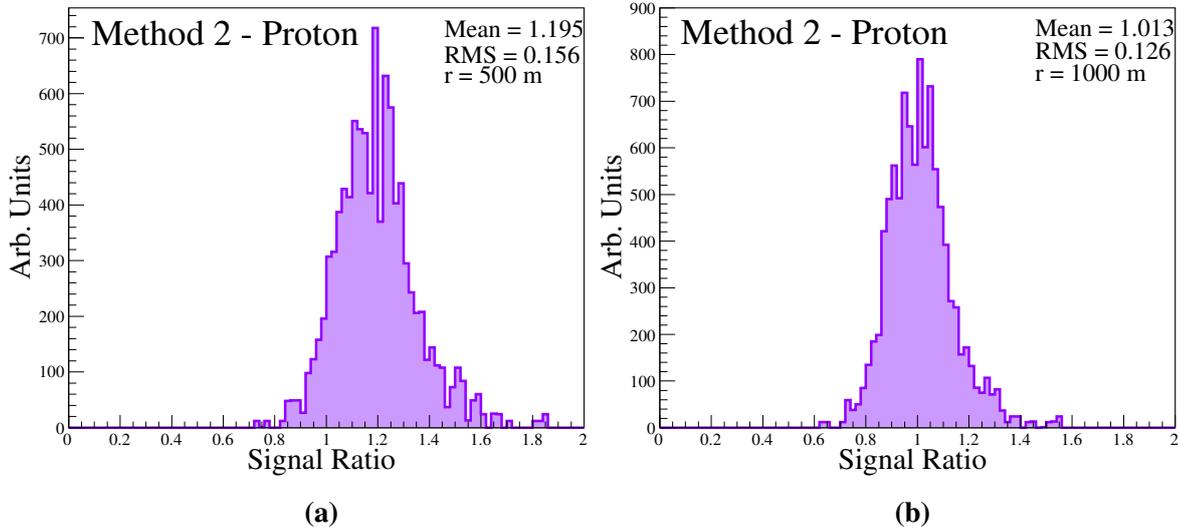
Figure 8.12: Ratio of signal distributions from calculating the signal ratio once per event, using a LDF fit to the calculated signal ratios of each station. Ratio of signal at (a) 500 m and (b) 1000 m for 1000 proton showers with primary energy of 10^{19} eV, zenith angle of 38° and random X_{max}.

suitable method to use in the future, when a significant amount of data becomes available for use.

8.2.5 Conclusion

Figure 8.15 shows the signal ratio for the best case scenario, and the two different methods as a function of distance to the shower axis. This figure clearly shows any potential bias in the magnitude of the evaluated signal ratios. It can be seen that method 1, which uses existing LDFs to calculate the ratio, has a consistent positive bias in the signal ratio with respect to the truth. However, it is also important to keep in mind that this bias is larger for proton as against iron showers, and, hence causes a better figure of merit value (Equation 8.1). The figure of merit has been calculated for each of the three cases as a function of distance to the shower axis, and the results are shown in Figure 8.16.

Method 1 and method 2 achieve similar shape for the figure of merit, it does not significantly decrease as distance to the shower axis increases. Both of these methods utilise parameterised LDF fits, assuming a fixed shape which has been derived from an ensemble of air shower simulations with 50% proton and 50% iron primary mass composition. As
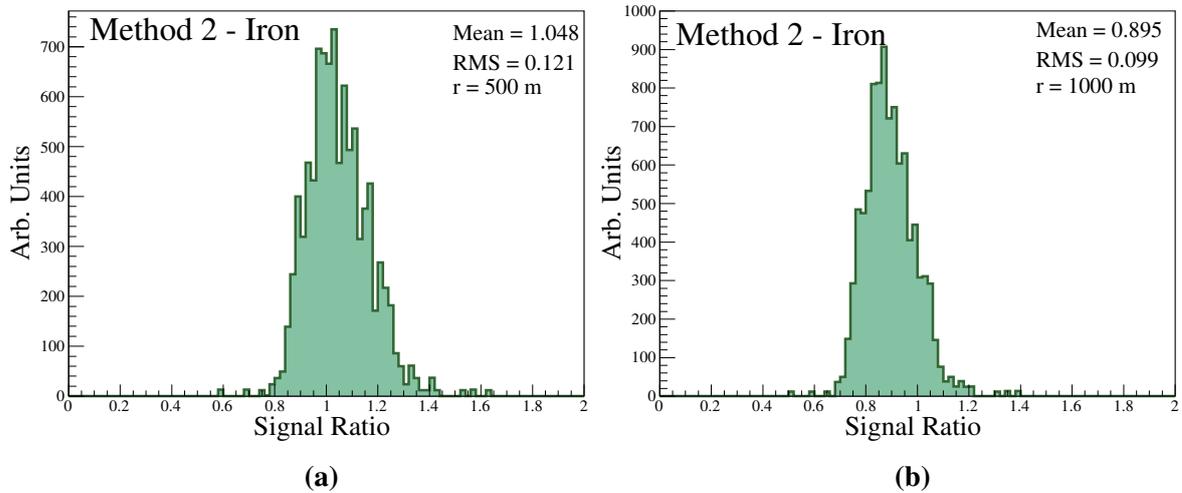
Figure 8.13: Ratio of signal distributions from calculating the signal ratio once per event, using a LDF fit to the calculated signal ratios of each station. Ratio of signal at (a) 500 m and (b) 1000 m for 1000 iron showers with primary energy of 10^{19} eV, zenith angle of $38°$ and random X_{max}.
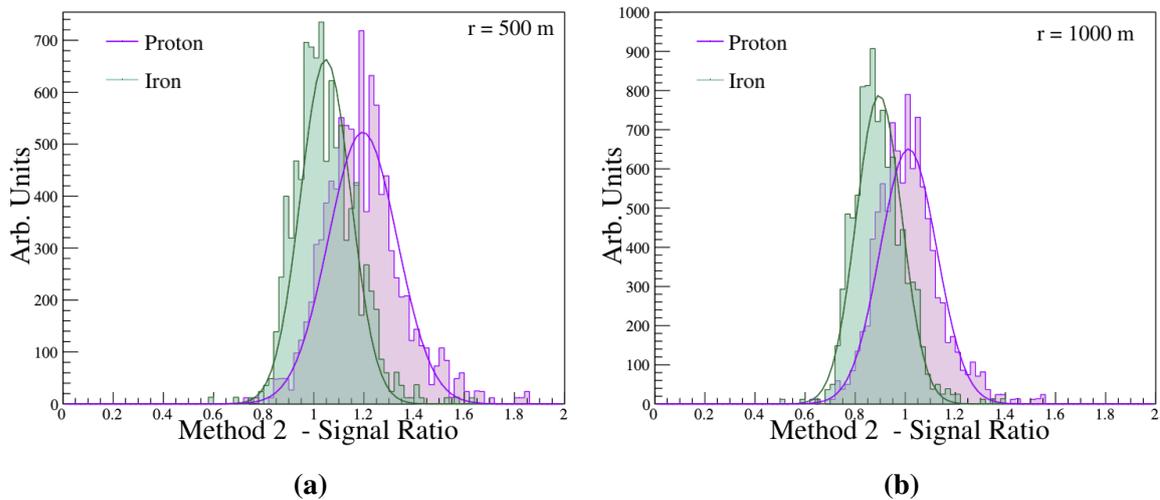
Figure 8.14: Ratio of signal distributions from calculating the signal ratio once per event, using a LDF fit to the calculated signal ratios of each station. Ratio of signal at (a) 500 m and (b) 1000 m for the 1000 proton showers and 1000 iron showers, both with primary energy of 10^{19} eV, zenith angle of $38°$ and random X_{max}.
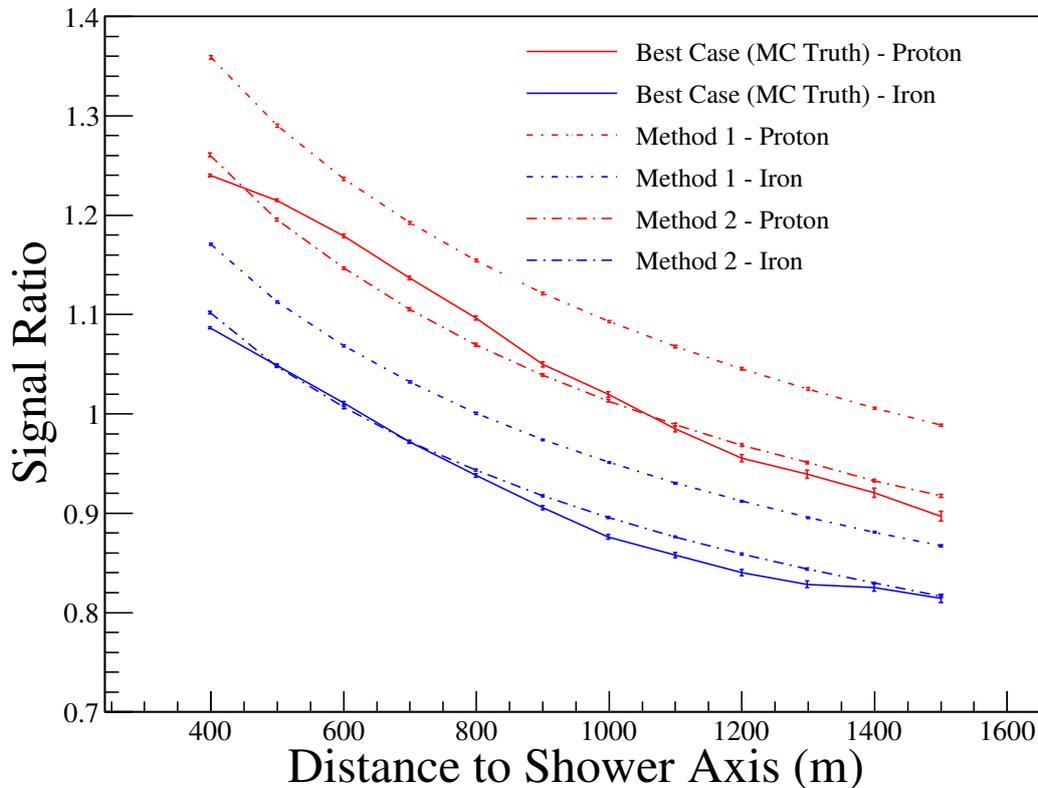
Figure 8.15: Signal ratios as a function of distance to the shower axis. The red and blue colours represent proton and iron showers, respectively. The different line styles represent the different methods used, with the solid line reflecting Monte Carlo truth. Errors are shown as the standard error of the mean. This figure shows that method 1 has a consistent positive bias on the signal ratio, however it is noted that this method also has a larger figure of merit, which is shown in Figure 8.16. Values for distances past 1500 m form the shower axis have been removed as the signals are commonly less than 1 MIP or VEM for an air shower with primary energy 10^{19} eV and zenith angle 38 degrees.

LDFs utilise information from stations across all distances to the shower axis, there is no sharp decline in the figure of merit. It should be noted that the fixed LDF shape may limit the potential of the figure of merit since differences in LDF shape play a crucial role in determining mass composition information. Method 1 performs slightly better than method 2 when it comes to the figure of merit, but a noticeable bias is present in the mean value of the ratio. The mean value calculated by the best case scenario with dense rings is considered
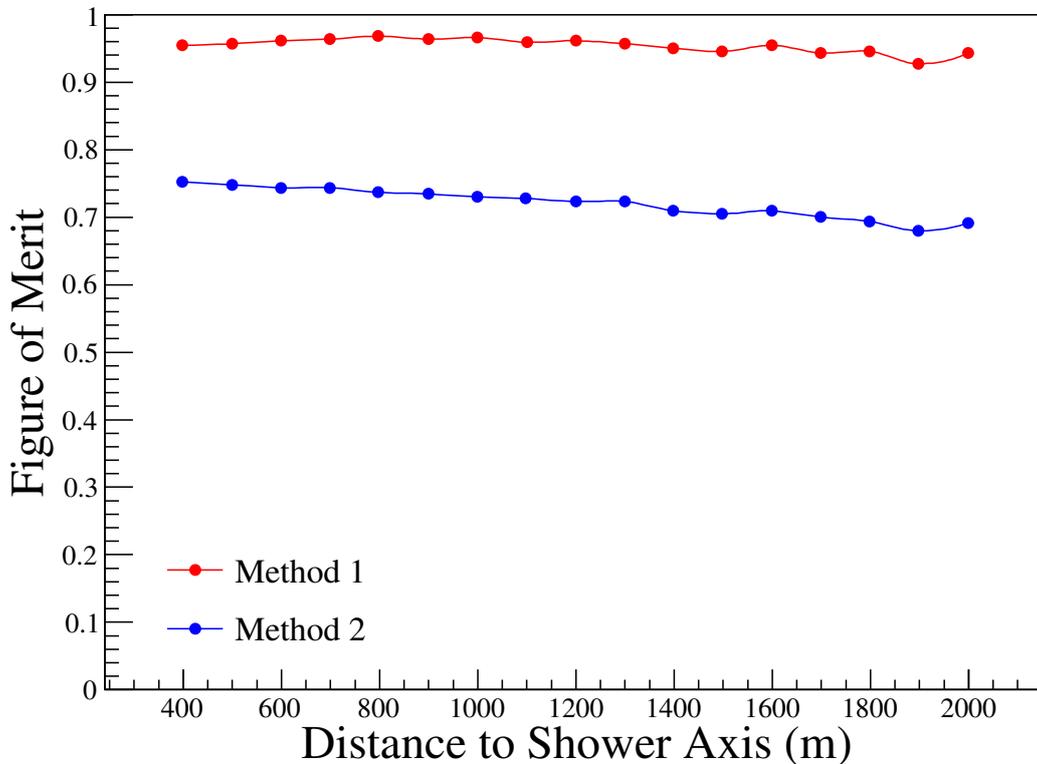
Figure 8.16: Figure of merit as a function of distance to the shower axis. The red line corresponds to method 1, where the ratio is calculated as the ratio of the SSD LDF and WCD LDF. The blue line represents method 2, where the ratio is calculated from evaluating a lateral distribution function fit to the ratio station SSD and WCD signals.

to be the correct value, as it reflects a signal ratio calculated in the purest way. This bias favours the figure of merit, and separates the mean values of the proton and iron distributions (as shown in Figure 8.15). The bias is in the same direction for both mass compositions, however it is much more significant for the proton composition. This bias may stem from fitting an incorrect LDF form for the simulation data. It is worth noting that both method 1 and method 2 can be substantially improved in the future by allowing more flexible fitting of the corresponding LDFs. It is suspected that the fixed shape is of the LDF in method 1 is responsible for the positive bias. The improvement for flexible fitting will require a large high-energy dataset which will take time to obtain.

8.3 Matrix Formalism

The matrix formalism method is an algorithm formulated using matrices to determine a parameter that is correlated with the number of muons that reach ground in an EAS. The muon content of an EAS is known to be correlated with the mass of the primary particle that initiated the shower, making it useful for mass determination. The matrix formalism is an adaptation of the original formalism designed for layered water-Cherenkov detectors [79]. In this study, the matrix formalism method is tested on very specific datasets to calculate the figure of merit. Similarly to the previous section on mass determination with signal ratios, the figure of merit is used to provide insights on the effectiveness of mass determination.

The first step of the formalism involves the separation of the flux of secondary particles into two distinct components; the electromagnetic component and the muonic component. The electromagnetic component is comprised of electrons, positrons and photons, whereas the muonic component is comprised of muons and anti-muons. The separation into two components allows for an analytical solution for the muon signal from either the SSD or WCD detectors, when combined with the total signal from both detectors. The total energy deposited by both components in both detectors is written as $E^{\gamma,e^{\pm}}$ for the electromagnetic component and $E^{\mu^{\pm}}$ for the muonic component. These total energy deposits can be further divided and expressed as the sum of the energy deposits in each detector,

$$E^{\gamma,e^{\pm}} = E^{\gamma,e^{\pm}}_{\text{SSD}} + E^{\gamma,e^{\pm}}_{\text{WCD}} \tag{8.5}$$

$$E^{\mu^{\pm}} = E^{\mu^{\pm}}_{\text{SSD}} + E^{\mu^{\pm}}_{\text{WCD}} \tag{8.6}$$

where $E^{\gamma,e^{\pm}}_{\text{SSD}}$ represents the total energy deposited in the SSD, from particles within the electromagnetic component of the air shower. $E^{\mu^{\pm}}_{\text{SSD}}$ represents the portion of total energy deposited in the SSD by particles of the muonic component, and similarly, $E^{\gamma,e^{\pm}}_{\text{WCD}}$ and $E^{\mu^{\pm}}_{\text{WCD}}$ represent energy deposits in the WCD for the respective air shower components. The fraction of energy deposited in the SSD out of the total energy deposited in both detectors for the electromagnetic component is then given by

$$m = \frac{E^{\gamma,e^{\pm}}_{\text{SSD}}}{E^{\gamma,e^{\pm}}_{\text{SSD}} + E^{\gamma,e^{\pm}}_{\text{WCD}}} \tag{8.7}$$

and similarly, for the muonic component,

$$n = \frac{E_{\text{SSD}}^{\mu^{\pm}}}{E_{\text{SSD}}^{\mu^{\pm}} + E_{\text{WCD}}^{\mu^{\pm}}} \tag{8.8}$$

These two additional variables allow for the matrix expression of the observables (SSD and WCD signals) and reconstruction quantities ($E^{\gamma,e^{\pm}}$ and $E^{\mu^{\pm}}$) to be formed as

$$\begin{pmatrix} E_{\text{SSD}} \\ E_{\text{WCD}} \end{pmatrix} = \begin{pmatrix} m & n \\ 1-m & 1-n \end{pmatrix} \begin{pmatrix} E^{\gamma,e^{\pm}} \\ E^{\mu^{\pm}} \end{pmatrix} \tag{8.9}$$

The total energy deposit of either the electromagnetic component or the muonic component of an air shower can be obtained as a function of the total energy deposit in the SSD and the WCD through matrix inversion.

For the electromagnetic component, the total energy deposit is given by

$$E^{\gamma,e^{\pm}} = \frac{1}{m-n}[(1-n)E_{\text{SSD}} - nE_{\text{WCD}}], \tag{8.10}$$

and for the muonic component, the total energy deposit is given by

$$E^{\mu^{\pm}} = \frac{1}{m-n}[(m-1)E_{\text{SSD}} + mE_{\text{WCD}}]. \tag{8.11}$$

To determine the total energy deposit of the electromagnetic component within a specific detector, Equation 8.7 can be substituted into Equation 8.10. For example, the total energy deposit of the electromagnetic component within the WCD can be solved as

$$E_{\text{WCD}}^{\gamma,e^{\pm}} = \frac{(1-m)}{m-n}[(1-n)E_{\text{SSD}} - nE_{\text{WCD}}]. \tag{8.12}$$

Similarly, through substitution of Equation 8.8 into Equation 8.11, the total energy deposit of the muonic component within a specific detector can be found. For example, the total energy deposit of the muonic component within the WCD can be solved as

$$E_{\text{WCD}}^{\mu^{\pm}} = \frac{(1-n)}{m-n}[(m-1)E_{\text{SSD}} - mE_{\text{WCD}}]. \tag{8.13}$$

The matrix formalism method allows for the calculation of the energy deposit due to either the electromagnetic component or muonic component of an air shower, based solely on the total energy deposited in the SSD and WCD, along with parameters m and n. For practical use, it is necessary to establish a relationship between the measured calibrated signal and the energy deposit. This relationship is assumed to be a constant proportionality, which can vary for each detector and is independent of component type. The proportionality between the signal and the energy deposit in each detector can be expressed with proportionality constant C as

$$S_{\mathrm{WCD}} = C_{\mathrm{WCD}} E_{\mathrm{WCD}}, \tag{8.14}$$

$$S_{\mathrm{SSD}} = C_{\mathrm{SSD}} E_{\mathrm{SSD}}. \tag{8.15}$$

where the units of S_{SSD} and S_{WCD} are MIP and VEM, respectively. By substituting Equations 8.14 and 8.15 into the equations for m and n, as well as Equation 8.12, it is possible to rewrite the result of the matrix inversion in terms of detector signal.

For example, here is Equation 8.12 expressed in terms of detector signal,

$$\frac{S_{\mathrm{WCD}}^{\gamma,e^{\pm}}}{C_{\mathrm{WCD}}} = \frac{(1-m)}{(m-n)} \left[(1-n)\frac{S_{\mathrm{SSD}}}{C_{\mathrm{SSD}}} - n\frac{S_{\mathrm{WCD}}}{C_{\mathrm{WCD}}} \right] \tag{8.16}$$

and here are the parameters m and n represented in terms of detector signal

$$m = \frac{C_{\mathrm{WCD}} S_{\mathrm{SSD}}^{\gamma,e^{\pm}}}{(C_{\mathrm{WCD}} S_{\mathrm{SSD}}^{\gamma,e^{\pm}} + C_{\mathrm{SSD}} S_{\mathrm{WCD}}^{\gamma,e^{\pm}})}, \tag{8.17}$$

$$n = \frac{C_{\mathrm{WCD}} S_{\mathrm{SSD}}^{\mu^{\pm}}}{(C_{\mathrm{WCD}} S_{\mathrm{SSD}}^{\mu^{\pm}} + C_{\mathrm{SSD}} S_{\mathrm{WCD}}^{\mu^{\pm}})}. \tag{8.18}$$

Now the matrix formalism method from Equation 8.9 can be rewritten in terms of detector signal and proportionality constant C,

$$\begin{pmatrix} \frac{S_{\mathrm{SSD}}}{C_{\mathrm{SSD}}} \\ \frac{S_{\mathrm{WCD}}}{C_{\mathrm{WCD}}} \end{pmatrix} = \begin{pmatrix} m & n \\ 1-m & 1-n \end{pmatrix} \begin{pmatrix} \frac{S_{\mathrm{SSD}}^{\gamma,e^{\pm}}}{C_{\mathrm{SSD}}} + \frac{S_{\mathrm{WCD}}^{\gamma,e^{\pm}}}{C_{\mathrm{WCD}}} \\ \frac{S_{\mathrm{SSD}}^{\mu^{\pm}}}{C_{\mathrm{SSD}}} + \frac{S_{\mathrm{WCD}}^{\mu^{\pm}}}{C_{\mathrm{WCD}}} \end{pmatrix} \tag{8.19}$$

By expressing the matrix inversion from this formalism in terms of the detector signals and solving it with the substituation of m and n, the constants of proportionality (C_{SSD} and C_{WCD}) will cancel out. As a result, the formalism no longer explicitly depends on these parameters. With this in mind, the matrix formalism method can be further rewritten and simplified for use with detector signals,

$$\begin{pmatrix} S_{\mathrm{SSD}} \\ S_{\mathrm{WCD}} \end{pmatrix} = \begin{pmatrix} a & b \\ 1-a & 1-b \end{pmatrix} \begin{pmatrix} S^{\gamma,e^{\pm}} \\ S^{\mu^{\pm}} \end{pmatrix} \tag{8.20}$$

with parameters a and b replacing m and n,

$$a = \frac{S_{\mathrm{SSD}}^{\gamma,e^{\pm}}}{S_{\mathrm{SSD}}^{\gamma,e^{\pm}} + S_{\mathrm{WCD}}^{\gamma,e^{\pm}}} \tag{8.21}$$

$$b = \frac{S_{\mathrm{SSD}}^{\mu^{\pm}}}{S_{\mathrm{SSD}}^{\mu^{\pm}} + S_{\mathrm{WCD}}^{\mu^{\pm}}} \tag{8.22}$$

Parameters a and b are written purely in terms of detector signals, and can be determined with simulated signals. Currently, the values of a and b are obtained from simulating the responses of the WCD and the SSD to simulated air showers; more specifically, they are the average values of a and b from all simulated data for signals 500 m from the shower axis. The distributions of a and b at 500 m and 1000 m are shown later in Section 8.3.3. It is crucial that a and b yield similar values for cosmic ray primaries of different mass, and ideally, they should be independent of the chosen hadronic model. If these conditions are not met, then a and b will not provide unbiased estimates in electromagnetic and muonic signal in either detector. The remainder of this study references the WCD reconstructed muon signal which has been calculated by performing matrix inversion on Equation 8.20 to yield

$$S_{\mathrm{WCD}}^{\mu^{\pm}} = \frac{(1-b)}{a-b}[(a-1)S_{\mathrm{SSD}} - aS_{\mathrm{WCD}}]. \tag{8.23}$$

8.3.1 Understanding the Uncertainty of the WCD Reconstructed Muon Signal

Previous work has analytically demonstrated [72], under the assumption of a Gaussian distribution, that the uncertainty in the WCD reconstructed muon signal is defined as

$$\sigma_{\text{WCD}}^{\mu^\pm} = \sqrt{\left(\frac{\partial S_{\text{WCD}}^{\mu^\pm}}{\partial S_{\text{SSD}}}\sigma_{\text{SSD}}\right)^2 + \left(\frac{\partial S_{\text{WCD}}^{\mu^\pm}}{\partial S_{\text{WCD}}}\sigma_{\text{WCD}}\right)^2 + 2\text{COV}\frac{\partial S_{\text{WCD}}^{\mu^\pm}}{\partial S_{\text{SSD}}}\frac{\partial S_{\text{WCD}}^{\mu^\pm}}{\partial S_{\text{WCD}}}} \qquad (8.24)$$

where σ_{SSD} and σ_{WCD} are the SSD and WCD signal uncertainties, respectively. The two partial derivatives simplify to constants dependent on the values of a and b. However, the calculated figure of merit, obtained by comparing the WCD reconstructed muon signal distributions from two different cosmic ray masses under very specific conditions, is not sufficiently large enough to easily distinguish between the two masses. Here, approaches are taken to understand the major contributing factors to the variance of the reconstructed muon signal distribution, and outline possible approaches or strategies that could be utilised in future studies to improve the figure of merit, and enhance the ability to separate information on cosmic ray mass.

Several variations of the same simulated datasets are used to create generate unique scenarios, where each scenario results in a distribution of reconstructed WCD muon signal. These scenarios are designed to facilitate a clear separation and understanding of the contributing factors to the variance observed in the signal.

The distributions can be analysed in conjunction with the characteristics of their respective datasets to determine the level of uncertainty arising from various physics processes. To achieve this, four unique datasets have been created, each customised to isolate and investigate three major contributions to the variance: shower to shower fluctuations, signal uncertainty and azimuthal asymmetry.

- **Dataset 1:** 1000 simulations of just one unique proton CORSIKA shower. Signals are taken as the average of all twelve stations from a simulated dense ring. This set aims to remove any resolution effects due to shower development (X_{max}) and detector signal asymmetry.

- **Dataset 2:** 1000 simulations, each with a different proton CORSIKA shower (1000 different showers). Signals are taken as the average of all twelve stations from the

simulated dense rings. This set will remove resolution effects due to detector signal asymmetry. Shower development will affect the variance in this set.

- **Dataset 3:** 1000 simulations of just one unique proton CORSIKA shower. All twelve stations in each ring are used separately. This set will not include any resolution effects due to shower development, leaving the set with shower particle sampling, signal uncertainty and signal asymmetry effects in the data.

- **Dataset 4:** 1000 simulations each with a different proton CORSIKA shower (1000 different showers). All twelve stations in each ring are used. This last set will include all possible effects of uncertainty from the CORSIKA library.

As outlined above, each dataset contains different contributing factors to the variance of the WCD reconstructed muon signal. Dataset 1, and dataset 2 differ solely in terms of shower-to-shower fluctuations. Consequently, the uncertainty in dataset 1 and the uncertainty contribution due to shower-to-shower fluctuations added in quadrature will result in the total uncertainty of dataset 2. It is important to note that the covariance of each dataset varies greatly, and it must be recalculated whenever input data is altered.

The contribution from shower-to-shower fluctuations results in an increase of approximately ± 30 VEM to signal uncertainty (from a typical ≈ 400 VEM signal at 500 m from the shower axis for proton air showers with primary energy of 10^{19} eV, and shower zenith of $38°$). A similar approach could be taken to approximate the uncertainty contribution due to signal asymmetry, sampling and signal uncertainty using dataset 2 and dataset 3.

It should be noted that the values chosen for a and b affect the variance of the reconstructed muon signal, and any uncertainties associated with a and b need to be properly propagated. In this study, the values of a and b are fixed at $a = 0.5802$ and $b = 0.3225$, respectively. These values were determined as the averages from the datasets used, for signals at 500 m from the shower axis. The dataset contains an ensemble of all showers within the simulation library, which consists of 1000 unique proton and iron showers, with primary energy of 10^{19} eV, random X_{max} and zenith angles of 38 degrees. While the a and b parameters display complexity and potential dependencies, this study still opt to fix these parameters. The decision to fix a and b is driven by the focus of this study, which is to investigate the potential of the matrix formalism method. The figure of merit (Eqn 8.1) has been calculated as a function of distance to the shower axis utilising the matrix formalism method, and the results are shown in Figure 8.17.
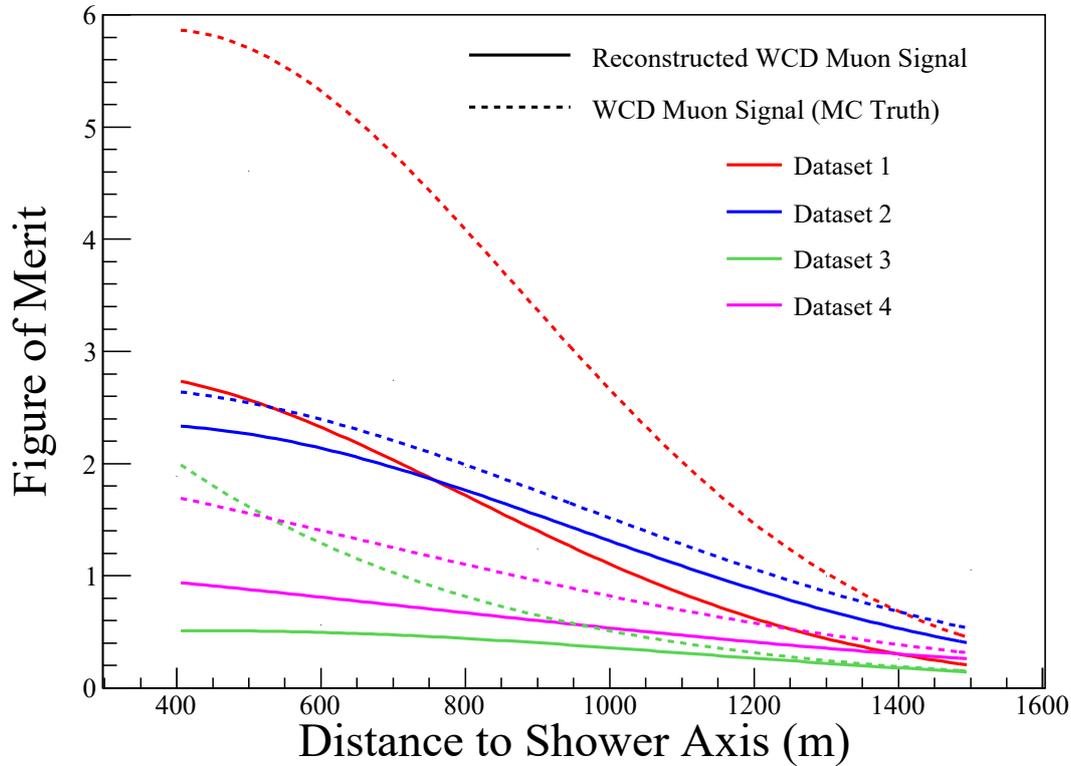
Figure 8.17: Figure of merit (Eqn 8.1 as a function of distance to the shower axis, for reconstructed WCD muon signal and the simulated WCD muon signal (Monte-Carlo truth). The figure of merit (Eqn 8.1 has been calculated for four different simulated datasets that have been carefully crafted to represent potential goals to work for in real data. Dataset 4 represents is the closest representation to the current state of real data. Dataset 2 represents a case where signal asymmetry and signal statistical fluctuations have been minimised. Solid and dashed lines represent the reconstructed WCD muon signal and Monte-Carlo WCD muon signal, respectively. The different colours correspond to different datasets. Dataset 1 contains no signal asymmetry, no shower-to-shower fluctuations and minimised signal statistical fluctuations. Dataset 2 contains no signal asymmetry, but it does contain shower-to-shower fluctuations and minimised signal statistical fluctuations. Dataset 3 contains signal asymmetry, but no shower-to-shower fluctuations. Dataset 4 contains both signal asymmetry and shower-to-shower fluctuations.

8.3.2 Matrix Formalism Figure of Merit

The figure of merit relies on the accuracy of the reconstructed WCD muon signal. It is known that the WCD muon content is dependent on the mass of the cosmic ray primary

particle, where heavier masses result in larger amounts of WCD muon signal due to an in-creased number of muons generated during shower development. In addition to calculating the figure of merit for the four different datasets mentioned earlier, it has also been computed for the Monte Carlo Truth. The Monte Carlo truth here refers to the figure of merit derived from the true simulated muon signal in the WCD, which does not require use of the matrix formalism method. By definition, if the "correct" value of a and b was chosen for every air shower, then the reconstructed WCD muon signal will equal the true WCD muon signal. The difference between the Monte Carlo truth and the reconstructed WCD muon signal in dataset 1 differ only due to the difference between the fixed a and b values, and the "correct" a and b value for that specific air shower. Dataset 1, which has been manipulated to exclude shower-to-shower fluctuations and signal asymmetry, yields the highest figure of merit values. Although dataset 1 yields the largest figure of merit values which are well above the target of approximately 1.5, it would require a significantly large amount of data to achieve similar results with real data.

Dataset 2, which contains shower-to-shower fluctuations, no signal asymmetry, and re-duced signal statistical fluctuations, results in the second highest figure of merit values. As there are no dense rings of stations in real data, signal asymmetry could be reduced by us-ing signal asymmetry parameterisations for both the SSD and WCD. This method could be practical for use with real data, as the figure of merit values are still above the target of 1.5. Dataset 4 returns larger figure of merit values than dataset 3. This is because the shower-to-shower fluctuations are a small contribution to the total signal uncertainty compared to the contribution from signal asymmetry, and, shower-to-shower fluctuations also result in a sig-nificant shift in the mean amount of signal in each detector, which is in favour of the figure of merit. Dataset 3 and dataset 4 do not yield sufficiently high enough values for the figure of merit to be practical. All the air showers used in this study have a zenith angle of 38 deg, and primary energy of 10^{19} eV. The presented figure of merit values serve as an upper limit of potential for showers of these properties. It is apparent that real data will face challenges in reaching this limit without further effort. At a bare minimum, significant reductions in signal asymmetry will be necessary for both detectors in the future.

8.3.3   The Effect of a and b Accuracy on Reconstructed Signal

The values of a and b are crucial as they indirectly impact the uncertainty in the recon-structed muon signal by adjusting the values of $\frac{\delta S^{\mu}}{\sigma_{\text{SSD}}}$ and $\frac{\delta S^{\mu}}{\sigma_{\text{WCD}}}$. If the values of a and b deviate
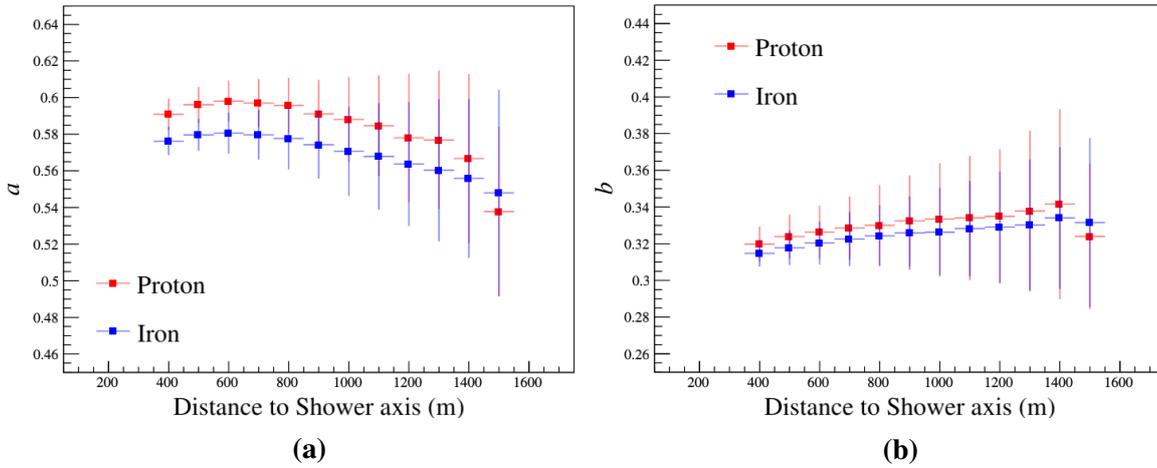
Figure 8.18: (a) The value of a as a function of distance to the shower axis. (b) The value of b as a function of distance. Blue and red points represent values of a or b from simulated iron and proton QGSJET-II.04 air showers, respectively. The error bars represent the standard error of the mean.

from their correct values for a given shower, the reconstructed muon signal will deviate further from the true value. The correct values are a and b calculated from Monte Carlo truth. In some cases, the reconstructed muon signal can result in negative signal. In known cases where the reconstructed muon signal has been negative, it can be attributed to the air shower X_{max} value being far from the mean X_{max} of the CORSIKA dataset. This means that a and b have a small mass dependence. Figures 8.21a and 8.21b show the dependency and spread of a and b as a function of distance to the shower axis, whereas Figures 8.19a and 8.19b show the dependency of a and b as a function of shower X_{max}. Since a and b are not constant, future studies will benefit from creating parameterisations for their values as a function shower energy and zenith angle, and potentially distance to the shower axis. It is evident that a and b have a dependency on distance to the shower axis as well as shower X_{max}. A dependency on distance to the shower axis can be thought of as dependency on shower development, which reinforces the fact that a and b have a dependency on shower X_{max}.

Parameters a and b depend on various factors, including but not limited to the distance to the shower axis and shower X_{max}. As stated previously, a and b are fixed at the average values obtained from all simulated data for signals 500 m from the shower axis. The distributions of a and b for at 500 m and 1000 m are shown in Figure 8.20 and Figure 8.21, respectively.
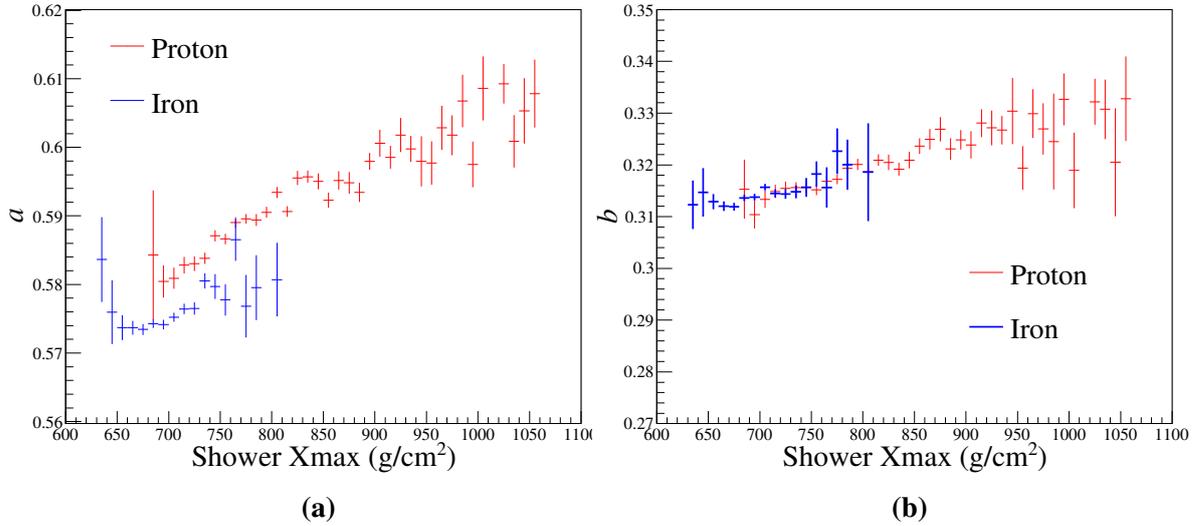
Figure 8.19: (a) The value of a as a function of shower X_{max}. (b) The value of b as a function of shower X_{max}. Blue and red points represent values of a or b from simulated iron and proton QGSJET-II.04 air showers, respectively. The error bars represent the standard error of the mean.

These distributions show that the possible range of values for a and b will increase as distance to the shower axis increases; this will translates into a larger uncertainty for the reconstructed WCD muon signal. Additionally, it has been found that small proportional shifts in a and b have a significant outcome in the accuracy of the reconstructed WCD muon signal.

It has also been acknowledged that the accuracy of a and b affects the accuracy of the reconstructed WCD muon signal. The impact of the accuracy is studied at two distances, namely 500 m and 1000 m, by applying small percentage shifts to the a and b parameters. Figure 8.22 and Figure 8.23 show how much a given reconstructed WCD muon signal changes from shifting either parameter by $\pm$ 2.5%. The $\pm$ 2.5% range corresponds to the following values for a and b:

$$a = 0.5802 \pm 0.145, \tag{8.25}$$

$$b = 0.3225 \pm 0.008. \tag{8.26}$$

In the context of X_{max}, a significant issue only arises when a given shower has a value of X_{max} which is on a tail of the X_{max} distribution. For example, if a proton shower had an X_{max} value of 1000 g/cm^2, then the *correct* values of a and b (as shown in to figure 8.19a and figure
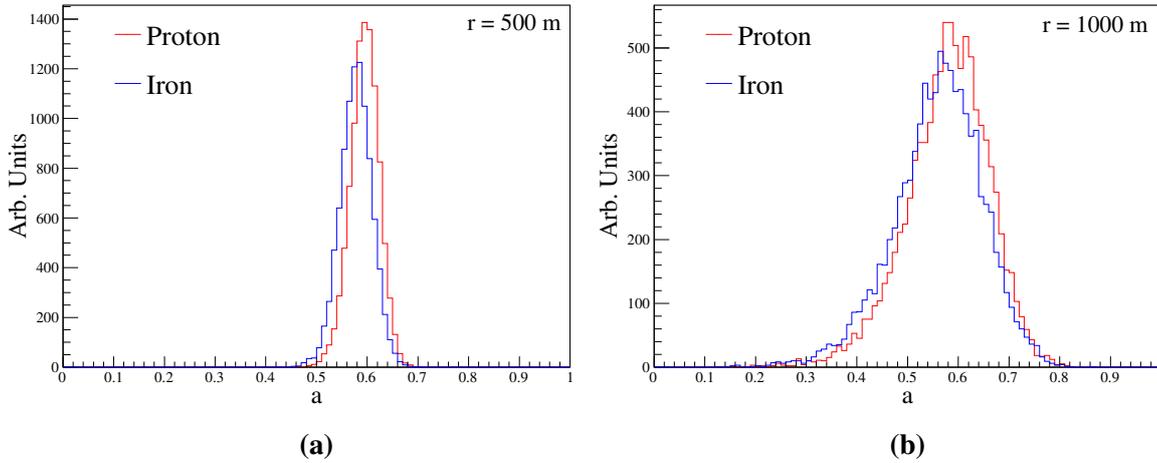
Figure 8.20: (a) The distribution of a at 500 m from the shower axis, for proton and iron air shower dataset used in this study. (b) The distribution of a at 1000 m from the shower axis, for proton and iron air shower dataset used in this study.
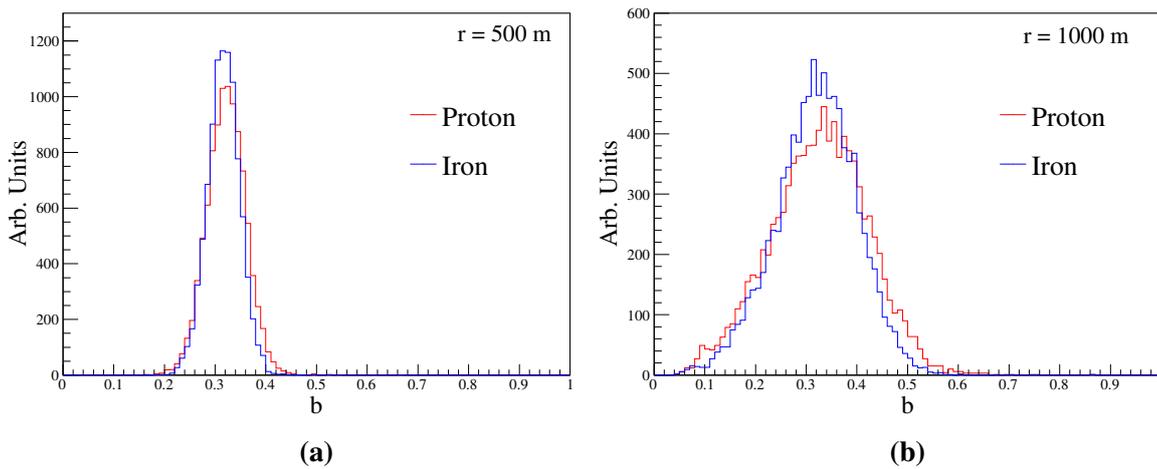
Figure 8.21: (a) The distribution of b at 500 m from the shower axis, for proton and iron air shower dataset used in this study. (b) The distribution of b at 1000 m from the shower axis, for proton and iron air shower dataset used in this study.

8.19b) should be around 0.61, and 0.33, respectively. These values are well above the 2.5% examples given before, and will result in the reconstructed WCD muon signal increasing by upwards of 60% (at 500 m), or 20% (at 1000 m) for proton showers. On the other end of the range, there may be large decreases in reconstructed signal of up to -60%. There are some
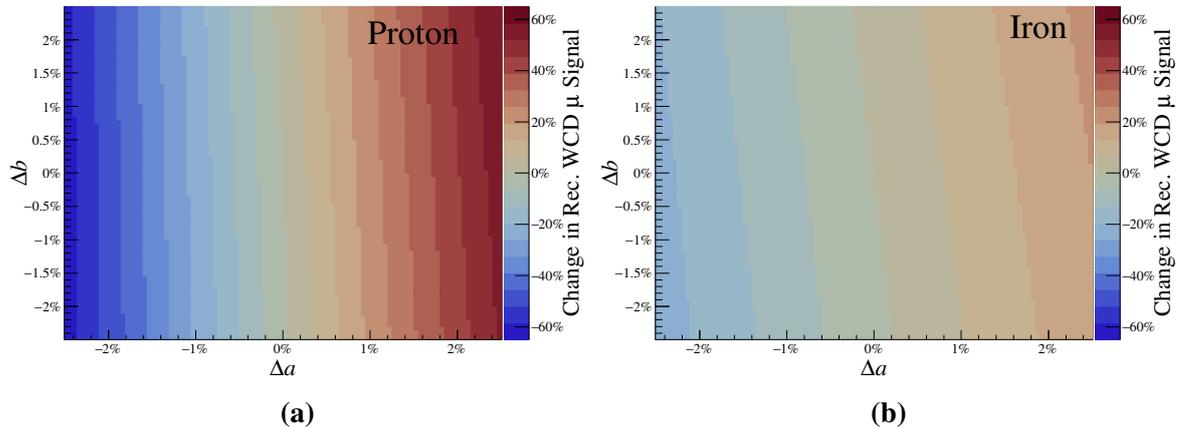
Figure 8.22: The corresponding percentage shift in a given reconstructed WCD muon signal at 500 m from the shower axis, resulting from very small shifts in a and b for (a) proton air showers and (b) iron air showers with an energy of 10^{19} eV and zenith angle of $38°$.
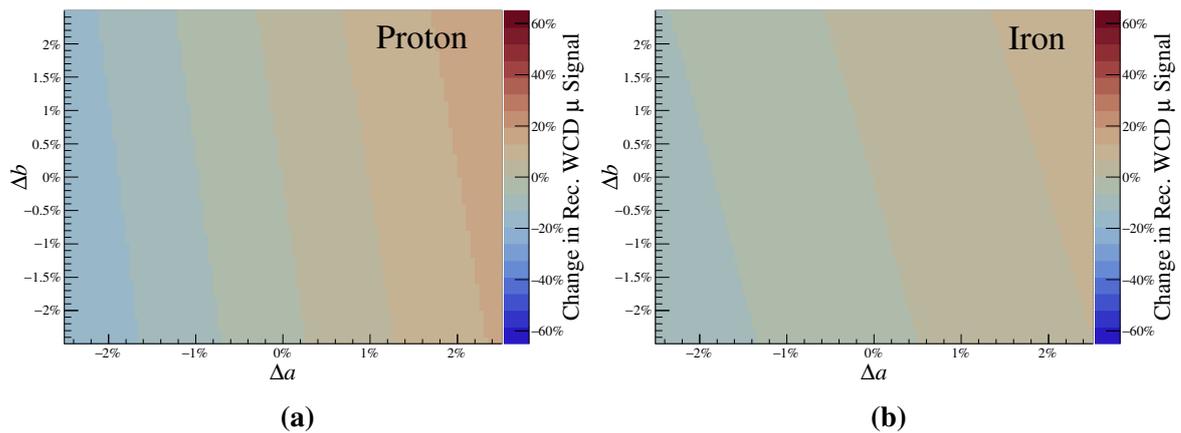
Figure 8.23: The corresponding percentage shift in a given reconstructed WCD muon signal at 1000 m from the shower axis, resulting from very small shifts in a and b for (a) proton air showers and (b) iron air showers with an energy of 10^{19} eV and zenith angle of $38°$.

extreme cases where which may result in reconstructed signals less than zero, for decreases greater than -100%. This expresses an important concern, where a very small relative change in a and b can result in a significant relative change in reconstructed signal.

The accuracy of a has a noticeably larger impact on the reconstructed WCD muon signal compared to the accuracy of b at smaller distances from the shower axis. This is related to

the proportion of electromagnetic and muonic signal within the total signal. When the total signal is primarily electromagnetic, any variation in a will have a substantial impact on the reconstructed signal. Conversely, when the muonic signal dominates, variations in b will be more influential. Furthermore, the proportion of electromagnetic and muonic signal also depends on the air shower primary particle mass, with heavier nuclei resulting in a larger quantity of muons. Consequently, it is not surprising to see that iron showers have a larger dependency on b than proton showers. While it is impossible to parameterise a perfect value for a and b for any given event without prior knowledge of the muon signal in each detector, the impact of incorrectness can be significantly reduced by parameterising a and b. In future studies, the parameters a and b should potentially be parameterised as a function of distance to the shower axis (potentially involving a parameterisation based on shower zenith angle and energy). This approach would minimise the uncertainty propagated to the reconstructed WCD muon signal. Additionally, all approaches here rely on EAS simulations, which may be deficient themselves.

8.3.4 Conclusion

This study aimed to determine the mass of primary cosmic ray particles using the reconstructed WCD muon signal. The figure of merit has been used as a way to grade and understand the limitations and challenges that need to be overcome to potentially identify cosmic ray mass with real data.

The results show that the accuracy of the a and b parameters, which are crucial in the calculation of the reconstructed muon signal, significantly impacts the accuracy of calculation. It has been shown that small deviations in these parameters lead to substantial changes in the reconstructed WCD muon signal, sometimes even resulting in negative values. For fixed values of a, and b, this study shows that calculating the reconstructed WCD muon signal with reasonable accuracy to allow for an acceptable figure of merit value is already quite difficult. The accuracy of a and b hinges on proper consideration of the distance to the shower axis (r), X_{max}, energy, and zenith angle. This study has highlighted the importance of parameterising a and b as functions of these factors. This approach could help minimise the uncertainties in the reconstructed WCD muon signal, providing more accurate results in future studies. As a result, it has been shown that there is still a large amount of development to occur in this area before the matrix formalism methodology can be used with confidence.

Acronyms

ADC analogue-to-digital converter.

Auger the Pierre Auger Observatory.

CDAS central data acquisition system.

CIC Constant Intensity Cut.

CORSIKA COsmic Ray SImulations for KAscade.

EA engineering array.

EAS extensive air shower.

FADC flash analog to digital converter.

FD fluorescence detector.

FNAL Fermi National Accelerator Laboratory.

ISM interstellar medium.

KIT Karlsruhe Institute of Technology.

LDF lateral distribution function.

LIP the Laboratory for Instrumentation and Experimental Particle Physics.

MARTA Muon Auger RPC for the Tank Array.

MC Monte-Carlo.

MIP minimum ionising particle.

NKG Nishimura-Kamata-Greisen.

PDF probability distribution function.

PMT photomultiplier tube.

RPC resistive plate chamber.

SD surface detector.

SD-1500 array with 1500m spacing.

SD-433 array with 433m spacing.

SD-750 array with 750m spacing.

SSD scintillator surface detector.

SSD PPA scintillator surface detector pre-production array.

SUGAR Sydney University Giant Air-shower Recorder.

TA the Telescope Array Experiment.

UB unified board.

UHECR ultra high energy cosmic ray.

UMD Underground Muon Detector.

UUB upgraded unified board.

VEM vertical equivalent muon.

WCD water-Cherenkov detector.

WLS wavelength-shifting.

Appendix A

Scintillator LDF Residuals for Pure Iron and Proton Simulations

As an extension of the plots shown in Chapter 7. The residuals and relative differences for pure iron and proton showers are shown in the Figures under this appendix. Similarly to the results discussed in the relevant chapter, the proton and iron residuals all follow similar patterns. The scintillator surface detector (SSD) lateral distribution function (LDF) parameterisation performs better for distances closer to the shower axis, and for larger signals. The bias for reconstructed energy estimator S(1000) also follows the same trend, it is often overestimated for smaller energy air showers, and then slightly underestimated for higher energy air showers. The figures in chapter 7 represent the average of the plots here, as the mixed configuration is 50/50 of proton/iron.
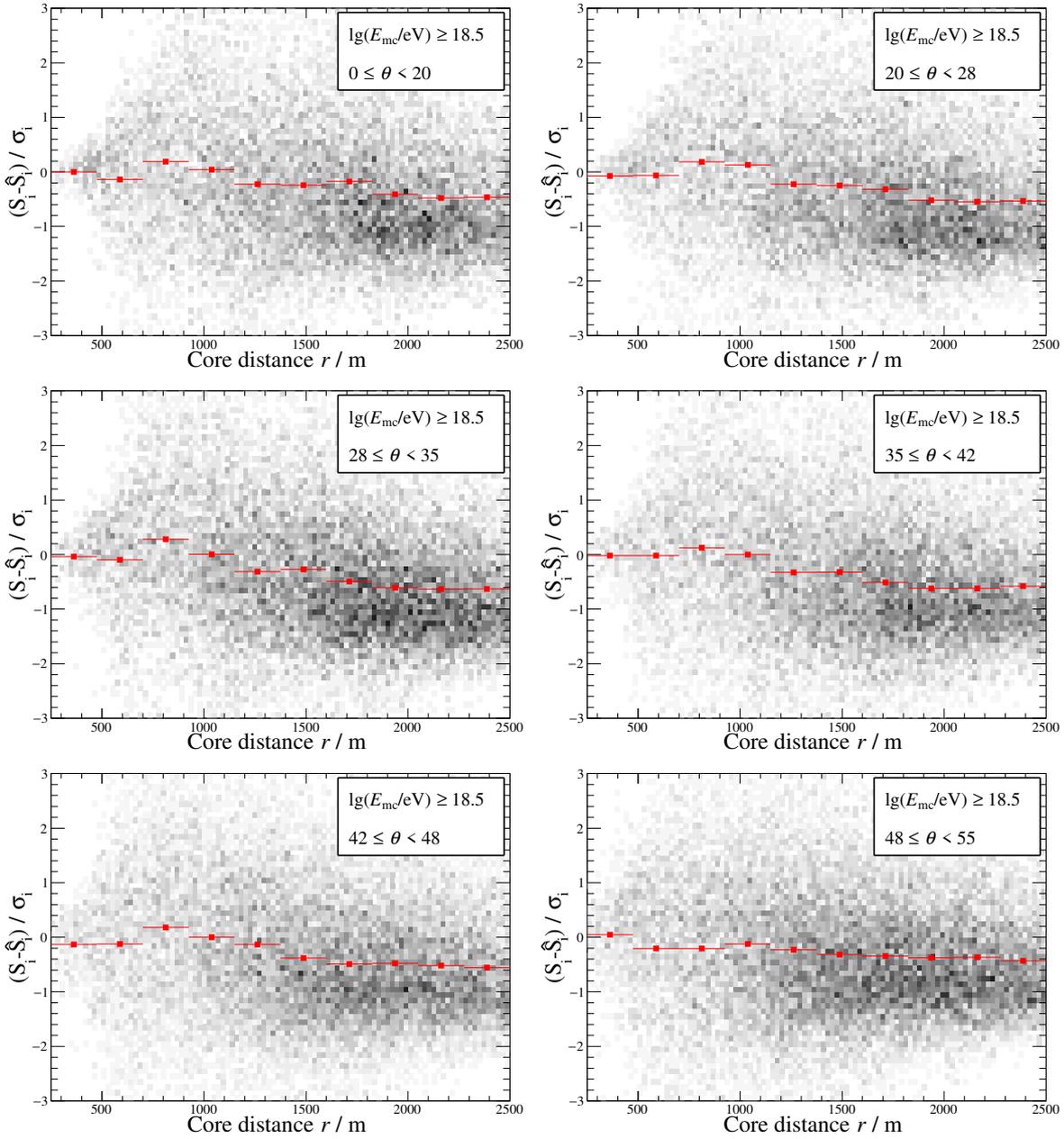
Figure A.1: Residuals for simulated proton data and the 50/50 proton/iron LDF parameterisation as a function of distance to the shower axis. The ensemble of data is described in Table 7.1. There are 12000 pure proton air showers, with energies distributed uniformly, and zenith angles distributed uniformly in $\sin^2(\theta)$. σ is defined from the SSD signal uncertainty model described in 6.
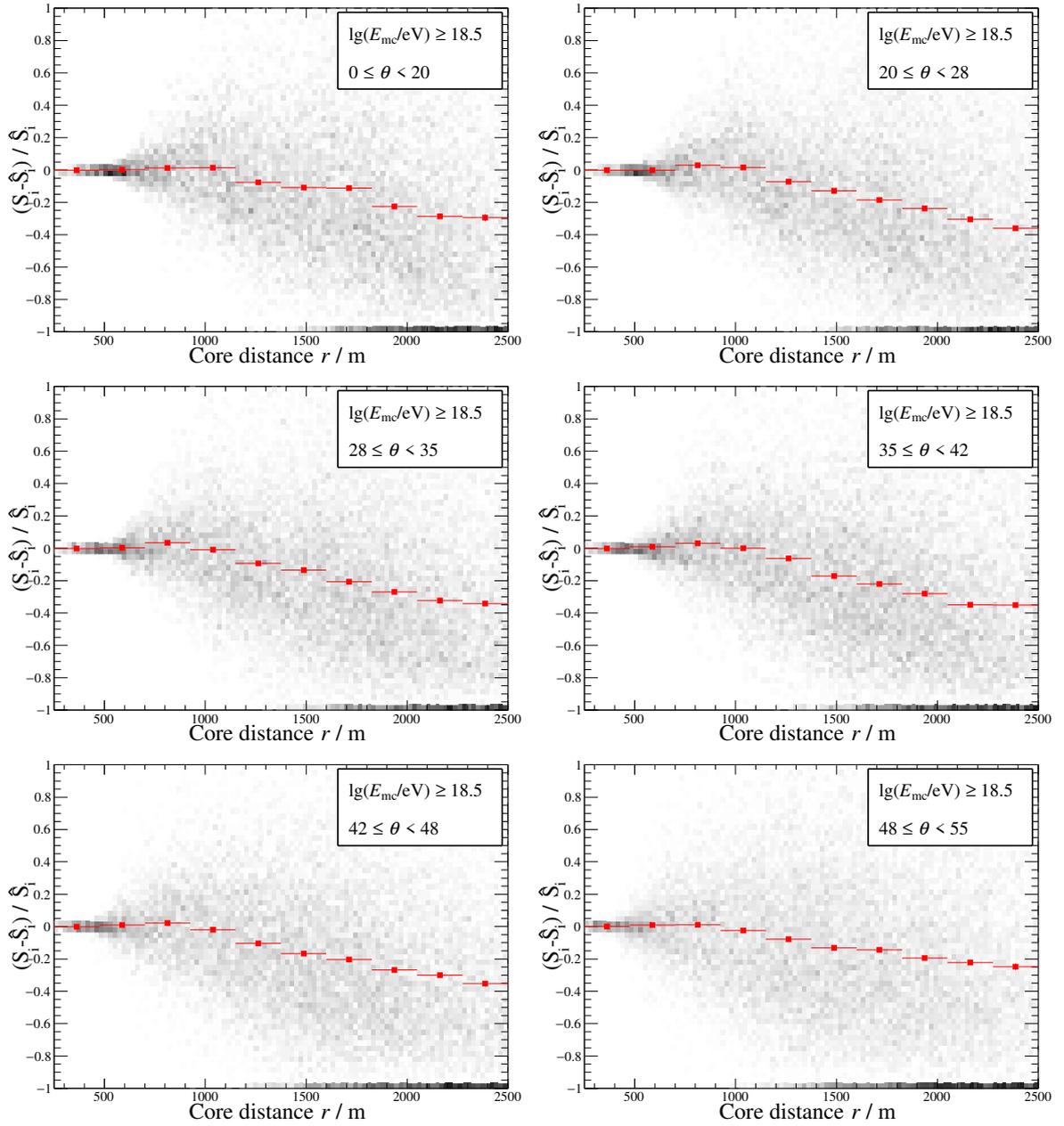
Figure A.2: Relative differences between measured and expected signal for simulated proton data and the 50/50 proton/iron LDF parameterisation as a function of distance to the shower axis. The ensemble of data is described in Table 7.1. There are 12000 pure proton air showers, with energies distributed uniformly, and zenith angles distributed uniformly in $\sin^2(\theta)$.
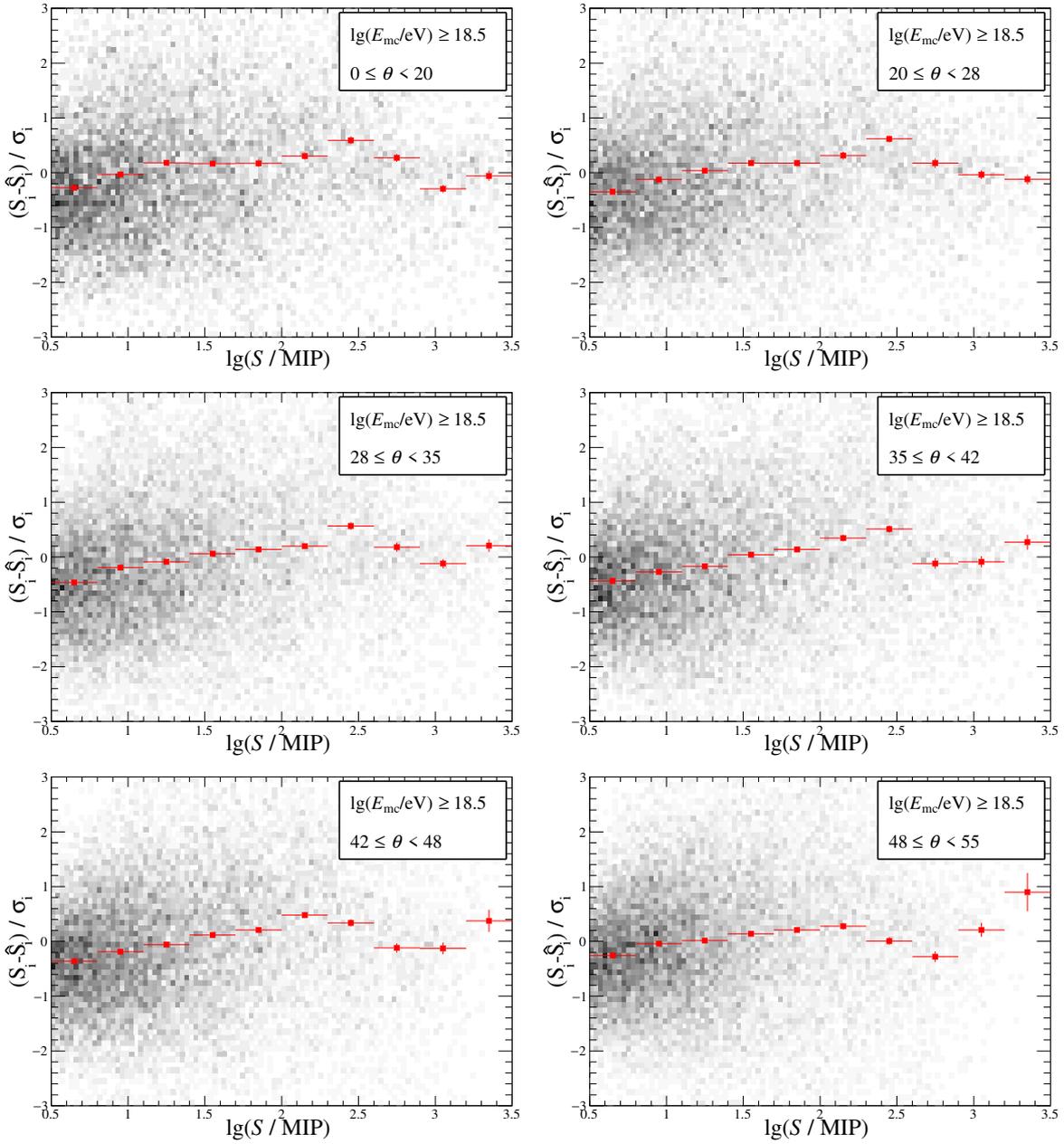
Figure A.3: Residuals for simulated proton data and the 50/50 proton/iron LDF parameteri-
sation as a function of SSD signal. The ensemble of data is described in Table 7.1. There are
12000 pure proton air showers, with energies distributed uniformly, and zenith angles dis-
tributed uniformly in $\sin^2(\theta)$. σ is defined from the SSD signal uncertainty model described
in 6.
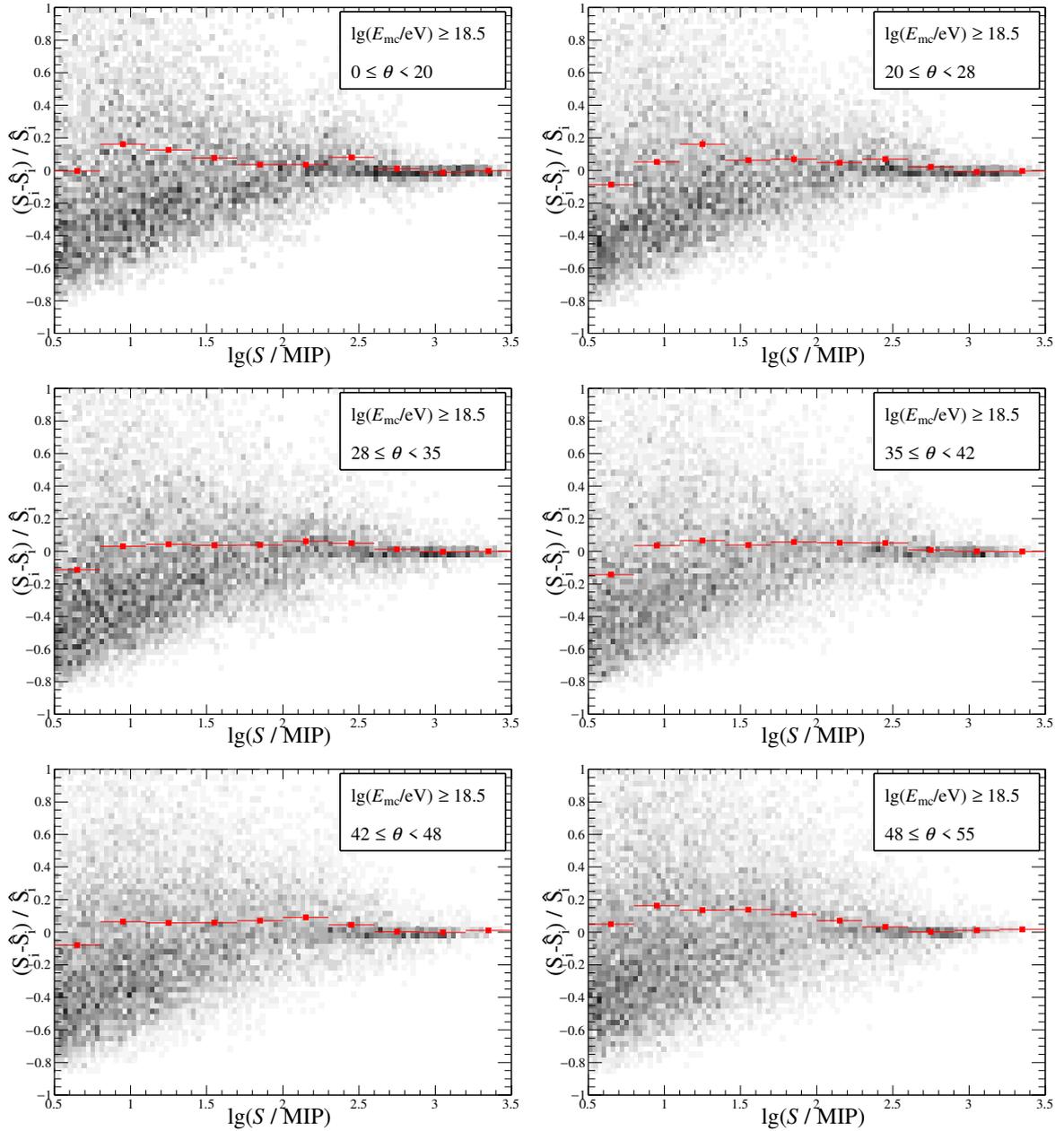
Figure A.4: Relative differences between measured and expected signal for simulated proton data and the 50/50 proton/iron LDF parameterisation as a function of SSD signal. The ensemble of data is described in Table 7.1. There are 12000 pure proton air showers, with energies distributed uniformly, and zenith angles distributed uniformly in $\sin^2(\theta)$.
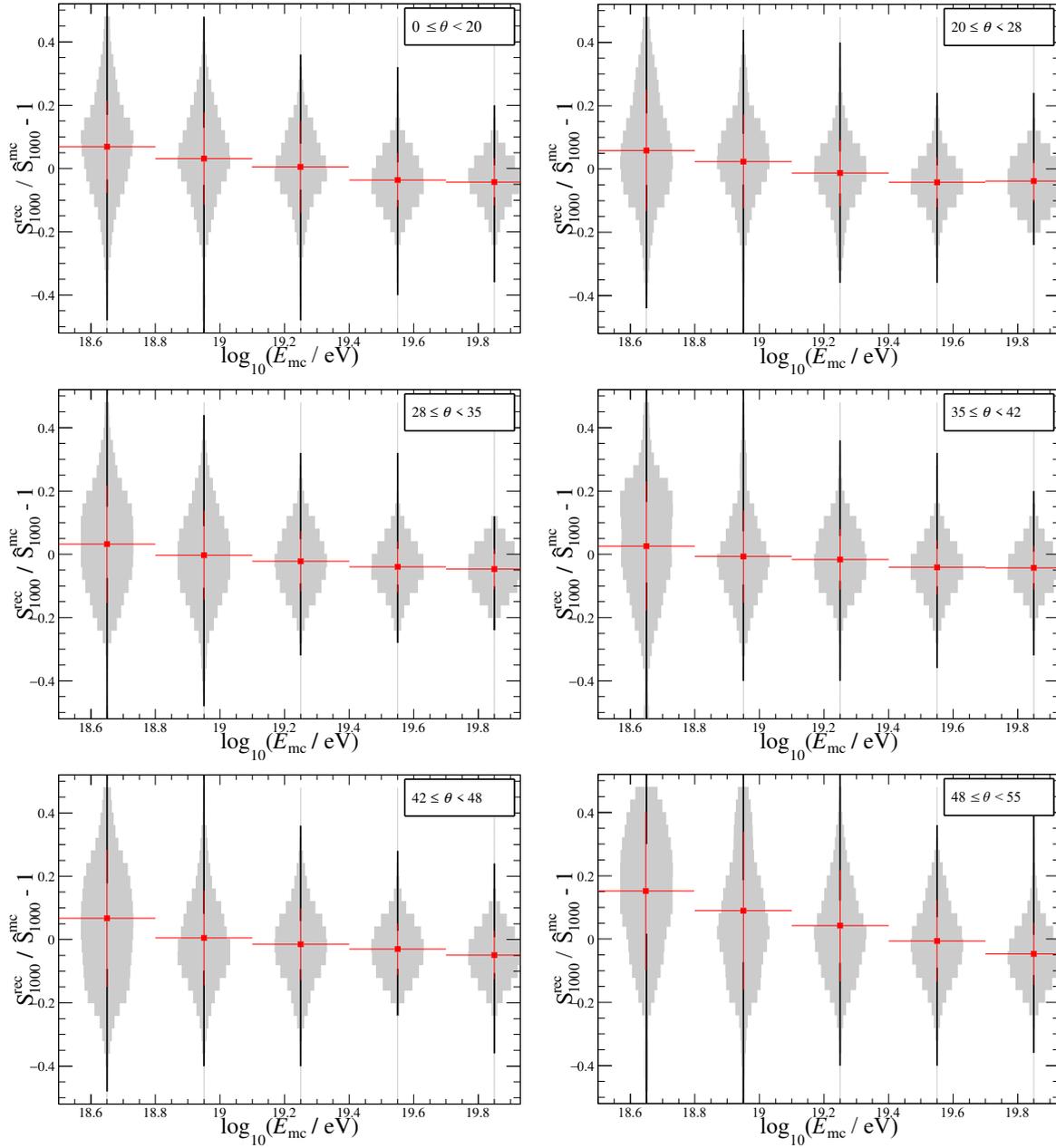
Figure A.5: Bias in the reconstructed energy estimator S(1000) for proton showers, in different bins of shower primary particle energy. The ensemble of data is described in Table 7.1. There are 12000 pure proton air showers, with energies distributed uniformly, and zenith angles distributed uniformly in $\sin^2(\theta)$.
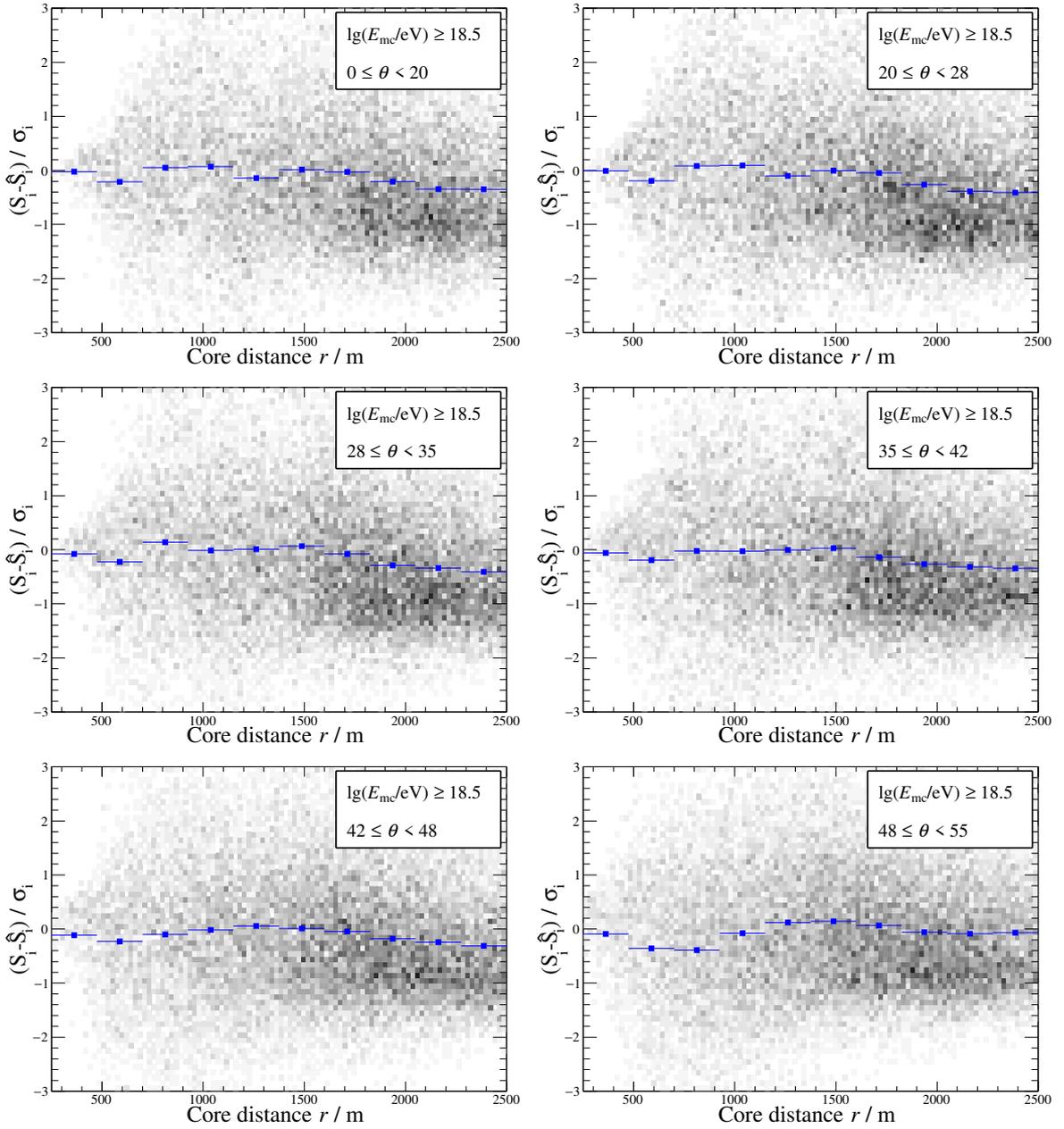
Figure A.6: Residuals for simulated iron data and the 50/50 proton/iron LDF parameterisation as a function of distance to the shower axis. The ensemble of data is described in Table 7.1. There are 12000 pure iron air showers, with energies distributed uniformly, and zenith angles distributed uniformly in $\sin^2(\theta)$. σ is defined from the SSD signal uncertainty model described in 6.
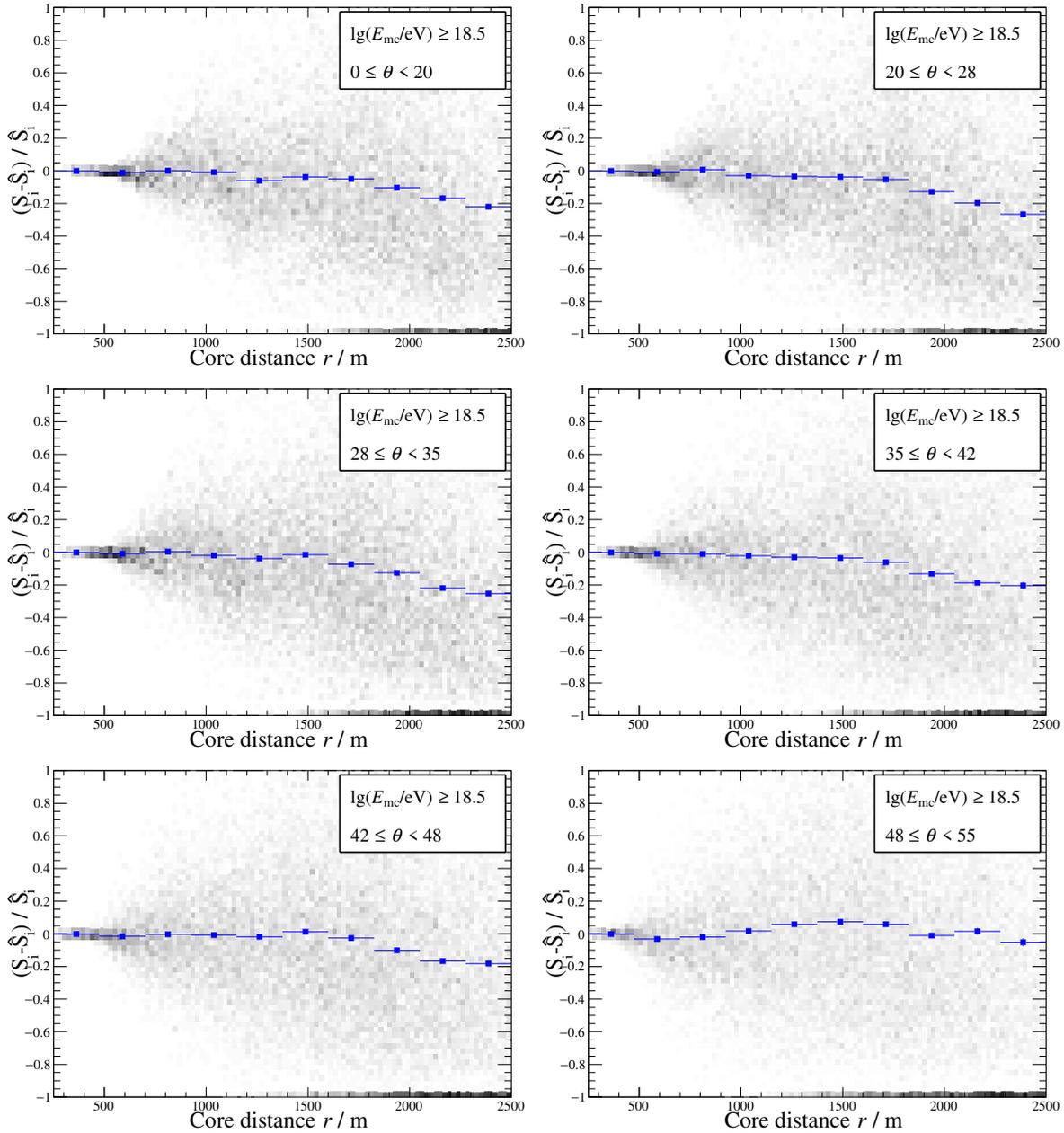
Figure A.7: Relative differences between measured and expected signal for simulated iron data and the 50/50 proton/iron LDF parameterisation as a function of distance to the shower axis. The ensemble of data is described in Table 7.1. There are 12000 pure iron air showers, with energies distributed uniformly, and zenith angles distributed uniformly in $\sin^2(\theta)$.
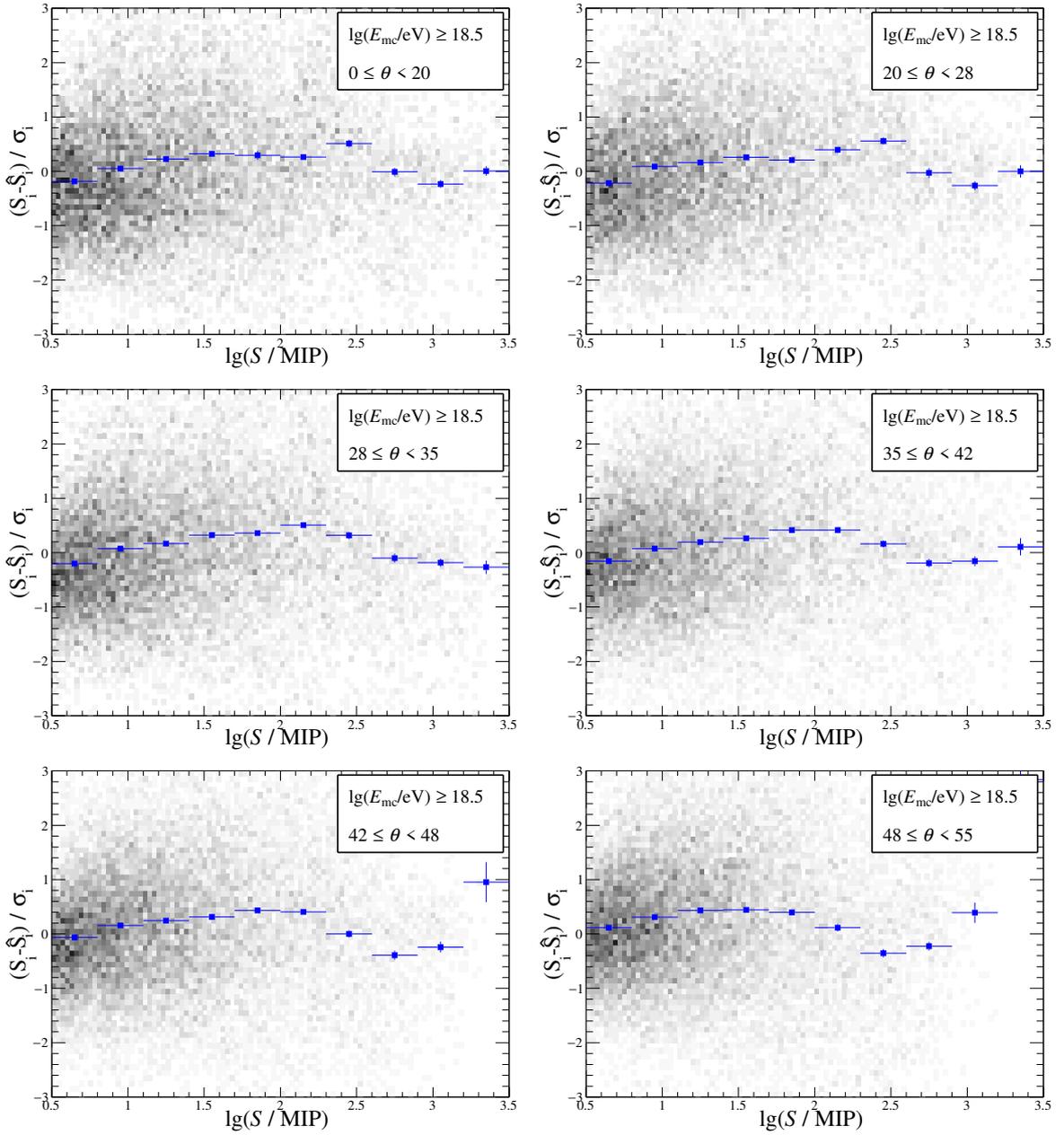
Figure A.8: Residuals for simulated iron data and the 50/50 proton/iron LDF parameterisation as a function of SSD signal. The ensemble of data is described in Table 7.1. There are 12000 pure iron air showers, with energies distributed uniformly, and zenith angles distributed uniformly in $\sin^2(\theta)$. σ is defined from the SSD signal uncertainty model described in 6.
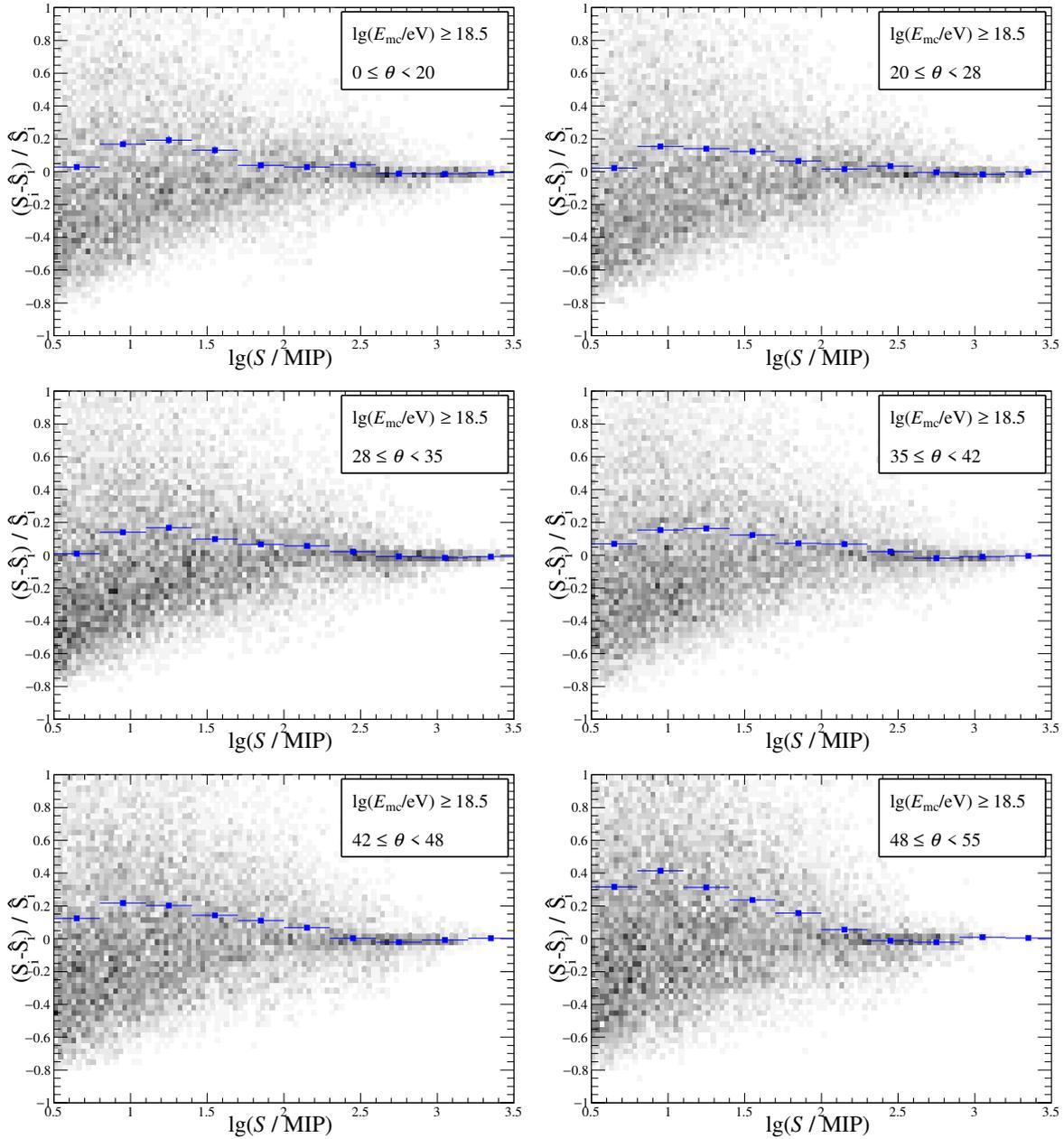
Figure A.9: Relative differences between measured and expected signal for simulated iron
data and the 50/50 proton/iron LDF parameterisation as a function of SSD signal. The en-
semble of data is described in Table 7.1. There are 12000 pure iron air showers, with energies
distributed uniformly, and zenith angles distributed uniformly in $\sin^2(\theta)$.
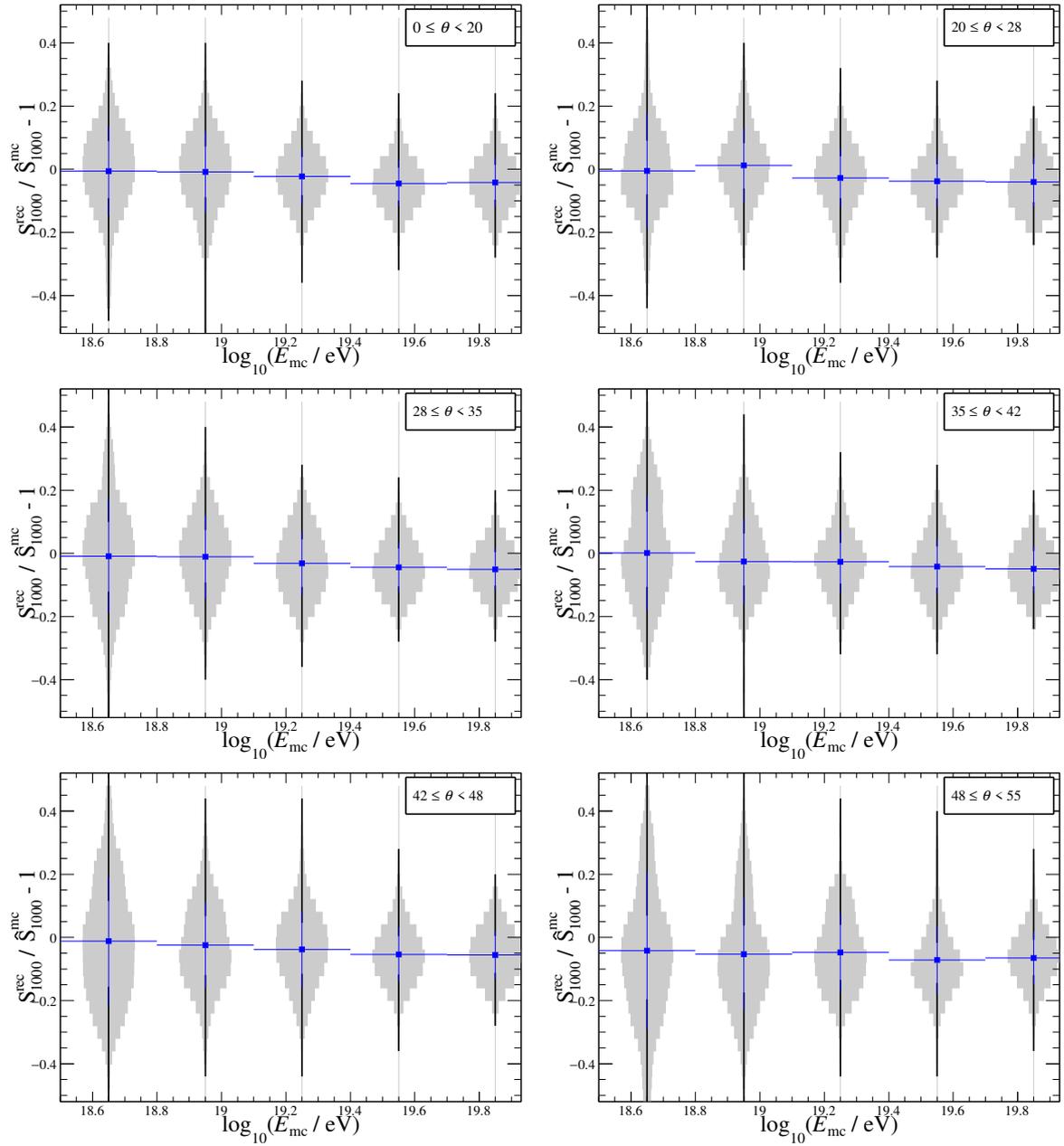
Figure A.10: Bias in the reconstructed energy estimator S(1000) for iron showers, in different bins of shower primary particle energy. The ensemble of data is described in Table 7.1. There are 12000 pure iron air showers, with energies distributed uniformly, and zenith angles distributed uniformly in $\sin^2(\theta)$.

Bibliography

[1] A. de Angelis. Atmospheric Ionization and Cosmic Rays: Studies and Measurements before 1912. Astroparticle Physics, 53:19 – 26, 2014. Centenary of cosmic ray discovery. (Cited on page 1.)

[2] L. Bonolis. From Cosmic Ray Physics to Cosmic Ray Astronomy: Bruno Rossi and the Opening of New Windows on the Universe. Astroparticle Physics, 53:67 – 85, 2014. Centenary of cosmic ray discovery. (Cited on page 2.)

[3] J. Blümer et al. Cosmic Rays from the Knee to the Highest Energies. Progress in Particle and Nuclear Physics, 63(2):293 – 338, 2009. (Cited on pages 3, 4, and 9.)

[4] A. Aab et al. Measurement of the cosmic-ray energy spectrum above 2.5×10^{18} eV using the pierre auger observatory. Phys. Rev. D, 102:062005, Sep 2020. (Cited on pages 3, 42, and 43.)

[5] C. Patrignani et al. Review of Particle Physics. Chinese Physics C, 40(10):100001, October 2016. (Cited on page 5.)

[6] Y. Tsunesada. Joint analysis of the energy spectrum of ultra-high-energy cosmic rays measured at the Pierre Auger Observatory and the Telescope Array. Proceedings of Science, 395(337), 2021. (Cited on page 4.)

[7] K. Greisen. End to the cosmic-ray spectrum? Phys. Rev. Lett., 16:748–750, Apr 1966. (Cited on page 5.)

[8] G. T. Zatsepin and V. A. Kuz'min. Upper Limit of the Spectrum of Cosmic Rays. Soviet Journal of Experimental and Theoretical Physics Letters, 4:78, August 1966. (Cited on page 5.)

[9] J. W. Cronin. The Highest-Energy Cosmic Rays. Nuclear Physics B - Proceedings Supplements, 138:465 – 491, 2005. Proceedings of the Eighth International Workshop on Topics in Astroparticle and Undeground Physics. (Cited on page 6.)

[10] J. Abraham et al. Upper Limit on the Cosmic-Ray Photon Fraction at EeV Energies from the Pierre Auger Observatory. Astroparticle Physics, 31(6):399 – 406, 2009. (Cited on page 7.)

[11] R. J. Protheroe. Acceleration and interaction of ultrahigh-energy cosmic rays. pp. 259–297, 12 1998. (Cited on page 8.)

[12] A. M. Hillas. The Origin of Ultra-High-Energy Cosmic Rays. Annual Review of Astronomy and Astrophysics, 22(1):425–444, 1984. (Cited on pages 9 and 10.)

[13] A. Aab et al. Observation of a Large-Scale Anisotropy in the Arrival Directions of Cosmic Rays above 8×10^{18} eV. Science, 357(6357):1266–1270, 2017. (Cited on pages 11 and 12.)

[14] A. Aab et al. An Indication of Anisotropy in Arrival Directions of Ultra-high-energy Cosmic Rays through Comparison to the Flux Pattern of Extragalactic Gamma-Ray Sources. The Astrophysical Journal, 853(2):L29, feb 2018. (Cited on page 11.)

[15] L. Caccianiga. Anisotropies of the Highest Energy Cosmic-ray Events Recorded by the Pierre Auger Observatory in 15 years of Operation. PoS, ICRC2019:206, 2019. (Cited on page 11.)

[16] A. Yushkov. Mass Composition of Cosmic Rays with Energies above $10^{17.2}$ eV from the Hybrid Data of the Pierre Auger Observatory. PoS, ICRC2019:482, 2019. (Cited on page 13.)

[17] K. Greisen. The Extensive Air Showers, volume 3:1. 1956. (Cited on page 15.)

[18] J. Matthews. A Heitler Model of Extensive Air Showers. Astroparticle Physics, 22:387–397, 01 2005. (Cited on pages 16, 17, 18, and 19.)

[19] Pierre Auger Collaboration. The Pierre Auger Cosmic Ray Observatory. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 798:172 – 213, 2015. (Cited on page 20.)

[20] L. Anchordoqui et al. Ultrahigh energy cosmic rays: The state of the art before the auger observatory. International Journal of Modern Physics A, 18(13):2229–2366, 2003. (Cited on page 21.)

[21] G. Bellini. Neutrino Physics and Astrophysics. Proceedings of the International School of Physics "Enrico Fermi" Neutrino physics and astrophysics. IOS Press, Amsterdam, 2012. (Cited on page 21.)

[22] J. Abraham et al. The Fluorescence Detector of the Pierre Auger Observatory. Nuclear Inst. and Methods in Physics Research, A, 620(2):227–251, 2010. (Cited on pages 22 and 23.)

[23] L. Bonolis. From cosmic ray physics to cosmic ray astronomy: Bruno rossi and the opening of new windows on the universe. Astroparticle Physics, 53:67–85, 2012. (Cited on page 23.)

[24] J. Linsley et al. Energy Spectrum and Structure of Large Air Showers. Journal of the Physical Society of Japan Supplement, 17:91, January 1962. (Cited on pages 23 and 24.)

[25] M. Nagano and A. A. Watson. Observations and implications of the ultrahigh-energy cosmic rays. Rev. Mod. Phys., 72:689–732, Jul 2000. (Cited on page 23.)

[26] J. Linsley. Evidence for a primary cosmic-ray particle with energy 10^{20} ev. Phys. Rev. Lett., 10:146–148, Feb 1963. (Cited on page 24.)

[27] M. Ave et al. The energy spectrum of cosmic rays in the range 3×10^{17}–4×10^{18} ev as measured with the haverah park array. Astroparticle Physics, 19(1):47–60, Apr 2003. (Cited on page 24.)

[28] P. R. Blake et al. Current studies of extensive air showers by the nottingham group at haverah park. Journal of the Physical Society of Japan, 33(5):1197–1206, 1972. (Cited on page 25.)

[29] M. Nagano and A. A. Watson. Observations and implications of the ultrahigh-energy cosmic rays. Reviews of Modern Physics, 72(3):689–732, July 2000. (Cited on page 25.)

[30] A. Ivanov. Some results from the Yakutsk array experiment and the status of the modernization program. In European Physical Journal Web of Conferences, volume 208 of European Physical Journal Web of Conferences, pp. 08005, May 2019. (Cited on page 26.)

[31] A. A. Ivanov et al. Measuring extensive air showers with cherenkov light detectors of the yakutsk array: the energy spectrum of cosmic rays. New Journal of Physics, 11(6):065008, jun 2009. (Cited on page 26.)

[32] D. J. Bird et al. Results from the fly's eye experiment. AIP Conference Proceedings, 338(1):839–854, 1995. (Cited on pages 26 and 27.)

[33] S. Yoshida and H. Dai. The extremely high energy cosmic rays. Journal of Physics G: Nuclear and Particle Physics, 24(5):905–938, May 1998. (Cited on page 27.)

[34] R. Baltrusaitis et al. The utah fly's eye detector. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 240(2):410–428, 1985. (Cited on page 27.)

[35] CERN Timelines, The Fly's Eye Mirrors, 1991. (Cited on page 27.)

[36] T. Abu-Zayyad et al. CORSIKA Simulation of the Telescope Array Surface Detector. 2014. (Cited on page 28.)

[37] T. Abu-Zayyad et al. The Surface Detector Array of the Telescope Array Experiment. 689:87–97, October 2012. (Cited on pages 28, 29, 30, 54, and 56.)

[38] I. Allekotte et al. The Surface Detector System of the Pierre Auger Observatory. Nuclear Inst. and Methods in Physics Research, A, 586(3):409–420, March 2008. (Cited on page 31.)

[39] I. Allekotte et al. The Surface Detector System of the Pierre Auger Observatory. Nucl. Instrum. Meth., A586:409–420, 2008. (Cited on page 32.)

[40] D. Barnhill et al. Testing of Photomultiplier Tubes for use in the Surface Detector of the Pierre Auger Observatory. Nuclear Inst. and Methods in Physics Research, A, 591(3):453–466, 2008. (Cited on page 32.)

[41] J. Abraham et al. Trigger and Aperture of the Surface Detector Array of the Pierre Auger Observatory. Nucl. Instrum. Meth. A, 613:29–39, 2010. (Cited on pages 32 and 35.)

[42] S. Messina. Extension to lower energies of the cosmic-ray energy window at the Pierre Auger Observatory. PhD thesis, University of Groningen, 2016. (Cited on page 33.)

[43] K.-H. Kampert and A. A. Watson. Extensive air showers and ultra high-energy cosmic rays: a historical review. The European Physical Journal H, 37(3):359–412, Jul 2012. (Cited on page 34.)

[44] X. Bertou et al. Calibration of the surface array of the pierre auger observatory. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 568(2):839–846, Dec 2006. (Cited on pages 36 and 88.)

[45] C. Bonifazi et al. A model for the time uncertainty measurements in the auger surface detector array. Astroparticle Physics, 28(6):523–528, Jan 2008. (Cited on page 37.)

[46] A. Aab et al. Reconstruction of events recorded with the surface detector of the pierre auger observatory. Journal of Instrumentation, 15:P10021–P10021, 10 2020. (Cited on pages 38 and 39.)

[47] The Pierre Auger Collaboration. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 798:172–213, Oct 2015. (Cited on pages 39, 43, and 44.)

[48] K. Kamata and J. Nishimura. The Lateral and the Angular Structure Functions of Electron Showers. Progress of Theoretical Physics Supplement, 6:93–155, 02 1958. (Cited on page 40.)

[49] G. Cataldi. Towards AugerPrime: The Upgrade of the Pierre Auger Observatory. Nuclear and Particle Physics Proceedings, 291-293:96–101, 2017. (Cited on page 44.)

[50] A. Dyshkant et al. About NICADD extruded scintillating strips. 4 2005. (Cited on pages 46 and 47.)

[51] Pierre Auger Collaboration. The Pierre Auger Observatory Upgrade - Preliminary Design Report. ArXiv e-prints, April 2016. (Cited on pages 46, 47, and 55.)

[52] Wavelength Shifting Fibers. http://kuraraypsdf.jp/psf/ws.html. (Cited on pages 47 and 48.)

[53] D. S. Ayres et al. The NOvA Technical Design Report. 10 2007. (Cited on page 48.)

[54] A. Castellina. The dynamic range of the AugerPrime Surface Detector: technical solution and physics reach. PoS, ICRC2017:397, 2017. (Cited on page 48.)

[55] P. Abreu et al. Update of the offline framework for augerprime. pp. 250, 07 2021. (Cited on page 49.)

[56] J. G. Gonzalez. The offline software of the pierre auger observatory: Lessons learned, 2012. (Cited on page 49.)

[57] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. Nucl. Instrum. Meth. A, 389:81–86, 1997. (Cited on page 49.)

[58] D. Heck et al. CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers. February 1998. (Cited on pages 49 and 59.)

[59] T. Pierog et al. First results of fast one-dimensional hybrid simulation of EAS using conex. Nuclear Physics B - Proceedings Supplements, 151(1):159–162, jan 2006. (Cited on page 49.)

[60] S J Sciutto. Aires a system for air shower simulations. user's guide and reference manual. 2019. (Cited on page 49.)

[61] T. Huege et al. Simulating radio emission from air showers with CoREAS. In AIP Conference Proceedings. AIP, 2013. (Cited on page 49.)

[62] J. Alvarez-Muñiz et al. Monte carlo simulations of radio pulses in atmospheric showers using ZHAireS. Astroparticle Physics, 35(6):325–341, jan 2012. (Cited on page 49.)

[63] V. Ivanchenko. Geant4 toolkit for simulation of hep experiments. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors

and Associated Equipment, 502(2):666–668, 2003. Proceedings of the VIII International Workshop on Advanced Computing and Analysis Techniques in Physics Research. (Cited on page 49.)

[64] R. Abbasi et al. Joint analysis of the energy spectrum of ultra-high-energy cosmic rays as measured at the Pierre Auger Observatory and the Telescope Array. PoS, ICRC2021:337, 2021. (Cited on pages 51 and 52.)

[65] D. Ivanov. Energy spectrum measured by the telescope array surface detector. PhD thesis, Rutgers U., Piscataway, 2012. (Cited on page 53.)

[66] S. Saffi. Analysis of Simulation and Reconstruction Methods used by the Pierre Auger Observatory. PhD thesis, University of Adelaide, 2020. (Cited on pages 60 and 61.)

[67] P. Billoir. A sampling procedure to regenerate particles in a ground detector from a "thinned" air shower simulation output. Astroparticle Physics, 30:270–285, 12 2008. (Cited on page 61.)

[68] D. Ivanov. Energy Spectrum Measured by the Telescope Array. PoS, ICRC2019:298, 2019. (Cited on page 78.)

[69] P. Abreu et al. Marta: A high-energy cosmic-ray detector concept with high-accuracy muon measurement. The European Physical Journal C, 78, 12 2017. (Cited on pages 85 and 95.)

[70] X. Bertou and Pierre Auger Collaboration. Calibration and Monitoring of the Pierre Auger Surface Detectors. Proc. 28th International Cosmic Ray Conference, 2:813, July 2003. (Cited on page 88.)

[71] A. Aab et al. Studies on the response of a water-cherenkov detector of the pierre auger observatory to atmospheric muons using an rpc hodoscope. Journal of Instrumentation, 15(09):P09002–P09002, September 2020. (Cited on page 88.)

[72] D. Schmidt. Sensitivity of AugerPrime to the masses of ultra-high-energy cosmic rays. PhD thesis, Karlsruhe Institute of Technology, 2018. (Cited on pages 97 and 166.)

[73] M. Ave et al. The accuracy of signal measurement with the water cherenkov detectors of the pierre auger observatory. Nuclear Instruments and Methods in

Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 578(1):180 – 184, 2007. (Cited on pages 99 and 100.)

[74] T. Schulz. Enhanced UHECR Event Reconstruction by means of Sampling Lateral Distributions with Multiple Surface Sub-Detectors. Auger internal note GAP-2018-025, 2018. (Cited on page 101.)

[75] A. Taboada. Analysis of the First Data of the AugerPrime Detector Upgrade. PhD thesis, Karlsruhe Institute of Technology, 2019. (Cited on pages 101, 114, 120, 122, 127, 131, and 144.)

[76] M. R. Roman Hiller. An update on the signal accuracy using the infill array. Auger internal note GAP-2012-012, 2012. (Cited on pages 102, 103, and 108.)

[77] A. Streich. Performance of the ssd pre-production array. Pierre Auger Collaboration Meeting, November 2019. (Cited on page 103.)

[78] D. Newton et al. The optimum distance at which to determine the size of a giant air shower. Astroparticle Physics, 26(6):414–419, jan 2007. (Cited on page 121.)

[79] A. Letessier-Selvon et al. Layered water cherenkov detector for the study of ultra high energy cosmic rays. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 767:41–49, 2014. (Cited on page 162.)