THE UNIVERSITY
of ADELAIDE

Extracting a Less Model Dependent Cosmic Ray Composition from Xmax Distributions

Simon Blaess

School of Physical Sciences

University of Adelaide

A thesis submitted for the degree of
Doctor of Philosophy.

January 2018

Contents

Abstract

Knowledge of the acceleration mechanisms, possible sources and arrival direction distribution of ultra high energy cosmic rays can be furthered by increasing our understanding of the cosmic ray mass composition. The Pierre Auger Observatory measures the depth of shower maximum (X_{max}) of ultra high energy cosmic rays, an observable which is sensitive to the mass of the cosmic ray, and often analysed in mass composition studies. At ultra high energies, the results of mass composition studies depend strongly on the particle interaction predictions of the assumed hadronic interaction model. A novel approach to reduce the model dependency in the interpretation of the mass composition will be outlined in this thesis. The X_{max} predictions of the models were parameterised in terms of the primary mass. The mass composition fractions and some of the parameterisation coefficients are then fitted to the data, resulting in a mass composition estimate which is of reduced model dependency. The results of this approach applied to Pierre Auger Observatory data are presented.

Declaration of Originality

I, Simon Garry Blaess, 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 consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the Universitys 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.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed: **Date:** 2/1/2018

Acknowledgements

I would like to thank my supervisors, Jose Bellido and Bruce Dawson, for the knowledge, instruction and enthusiasm they provided throughout my Ph.D. Their guidance was extremely helpful during my research. Additionally, I would like to thank Gary Hill and Roger Clay for the assistance and input they provided in weekly discussions.

I would also like to thank the other members of the Adelaide High Energy Astrophysics Group, past and present. It was a helpful and welcoming environment that was a pleasure to be involved in. In particular, I single out James, Patrick and Phong, who have shared the same adventures as I since we enrolled in Space Science and Astrophysics over seven years ago. Their advice, not just restricted to study, has been invaluable, and I suppose their company has also been a lot of fun.

The Pierre Auger Collaboration was great to be a part of, and I am grateful for the opportunity to work with such an exceptional experiment. I thank the collaboration members for their input and the resources provided during my research.

Lastly, thank you to my family and friends for their support. In particular, I thank my parents, Garry and Anna, as without their support I wouldn't have made it this far.

Really activates the brain almonds,
really massages the brain aneurysms,
I mean, I mean, I mean,
there's just no accountability.

A VISIONARY

There's giant weather modification systems and
there's weather wars between different governments...
Bill Gates owns the weather machines!

ALEX JONES

Introduction

First discovered over a century ago, cosmic rays have been intensively studied, but major questions remain unanswered. Our knowledge of the atomic mass composition, origin and production mechanisms of cosmic rays above an energy of 10^{17} eV remains uncertain. Hindering the resolution of these questions is the incredibly low flux of these charged particles at Earth. At energies of 10^{20} eV, less than 1 particle per km^2 per century reaches Earth. To measure enough of these particles, detector arrays observing vast areas were constructed. These arrays measure cosmic rays indirectly by detecting extensive air showers, the huge cascade of secondary particles produced when a very high energy cosmic ray interacts with a nucleus in the atmosphere.

The biggest cosmic ray experiment thus far is the Pierre Auger Observatory which, located in Argentina and covering a massive surface area of 3000 km^2, measures the properties of cosmic rays above 10^{17} eV with exceptional precision. The Pierre Auger Observatory incorporates 1600 ground based particle detectors, which sample the secondary particles of the extensive air showers, with 27 fluorescence telescopes, which overlook the ground array from four sites on the boundary. The fluorescence detectors measure the longitudinal development of the extensive air showers by detecting fluorescence light emitted from atmospheric nitrogen molecules excited by interactions with the extensive air shower cascades. This hybrid design of the array allows properties of the cosmic ray air shower to be studied in great detail.

Resolving the mass composition of these cosmic rays is pivotal for furthering our understanding of their acceleration mechanisms, possible sources and the arrival direction distribution. Measured with unprecedented sensitivity by the Pierre Auger Observatory is an observable commonly utilised in mass composition investigations, X_{max}, the atmospheric depth in g/cm^2 from the top of the atmosphere where the longitudinal development of an air shower reaches the maximum number of particles. Typical mass composition estimates using X_{max} depend considerably on the predictions of hadronic interaction models, predictions which are uncertain at these very high cosmic ray energies and therefore vary between models. This introduces significant uncertainty in the estimated mass composition, and the mass composition interpretation from different models is often inconsistent. Presented in this thesis is a novel approach which reduces the reliance on hadronic models in determining the mass composition of very high energy cosmic rays. By fitting some of the parameters arising from the model predictions, the mass composition estimated from this new method depends less on the assumed hadronic interaction model. The performance and limitations of the method are presented, and it will be shown that with the current models, applying this approach to X_{max} data measured by the Pierre Auger Observatory results in a mass composition that is consistent for the three models used.

This thesis is organised as follows:

Chapter 1 A brief historical overview of the study of cosmic rays, a detailed description of extensive air showers, and an explanation of the current techniques used to detect them.

Chapter 2 The current knowledge of cosmic rays is reviewed, and observational results are detailed.

Chapter 3 A summary of previous and current cosmic ray detectors, with a detailed description of the Pierre Auger Observatory.

Chapter 4 The motivation for analysing X_{max} data to infer the mass composition of cosmic rays, a description of the parameterisation of the expected X_{max} distributions, and a detailed description of the new approach applied in this work to analyse X_{max} data.

Chapter 5 The performance of this new X_{max} analysis method for estimating the mass composition of simulated X_{max} data.

Chapter 6 Scan of the mass composition solutions for the X_{max} data measured by the Pierre Auger Observatory.

Chapter 7 Detailing the methods applied to determine the statistical errors, systematic errors and p-values of the fits.

Chapter 8 A presentation and discussion of the mass composition and hadronic property results from the analysis of the Pierre Auger Observatory X_{max} data.

Chapter 9 Concluding remarks.

Chapter 1

Cosmic Rays and Extensive Air Showers

1.1 Historical Overview

Cosmic rays are extraterrestrial, high-energy charged particles, and those that reach Earth cause ionisation in our atmosphere. Composed of mainly protons and atomic nuclei, this extraterrestrial radiation has been a chosen matter of investigation for over a century. The pioneering balloon flights of Victor Hess in 1912 [1] initiated the specific field of cosmic ray (CR) science, however the existence of cosmic rays can be found in experiments carried out more than a century earlier.

Coulomb may have been the first to scientifically observe the presence of cosmic rays. In Paris during 1785, he was carrying out experiments on electro-statically charged bodies when he noticed the charged bodies would gradually leak charge, even if they were insulated. Coulomb was unable to explain the reason [2].

In the late 1800s, J. Elster, H.F. Geitel and C.T.R Wilson were investigating 'dark current' [3, 4], a different name for this observation. Despite isolating electroscopes from all known forms of radiation by placing them in airtight containers, discharge was still observed. Rutherford attempted to shield a detector with lead yet discharges were still observed, albeit at a reduced rate [5]. It was realised that ions must be forming in the air around the electroscopes, resulting in the air acting as a conductor. However, the reason for the ionisation was unknown.

Wilson posed that the radiation could originate from outside the Earth's atmosphere [2,4], but the experiments he conducted in railway tunnels compared to a normal room did not produce an observable difference in the rate of ionisation. Geitel and Wilson reasoned that the ionisation was an inherent property of the atmosphere [2, 5], and thus the ionisation was termed 'spontaneous' [2, 4]. Many factors confused those researching the origin of this radiation. Not only were the measurements from early experiments inaccurate, but the radiation being measured was not solely of extraterrestrial origin. Terrestrial radiation was convoluting the observations, and atmospheric effects and environmental factors also added to the puzzle. Some experiments pointed to an extraterrestrial source of the radiation, as a diurnal variation of the ionisation was measured, but these results were unable to be reproduced.

In 1912, Victor Hess carried out the crucial experiment. He conducted high-altitude balloon flights [1], carrying three ionisation detectors designed by T. Wulf. Hess recorded the amount of ionisation measured during each flight, and noticed there was less ionisation at the beginning of each flight. This was due to less terrestrial radiation reaching his detectors. Upon reaching higher altitudes though, there was a significant increase in the measured ionisation, which Hess concluded was due to radiation of an extraterrestrial origin [1, 2, 5, 6]. Hess was the first to show conclusively that 'spontaneous' ionisation was of a cosmic nature, earning him the 1936 Nobel Prize [2, 4, 7].

The experiments conducted by Hess had reached altitudes beyond 5 km. In 1913 and 1914,

Werner Kolhörster's flights reached altitudes of over 9 km, measuring that the amount of ionisation increased further at higher altitudes [2, 5]. Robert Millikan however was skeptical of the cosmic explanation, so in 1925 Millikan and Harvey Cameron attempted to measure the absorption coefficient of the radiation in an environment of low background ionisation by lowering detectors into lakes. Millikan's measurements were consistent with the attenuation of extraterrestrial radiation with distance from the top of the atmosphere, and in conceding he coined the term 'cosmic rays' to describe the ionisation source [2, 5, 6].

In 1938 Pierre Auger and Roland Maze deployed multiple Geiger counters at different distances with the aim of observing temporal coincidences of particle hits. They observed the coincidences to be apparent out to a distance of 20 m [6], and with that the phenomenon of extensive air showers (EAS) was discovered. Experiments at higher altitudes followed with coincidences being observed at 300 m. The broad hypothesis was that multiplicative cascades of particles were arriving from a common source resulting in these coincident measurements. Auger concluded that the particles initiating these extensive air showers were of energies greater than 10^{15} eV [8], energies which were beyond the maximum known particle energies at the time and 10^8 greater than the energies accelerators of that period could generate.

Since the discovery of cosmic rays, significant strides have been made in understanding these particles. Larger and more sensitive detectors have been constructed over time, some reaching thousands of square kilometres. The world's largest cosmic ray detector, the Pierre Auger Observatory (see Section 3.9), spans 3000 km^2. These detectors have made important observations and answered many questions, but crucial questions such as the origin and atomic mass of the highest energy cosmic rays remain unanswered. The goal of current and future cosmic ray detectors is to have the accuracy and statistical gathering power to discover the answers to the fundamental questions that remain.

1.2 Extensive Air Showers

Extensive air showers (EAS) are initiated by an incoming cosmic ray (the primary particle) interacting with an air molecule in the Earth's atmosphere. A sequence of particle interactions and decays leads to the development of a cascade of secondary particles, travelling along the axis of the primary particle's motion. At the leading edge of the cascade, called the shower front, a thin disc-like structure of particles moves along the shower axis at close to the speed of light, spreading laterally through particle interactions and Coulomb scattering. From the first interaction of the primary particle with an air molecule, the number of particles (shower size) in an EAS multiplies rapidly, reaching a maximum number of particles after which energy losses to the surrounding environment result in the shower size decreasing [9]. The extensive air shower is detected by measuring these interactions and energy losses.

The longitudinal development of an EAS, measured in dimensions of length, depends on the local atmospheric density. Therefore, knowledge of the atmospheric conditions through which the EAS propagates is needed to describe the EAS in terms of length. Atmospheric depth, X, is a more convenient quantity for describing shower development. Atmospheric depth is the integral of the atmospheric density from the altitude of interest, h, to the top of the atmosphere (defined to be at infinity) along the axis of shower propagation:

$$X\ [\mathrm{gcm}^2] = \int_h^\infty \rho(l)dl. \tag{1.1}$$

Using atmospheric depth allows showers observed in different locations or interacting with different atmospheric conditions to be more easily compared. For some particle species such as pions, the relative likelihood of particle decay or interaction depends on the atmospheric density, therefore shower development is not completely decoupled from atmospheric density by using

atmospheric depth [7]. Despite this, atmospheric depth is a handy quantity that is frequently applied.

1.2.1 Heitler's Model of Electromagnetic Cascades

The Heitler model is a basic theory describing the development of electromagnetic cascades [10]. The Heitler model consists of two main processes. These are energy loss by bremsstrahlung:

$$N + e \rightarrow N + e + \gamma \tag{1.2}$$

and pair production:

$$N + \gamma \rightarrow N + e^- + e^+ \tag{1.3}$$

where N represents a nucleus in the medium in which the cascade develops.

The Heitler model assumes that particles interact upon travelling a distance $d = \lambda_r \ln 2$, where d is called the interaction length and λ_r is the particle's radiation length (for electrons in air, $\lambda_r \approx 37\,\mathrm{gcm}^{-2}$ which is in units of atmospheric depth [9]). The bremsstrahlung and pair production processes are both assumed to have equal interaction lengths. After each interaction, the two particles that result (either e and γ for bremsstrahlung or e^+ and e^- for pair production) each have half the energy of the interacting particle.

In the conditions of the Heitler model, consider an electron of energy E_0 travelling through a medium. At a depth of $\lambda_r \ln 2$, a photon is created through bremsstrahlung, with the energy of the initial electron shared amongst the secondary particles, thus $E_e = E_\gamma = \frac{1}{2}E_0$.

Subsequently, at a depth of $x = 2\lambda_r \ln 2$ ($2 \ln 2$ radiation lengths), the electron will undergo bremsstrahlung and the photon will undergo pair production. The number of particles at each interaction point doubles while the energy of the particles after each interaction point is half the energy of the particles before the interaction. Thus, as these energy loss processes continue the cascade proceeds to grow.

After n interactions, which is a radiation length interaction level of $t = n \ln 2$ (t is measured in units of radiation lengths) and occurs at a depth of

$$X = n\lambda_r \ln 2, \tag{1.4}$$

the number of particles in the shower, N_p, is:

$$N_p = 2^n = e^t, \tag{1.5}$$

and the energy of each particle is:

$$E = \frac{E_0}{N_p} = E_0 e^{-t} \tag{1.6}$$

The cascade continues growing until the particle energy drops below a critical energy E_c, the energy at which collisional energy losses dominate over radiative energy losses. Below the E_c threshold, the Heitler model assumes that the particles stop multiplying and they dissipate their energy rapidly through ionisation. The critical energy in air is approximately 80 MeV [9].

Consequently, when $E_c = E_0 e^{-t_{max}}$, the maximum number of energetic particles has been created, and so the atmospheric depth at which the shower reaches its maximum size is:

$$X_{\max} = \lambda_r t_{max} = \lambda_r \ln \frac{E_0}{E_c} \tag{1.7}$$

An illustration of the development of a shower according to the Heitler model is shown in Figure 1.1. The two main characteristics of an electromagnetic cascade are defined by the Heitler model, these being the number of particles at shower maximum is proportional to the primary energy, and the depth of shower maximum is proportional to the logarithm of the primary energy [7, 10]. This simple model does not express the finer details of shower development however.

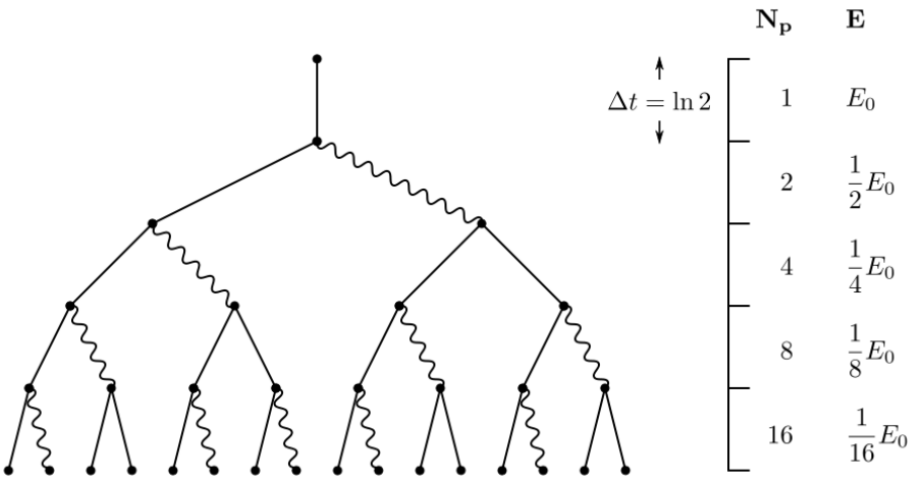


Figure 1.1: A visualisation of Heitler's toy model for cascade development. The legend shows the number of particles, N_p, and the energy E of each particle, at various stages of shower development. Electrons are depicted by straight lines and photons are depicted by wavy lines. From [11].

1.2.2 Hadronic Air Showers

Hadronic air showers are initiated by the interaction of a single nucleon or nucleus with an air molecule. The interaction converts approximately half of the primary particle's energy into mesons. For simplicity, we will assume the primary particle interaction produces only pions.

$$p + N \rightarrow p + \pi^0 + \pi^- + \pi^+ + ... \tag{1.8}$$

where in this case the nucleon is a proton interacting with an air molecule N. Of the pions produced, approximately two-thirds are charged pions and one-third are neutral pions. Pions can either interact to produce subsequent generations of pions, or they can decay. The mean laboratory lifetime of neutral pions is $\sim 8 \times 10^{-17}$ s. Therefore, unless the neutral pions are very high energy, they decay rapidly, with the main decay mode ($\sim 99\%$ chance) being the production of two photons [12]:

$$\pi^0 \rightarrow \gamma + \gamma \tag{1.9}$$

If the produced photons have high enough energy, they undergo pair production, initiating electromagnetic sub-showers.

Charged pions have a mean lifetime of approximately 3×10^{-8} s, thus their likelihood of interaction is high, especially at higher energies. As pion decay and interaction compete with each other, the development of individual hadronic showers fluctuates greatly [7]. This is especially true in an atmosphere of changing density. The interaction of pions produces subsequent generations of pions, with the same ratio of approximately two-thirds charged to one-third neutral. Pion interactions proceed until the pion energy threshold is reached, below which decay almost always occurs as low energy charged pions are less time dilated [7]. With almost 100% probability, the decay of charged pions produces muons and neutrinos [12]:

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \tag{1.10}$$

$$\pi^- \rightarrow \mu^- + \overline{\nu}_\mu \tag{1.11}$$

As a result of the decays of charged pions, the hadronic air shower contains a significant muon component. This is a noticeable difference between air showers of hadronic origin and electromagnetic

showers, in addition to the considerable fluctuation in the development of individual hadronic showers compared to electromagnetic showers. The produced muons deeply penetrate the atmosphere because they lose energy slowly ($\sim 2\,\mathrm{MeV/gcm^{-2}}$) to the ionisation of air molecules. Consequently, most of the muon content reaches ground level, unlike the electromagnetic component which will rapidly attenuate after the shower maximum is reached [9].

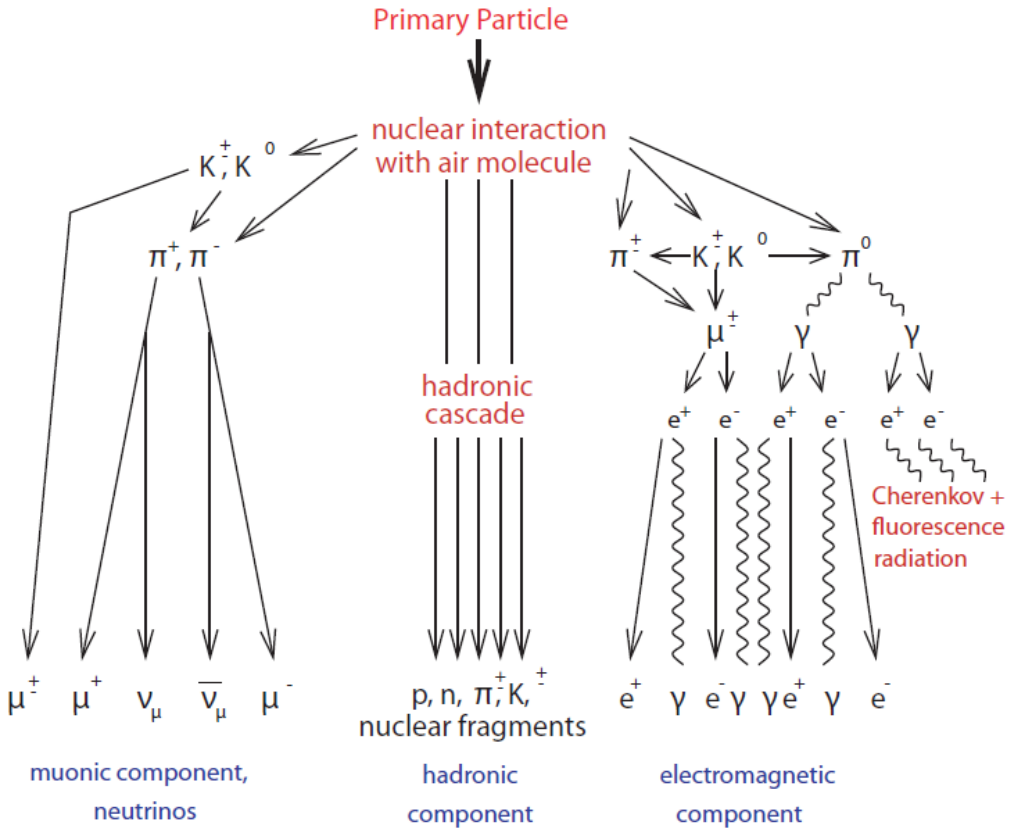


Figure 1.2: A visualisation of the Heitler model applied to an EAS initiated by a hadronic primary. From [13].

A hadronic EAS can be thought of as containing an electromagnetic component, which consists of numerous independent electromagnetic cascades overlapping each other, initiated at different levels of the atmosphere. Also contained in the hadronic EAS is a muon component carrying information on the hadronic interactions that occured within the shower. A pure hadronic component is also present in the shower, which consists of significantly fewer particles than the number of electrons, photons and muons present [7]. The majority of the primary particle's energy is transferred to the electromagnetic component of the EAS [9]. For a primary particle of energy $10^{20}\,\mathrm{eV}$, about 90% of its energy is channelled to the electromagnetic component [14].

To describe the size of an EAS as a function of atmospheric depth, a parameterisation called the Gaisser-Hillas profile is used:

$$N_e(X) = N_e^{max}\left(\frac{X - X_0}{X_{\max} - X_0}\right)^{\frac{X_{\max}-X_0}{\lambda}} e^{-\frac{X_{\max}-X}{\lambda}}, \tag{1.12}$$

where four parameters are fitted to define the average longitudinal development, these being the shape parameters X_0 and λ, the depth of shower maximum $X_{\max}$, and N_e^{max} which is the number of electrons at $X_{\max}$ [15–17]. By measuring an EAS through the atmospheric fluorescence emission it causes, this profile can be fitted to the observed data to determine the shower energy and $X_{\max}$ (see Section 1.3.2).

While the shower front propagates through the atmosphere, the muons define the shower front's leading edge because their heavier mass means they are deflected less by Coulomb scattering. Conversely, electrons are significantly scattered and therefore take longer to reach ground level. Consequently, near the shower core, the shower front is thin with large particle densities because the electrons and photons greatly outnumber the muons. Away from the shower axis, the particle densities rapidly decrease and the shower front is thick, with the relative muon content greater because the muon lateral distribution is flatter compared to the electromagnetic one. For particle detectors located at core distances (the perpendicular distance from the shower axis) of several hundred metres, the width of the shower front results in measured signals which may be spread over hundreds of nanoseconds [7, 18]. The Molire radius, r_M, is the spread of low energy particles due to Coulomb scattering, hence it can quantify the lateral spread of the electromagnetic component of an EAS. For low energy electrons at sea level, r_M is approximately 80 m [9]. The total footprint of high energy showers at ground level can be many square kilometres.

To understand shower development from hadronic primaries, Matthews has utilised the Heitler model of electromagnetic cascades and extended it to include hadronic processes [19]. Similar to the Heitler model, Matthews' model is a simplification but its predictions agree well with both observations and detailed Monte Carlo simulations.

The following scenario applies Matthews' model. Consider a proton of energy E_0 propagating through the Earth's atmosphere. Analogous to the Heitler model, the hadronic component interaction points have a spacing of $\lambda_I \ln 2$, where λ_I is the interaction length ($\sim 120\,\mathrm{gcm}^{-2}$ for pions in air). At each interaction point, N_{ch} charged pions and $\frac{1}{2}N_{ch}$ neutral pions of equal energy are produced, until the energy of the charged pions is below the critical energy $E_{\pi,c}$, the energy at which the decay length equals the interaction length and so instead the charged pions decay into muons. $E_{\pi,c}$ is the energy where the decay length becomes less than $\lambda_I \ln 2$ (this value is dependent on the atmospheric conditions, but in [19] a value of 20 GeV is used). Any neutral pions which are created at an interaction point immediately decay, producing photons which initiate electromagnetic cascades. The development of a shower according to Matthews' model is illustrated in Figure 1.2.

The multiplicity N_{ch} of charged particles for proton-proton and pion-proton collisions is approximated to be equal, so after n interactions the number of charged pions in the shower is $(N_{ch})^n$. The total energy of charged pions in the shower is $(\frac{2}{3})^n E_0$, therefore the energy per charged pion is

$$E_\pi = \frac{E_0}{\left(\frac{3}{2}N_{ch}\right)^n} \tag{1.13}$$

The remaining energy of the primary particle is channelled into electromagnetic cascades initiated by neutral pions. Defining n_c as the number of interactions after which $E_\pi \leq E_{\pi,c}$, we obtain:

$$n_c = \frac{\ln\left(\frac{E_0}{E_{\pi,c}}\right)}{\ln\left(\frac{3}{2}N_{ch}\right)} \tag{1.14}$$

by rearranging Equation (1.13).

In the Matthews model, all energy is contained in the electromagnetic component and the muonic component, therefore the primary particle's energy is:

$$E_0 \approx 0.85\,\mathrm{GeV}(N_e + 24N_\mu) \tag{1.15}$$

where N_e and N_μ are the number of electrons and muons at shower maximum respectively, 0.85 GeV is the energy of each electron, and the factor of 24 describes the energy contribution of each muon relative to an electron. Equation (1.15) shows that at shower maximum, if the number of charged particles is known, the energy of the primary can be estimated. The relative weighting between the electron and muon contributions is governed by the ratio of the electron and muon critical energies. This relationship has been corrected for an oversupply of photons in the

electromagnetic component according to the Heitler model, verified by Monte Carlo simulations [19].

Matthews' model can be used to calculate the number of muons in a shower. Assuming that every charged pion eventually decays to muons, the number of muons is $N_\mu = (N_{ch})^{n_c}$. From Equation (1.14):

$$\ln N_\mu = n_c \ln N_{ch} = \beta \ln \left(\frac{E_0}{E_{\pi,c}} \right) \tag{1.16}$$

therefore

$$N_\mu \propto E_0^\beta \tag{1.17}$$

where β is the ratio in each hadronic interaction of the logarithms of the charged and total particle multiplicities:

$$\beta = \frac{\ln N_{ch}}{\ln \left(\frac{3}{2} N_{ch} \right)} \tag{1.18}$$

Different hadronic models use different values for β, but $\beta = 0.85$ is a fair assumption [14, 19].

Matthews' model can also be used to calculate the average value of the depth of shower maximum for protons, $X^p_{\max}$, at different energies. To derive this, only the first generation of electromagnetic showers is considered, therefore there is a systematic underestimation of $X^p_{\max}$, but the extent of this systematic with energy can be estimated [19]. The depth of shower maximum can be represented by the depth of the first interaction, X_1, plus the depth of maximum of the electromagnetic showers:

$$X^p_{\max} = X_1 + \lambda_r \ln \left(\frac{E_0}{3 N_{ch} E_{c,e}} \right) \tag{1.19}$$

where λ_r is the electron radiation length and $E_{c,e}$ is the electron critical energy. The second term of Equation (1.19) is from Heitler's model of the depth of maximum of an electromagnetic shower initiated by a photon with energy $E_\gamma = \frac{E_0}{3N_{ch}}$, applying the assumptions of the first interaction to obtain E_γ. To compare with electromagnetic showers, Equation (1.19) is rewritten:

$$X^p_{\max} = X^\gamma_{\max} + X_1 - \lambda_r \ln (3 N_{ch}) \tag{1.20}$$

where $X^\gamma_{\max}$ is the electromagnetic shower depth of maximum. We can see that a purely electromagnetic shower develops deeper in the atmosphere than a shower initiated by a proton primary. The energy dependence of $X^p_{\max}$ can be represented by the change in average $X^p_{\max}$ over a decade of primary energy, referred to as the 'elongation rate' Λ^p:

$$\Lambda^p = \frac{dX^p_{\max}}{d \log_{10} E_0} = \Lambda^\gamma + \frac{d}{d \log_{10} E_0} \left[X_1 - \lambda_r \ln (3 N_{ch}) \right] \tag{1.21}$$

where Λ^γ is the elongation rate of electromagnetic showers. The value of X_1 decreases with energy as the proton-proton collision cross-section increases with energy, and N_{ch} increases with energy. Consequently, these terms reduce the elongation rate for protons from Λ^γ. The prediction of $\Lambda^p = 58\,\mathrm{gcm}^{-2}$ per decade of energy from Matthews' model is consistent with simulation results [19].

These proton derivations can be applied to heavier nuclei by invoking the 'superposition principle', which approximates a nucleus of A nucleons to be A independent protons arriving together, each of energy $\frac{E_0}{A}$. These nucleons each initiate their own independently developing shower. By applying the preceding derivations to these conditions we obtain:

$$N^A_\mu \propto A \left(\frac{E_0}{A} \right)^{0.85} = A^{0.15} N^p_\mu \tag{1.22}$$

$$X^A_{\max} = X_1 + \lambda_r \ln \left(\frac{E_0}{3 A N_{ch} E_{c,e}} \right) = X^p_{\max} - \lambda_r \ln A \tag{1.23}$$

These equations show that on average, showers initiated by heavier nuclei produce more muons and that their shower maxima occur at a shallower depth than proton initiated showers of the same energy. An iron shower for example ($A = 56$) contains around 80% more muons, and assuming $\lambda_r = 37\,\mathrm{gcm}^{-2}$ it reaches shower maximum about $150\,\mathrm{gcm}^{-2}$ earlier than a proton shower of the same energy. For nuclei there is also less fluctuation in $X_{\max}$ compared to protons, because the A individual showers superimpose. This reduction however is less than a factor of $\frac{1}{\sqrt{A}}$ as the interactions of the nucleons are not completely independent [20]. If the composition of the CR flux changes with energy, then A is energy dependent and so a term of the form $\frac{d\ln A}{d\ln E}$ will be added to the elongation rate [21].

The inelasticity, κ, is the fraction of the total energy converted into charged and neutral pions at each interaction. Therefore if $\kappa < 1$, the primary particle retains some energy. To account for a changing inelasticity, the model can be altered so that at each interaction, instead of the primary particle energy being transferred entirely into charged and neutral pions, a fraction of the primary energy is kept by the primary particle. By adjusting for inelasticity, decreasing κ increases $X_{\max}$. If β in Equation (1.17) is altered for inelasticity, the number of muons produced also increases if κ decreases [19].

The basic properties of hadronic air showers are explained with good accuracy by the previously described model. Investigating EAS development in more detail requires complicated simulations, information of which can be found in [7, 9, 22] and other references therein.

1.2.3 Atmospheric Light Production

Atmospheric nitrogen fluorescence is produced when charged particles travel through air, while Čerenkov light is produced by charged particles travelling through air or water. By deploying detectors to collect this light, an analysis and reconstruction of the primary particle's properties can be achieved.

1.2.3.1 Atmospheric Fluorescence

As an EAS propagates through the atmosphere, the charged particles of the shower excite nitrogen molecules. The de-excitation of these molecules results in the isotropic emission of light of wavelengths between $\sim 300 - 430\,\mathrm{nm}$. Figure 1.3 shows the measured spectrum of this light. The energy deposited in the atmosphere by the charged shower particles is proportional to the energy of the fluorescence photons emitted. For example, for every MeV of energy deposited by the shower particles at room temperature and atmospheric pressure, in the main emission band of 337 nm the absolute fluorescence yield is about 5 photons [23]. This fluorescence yield is small, but at very high cosmic ray energies enough fluorescence light is produced to be distinguishable from background sources during favourable viewing conditions. Fluorescence detectors are discussed in Section 1.3.2.

If the expected photon yield is well known, the number of photons collected by a detector can be used to calculate the energy deposited electromagnetically in the atmosphere by the shower particles (approximately 90% of the primary particle's energy is deposited electromagnetically [24]). In laboratory conditions the yield can be accurately measured, therefore estimating cosmic ray energies through fluorescence measurements is almost calorimetric. Some energy is not deposited in the atmosphere, for example energy carried by muons and neutrinos that is deposited into the ground [23, 25]. This energy is not detected by looking for fluorescence emission, and as such is referred to as 'invisible energy' which is corrected for in the estimation of the primary energy.

As a charged particle moves through a dielectric medium, molecules surrounding the particle are polarised which, upon returning to an unpolarised state will emit radiation. If a charged particle is moving faster than the local speed of light in the medium, $v > \frac{c}{n}$ for a refractive index n, then the emitted radiation is coherent and termed Čerenkov light. At sea level, the

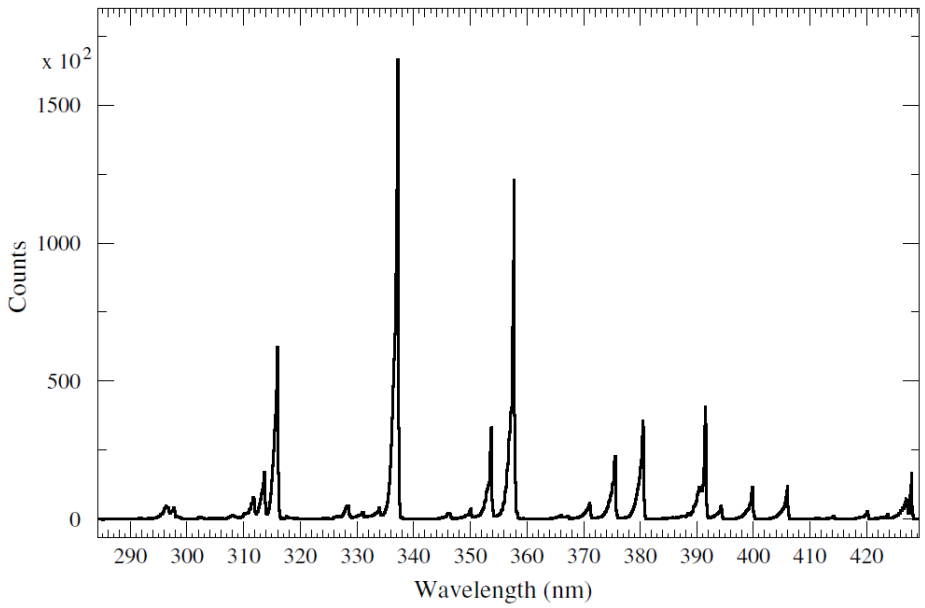


Figure 1.3: The air fluorescence spectrum measured in dry air at 293 K and 800 hPa. From [26].

refractive index of air is 1.0003, so the minimum energy for electrons to produce Čerenkov light is ~ 21 MeV, whereas in water the threshold energy is ~ 1 MeV [7].

The Čerenkov light emitted in air is mainly in the visible and UV range [7]. Čerenkov light is emitted in a cone around the charged particle's axis of motion. The opening angle θ is:

$$\cos\theta = \frac{1}{\beta n} \tag{1.24}$$

and its intensity is

$$I \propto q^2 \left[1 - \frac{1}{(\beta n)^2}\right] \tag{1.25}$$

for a particle of charge q and $\beta = v_p/c$, where v_p is the particle speed [5].

Čerenkov light is taken advantage of in numerous CR detectors, including the Pierre Auger Observatory which uses water tanks as a detection medium (see Chapter 3).

1.3 Extensive Air Shower Detection Techniques

Cosmic rays below an energy of 10^{14} eV can be directly detected, as the CR flux at Earth is large enough that instruments on balloons and satellites can measure a significant number of events. These detectors contain components that determine the energy, mass and charge of a cosmic ray. Measuring these fundamental cosmic ray properties allows us to determine the particle species of a cosmic ray. To detect cosmic rays on an event-by-event basis like this is the ideal situation. The energy range of these detectors however is limited by the physical dimensions of the detector [7].

The flux of cosmic rays above 10^{14} eV at Earth is too small for direct detection. Instead we measure these cosmic rays indirectly [27]. A high energy primary particle propagating through the atmosphere can produce an EAS with a footprint at ground level of many square kilometres. Indirect detection requires the detector lie within the EAS footprint, therefore the effective aperture of the detector is increased. Light is also produced by the interaction of the EAS with the atmosphere, allowing the atmosphere to be used as a detection medium if detectors can observe the light from a distance. Provided the detector views a large enough area, a sufficient number of high energy cosmic ray events can be observed. By measuring the properties of extensive air showers, fundamental properties of the primary particle can be deduced.

This section will discuss two common strategies for the detection of ultra-high energy cosmic rays: ground arrays and atmospheric fluorescence detection.

1.3.1 Ground Arrays

The same general approach is used by different ground array experiments to measure and reconstruct showers. By distributing a number of particle detectors over an area, as the shower front passes over, the ground array measures the signals produced in the detectors at each location. The number of particles passing each detector can be reconstructed by using these measured signals and a knowledge of the detector response induced by different particle species. Fundamental shower properties such as the primary energy, arrival direction and even particle composition can be reconstructed from the measured signals and their time of measurement.

The size, spacing between individual detectors, and atmospheric depth at the location of the array dictate the ability of a ground array to measure a particular energy range. Above energies of 10^{19} eV, the cosmic ray flux at Earth is approximately one particle per square kilometre per year, with a shower footprint at ground level of several square kilometres [28, 29]. If the detector array is to collect sufficient statistics, the array must span many square kilometres. Additionally, to achieve an accurate shower reconstruction, the spacing between individual detectors must not be larger than the size of the shower footprint. At these energies the average depth of shower maximum is $\sim 750\,\mathrm{gcm}^{-2}$, consequently the ground array must be at an atmospheric depth greater than $\sim 800\,\mathrm{gcm}^{-2}$ to effectively measure these showers. Detection close to the shower maximum reduces the effect of shower development fluctuations [18].

The advantage of ground arrays is they can operate constantly, as opposed to fluorescence detectors which have a duty cycle of approximately 15% [30] (see Section 1.3.2). Therefore, ground arrays can gather very large data sets. Ground arrays also have an almost uniform exposure in right ascension, with only minor corrections required for detector downtime [28].

Typically, the first step in the reconstruction of an event detected by a ground array is the calculation of its arrival direction. To determine the shower axis, the arrival times of the shower in each detector are fitted with a shower front model that assumes the shower front moves at the speed of light. Close to the shower core a planar model may be sufficient, but for signals measured at large core distances a curved shower front model must be applied [7, 31]. To achieve a precise arrival direction reconstruction, the timing resolution of individual detectors and the synchronisation between detectors must be accurate [18, 32]. The angular resolution of ground arrays is usually several degrees, but if an event is measured by many detectors the angular resolution may be less than 1°.

Once the shower axis has been determined, the core location, which is the intersection of the shower axis with the ground plane, can be reconstructed. This is achieved by fitting a lateral distribution function (LDF) to the signals measured by each detector. Across different experiments, the LDF quantifies either the shower particle density, signal density produced by shower particles, or signal recorded by a given detector, as a function of core distance (the perpendicular distance from the shower axis). The LDF depends on the distribution of particles in the shower as well as the type of detector utilised, and therefore differs between experiments [7, 18, 33].

Upon fitting an LDF, it is used to estimate the primary energy by using a 'ground parameter'. This parameter is extracted from the LDF fit or from an interpolation of the measured signal between detectors, to provide the observed value of the LDF if the detector was at a particular distance from the shower core, typically a distance between $500 - 1000$ m. A ground parameter is used because of the research of Hillas. He found that the signal at core distances of $500 - 1000$ m was less susceptible than the estimation of the total number of particles at ground level to large changes resulting from differences in the primary particle species, shower development fluctuations and uncertainties in the true LDF [34]. This is because at these core distances the particles have been produced near shower maximum, and thus the number of these particles is more consistent

for a given energy. The ideal core distance value depends on the detector spacing, atmospheric depth and energy range of interest, but for the simulations considered by Hillas the best value was 500 m [25, 29].

A correction is applied to the ground parameter to compensate for the attenuation of shower particles at different zenith angles [25]. The relationship between the ground parameter and primary energy E is then used to estimate the energy of the initial particle that initiated the shower. This relationship is inferred from air shower simulations and is usually of the form $E = \beta P_{ground}^{\alpha}$, where P_{ground} is the ground parameter and α and β are fitted coefficients [14,35,36].

A consequence of the reliance on air shower simulations to formulate the relationship between a ground parameter and primary energy is that hadronic interaction models are relied upon at energies beyond which experimental verification of their predictions can be achieved [28]. As a result, the energy scales of the ground arrays at ultra high energies have significant systematic uncertainties, and therefore comparing results between detectors is challenging, unless energy estimation methods that are independent of air shower simulations are applied (see Section 1.3.3) [18, 25, 37].

The most difficult facet of the CR flux for ground arrays to measure is typically the composition at a given energy. Ground arrays sample the shower front at a single altitude, as opposed to fluorescence detectors that reconstruct the EAS longitudinally. Consequently, ground arrays can not directly measure $X_{\max}$ and so other composition observables need to be considered. The muon content is an observable that is sensitive to the primary particle composition. Measuring the muon densities at ground level is one way to study the mass composition. Another useful quantity is the calculation of the 'risetime' of detector signals - the time taken for the integrated signal to increase from 10% to 50% of its final value - providing information on the ratio of the muon and electromagnetic components. Studies involving inclined showers also probe the primary composition, because the longitudinal development of a shower is sensitive to the primary composition. For inclined showers, the amount of atmosphere shower particles travel through changes with angle around the shower axis, therefore these 'asymmetry' studies essentially determine shower development at different stages [18, 38–40]

Ground arrays typically detect showers by using scintillators or water-Čerenkov tanks. A scintillator detector emits pulses of light as a charged particle passes through it, and then the light is collected by a photomultiplier tube (PMT). Scintillators count the number of incident charged particles and so they can not distinguish muons from electrons. To overcome this, shielding is utilised to absorb the electromagnetic component of the shower before it can enter the detector [9]. Scintillator detectors do not efficiently absorb photons [18].

Water-Čerenkov detectors use a pool of water enclosed within the detector to measure the Čerenkov emission initiated by shower particles propagating through the water. Water-Čerenkov arrays are not only cheaper but have a larger aperture than scintillator arrays of equal size, as the zenith angle acceptance of water-Čerenkov detectors is larger. The depth of water-Čerenkov detectors allows several radiation lengths to be presented to incoming photons, thus photons can undergo pair-production and the subsequent electron-positron pair can be detected. This provides a consistent sampling of the dominant shower components, reducing reconstruction uncertainties from statistical fluctuations in the number of particles detected [33]. The signals produced by muons have a different size and shape compared to the signals produced by the electromagnetic component in these detectors [7, 41].

The approaches described in this section generally do not apply to highly inclined showers, which are instead reconstructed with different techniques [36]. This is because at large zenith angles the attenuation of shower particles may be so great that the applied LDF does not adequately describe the particle distribution at ground level, and zenith angle ground parameter corrections can no longer suffice.

1.3.2 Fluorescence Detection

Fluorescence detectors (FDs) longitudinally measure the nitrogren emission induced by an EAS, usually from a single ground location. Unlike ground arrays which sample the shower at a single altitude using detectors covering a large area, FDs can study an EAS over a range of shower ages in great detail.

Fluorescence detectors rely on clear nights and low levels of background illumination to be able to view fluorescence emission. Therefore FDs are not operational during day time, bad weather or periods of significant moonlight, resulting in an FD duty cycle of around 15%, compared to ground arrays with a possible duty cycle of 100%. As many factors affect the uptime of the FD, meticulous monitoring is necessary to determine the detector exposure [42]. Hybrid detectors, discussed in the following section, combine the operation of FDs with the ground array to provide a solution to the low duty cycle of the FD array.

A typical fluorescence detector uses a collection of pixels that observe separate regions of the sky to image the atmospheric nitrogen fluorescence. Light is directed by optical instruments to the surface of each pixel, which is essentially the photocathode of a photomulitplier tube (PMT). As an EAS propagates through the atmosphere, the FD observes the EAS as a small spot of light moving across the photomultiplier camera. The geometry of the shower is reconstructed by measuring the sequence of pulse times from each pixel, and an estimate of the primary energy is obtained by integrating the measured pulses.

The first step in determining the energy of a measured shower is to determine which observed photons were induced by the EAS. Hence, the 'shower axis' along which the EAS developed must be determined. This axis lies within the 'shower-detector plane' (SDP). The SDP is illustrated in Figure 1.4.

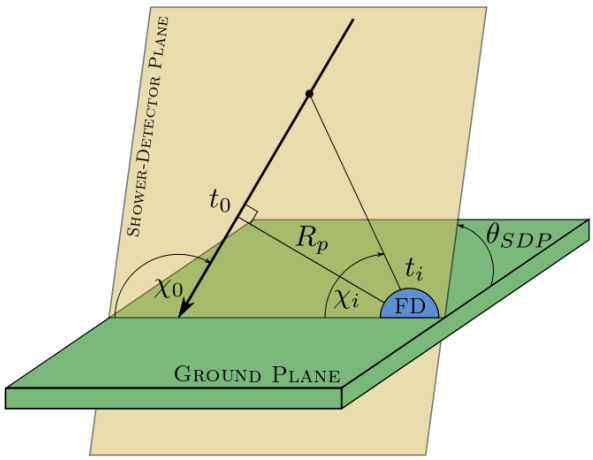


Figure 1.4: A diagram displaying the parameters used to define the geometry of an EAS observed by a fluorescence detector. From [11].

To reconstruct the shower geometry, the first step is a fit to the pointing directions of each pixel which measured a signal, with each pixel weighted by the size of that signal [43]. The result is a plane that contains the shower axis and intersects the detector. The angle between the ground plane and the SDP is defined as θ_{SDP}. The accuracy of the SDP derived from the fit depends on the angular size of each pixel and on the rejection of accidently triggered pixels.

The next step is to use pixel timing information to determine the shower axis. The distance of closest approach of the shower to the detector, defined as R_p, is referred to as the impact parameter. The time at which the shower crosses this point is t_0, and the angle that the shower

axis makes with the horizontal within the SDP is χ_0. Combining these parameters gives the expected arrival time of light in the i^{th} pixel:

$$t_i = t_0 + \frac{R_p}{c} \tan\left(\frac{\chi_0 - \chi_i}{2}\right) \tag{1.26}$$

where c is the speed of light and χ_i is the angle of elevation in the SDP of the i^{th} pixel [24]. If a fluorescence detector is working in isolation ('monocular mode'), by considering t_0, R_p and χ_0 as the parameters to be determined, the axis can be determined by minimising the sum:

$$\chi^2 = \sum_{i=1}^{n} \frac{(t_i - t_{i,obs})^2}{\sigma_i^2} \tag{1.27}$$

where $t_{i,obs}$ is the observed signal time in the i^{th} pixel and σ_i is the uncertainty in its value. If there is minimal variation in the shower's angular velocity $\frac{d\chi}{dt}$ as seen by the FD, then difficulties may arise due to degeneracy between the fitted parameters R_p and χ_0, resulting in no uniquely determined shower geometry. Using timing information from the ground array though can solve this degeneracy and therefore determine the shower axis (see Section 1.3.3).

With the shower axis determined, the primary particle's energy can be estimated. This involves converting the light collected at the PMT photocathodes into the energy deposited in the atmosphere by the shower as a function of atmospheric depth. To accurately achieve this, the altitude profiles of atmospheric density, air pressure, temperature and humidity must all be well known, as these atmospheric properties affect the nitrogen fluorescence yield [44]. Additionally, the attenuation of fluorescence light from its emission point to the detector must be considered, therefore the prevailing atmospheric conditions, including the presence of clouds and aerosols, must be known as these impede the path of light towards the detector [28]. Other sources that add to the flux of photons measured must also be taken into account. These sources include background light from stars and terrestrial sources which contribute to the measured photon flux, Čerenkov light from the shower that can either directly reach or scatter into the detector, and fluorescence photons that otherwise should not have reached the detector but managed to be scattered into the optics. All of these factors can sway the energy estimation, and also influence the shape of the measured profile which affects the calculation of $X_{\max}$ [45].

Upon calculating the energy deposit profile, a Gaisser-Hillas function is fitted to calculate $X_{\max}$. Integrating the fitted function gives the energy of the event. A correction determined through air shower simulations is applied to compensate for the invisible energy component, which at high energies is minimal [7]. The energy deposited in the atmosphere by the shower is proportional to the energy of the emitted fluorescence photons, therefore this energy estimation is calorimetric, as opposed to the energy estimation derived from ground arrays which relies significantly on air shower simulations. As a result the systematic uncertainties between different types of energy estimation differ.

Fluorescence detectors directly measure $X_{\max}$, with measurement uncertainties of around $20\,\mathrm{gcm}^{-2}$, allowing useful studies of the primary mass composition. Finding the width of the measured $X_{\max}$ distributions as a function of energy, and measuring the elongation rate, are two approaches which use $X_{\max}$ to infer the mass composition. Studies utilising $X_{\max}$ match the measured results with air shower simulations, typically comparing the expectation for pure proton or pure iron primaries, giving an indication of how 'heavy' the CR flux is at particular energies [18, 20, 46].

In ideal atmospheric conditions, the aperture of a single fluorescence detector can be thousands of $\mathrm{km}^2\mathrm{sr}$. The aperture increases with energy because the amount of fluorescence light emitted is proportional to the primary energy [47]. The FD reconstruction quality can be improved by using multiple, independent fluorescence detectors at different locations. This increases the aperture of the experiment, and allows some events to be detected by multiple FDs. Events seen

by two different FDs, referred to as 'stereo' events, can have their axis reconstructed without using a timing fit but instead by finding the intersection of the calculated SDPs from each detector [18]. Furthermore, by fitting stereo events independently with each detector and then comparing the parameters from each reconstruction, statistical reconstruction uncertainties can be directly calculated from data [37, 46].

1.3.3 Hybrid Detection

Ground arrays and fluorescence detection can be synergised to effectively operate as a single detector, referred to as a 'hybrid' [7, 18, 30, 48, 49]. A hybrid detector is advantageous as the weaknesses of either the ground array or fluorescence detector can be compensated for by the other detector. The ground array has a large duty cycle, allowing it to be used for the majority of the data collection. The fluorescence detector operates for a fraction of the time, observing a subset of the events that the ground array observes. The events in this subset are called hybrid events, which are reconstructed with greater accuracy than events observed with a single type of detector [37,50]. Hybrid data sets have limited statistics, but they provide much better reconstructions. In Chapter 2 some studies which take advantage of these high quality reconstructions are discussed.

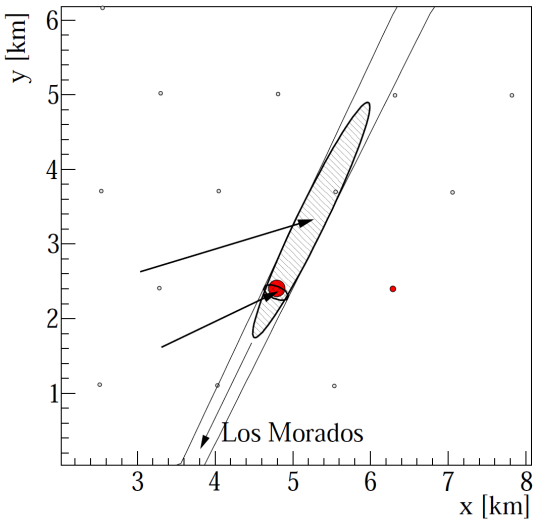


Figure 1.5: This image shows the extra precision obtained by using a hybrid reconstruction. An event measured by the Pierre Auger Observatory was reconstructed using monocular and hybrid methods, resulting in different uncertainties in the core location, which are represented by the two ellipses. The larger ellipse represents the uncertainty in the core location from the monocular reconstruction, the smaller ellipse is from the hybrid reconstruction. The arrows pointing to the centre of the ellipses displays the reconstructed shower axis for each reconstruction. The ground array detectors are represented by the circles, with open ones being untriggered and filled ones being triggered detectors. The size of the filled circles is proportional to the logarithm of the signal measured. The two solid lines represent the uncertainty in the intersection of the SDP with the ground. The arrowing pointing towards the bottom of the image shows the direction to the fluorescence detector (named Los Morados) that measured the event. Image from [50].

A hybrid detector can meld timing information from the ground array with the shower track observed by the fluorescence detector [24]. If the angular speed $\frac{d\chi}{dt}$ of a shower observed by a single fluorescence detector displays minimal variation from linearity, then the fit parameters R_p and χ_0 may be highly degenerate. This usually occurs only if a small track segment is observed. Consequently, a large range of possible (R_p, χ_0) solutions which are highly correlated can fit the data points, resulting in a poorly constrained shower geometry. If the shower triggered ground-

based detectors, the time of ground impact is known, and due to the time difference between the fluorescence light emission (i.e. light arriving to the FD) and the shower front crossing the ground array, this larger range of timing information allows the curvature in $\frac{d\chi}{dt}$ to be determined. This constrains the fit, allowing the shower geometry to be accurately calculated [43]. The (R_p, χ_0) degeneracy can be broken with timing information from a single ground array detector, allowing events that would not trigger the ground array to be detected and allowing good-quality reconstruction of low energy events [24]. It is important that the timing offset between the two detectors are well known if a hybrid reconstruction is performed [37, 51]. Figure 1.5 displays an example of the core location uncertainties of an event observed by the Pierre Auger Observatory, reconstructed with the hybrid method compared to reconstruction with the monocular method (see Section 3.9.4).

If the geometry is well-constrained, the integration of the reconstructed energy deposit profile of the event reduces the statistical uncertainty in the estimation of the primary particle's energy [50]. In the case of hybrid events, the ground array parameter used for energy estimation can be compared with the more accurate calorimetric estimate from the fluorescence detector, E_{FD}. Combining the information from hybrid events allows a parameterisation to be formulated that describes the ground parameter energy estimate as a function of E_{FD}. This 'calibration curve' eases reliance on air shower simulations which introduce significant systematic uncertainties in the ground array energy estimation. Applying the calibration curve to any event measured by the hybrid detector allows an internally-consistent energy estimation to be obtained, irrespective of the event detection method. Therefore, the ground array energy scale acquires the systematic uncertainties of the FD energy scale.

In Section 1.3.1, it was explained that composition measurements are difficult with ground arrays. Hybrid detectors allow composition estimators from ground array measurements to be related with the direct measurements of $X_{\max}$ from the fluorescence detectors, similar to the energy estimator calibration. As before, the advantage of this comparison is that the significant data gathering power of the ground array can be fully utilised, whilst air shower simulations to interpret the ground array results are no longer necessary [39].

Chapter 2

Theoretical Models and Experimental Observations of Cosmic Rays

The three fundamental aspects of the cosmic ray flux at Earth are its energy spectrum, composition and anisotropy. The individual study of these three properties can provide information regarding the sources of cosmic rays, and the processes cosmic rays undergo as they propagate to Earth. A combination of these fundamental areas is required to comprehensively understand CR astrophysics.

Cosmic ray observatories endeavour to measure each of these cosmic ray properties at the energy range of interest. This chapter will mainly discuss experimental observations and theoretical models relating to ultra-high energy cosmic rays (UHECRs), meaning energies of at least 1 EeV (10^{18} eV). Information on the detectors mentioned in this chapter can be found in Chapter 3. Detailed discussion of measurements at lower cosmic ray energies can be found in [7, 14, 18, 52] and references therein.

2.1 Cosmic Ray Production and Sources

Cosmic rays with lower energies are believed to originate from Galactic sources. The magnetic fields of the Milky Way are not considered strong enough to confine particles of energies above the knee ($\approx 10^{16}$ eV for protons). It is postulated that these higher energy particles originate from extra-galactic sources, otherwise they should point back towards their galactic sources.

2.1.1 Cosmic Ray Production

There are two main types of cosmic ray production models - 'bottom-up' or 'top-down'. Bottom-up models describe the acceleration of a set of charged particles from low to high energies, whereas top-down models consider UHECRs as originating from exotic particles.

2.1.1.1 Bottom-up Models

Bottom-up models theorise that UHECRs are produced from the acceleration of a population of particles within a source region, attaining a large enough energy that they may escape. There are two main bottom-up production theories.

The first theory is that an extended electric field directly accelerates particles to a high energy, a mechanism that was first proposed in 1933 by W.F. Swann [53]. Changing magnetic fields near the surface of stars prompts this acceleration. Our Sun produces sunspots which come and go over a matter of days or weeks, generating magnetic fields of up to several kilo-Gauss. This

acceleration is termed a "one-shot" mechanism, and is generally attributed to pulsars or active galactic nuclei (AGN) which are highly magnetised and rapidly rotate. One-shot mechanisms are currently disfavoured, as these processes occur at very high density astrophysical sites where energy loss processes occur with high probability. The observed power law energy spectrum is also not explained naturally by these mechanisms.

The second theory is that particles are accelerated stochastically, a theory which originates in 1949 from E. Fermi who proposed a process in which particles gradually gain energy through continued interactions with moving magnetised plasma [54]. Compared to electric field acceleration, stochastic acceleration is slow, therefore the source must be able to accelerate particles for extended periods of time if the particle is to reach a high energy [55]. This requires the particles remain confined to the acceleration region for extended periods. By considering the probability of escaping the acceleration region, an upper limit on the particle energy that sources can accelerate particles to can be estimated. The maximum energy possible (E_{max}) for a particle of charge Z is:

$$E_{max} \approx 10^{18}\,\mathrm{eV} Z \beta_s \left(\frac{R}{kpc}\right)\left(\frac{B}{\mu G}\right) \tag{2.1}$$

for an acceleration region of size R, magnetic field strength B and shock velocity β_S in units of c [14].

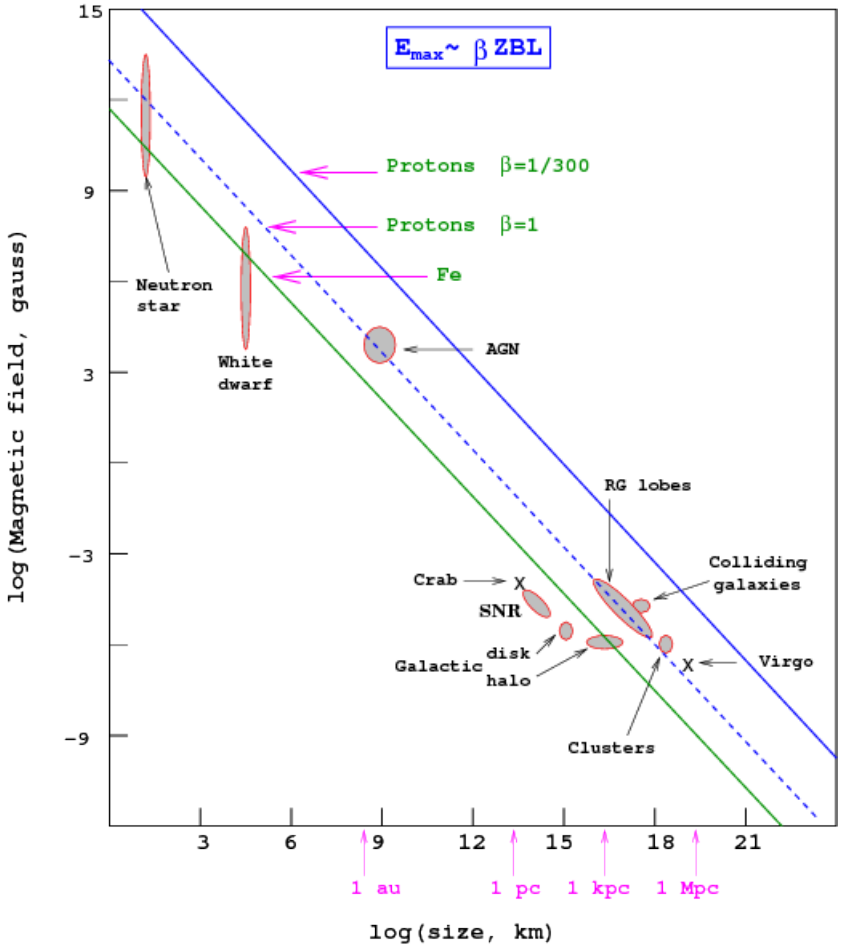


Figure 2.1: The 'Hillas plot', displaying the dimensions and magnetic field required of astrophysical sources for the containment of a 10^{20} eV proton and iron nucleus, respectively. The shaded region shows possible source candidates for these particles. Objects below the corresponding blue (green) diagonal lines cannot accelerate protons (iron nuclei) to 10^{20} eV. From [56].

The plot shown in Figure 2.1 was produced by Hillas based on the containment requirements

from Equation (2.1). The 'Hillas plot' shows the magnetic field and dimensions required of an astrophysical object to accelerate particles of energy 10^{20} eV according to Equation (2.1). The plot indicates objects such as active galactic nuclei (AGN), radio galaxies and galactic clusters are capable of producing the highest energy CRs. The process of acceleration however is not well-known. It is a mystery how CRs can be accelerated to the highest energies when at these energies there are predicted to be major energy losses, for example due to interactions with ambient photon fields in the source. These complications are discussed in [7, 57].

Fermi's acceleration mechanism is discussed in further detail below.

2.1.1.1.1 Fermi mechanism The Fermi mechanism theorises that cosmic ray particles propagate through interstellar space and collide with large objects moving with a random velocity and direction (such as magnetised clouds), and interact with astrophysical shocks. The cosmic ray will either lose or gain energy depending on the relative motion between the cosmic ray and the object.

Consider a test particle which increases its energy E by an amount $\Delta E = \xi E$ proportional to its energy per encounter with a magnetic cloud. After n encounters, the energy E_n is

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

$$n = \frac{\ln\left(E_n/E_0\right)}{\ln\left(1+\xi\right)} \tag{2.3}$$

where E_0 is the energy of injection.

P_{esc} is the probability of a particle to escape after one encounter from the acceleration region occupied by magnetic clouds. After n encounters, the probability of a particle to remain in the accleration region is $(1-P_{esc})^n$. Particles which remain in the acceleration region will be accelerated to higher energies, therefore the proportion of particles accelerated to energies greater than E_n is

$$N(>E_n) = N_0 \sum_{m=n}^{\infty} (1-P_{esc})^m \propto \frac{1}{P_{esc}} \left(\frac{E_n}{E_0}\right)^{-\gamma} \tag{2.4}$$

with

$$\gamma = \frac{\ln\left(1/\left(1-P_{esc}\right)\right)}{\ln\left(1+\xi\right)} \approx \frac{P_{esc}}{\xi} \tag{2.5}$$

The equations demonstrate that stochastic acceleration results in a power law energy spectra if P_{esc} and ξ are independent of energy.

2.1.1.1.2 Second order Fermi acceleration Second order acceleration concerns the amount of energy gained by a charged particle during its motion in the vicinity of randomly moving magnetised clouds, referred to as 'magnetic mirrors'. Fermi posed that the probability of a head-on particle and cloud collision is greater than a head-tail particle and cloud collision, therefore on average particles would be accelerated.

Consider a cosmic ray particle of energy E_i entering a single cloud with incident angle θ_i with respect to the cloud's direction. The particle experiences diffuse scattering due to magnetic field irregularities, eventually exiting the cloud at an angle of θ_f with energy E_f. Using Lorentz transformations between the laboratory frame (unprimed) and the cloud frame (primed), the energy gained by the particle can be calculated:

$$E_i^{'} = \Gamma E_i \left(1 - \beta\cos\theta_i\right) \tag{2.6}$$

$$E_f = \Gamma E_f^{'} \left(1 - \beta\cos\theta_f\right) \tag{2.7}$$

where Γ is the Lorentz factor and $\beta = V/c$ is the velocity of the magnetic cloud in units of the speed of light. In the cloud frame, the particle's energy is conserved, the elastic scattering meaning $E_i^{'} = E_f^{'}$. Therefore, the fractional energy change is

$$\xi = \frac{\Delta E}{E} = \frac{E_f - E_i}{E_i} \tag{2.8}$$

Averaging over $\cos\theta_i$ (depending on the relative velocity between the cloud and the particle) it can be shown that the fractional energy change is proportional to $\frac{4}{3}\beta^2$ [9].

2.1.1.1.3 First order Fermi acceleration (diffusive shock acceleration) Second order Fermi acceleration is a very slow process as $\beta \lll 1$. A more efficient acceleration process for cosmic ray encounters with planar shock fronts was proposed in the late 1970s [58].

Assume a large shock wave propagating with velocity $-\mu_1^{\rightarrow}$, and relative to the shock front the downstream shocked gas is receding with velocity $\mu_2^{\rightarrow}$, where $|\mu_2| < |\mu_1|$. Therefore, in the laboratory frame the shocked gas is moving in the direction of the front with velocity $V^{\rightarrow} = \mu_2^{\rightarrow} - \mu_1^{\rightarrow}$. Classifying the magnetic irregularities on either side of the shock as the magnetised plasma clouds, the energy gained per crossing and the total energy gained can be calculated by applying Fermi's original idea. For the rate at which cosmic rays cross the shock from downstream to upstream, and upstream to downstream, $\langle\cos\theta_i\rangle = -2/3$ and $\langle\cos\theta_f^{'}\rangle = -2/3$, cf. [9]. The fractional energy change is proportional to $\frac{4}{3}\beta$ [9].

As the energy gain per shock crossing is proportional to β, which is the velocity of the shock divided by the speed of light, the term 'first order' is used. This first order mechanism is more efficient than Fermi's original mechanism as the collisions are head on. Regardless of which side of the plasma the particles are, if the particles are moving with the plasma, the plasma on the other side is approaching them and so the particles gain energy from encounters with approaching magnetic irregularities. In the first order mechanism, γ only depends on the ratio of the upstream and downstream velocities. It is independent of the magnitude of the velocity of the plasma. This acceleration mechanism for strong shocks leads naturally to an E^{-2} particle energy spectrum [59].

2.1.1.2 Top-down Models

Top-down models consider UHECRs to be the decay products of supermassive elementary particles, called X particles, that may be remnants of the early stage of the universe.

These X particles may have originated from topological defects, caused by phase transitions in the early Universe and associated with the spontaneous symmetry breaking [60–65]. Examples of topological defects include magnetic monopoles, cosmic strings, domain walls, skyrmions and textures. Topological defects are stable and can last forever. Supermassive X particles of energy around 10^{16} to 10^{19} GeV can be trapped inside topological defects. These topological defects can be destroyed through processes such as annihilation or collapse, resulting in the release of energy in massive quanta that decay into quarks and leptons.

Another possible source of X particles are long-lived metastable super-heavy relic particles produced in the inflationary stage of the Universe through vacuum fluctuations [66–69]. These relic particles can also be destroyed through similar processes, decaying into quarks and leptons, thereby producing cosmic rays of maximum energy m_X. An alternative theory is that the relic topological defects are actually the cosmic ray primaries [70, 71].

Topological defects or super-heavy relic particles would exist in the galactic halo region. Some top-down models theorise that the decay of these particles results in a large fraction of photons at high energies. A different model called the Z-burst model predicts that interactions of ultra high energy neutrinos with relic neutrinos produces a large number of photon primaries. Recently, these models have been disfavoured as results from a number of experiments including the Pierre Auger Observatory have put strong limits on the photon fraction over a large energy range [20, 72, 73]. For more information on top-down models refer to [56, 74–77].

2.1.2 Cosmic Ray Sources

Several source models explain the observed high energy cosmic rays, predicting different mass composition distributions at Earth [78–81]. However, a specific astrophysical source of UHECRs has not been identified. Acceleration models for specific astrophysical objects have been developed, some of which will be discussed briefly here. For a more detailed discussion, see [82].

- Active Galactic Nucleus (AGN). An AGN is a compact region at the centre of a galaxy consisting of a supermassive black hole (SMBH). The accretion of matter by a SMBH results in confined, highly relativistic jets of particles. AGNs are the brightest known objects in the Universe. The lobes of the jets contain intense, turbulent fields which provide a strong shock that accelerates particles. The advantage of this mechanism is acceleration in the jet frame could have maximum energies that are smaller than the observed energy of the cosmic rays. However, the main problem with AGNs as UHECR sources is that as the jet velocity starts to slow, adiabatic deceleration of particles occurs. [83]

- Gamma Ray Bursts (GRBs). GRBs are bright flashes of gamma rays from immensely energetic explosions in distant galaxies, and are the most luminous electromagnetic events known. Depending on the proposed model, either the inner [84] or outer [85] GRB shock is the acceleration site, both using diffusive shock acceleration. The GRBs with measured redshifts are mostly at $z > 1$, therefore high GRB activity is required to explain the observed flux of cosmic rays.

- Galaxy clusters. Galaxy clusters have dimensions typically within $1 - 10\,\mathrm{Mpc}$ and magnetic fields of several μG. Shock acceleration between accretion shocks which are predicted to form in large clusters of galaxies is the proposed acceleration mechanism. Lower energy cosmic rays would be confined to the source by magnetic fields, whereas higher energy particles are able to escape. Particles could be accelerated up to $10^{20}\,\mathrm{eV}$. [86]

- Magnetars. These are young, magnetised neutron stars that have strong magnetic fields up to $10^{15}\,\mathrm{G}$ which accelerate CR through a magneto-hydrodynamic process rather than a stochastic process. The result is an energy spectrum proportional to $1/E$ [87]. Magnetars are capable of accelerating iron nuclei to energies of $10^{20}\,\mathrm{eV}$ [88].

- Giant Radio Galaxies. A proposed model is that cosmic rays are accelerated at the termination shocks of the jets, which extend to over $100\,\mathrm{kpc}$ [89]. The shocks are in extragalactic space therefore no adiabatic deceleration is predicted. The magnetic fields are capable of accelerating particles to $10^{20}\,\mathrm{eV}$. Potential sources are Centaurus A and M87 in the Virgo cluster.

- Starburst Galaxies. Starburst galaxies have a very high rate of star formation, which is typically triggered by the starburst galaxy merging with another galaxy or tidal interactions with a nearby galaxy [90]. The increased rate of short-lived star deaths, in addition to the winds of these galaxies, suggests starburst galaxies are a promising source of UHECRs.

2.2 Cosmic Ray Propagation

Charged cosmic rays propagating through space undergo a number of interactions. These processes determine the energy spectrum, composition and arrival direction distribution of the cosmic rays measured at Earth.

2.2.1 Magnetic Fields

Galactic magnetic fields (GMF) and extragalactic magnetic fields (EGMF) alter the trajectory of a cosmic ray propagating through space. The lack of knowledge of the strength of magnetic fields, the location of sources and the charge of cosmic rays prevents quantitative estimations of deflection angles. A charged particle in a magnetic field will travel in a helical path around magnetic field lines. The radius of the path, referred to as the Larmor or gyroradius r_L, is given by

$$r_L = \frac{E}{cqB} \tag{2.9}$$

where E is the particle energy, q is the particle's charge, B is the magnetic field strength and c is the speed of light.

Figure 2.2 illustrates the effect of a 10^{-9} G magnetic field on the paths of protons of some energy originating from a point source. This is a weak magnetic field, which is a reasonable estimate of the extragalactic magnetic field strength [91]. As the primary energy increases from 1 to 100 EeV, we see the CR transitions from diffusive to rectilinear propagation.

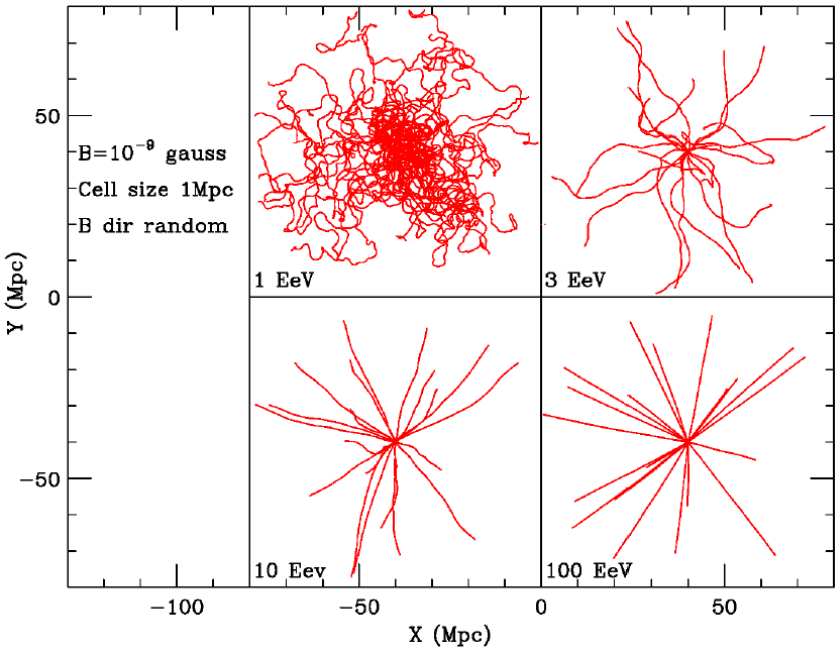


Figure 2.2: The projected view of the trajectories of 20 protons of some energy originating from a point source. The magnetic field is divided by cells of size 1 Mpc, with a mean value of 1 nG oriented randomly in each cell. From [92].

Our galactic magnetic field consists of two components, an organised component and a random component [93]. Despite the random component not being well quantified, current indirect measurements estimate the average total intensity of our magnetic field is 1 to 5 μG. Consequently, protons of energy 10^{20} eV would have a gyroradius of 10 to 50 kpc which is comparable to the size of the galaxy. As a strong UHECR anisotropy has not been observed, this estimated gyroradius suggests that these cosmic rays originate from outside the galaxy.

The galactic magnetic fields will cause some deflection of UHECRs. Excluding the Galactic Centre, different galactic field models propose that 10^{20} eV protons have an average deflection angle between 3.1° to 4.5°, whereas 2×10^{18} eV protons have an average deflection angle between 17.7° to 25.9° [94]. Iron particles have a deflection angle that is 26 times larger than protons, as the deflection angle magnitude is linearly dependent on charge.

2.2.2 Energy Losses

Heavy nuclei can experience losses due to interactions with the cosmic microwave background (CMB) radiation and extragalactic background light (EBL). For heavy nuclei of mass number A, CMB and EBL interactions can result in nuclear photo-disintegration and pair production [95,96]:

$$A + \gamma_{CMB,EBL} \rightarrow (A-1) + N, \tag{2.10}$$

$$A + \gamma_{CMB,EBL} \rightarrow (A-2) + 2N, \tag{2.11}$$

$$A + \gamma_{CMB} \rightarrow A + e^+ + e^- \tag{2.12}$$

where N is a nucleon.

For an iron nuclei of energy 2×10^{20} eV (energy at Earth), CMB and EBL interactions limit the distance to candidate sources to around 100 Mpc [95]. Protons can also undergo pair production, with the threshold being lower at $\approx 10^{18}$ eV. Despite pair production energy loss being minuscule overall compared to other mechanisms (see Section 2.2.3), some predict pair production is the major energy loss process for energies between the second knee (4×10^{17} eV) and ankle (5×10^{18} eV), resulting in the nature of the energy spectrum around this region [97]. A contrary theory is that the shape of the ankle is due to CR sources transitioning from galactic to extragalactic [42].

2.2.3 GZK limit

After the discovery of the CMB by Penzias and Wilson in 1965 [98], a flux suppression above an energy of 6×10^{19} eV was predicted by Greisen [99], and independently by Zatsepin and Kuz'min [100]. Referred to as the GZK (Greisen, Zatsepin and Kuz'min) cut-off, it is predicted to be caused by interactions between photons of the CMB and protons at energies around 10^{20} eV. If the proton's energy is greater than 6×10^{19} eV, the CMB photon has a blue-shifted energy of around 140 MeV in the proton's rest frame, which is above the threshold for pion production. There are two main ways pion photoproduction can occur via the Δ^+ resonance:

$$p + \gamma_{CMB} \rightarrow \Delta^+ \rightarrow p + \pi^0 \tag{2.13}$$

$$p + \gamma_{CMB} \rightarrow \Delta^+ \rightarrow n + \pi^+ \tag{2.14}$$

At higher energies a single reaction can produce additional pions such as a $\pi^-\pi^+$ pair. Figure 2.3 shows how successive CMB radiation interactions decrease the energy of a proton. When a proton undergoes pion photoproduction it loses approximately 20% of its energy. Above 3×10^{20} eV, the mean free path of this interaction is $\approx$ 5 Mpc, with both the mean free path and inelasticity being energy-dependent. As a result of the interactions with the CMB, the sources of significant numbers of CRs above the GZK cut-off must be located within 100 Mpc, hence there is an expectation of a cut-off assuming a uniform distribution of sources in the local universe [18,28,52]. Additionally, as proton primaries of energy greater than 6×10^{19} eV are expected to be minimally deflected by galactic and extragalactic magnetic fields, and these protons are not expected to travel more than 100 Mpc, the protons should point back to their sources.

Nuclei propagating through intergalactic space are also predicted to undergo pair production and photodisintegration through interactions with CMB, infra-red, optical and ultra-violet photons. Compared to protons, the energy threshold for this loss to occur is greater, however the energy loss length is shorter [7]. The spectra observed for heavier nuclei propagating to Earth would be similar to the spectra for protons impacted by the GZK effect [18, 102].

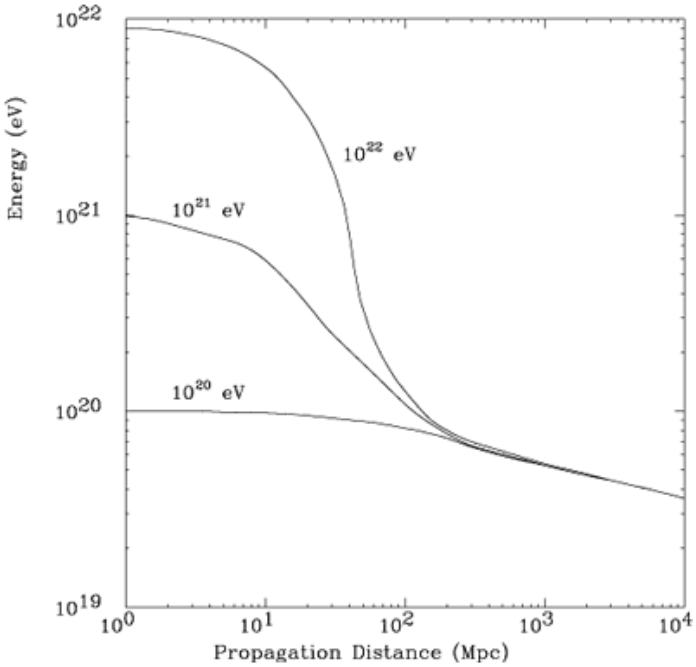


Figure 2.3: Energy of a proton as a function of its propagation distance through the CMB radiation for particular initial proton energies. From [101].

2.3 The Energy Spectrum of Cosmic Rays

The cosmic ray energy spectrum describes the number of particles arriving at Earth as a function of energy. The spectrum has the following form:

$$\frac{dN}{dE} \propto E^{-\gamma} \quad \mathrm{m}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}\,\mathrm{eV}^{-1} \tag{2.15}$$

where E is the particle energy and N is the number of particles arriving at Earth. From $\sim 10^{10}$ eV to $\sim 10^{20}$ eV, a value of $\gamma \sim 3$ describes the spectrum at Earth very well. The measured spectrum indicates a non-thermal origin of the cosmic rays. The spectrum falls steeply, therefore above 10^{15} eV, indirect detection methods are applied to measure significant numbers of these cosmic rays. At an energy of 10^{19} eV, approximately only one particle per square kilometre per year arrives at Earth [28].

Small changes in the observed value of γ are believed to indicate energy-dependent propagation effects, changing CR primary composition and/or different CR sources dominating. There are 3 main features of the spectrum. The 'knee' and 'ankle' are two of these features, named due to the spectrum resembling a human leg. The third spectral feature is a sharp cut-off at the end of the energy spectrum. The total energy spectrum above $\sim 10^{12}$ eV is displayed in Figure 2.4, along with the proton only energy spectrum below $\sim 10^{12}$ eV.

At an energy of $\sim 10^{15.5}$ eV, there is a steepening of the spectrum known as the knee. At this energy γ changes from ≈ 2.7 to ≈ 3.1 [14]. The steepening is due to the spectra of individual species successively falling off. The cut-off energy of each species is proportional to the charge of the species (see Section 2.4). Galactic sources are believed to be sites where diffusive shock acceleration occurs, such as in supernova remnants [7, 103]. It is theorised that past the point of the knee, Galactic sources no longer accelerate protons as these particles have exited the acceleration region [104]. Within the source region, nuclei are able to be accelerated to higher energies than protons as at a given energy a particle's rigidity is inversely proportional to its charge.

Conversely, if the knee-to-ankle region is theorised to consist of cosmic rays of Galactic origin,

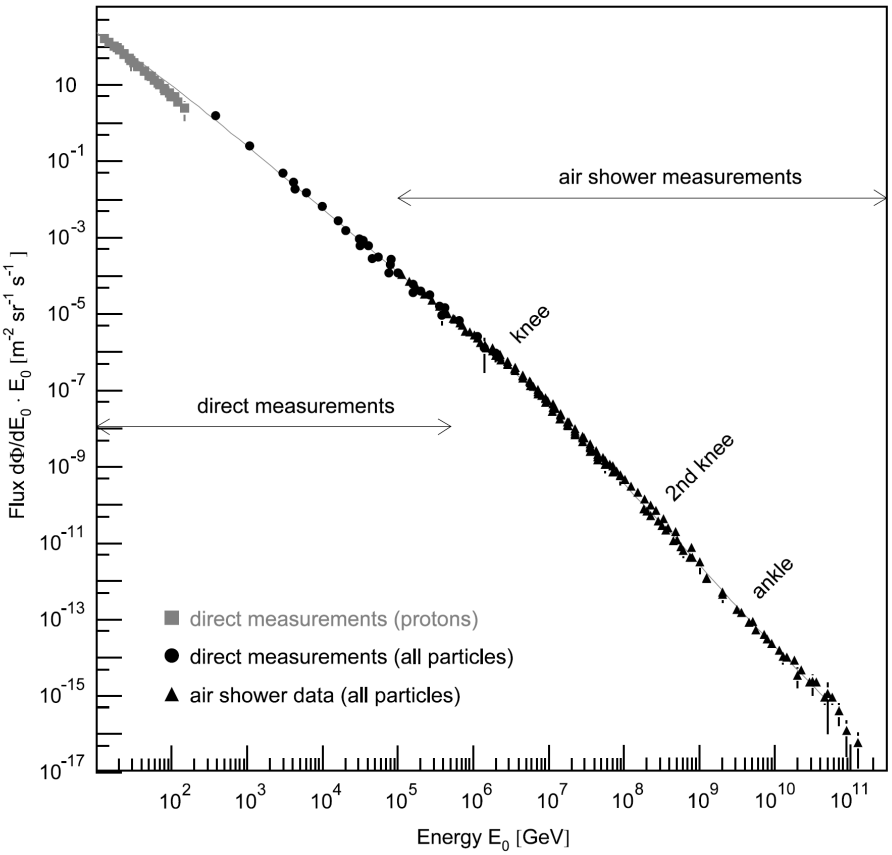


Figure 2.4: The measured cosmic ray energy spectrum, observed by a number of different experiments. The flux has been multiplied by E_0. From [14].

then the Galactic acceleration models must explain how particles can be accelerated to these high energies [27, 104, 105]. There are various rationales adopting this alternative assumption, such as the theory that the knee is due to the Milky Way being unable to confine higher rigidity particles within its magnetic field as opposed to an acceleration limit of the source, or a combination of various effects [106].

A spectral feature referred to as the 'second knee' is observed between energies of 10^{17} eV to 10^{18} eV, the exact energy of this feature is not well measured. This second knee is a further steepening of the spectrum after the knee [14, 52]. The cause of this feature is unclear. One theory is that it is due to the Galactic acceleration limit of nuclei heavier than iron, thereby predicting a heavy composition at this energy [107]. Another theory poses the feature is a signature of the transition between Galactic and extragalactic CRs [97, 106], this theory predicting a light composition. Determining the composition at the second knee can eliminate conflicting models.

The spectral feature referred to as the ankle is observed at an energy of $\approx 5 \times 10^{18}$ eV [108, 109]. The approximate value of γ here is about 2.7, with data from different experiments suggesting slightly varying values. Some observed values of γ are 2.68 ± 0.01, 2.68 ± 0.04 and 2.81 ± 0.03, from the Pierre Auger Observatory [36], Telescope Array [110] and HiRes [109] respectively. The ankle is typically considered to be the point at which the extragalactic CR flux dominates over the Galactic flux [42]. This is theorised because it is considered unlikely that there are galactic acceleration sites capable of producing cosmic rays at these energies, and regardless of their acceleration site, galactic cosmic rays of ankle energies are expected to diffuse out of the galaxy resulting in large scale anisotropies in the measured arrival directions at Earth. An opposing theory assumes an extragalactic flux of protons, resulting in a predicted transition to occur at an energy between 10^{17} eV to 10^{18} eV. The theory, referred to as the 'dip model',

predicts the second knee and ankle result from the energy spectrum flattening on either side of the dip, which has been observed by detectors, as a result of pair production of extragalactic protons interacting with the cosmic microwave background (CMB) [52, 97, 106].

Above the ankle, it is expected that there is a suppression of the flux above energies of $\sim 6 \times 10^{19}$ eV, due to protons undergoing pion photoproduction at energies around 10^{20} eV (see Section 2.2.3) and nuclei undergoing photo-disintegration and pair production (see Section 2.2.2).

A review of models explaining the spectra ranging from below the knee to above the ankle can be found in [14].

In 2005 the HiRes Collaboration reported on the energy spectrum observed using monocular fluorescence detectors. A spectral break was measured at a energy of $\log_{10}(E/\text{eV}) = 19.79 \pm 0.09$, above which the spectral index was $\gamma = 5.2 \pm 1.3$ [111]. The existence of a suppression above the break energy is consistent with the results, evidenced by a single power law being unable to describe the spectrum above the ankle. In 2008 the HiRes Collaboration published another analysis which found a spectral break at an energy of $\log_{10}(E/\text{eV}) = 19.75 \pm 0.04$, above which the spectral index was $\gamma = 5.1 \pm 0.7$. From comparisons of their measurements with theoretical models, the HiRes collaboration concluded a GZK cut-off exists in their observed spectrum, with a significance of 5σ [112]. Another study published in 2009 by the HiRes collaboration reported on the analysis of stereo data containing fewer statistics but better energy resolution. The results from the stereo data analysis were consistent with the monocular data analysis [46, 52, 109].

In 2008 the Pierre Auger Collaboration published their findings from the analysis of ground array data which adopted an energy scale that was calibrated using their fluorescence detectors. The hypothesis of a single power law describing the spectrum above 4×10^{18} eV was rejected with a significance greater than 6σ [113]. The spectral index measured was $\gamma = 4.2 \pm 0.4(stat) \pm 0.06(syst)$ above the break at 4×10^{19} eV . The systematic uncertainties in the energy scales of the Pierre Auger Observatory and HiRes data were 22% and 17%, respectively. The break energies measured by the Pierre Auger Observatory and the HiRes Collaboration are compatible within these energy scale systematic uncertainties.

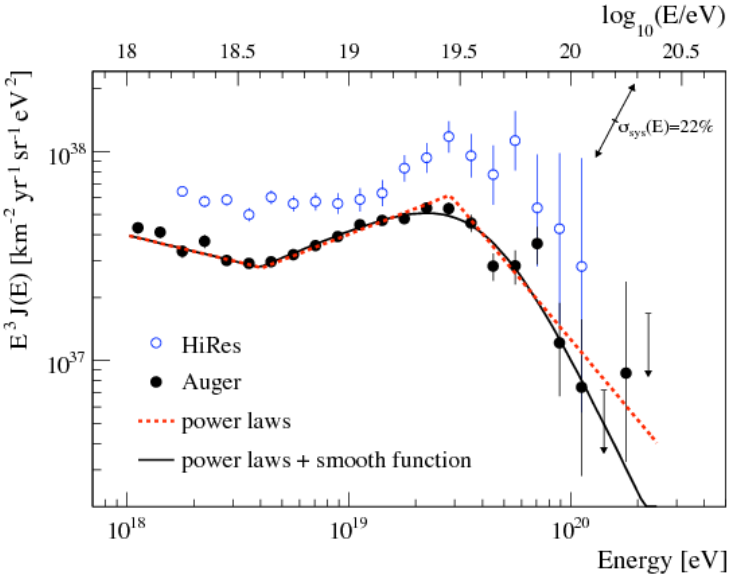


Figure 2.5: The Auger combined spectrum (hybrid and SD events) and the HiRes stereo spectrum. The flux has been multiplied by E^3 so that the ankle and energy suppression are clearly displayed. The Auger systematic uncertainty of the flux, scaled by E^3, is indicated by arrows. Within systematic uncertainties, the observations of the two experiments are consistent. From [42].

In 2010, after analysing fluorescence and ground array data, the Pierre Auger Observatory reported that above the spectral break at $\log_{10}(E/\text{eV}) = 19.46 \pm 0.03$, the spectral index was

$\gamma = 4.3 \pm 0.2$. These findings were consistent with previous results, and the existence of a flux suppression in the combined data was supported at a significance of at least 20σ [42]. Figure 2.5 displays the observed Auger and HiRes spectrums.

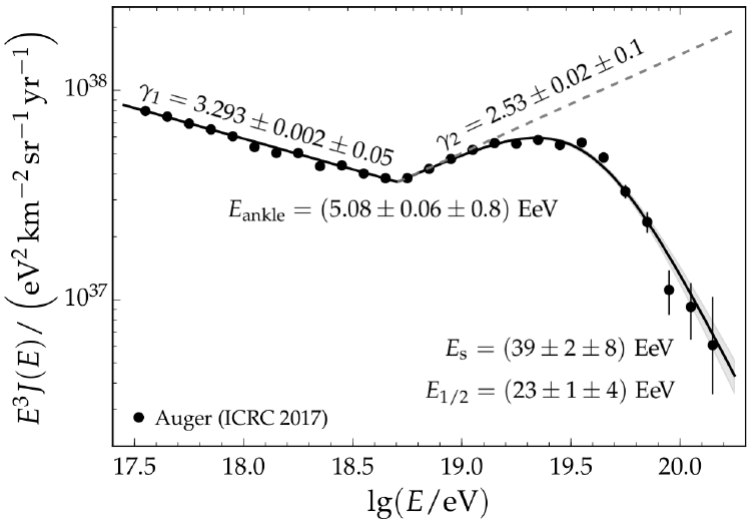


Figure 2.6: The UHECR energy spectrum reported by the Pierre Auger Collaboration in 2017. See the text for the functional form fitted to the spectrum. From [108].

Current results continue to illustrate a spectral cut-off above the GZK threshold. Shown in Figure 2.6 is the energy spectrum published by the Pierre Auger Observatory in 2017 [108]. The spectrum was fitted with the following function:

$$\begin{aligned} J_{\mathrm{unf}}(E; E \leq E_{\mathrm{ankle}}) &= J_0 \left(\frac{E}{E_{\mathrm{ankle}}}\right)^{-\gamma_1} \\ J_{\mathrm{unf}}(E; E > E_{\mathrm{ankle}}) &= J_0 \left(\frac{E}{E_{\mathrm{ankle}}}\right)^{-\gamma_2} \left[1 + \left(\frac{E_{\mathrm{ankle}}}{E_S}\right)^{\Delta\gamma}\right] \left[1 + \left(\frac{E}{E_S}\right)^{\Delta\gamma}\right]^{-1} \end{aligned} \tag{2.16}$$

where E_{ankle} is the energy of the ankle and E_S is the energy of the suppression. The energy $E_{1/2}$ at which the integral spectrum drops by a factor of two below what would be the expected with no steepening is also denoted in Figure 2.6.

It is important to realise that observations of a cut-off, although consistent with the predictions from GZK interactions, may be evidence of other phenomena such as the source spectrum changing [42]. Measurements of the cosmic ray mass composition at these energies can help explain the nature of this suppression.

A comparison of the energy spectrums measured by recent experiments is shown Figure 2.7. The flux measured by the Telescope Array (TA) appears larger than the flux measured by the Pierre Auger Observatory. A joint TA and Auger analysis of their measured energy spectrums was conducted, finding that within systematic uncertainties, the energy spectrums of TA and Auger were consistent below 2×10^{19} eV, but above this energy the flux measured by TA is significantly larger [114]. Given TA and Auger are in the Northern and Southern hemispheres respectively, this discrepancy could be a signature of anisotropy in the arrival directions of the UHECRs.

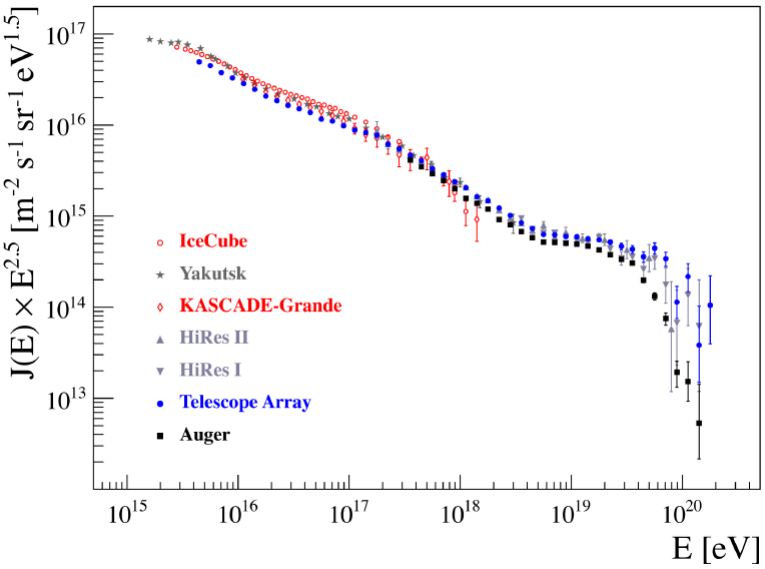


Figure 2.7: The energy spectrum measured by IceCube [115], Yakutsk [116], KASCADE-Grande [117], HiRes I and HiRes II [112], Telescope Array [118] and Auger [119]. From [114].

2.4 Composition of Cosmic Rays

Knowledge of the mass composition of cosmic rays as a function of energy is important for understanding their acceleration mechanisms, possible sources and arrival direction distribution.

Low energy cosmic rays (up to $\sim 10^{15}$ eV) can be measured directly by detectors with small collecting areas, as the flux of these cosmic rays is relatively large. These detectors, located on satellites or balloons, have measured the mass composition of low energy cosmic rays to consist of mainly protons ($\sim 87\%$) and alpha particles ($\sim 12\%$), with various heavier nuclei comprising the remaining 1% [120].

Figure 2.8 compares the relative abundances of cosmic rays and solar system elements. There are two groups of elements that are prolific in the cosmic ray composition compared to the solar system elements. These two element groups are Li, Be, B and Sc, Ti, V, Cr, Mn. This is due to spallation, a process in which propagating heavier nuclei collide with the interstellar medium and fragment into smaller nuclei. The Fe nuclei fragment into Sc, Ti, V, Cr and Mn, whereas C and O nuclei fragment into Li, Be and B.

The composition of higher energy cosmic rays is determined by measuring composition sensitive EAS parameters and comparing these measurements with predictions from hadronic interaction models. Mass composition studies at ultra high energies are inherently uncertain due to their reliance on particle interaction predictions at energies which cannot currently be produced by particle accelerator experiments.

From basic principles, a general prediction can be made of the cosmic ray composition. A cosmic ray species must be magnetically confined if it is to be accelerated in a source region, assuming the 'bottom-up' model for CR production (see Section 2.1.1.1). It follows that the maximum energy of a particle with charge Z is approximately Z times a proton's maximum energy. Thus, with increasing energy, the maximum energy of each species is exceeded, resulting in an increasing average mass of the cosmic ray flux. This trend continues until the extragalactic component of the flux starts to become significant. If the extragalatic flux consists of a significant abundance of protons, the emergence of the extragalactic component would result in a transition where the average mass decreases towards lighter masses [14, 104]. The transition is expected to occur between 10^{17} eV to 10^{18} eV, according to models like the pair production 'dip' model

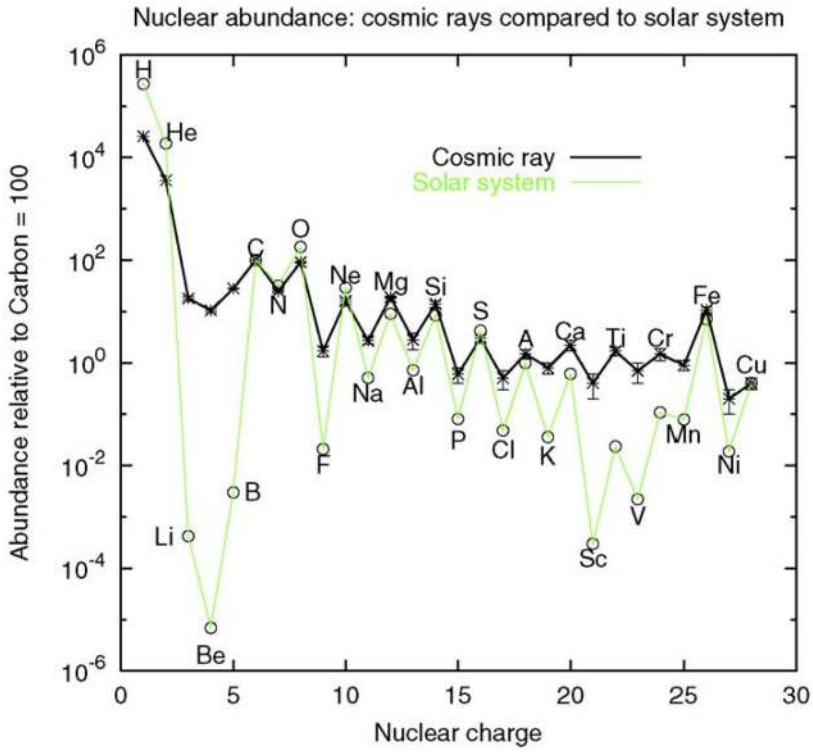


Figure 2.8: The relative abundances of solar [121] and cosmic ray material [9,122–124], normalised to carbon, for $E < 10^{14}$ eV cosmic rays. From [104].

[97,106]. Where the extragalactic flux dominates ($E > 5 \times 10^{18}$ eV), if there are no sources nearby, the primary flux is predicted to consist mainly of protons and iron, as intermediate-mass particles propagating to Earth will disintegrate relatively quickly [14]. Cosmic ray propagation models show a variety of source compositions are compatible with the energy spectrum observed [102].

Composition measurements between the knee and ankle find that the proportion of light elements decreases with increasing energy. The cut-off energies of the measured spectra of individual species are proportional to the charge of the species, a finding which is compatible with a rigidity dependence of source acceleration limits or CR confinement in the Galactic magnetic field (GMF) [103, 125].

Ultra high energy cosmic ray detectors such as the Pierre Auger Observatory measure a number of different EAS observables, some of which are sensitive to the mass of the primary cosmic ray. The following sections contain a discussion of some of these mass sensitive observables and the results of their study.

2.4.1 Optical Detector Observables

Discussed in Section 1.2.1, the depth of shower maximum, $X_{\max}$, is sensitive to the primary mass of the cosmic ray. Due to fluctuations in the first hadronic interactions of a primary particle with the atmosphere and fluctuations in the shower evolution, the primary mass can not be determined on an event-by-event basis by measuring $X_{\max}$. Instead, the primary mass is estimated by examining the energy binned distributions of $X_{\max}$. The mean depth of shower maximum of a distribution, $\langle X_{\max} \rangle$, is used to determine whether the measured cosmic rays consist of predominantly light, mixed or heavy elements. The change in $\langle X_{\max} \rangle$ per decade of energy is referred to as the elongation rate. A changing elongation rate indicates a changing composition.

RMS($X_{\max}$) - the root mean square of the $X_{\max}$ distribution of an energy bin (generally

denoted as $\sigma(X_{\max})$) - quantifies the degree of shower-to-shower fluctuations of the depth of maximum. The superposition theory (see Section 1.2.2) predicts that a primary nucleus of energy E_0 and A nucleons, is equivalent to A protons with energy E_0/A, thus showers from heavier nuclei will have fewer shower to shower fluctuations than proton showers of the same energy [17]. Consequently, $\sigma(X_{\max})$ is a useful quantity for extracting composition information. However, changes in the atmosphere can cause significant fluctuations in shower development, fluctuations which could be incorrectly credited to intrinsic $X_{\max}$ fluctuations. Atmospheric density profile information is required to account for the effect of atmospheric fluctuations, so that the measurement resolution of $X_{\max}$ is small compared to $\sigma(X_{\max})$.

The $\sigma(X_{\max})$ is sensitive to the average composition, in addition to the composition dispersion. Depending on the mass composition mixture and the separation between the components of the mixture, a value of $\langle X_{\max}\rangle$ can be associated with a range of different $\sigma(X_{\max})$ values. Similarly, an $\sigma(X_{\max})$ value can be associated with different $\langle X_{\max}\rangle$ values. Consequently, a significant shift in $\sigma(X_{\max})$ will only be observed if there is a considerable reduction of the relative proton abundance.

There is significant variation in the particle interaction predictions of the different hadronic interaction models used in mass composition studies. As a result, different hadronic interaction models predict different $X_{\max}$ distributions for a particular primary. At a given energy, a larger cross section prediction corresponds to a shorter interaction length, resulting in a shallower $X_{\max}$ (and consequently $\langle X_{\max}\rangle$) and smaller $\sigma(X_{\max})$. A larger inelasticity prediction means the primary particle retains a smaller fraction of its energy in an interaction, resulting in a smaller $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$. A larger multiplicity corresponds to a greater number of particles being produced at each interaction and therefore a smaller $\langle X_{\max}\rangle$. The disagreement between the model predictions results in a mass composition interpretation between models that is often in disagreement.

Figure 2.9 provides a summary of the historical measurements of $\langle X_{\max}\rangle$ by different experiments. There is considerable spread between the data of different experiments, which in addition to the spread between different hadronic interaction models, results in various interpretations of the CR mass. A common interpretation is that below the knee ($E \approx 10^{15.5}\,\text{eV}$), the CR primary mass consists of a mixed composition. Above the knee there is a heavy composition up to $E \approx 10^{16.5}\,\text{eV}$, where the primary mass then transitions back to a mixed composition through to the ankle ($5 \times 10^{18}\,\text{eV}$). The transition above the knee may be due to supernova remnants (SNR) accelerating cosmic rays less efficiently as the Larmor radius of the particles reaches the SNR radius. In a SNR, the maximum energy a CR can be accelerated to is proportional to its charge, consequently iron nuclei will reach greater energies than lighter nuclei.

In 1993, an $X_{\max}$ analysis by Fly's Eye suggested a heavy composition below the ankle at $10^{17.5}\,\text{eV}$, transitioning to a lighter composition at energies up to $10^{19}\,\text{eV}$ [144]. Combined with the Fly's Eye anisotropy and energy spectrum measurements, the findings indicated that at the ankle the cosmic ray flux changes from predominantly Galactic to extragalactic.

In 2005 the HiRes collaboration published an analysis of $\langle X_{\max}\rangle$ data between $10^{18}\,\text{eV}$ and $10^{19.4}\,\text{eV}$. The $\langle X_{\max}\rangle$ was consistent with a light composition, and the elongation rate was calculated to be $54.5\pm6.5(stat)\pm4.5(syst)$ g/cm^2 per decade, suggesting a composition that is constant or slowly-changing. An elongation rate of approximately $90\,\text{g}\,\text{cm}^{-2}$ /decade between $10^{17}\,\text{eV}$ to $10^{18}\,\text{eV}$ was measured by the HiRes-MIA detector. These measurements were compatible with a heavy tail of the galactic flux, becoming dominated by a light extragalactic flux [17, 133]. The HiRes results, within errors, were consistent with the Fly's Eye findings [52]. In 2010, HiRes measured a constant elongation rate above 1.6 EeV of $47.9 \pm 6.0(stat) \pm 3.2(syst)$ g/cm^2/decade. Illustrated in Figure 2.10, the data and hadronic models suggested an almost pure proton composition [143].

In 2014 the Pierre Auger Collaboration published the results of their latest $X_{\max}$ analysis [145], shown in Figure 2.11. Below $10^{18.27}\,\text{eV}$, a $\langle X_{\max}\rangle$ elongation rate of $86.4\pm5.0(stat)^{+3.8}_{-3.2}(syst)$ g/cm^2

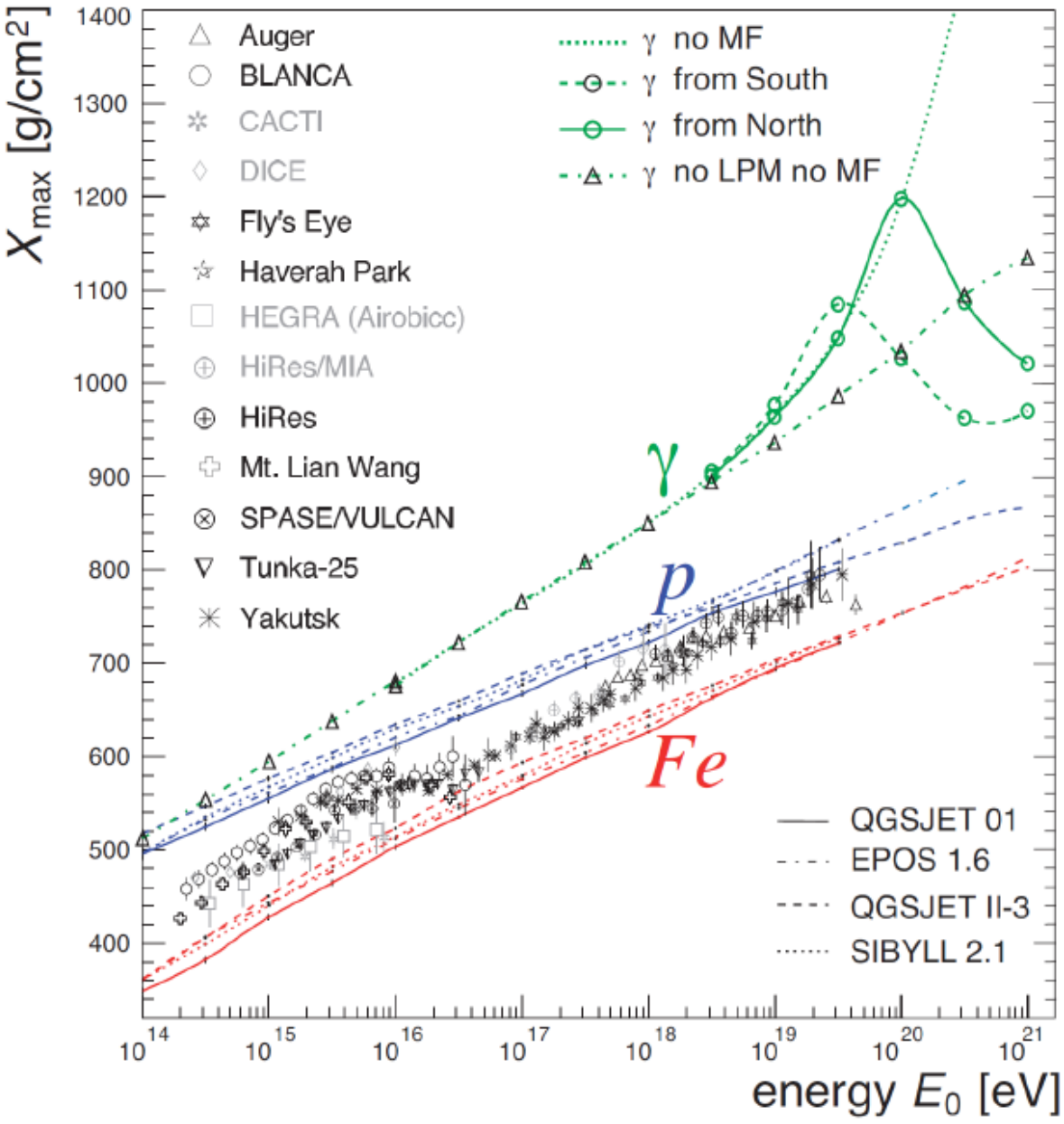


Figure 2.9: The average $\langle X_{\max} \rangle$ as a function of primary energy measured by the following experiments: Auger [126], BLANCA [127], CACTI [128], DICE [129], Fly's Eye [130], Haverah Park [131], HEGRA [132], HiRes/MIA [133], HiRes [17], Mt. Lian Wang [134], SPASE/VOLCAN [135], Tunka-25 [136] and Yakutsk [137]. Included is the expected average depth of shower maximum for primary photons (green), protons (blue) and iron (red) according to the following hadronic interaction models: QGSJET 01 [138], EPOS 1.6 [139], QGSJET II-3 [140] and Sibyll 2.1 [141]. From [142].

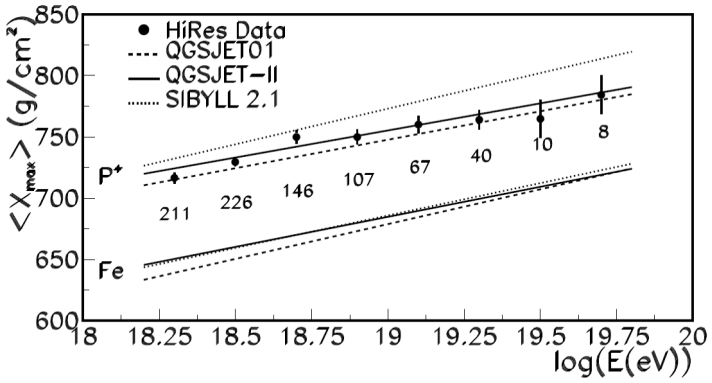


Figure 2.10: The $\langle X_{\max} \rangle$ measured by the HiRes detector. The expected values of $\langle X_{\max} \rangle$ for a pure proton or pure iron composition according to different hadronic interaction models is shown. Included is the number of events in each energy bin. From [143].

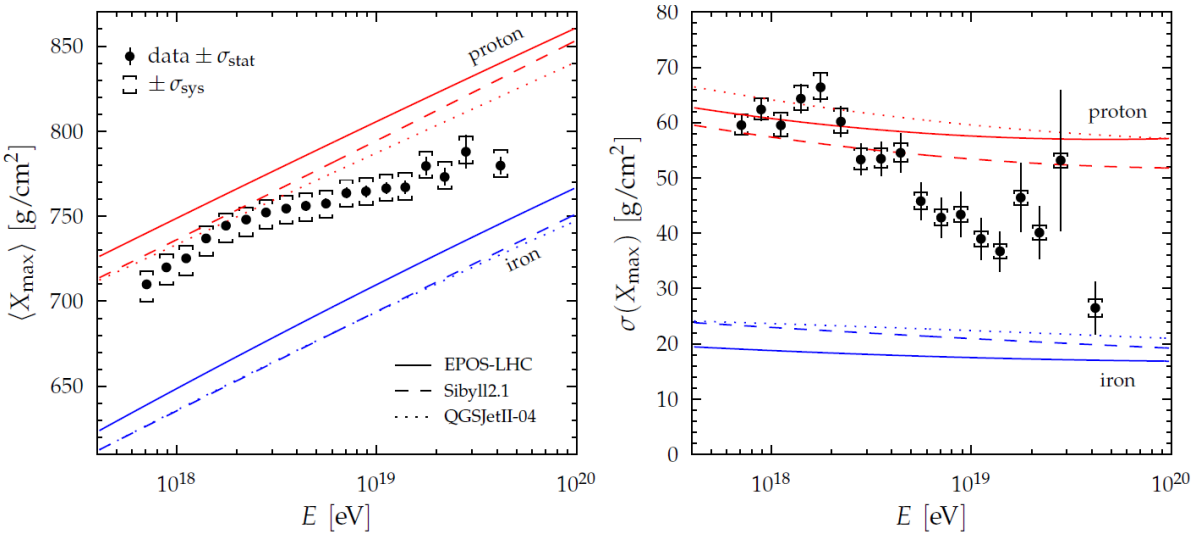


Figure 2.11: The $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$ measured by the Pierre Auger Observatory, published in 2014. The expected values for a pure proton (red lines) or pure iron (blue lines) composition according to different hadronic interaction models is shown. From [145].

/decade was fitted. An elongation rate of $26.4 \pm 2.5(stat)^{+7.0}_{-1.9}(syst)$ g/cm^2/decade was fitted above this energy. The $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$ measurements suggested that the cosmic ray composition transitions to lighter nuclei up to $\sim 10^{18.3}$ eV, upon which the trend reverses and the fraction of heavy nuclei gradually increases up to energies of $10^{19.6}$ eV. The $X_{\max}$ analysis published in 2011 by the Pierre Auger Collaboration suggested a similar transition [20].

In the 2014 publication, the $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$ measurements were converted into the first two moments of the $\ln A$ distribution (A is the cosmic ray atomic mass), providing a more quantitative examination of the mass composition and a test of the validity of the hadronic interaction models [146]. The results are shown in Figure 2.12. The QGSJetII-04 interpretation is disfavoured as it leads to unphysical variances of $V(\ln A) < 0$. The EPOS-LHC and Sibyll2.1 models indicate that the mass composition consists of a range of masses at low energies, but above $10^{18.7}$ eV the variance of $\ln A$ is close to zero, suggesting the mass composition is dominated by a single nuclei. The mass composition transition suggested by the $\ln A$ moments is consistent with

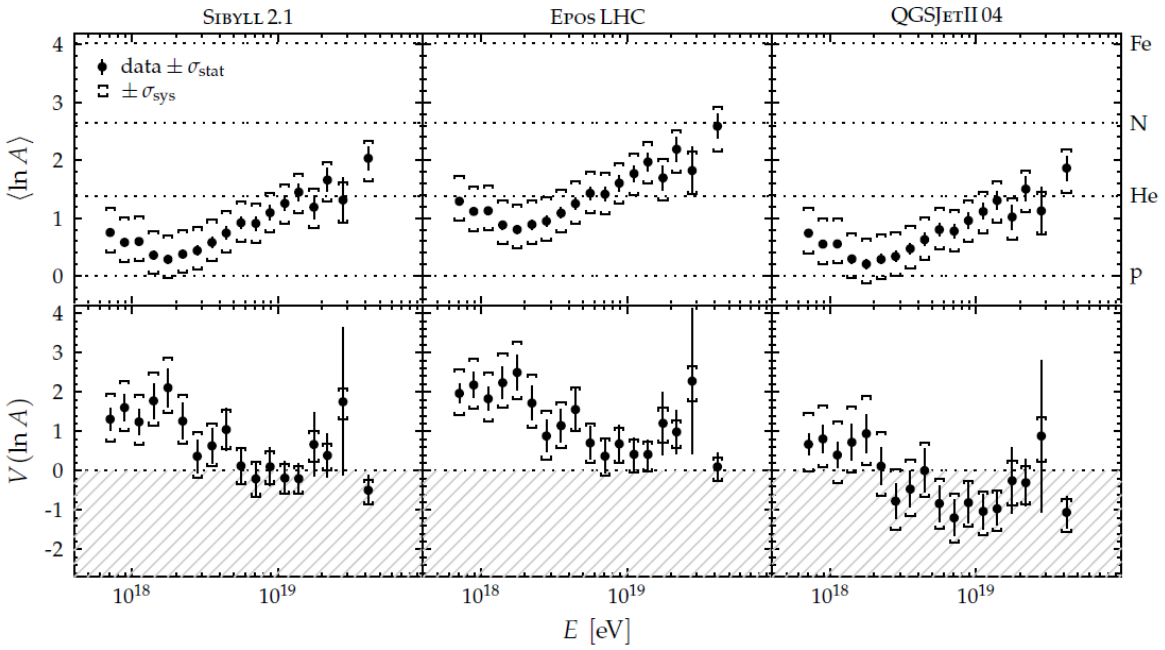


Figure 2.12: The $\langle \ln A \rangle$ and $V(\ln A)$ estimated from the first two $X_{\max}$ moments, using different hadronic interaction models. From [145].

that of the $X_{\max}$ moments. Despite the hadronic model differences, the three models suggest a similar trend with energy.

A fit of the full $X_{\max}$ distributions was performed in 2014 to estimate the fractions of groups of nuclei contributing to the cosmic ray flux [147]. Figure 2.13 shows the fractions of proton, helium, nitrogen and iron which resulted in the best fit of the $X_{\max}$ distributions. The fraction fits do not support a proton dominated mass composition, or a large iron nuclei contribution.

The results of the $X_{\max}$ analysis published by the Pierre Auger Collaboration in 2017 [148] are consistent with previous Pierre Auger Collaboration findings. The energy range of this data set begins at $10^{17.2}$ eV, the analysis indicating that the CR mass transitions to lighter nuclei up to $\sim 10^{18.3}$ eV. The 2017 analysis utilised a version of the fit method detailed in this thesis, and the $X_{\max}$ hadronic model parameterisations presented in this thesis (see Chapter 4).

The Pierre Auger Collaboration findings conflict with those of HiRes. Below the ankle however, both experiments indicate a light composition, which is consistent with the pair production dip model. Above the ankle, the Pierre Auger Observatory has measured a flatter elongation rate, suggesting a transition from a light or mixed composition towards heavier primaries, whereas HiRes has measured an elongation rate consistent with a pure proton composition. The main cause of the conflicting results may be the different analysis techniques. The Pierre Auger Observatory fits a broken line to the elongation rate as the data are not adequately described by a single line. HiRes applies a single line fit. Additionally, the Pierre Auger Observatory applies cuts to remove triggering and selection biases in the measured data, before comparing the measurements to Monte Carlo data. HiRes on the other hand applies cuts to only the tails of their distributions, comparing their biased data to biased Monte Carlo data that was passed through their detector simulation. Consequently, the $X_{\max}$ results from these different experiments can not be directly compared. Furthermore, the two experiments view different parts of the sky, with HiRes in the Northern hemisphere and the Pierre Auger Observatory in the Southern hemisphere.

Measurements of $X_{\max}$ by the Telescope Array (TA) also indicate a light composition [149], illustrated in Figure 2.14. To compare the measurements of the Pierre Auger Observatory and TA, a simulated hypothetical composition mixture which represents a good fit of the Pierre Auger Observatory $X_{\max}$ distributions (which are intended to be unbiased), was passed through the TA

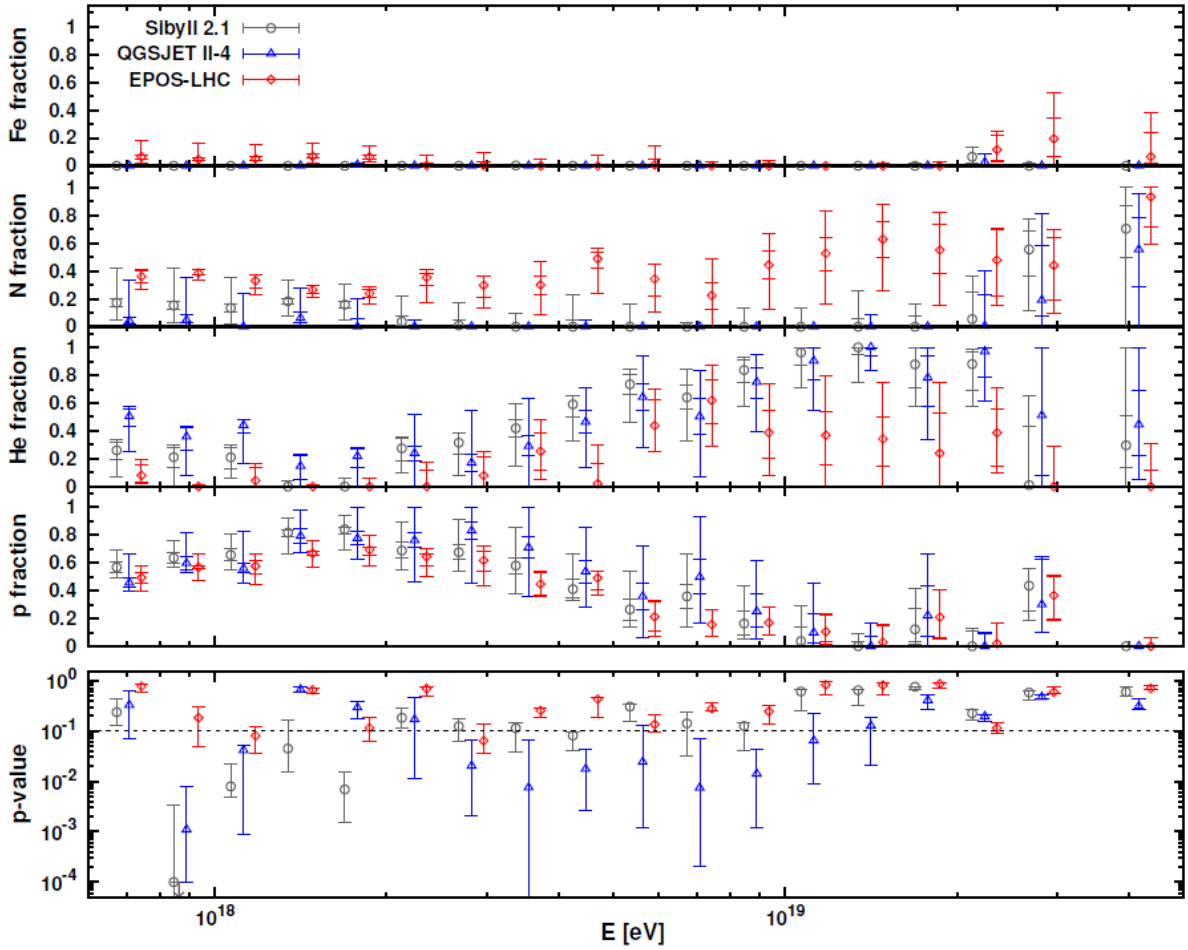


Figure 2.13: The fractions of proton, helium, nitrogen and iron fitted to the $X_{\max}$ distributions in [145]. The upper panels show the species fractions and the lower panel indicates the goodness-of-fit. From [147].

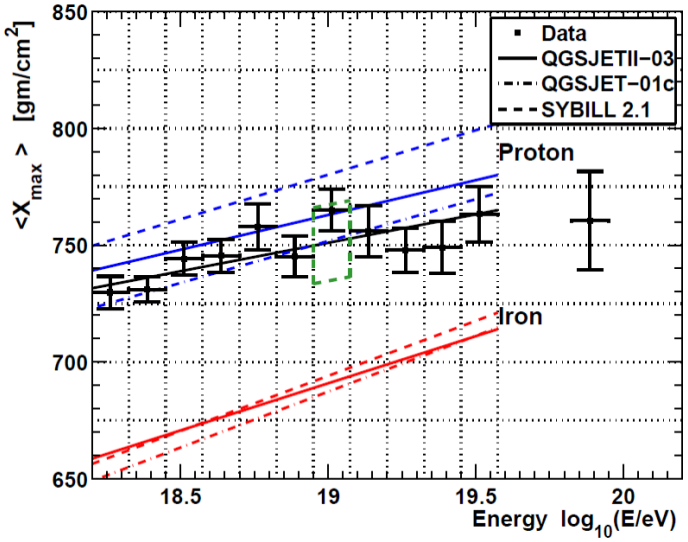


Figure 2.14: The $\langle X_{\max} \rangle$ measured by TA. The green hashed box indicates the total systematic error on $\langle X_{\max} \rangle$. From [149].

detector simulation and analysis chain. Comparisons of the Auger $X_{\max}$ distribution passed through the TA chain, and the original $X_{\max}$ distribution measured by TA, concluded that the $X_{\max}$ data measured by TA is compatible with the mixed composition which best describes the Auger $X_{\max}$ data [150].

2.4.2 Surface Detector Observables

Optical detectors have a limited up-time compared to surface detectors which can operate all day, consequently mass composition studies focusing on $X_{\max}$ are statistically limited compared to studies which examine surface detector observables. Surface detector measurements can also be used to perform independent cross checks. However, compared to $X_{\max}$, obtaining mass composition information from surface detector observables requires a greater reliance on hadronic interaction models, which introduces greater uncertainty in the results.

2.4.2.1 Muon content

At a given primary energy, heavier nuclei produce showers with a greater number of muons than lighter nuclei (see Section 1.2.2). Shower to shower fluctuations, detector sampling, and the varying predictions of different hadronic models at energies much greater than those achieved by accelerator experiments, complicates the interpretation of the muon content. Muon content studies can be used to estimate the CR mass composition and the validity of hadronic interaction models.

AGASA analysed shower muon densities and found that between $10^{17.5}$ eV and 10^{19} eV, the average mass decreased towards a dominant light composition at 10^{19} eV. An upper limit of 40% on the iron fraction above 10^{19} eV with a confidence level of 90% was determined, but this result relied heavily on QGSJET98. The AGASA findings are consistent with the Fly's Eye composition results [38, 151]. The muon array CASA-MIA found that the composition became lighter with increasing energy between 10^{17} to 10^{18} eV [133].

The water-Čerenkov detectors of the Pierre Auger Observatory do not distinguish between different particle types. Instead, the Pierre Auger Observatory measures muons indirectly. Some methods for measuring the muon content are described below [152]:

- Measuring highly inclined showers. Showers arriving at the ground with zenith angles greater than 60° are categorised as highly inclined showers. As these showers have propagated through a large atmospheric depth, much of the EM component has been absorbed, leaving predominantly secondary muons at the ground. Examining highly inclined showers can provide an almost direct measurement of the muon number. The muon number can be measured using universality principles. For a given shower direction, the shape of the muon distribution is universal. Only the normalisation of the muon distribution depends on the shower energy and primary mass. Different hadronic interaction models and air shower simulation software packages consistently reproduce the lateral shape of the muon density.

  The muon density $\rho_\mu(\vec{r})$ can be written as:

  $$\rho_\mu(\vec{r}) = N_{19}\rho_{\mu,19}(\vec{r};\theta,\phi) \tag{2.17}$$

  where N_{19} is the scale factor of a particular event with respect to a reference muon distribution $\rho_{\mu,19}$. The reference is typically chosen as the average muon density for 10^{19} eV primary protons of QGSJetII-03 [153].

- Examining the structure of surface detector signals as a function of time. The SD signal amplitude of a typical EM particle is much smaller than that of a muon particle. Additionally, the number of EM particles is typically an order of magnitude larger. The result is large muon signal spikes against a smooth background EM signal. The muon content can be estimated by separating these muon and EM signals.

- Examining hybrid events. For each event measured by both the SD and FD, the FD longitudinal profile of simulated showers of various compositions is compared to the FD longitudinal profile of the real event. The best matching simulated showers are then re-simulated with a lower thinning level to provide a more detailed simulation of the particles reaching the ground. The ground signal of these simulated showers is then compared to the ground signal of the real event. Regardless of the hadronic model applied, the model predictions do not match the observed signals. The simulated ground signal is rescaled by an independent factor R_E due to the energy scale uncertainty, and the signal from hadronically produced muons is rescaled by a factor of R_μ. Figure 2.15 displays the values of these rescaling factors that best reproduce the Pierre Auger Observatory hybrid data. Within systematic uncertainties, R_E is compatible with 1, suggesting the predicted energy is consistent with the observed energy. Depending on the assumed model and composition, the simulated muon size requires a rescaling of 1.3 to 1.6 to match the observed muon size, suggesting a muon deficit in the EPOS-LHC and QGSJetII-04 predictions.

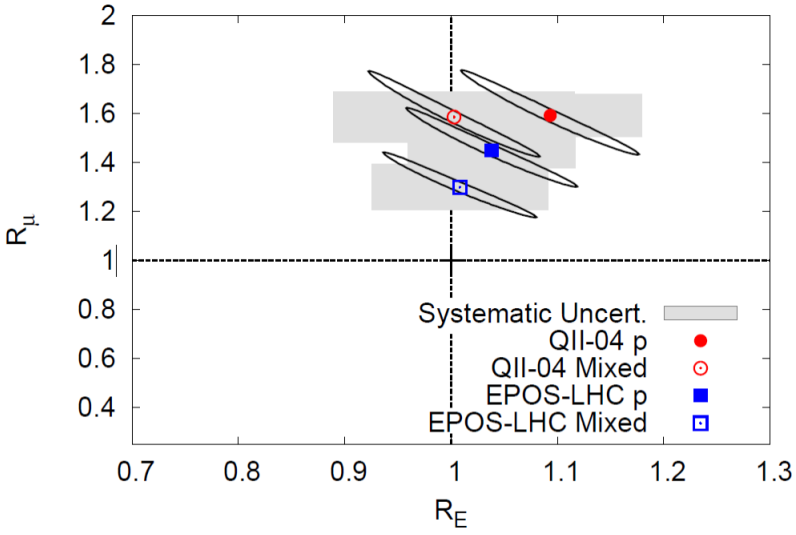


Figure 2.15: The energy (R_E) and muon (R_μ) rescaling factors that best represent hybrid data at 10^{19} eV. From [152].

2.4.2.2 Risetime $t_{1/2}$

In an EAS, the propagation of muons is almost straight, whereas the EM component is deflected by multiple Coulomb scattering. Scattering deflections in addition to other factors such as geometrical effects, means that these shower particles do not arrive simultaneously at a surface detector, thus the measured signal trace generally consists of a short duration early component dominated by the muon signal, and a long duration trace that is the electromagnetic signal. This feature is quantified by the risetime, $t_{1/2}$, the time for the integrated signal for each surface detector to increase from 10% to 50% of its final value.

The $t_{1/2}$ fluctuations of showers were discovered to be larger than expected from sampling uncertainties alone, suggesting that $t_{1/2}$ is sensitive to the CR primary mass [154]. Following this finding, the risetime was calibrated with $X_{\max}$, and using energies up to 10^{18} eV an elongation rate using the risetime was calculated [155]. An analysis of $t_{1/2}$ for events detected at Haverah Park with an average energy of 10^{19} eV indicated an iron fraction at this energy of approximately 80% [156]. The Pierre Auger Observatory uses $t_{1/2}$ as a mass composition indicator, applying a technique that studies the asymmetry of the risetime (see Section 2.4.2.3 and Figure 2.17 of Section 2.4.2.5), and a technique that corrects for the risetime asymmetry (see Figure 2.18 of

Section 2.4.2.5).

2.4.2.3 Azimuthal Asymmetry

Inclined showers with a difference in the particle distribution between the side underneath the shower axis and the side opposite the shower axis, have an azimuthal asymmetry. Figure 2.16 shows how a particle density azimuthal asymmetry between two sides of a shower can eventuate. Two parameters which can have an azimuthal asymmetry are $t_{1/2}$ and S_{1000} (the signal that would be measured 1000 m from the shower core).

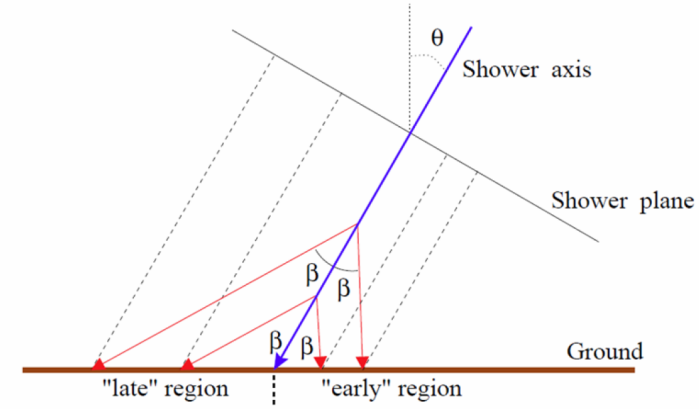


Figure 2.16: Diagram illustrating how a geometric particle density asymmetry appears. Particles propagate symmetrically from the shower axis with angle β. From [157].

The multiple scattering of electrons in air results in an EM component signal that is spread over a longer duration than the muonic component signal. Thus, the risetime of the surface detector signals is a proxy for the muon-to-electron ratio in a surface detector station. The zenith angle at which the corresponding risetime asymmetry is at a maximum, θ_{max}, can be used to study CR mass composition [158].

2.4.2.4 Lateral Distribution Function

A lateral distribution function (LDF) can be used to describe the decrease of air shower particle density as a function of distance from the shower axis (see Section 1.3.1). The slope of the LDF is correlated with $X_{\max}$, thus the LDF slope can provide mass composition information. Compared to shallower showers, showers which develop deeper into the atmosphere are at an earlier stage of development at ground level. Consequently, a greater number of the EM particles produced close to the core will be able to reach ground level before being attenuated, resulting in a larger LDF slope.

2.4.2.5 Other Observables

Other observables used for mass composition studies include:

- Muon production depth. This is measured by using the signal traces of surface detectors far from the shower core, for large zenith angle showers [20].
- Radius of curvature of a shower. This is measured by applying a curved shower front fit [72].
- Relative muon and electron numbers. These observables are used by KASCADE [159, 160].

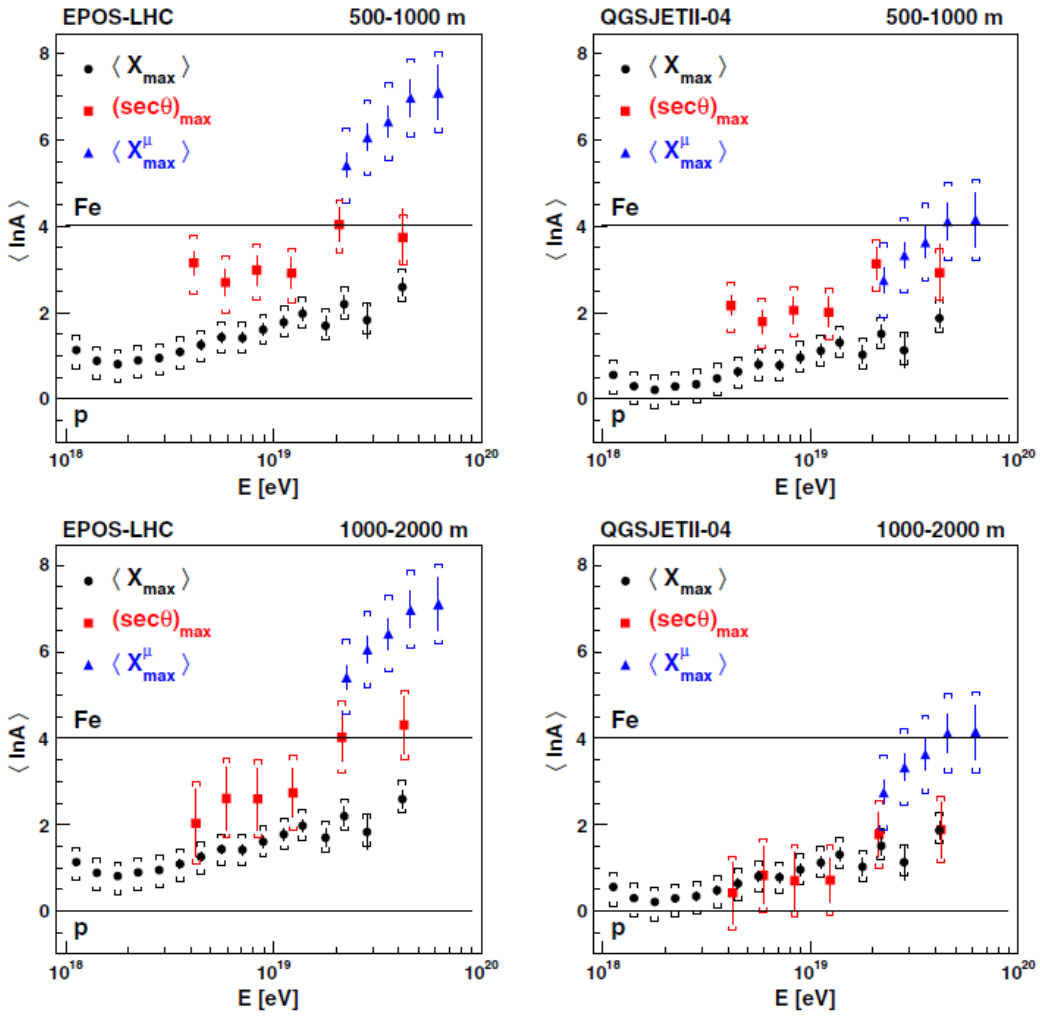


Figure 2.17: $\langle \ln A \rangle$ as a function of energy predicted by EPOS-LHC and QGSJetII-04. The asymmetry analysis results [158] (see $(\sec\theta)_{\max}$) for both distance intervals to the core are displayed, and compared with the elongation curve in [145] (see $\langle X_{\max} \rangle$) and the muon production depth (MPD) method in [161] (see $\langle X^{\mu}_{\max} \rangle$). From [158].

Figure 2.17 displays in terms of $\langle \ln A \rangle$ the results of the risetime asymmetry analysis published by the Pierre Auger Collaboration in 2016 [158], compared with the 2014 Auger analysis of $X_{\max}$ [145] and the 2014 muon production depth (MPD) analysis [161]. The three analyses are sensitive to different types of hadronic interactions, and therefore possess different systematic uncertainties. The three measurements suggest the mass composition transitions from lighter to heavier nuclei with increasing energy. The risetime asymmetry analysis however shows that according to the QGSJetII-04 model, the mass composition depends on the distance of the stations from the core, which is unphysical. The MPD results indicate that the hadronic models utilised incorrectly describe the muon component of showers.

Figure 2.18, published by the Pierre Auger Collaboration in 2017 [164], displays in terms of $\langle \ln A \rangle$ the results of the latest FD $X_{\max}$ analysis [148] (utilising a version of the fit method detailed in this thesis), the risetime analysis (not cross-calibrated with the FD) which corrects for the risetime asymmetry [162], and the updated average muon production depth analysis [163].

Significant progress is being made towards using surface detector measurements for mass composition studies. Achieving precise measurements of the number of muons will allow the CR mass composition to be estimated on an event-by-event basis. The ability of the Pierre Auger

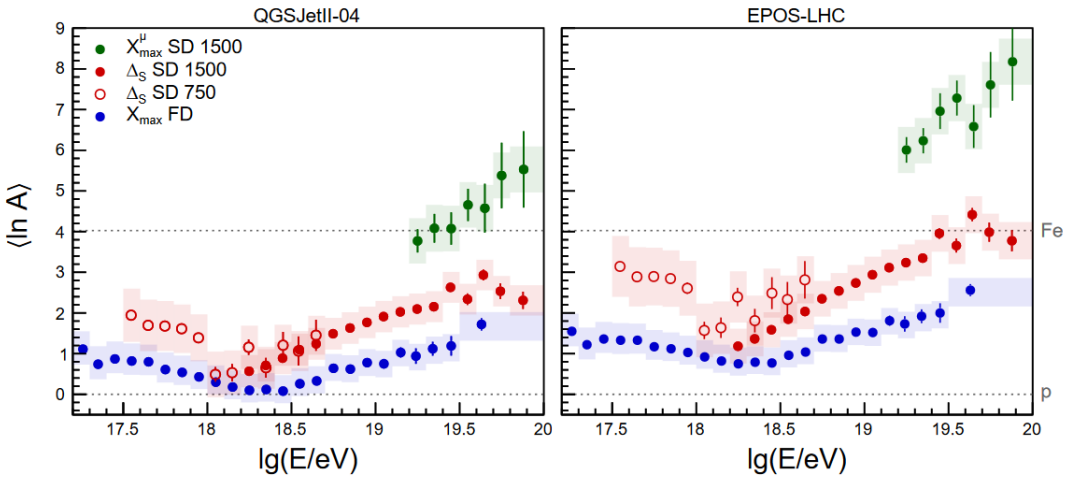


Figure 2.18: $\langle \ln A \rangle$ as a function of energy predicted by QGSJetII-04 and EPOS-LHC, according to FD $X_{\max}$ measurements (blue) [148], risetime measurements which correct for the risetime asymmetry (red) [162], and the updated measurement of the average muon production depth (green) [163]. The risetime measurements were obtained for two different SD spacings (750 m and 1500 m). From [164].

Observatory to measure the muon number will be enhanced by a planned surface detector upgrade (see Section 3.9.6.2).

As heavy primaries have a very high energy threshold for photo-pion production with the CMB (see Section 2.2.3), the observation of cosmogenic photons or neutrinos would support a light composition at ultra-high energies. The Pierre Auger Observatory has used the SD array to search for showers initiated by ultra-high energy neutrinos and photons, the results unsupportive of a light composition at ultra-high energies. A 2015 study discovered only 4 photon-like events and zero neutrino-like events, which is consistent with the predicted background [165].

2.5 Cosmic Ray Anisotropies

A deviation from a uniform distribution of CR arrival directions is referred to as an anisotropy. The discovery of an anisotropy can lead to CR source information. The deflection of the charged cosmic rays by magnetic fields, which are not well-known, as they propagate to Earth complicates anisotropy studies and their interpretation.

Depending on factors such as the number of events available, the detector resolution, and the physical expectations from models of CR production and propagation, different methods are used to examine CR data for anisotropies. These methods include harmonic analysis, autocorrelation analysis, and the correlation between CR data and a different data set such as a source distribution catalogue. These methods involve performing many statistical trials on the data. Many parameters may also be varied in the search, such as energy threshold, angular cut (for correlation studies), or the source catalogue that the data are compared to. By performing many trials, the likelihood of a positive detection due to a statistical fluctuation increases, instead of a positive detection due to a true signal contained in the data. If a true signal is detected in the data, estimating its statistical significance may be impossible, and therefore the results of the *a posteriori* analysis may be difficult to understand. When a large number of trials have been performed to extract a result, an identical *a priori* analysis of independent data should be performed to determine the statistical significance of the signal, if it reappears [28].

2.5.1 Recent Anisotropy Results at EeV energies

A discussion of lower energy anisotropy results may be found in [14]. This section details a number of findings for energies at and above 10^{18} eV. A popular target for CR anisotropy studies is the Galactic Centre (GC), which consists of a supermassive black hole that can potentially accelerate CRs to high energies. Furthermore, gamma ray observations of the GC suggest there may be high energy cosmic rays interacting with ambient matter (in [166] for example).

2.5.1.1 Large Scale Anisotropies

The transition of galactic to extragalactic UHECRs may produce an arrival direction dipole anisotropy [167]. Published in 2011, the Pierre Auger Observatory measured the phase and amplitude of the first harmonic modulation in the right ascension distribution. This was achieved by applying two techniques, the East-West method and Rayleigh analysis. The studies examined well-reconstructed events measured between 2004 and 2010 with zenith angles less than 60°, with corrections made due to weather effects. The Rayleigh analysis was applied to SD events above 10^{18} eV, and the East-West Method was applied to events below 10^{18} eV [167].

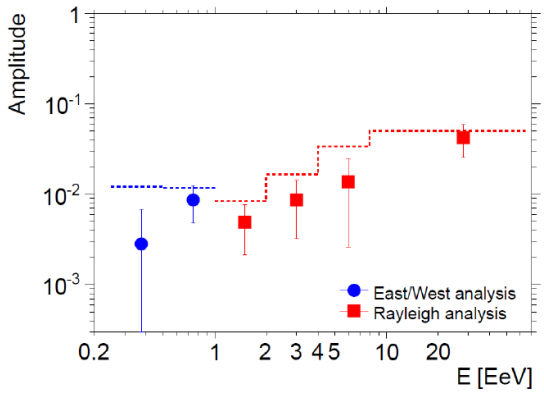


Figure 2.19: Amplitude of the first harmonic versus energy. The dashed line is the 99% C.L. upper bound on the amplitudes that could result from fluctuations of an isotropic distribution. From [168].

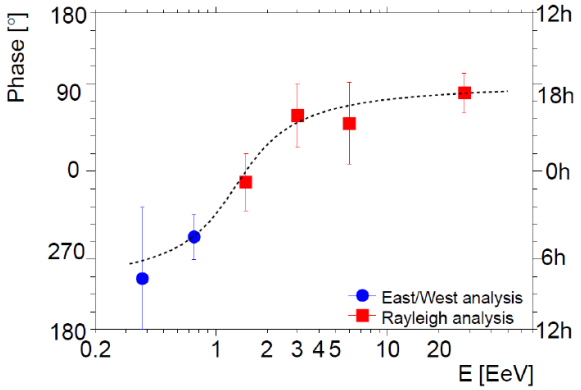


Figure 2.20: Phase of the first harmonic versus energy. The dashed line (from an empirical fit) is used in the likelihood ratio test. From [168].

If a dipole anisotropy exists, there is expected to be a modulation in the UHECR arrival distribution with a period of one sidereal day. The amplitude measurements in Figure 2.19 provide no evidence of an anisotropy. The phases are expected to be distributed randomly if the sources are isotropic, however the measurements in Figure 2.20 show a smooth transition with energy. With increasing energy, the phase transitions from $\approx 270°$ to $\approx 100°$, a transition away from the galactic centre, and consistent with a transition from galactic to extragalactic cosmic rays.

A likelihood ratio test is used to determine whether the phase measurements in adjacent energy intervals are better reproduced by a random arrival direction parent distribution, or an alternative dipolar parent distribution [167]. Applying the test to the measurements in Figure 2.20 results in a probability of $\sim 10^{-3}$ to accept the random distribution compared to the dipolar distribution. Confidence levels can not be determined as this was not an *a priori* search for this smooth transition in the phase measurements. To determine if the effect is real, an independent data set of similar statistics is needed.

In 2017, the Pierre Auger Collaboration published the discovery of a large-scale anisotropy for cosmic rays of $E \geq 8 \times 10^{18}$ eV, using a Rayleigh analysis on a dataset of 32187 events [169].

The anisotropy is described by a dipole of amplitude $6.5^{+1.3}_{-0.9}\%$ in the direction of right ascension $\alpha_d = 100 \pm 10^\circ$ and declination $\delta_d = -24^{+12}_{-13}{}^\circ$, detected with a level of significance greater than 5.2σ. The magnitude and direction of the anisotropy ($\sim 125^\circ$ from the Galactic centre) suggests an extragalactic origin for the highest-energy cosmic rays, as opposed to a Galactic origin. In the Galactic coordinate system, the cosmic ray dipole and the flux weighted dipole of the distribution of infrared-detected galaxies in the 2MRS (2 Micron All-Sky Redshift Survey) catalogue are separated by $\sim 55^\circ$. Taking into account CR deflections due to galactic magnetic fields reduces the disagreement. An energy range of $4 - 8$ EeV was also examined, discovering a dipole amplitude of $2.5^{+1.0}_{-0.7}\%$ at $\alpha_d = 80^\circ$ and $\delta_d = -75^\circ$ which was not statistically significant.

In 2015 the TA Collaboration confirmed the observation of a 20° excess in the northern sky of cosmic rays with energies above 57 EeV, the hotspot 19° off the super-galactic plane [170]. The post-trial significance of the hotspot in this seven year data set was 3.4σ.

2.5.1.2 Correlation Studies

To determine possible UHECR sources, correlation studies are performed comparing UHECR arrival direction data to source catalogues. A correlation was discovered in AGASA and Yakutsk data with the directions of BL Lacertae (BL Lac) objects ([171]), using an *a posteriori* analysis to maximise the correlation signal. Following this finding, a maximum likelihood analysis was applied to HiRes data to search for a correlation with the same subset of BL Lac objects, but no significant correlation was discovered [172]. However, a different subset of BL Lac objects was found to potentially correlate with the HiRes data. An analysis of an independent data set would be needed to claim a statistical significance. The same analysis was applied to Pierre Auger Observatory data, and no correlation was found [173]. This result however is complicated by the difference between known BL Lac populations in the northern and southern skies [28].

In 2007 the Pierre Auger Collaboration published the discovery of a correlation between UHECR arrival directions and the 12th edition of the Věron-Cetty and Věron (VCV) active galactic nuclei (AGN) catalogue [174–176]. This analysis used events measured between January 2004 and May 2006 as the exploratory data set, applying a parameter space scan of the minimum cosmic ray energy threshold (E_{th}), the maximum redshift of galaxies from the VCV catalogue (z_{max}), and the maximum angular separation between the UHECR arrival directions and AGN directions (ψ), to determine the parameter values which maximised the correlation signal. The parameters were scanned to minimise the cumulative binomial probability of the observed correlation occurring from an isotropic distribution, the results $E_{th} = 56$ EeV, $z_{max} = 0.018$ (corresponding to a distance of 75 Mpc), and $\psi = 3.1^\circ$, equating to a chance probability of $p = 0.21$ for a single event from an isotropic distribution correlating. These values were used as an *a priori* test for an independent data set (data collected after 27 May 2006) as a running prescription, applied until the hypothesis of an isotropic arrival direction distribution was rejected at the 99% level. This was achieved after 6 of the 8 detected events correlated with the VCV catalogue.

Upon passing the prescription confirming the existence of a correlation, the correlation signal of the entire data set was maximised by performing another parameter scan on the two data sets combined. This resulted in values of $E_{th} = 57$ EeV, $z_{max} = 0.017$ and $\psi = 3.2^\circ$, these parameters resulting in 20 of 27 events correlating with the VCV catalogue of AGN. The VCV catalogue traces the distribution of the local large-scale matter distribution, therefore this finding does not necessarily mean AGN are the source of UHECRs. However, as E_{th} is above the GZK threshold and z_{max} limits the AGN to within the GZK horizon, the finding does suggest the existence of a GZK suppression.

Following the original discovery of a correlation between Pierre Auger Observatory UHECR data and local AGN, a similar anisotropy was searched for in HiRes data [177], using the same parameters which maximised the correlation signal of the Pierre Auger Observatory's exploratory data set. A statistically significant correlation was not found, with 2 of the 13 events correlating with the VCV AGN catalogue. Additionally, a search to maximise the signal of the HiRes data

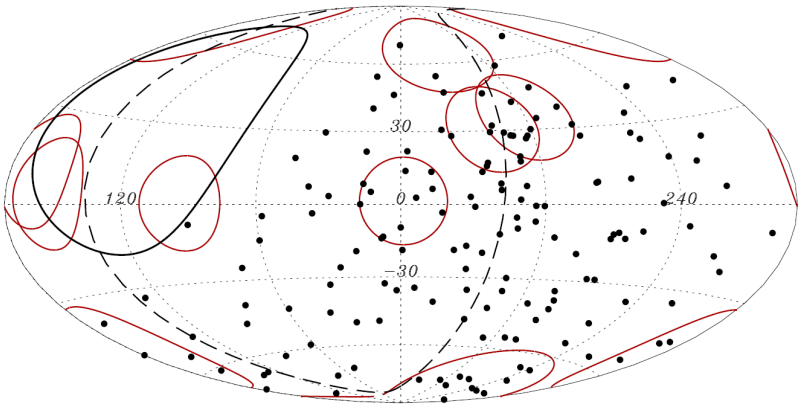


Figure 2.21: A map in Galactic coordinates of cosmic rays with $E \geq 58\,\text{EeV}$ (black circles), and Swift AGNs closer than 130 Mpc and brighter than $10^{44}\,\text{erg/s}$ (indicated by red circles of 18° radius). From [179].

produced results which were consistent with an isotropic arrival direction distribution with a chance probability of 24%. The AGN distribution is not uniform and the completeness of the VCV catalogue is different in each hemisphere, therefore the same degree of correlation is not necessarily expected [178]. An autocorrelation analysis of HiRes data above 56 EeV also did not find a significant small-scale clustering [177].

The Pierre Auger Collaboration has updated the VCV analysis over the years, each subsequent analysis with larger statistics finding the existence of a correlation above isotropic expectations, but with a smaller strength. The analysis in 2015 determined the fraction of correlating events with energy above 53 EeV is $28.1^{+3.8}_{-3.6}$%, with 41 of 146 events correlating. The correlation fraction is 2 standard deviations above the isotropic expectation, which is still 21% [179]. It is likely that the initial high correlation fraction was due to statistical fluctuation. The latest analysis does not find significant evidence of anisotropies above this energy threshold.

The Pierre Auger Observatory also searched for correlations with the Galactic Center, Galactic Plane, Super-Galacatic Plane, galaxies in the 2MRS catalog, AGNs detected by Swift-BAT, radio galaxies with jets and the Centaurus A galaxy [179]. These correlation studies did not find a statistically significant anisotropy. The strongest evidence of an anisotropy was for cosmic rays with $E \geq 58\,\text{EeV}$, in 18° radius windows around Swift AGNs closer than 130 Mpc and brighter than $10^{44}\,\text{erg/s}$ (see Figure 2.21), and in a 15° radius window around the direction of Centaurus A. The post-trial probability of these two correlations is $\sim 1.4\%$. This observed high degree of isotropy is not supportive of a light composition at the highest energies. If a large fraction of these cosmic rays were heavy nuclei, an anisotropic source distribution could be concealed due to these nuclei undergoing large deflections caused by magnetic fields. A large number of individual sources contributing to the cosmic ray flux could also result in the observed isotropy. The lack of a significant excess around the Galactic Center, Galactic Plane or Super-Galactic Plane, assuming the CR deflections are not too large, is supportive of the hypothesis that at these energies the CR sources are likely to be extragalactic.

A 2017 study by the Pierre Auger Collaboration investigated the compatibility of the detected UHECR arrival directions with flux models based on 17 AGNs detected by Fermi-LAT and 23 nearby starburst galaxies [180]. The post-trial significance of the deviation from anisotropy was 2.7σ for the gamma-ray AGNs above 60 EeV and 4σ for the starburst galaxies above 39 EeV. The starburst galaxy scenario resulted in a deviation at an angular scale of $\sim 13°$. Expected to be one of the dominant sources in the full-sky starburst model, the starburst galaxy M 82 is outside the exposure of the Pierre Auger Observatory, but is viewed by the Telescope Array. The hotspot observed by TA [170] (see Section 2.5.1.1) has some coincidence with the location of M

82 and the location of the blazar Mkn 421 [181].

Chapter 3

Ultra High Energy Cosmic Ray Detectors

In this chapter the Pierre Auger Observatory will be discussed in detail. Various other UHECR detectors will be briefly discussed.

3.1 Volcano Ranch

This detector was the first giant air shower array, operating from 1959 to 1963 [7, 18]. Located at Volcano Ranch, near Alberquerque, New Mexico, it was built to measure CRs of energy above 10^{17} eV. The array consisted of a triangular grid of 19 detectors, each consisting of a 3.3 m^2 plastic scintillator [2, 182] which was viewed by a 5 in photomultiplier tube [18]. During the operation of Volcano Ranch a 20^{th} detector was placed in various positions, and for a period was shielded by 10 cm of lead for the purpose of studying the muonic component of air showers [183].

Around 1000 showers were detected above an energy of 10^{18} eV [2]. A detection of particular note was the observation of an event with an estimated energy of 1×10^{20} eV [184]. This was the highest energy particle ever observed at the time [2]. The event was re-estimated to have an energy of 1.4×10^{20} eV, which is still one of the most energetic cosmic rays ever detected [7, 18]. Considering the low flux of CRs of this energy and the relatively low exposure of the array (the array reached a maximum area of 8 km^2 after a configuration change [18, 183]), the observation of such an energetic event is astonishing.

Volcano Ranch was responsible for the earliest measurements of the cosmic ray energy spectrum and arrival direction distribution at the highest energies. The experiment contributed significantly to the early knowledge of very high energy cosmic rays, and important inroads were made in describing the lateral distribution of a high energy EAS [7].

3.2 Haverah Park

Located in Haverah Park, England, the 12 km^2 ground array operated between 1968 to 1987 and was a collaboration between the Universities of Durham, Leeds and Nottingham, and Imperial College [7, 25, 35, 185, 186]. The array deployed water-Čerenkov detectors which was unorthodox at the time. Towards the end of its lifetime, eight scintillators were added to the centre of the array. Haverah Park proved that water-Čerenkov detectors were viable for detecting UHECRs. The experiment demonstrated that the particle densities of showers triggering the array could be measured in fine detail (especially near the shower core), allowing the formulation of functions which accurately describe the lateral distribution of showers [35, 187]. The Pierre Auger Observatory would adopt a similar design for its ground array.

3.3 SUGAR

SUGAR (The Sydney University Giant Airshower Recorder), located in Pilliga State Forest, New South Wales, Australia, collected data from 1968 to 1979 and was the only giant EAS array in the Southern Hemisphere prior to the Pierre Auger Observatory. [18,25,188,189]. The array spanned an area of 70 km^2, consisting of 47 detectors each comprising two liquid-scintillators buried 1.5 m underground and separated by 50 m in a North-South direction [190,191]. The scintillators were designed so that muons entering at any angle of incidence would produce the same light flux at the photomultiplier tube. The array applied innovative technology that would be utilised in future UHECR detectors, but the array's sparsity and its poor sensitivity to the electromagnetic EAS component would result in poor shower lateral distribution data and consequently poor energy estimations [18]. As a result, SUGAR concentrated on arrival direction studies (for example see [189,192]), which at the time was crucial because SUGAR viewed a region of the sky difficult for Northern Hemisphere detectors to observe.

3.4 Yakutsk

Located in Yakutsk, Russia, the array consists of upward-facing PMTs and scintillators. The PMTs measure Čerenkov light produced by the EAS in the atmosphere, facilitating a primary energy estimation that is calorimetric. The scintillator array ground parameter measurements are calibrated with respect to the Čerenkov array energy estimation, thus Yakutsk avoids relying significantly on air shower simulations to quantify the energy scale [18,193,194].

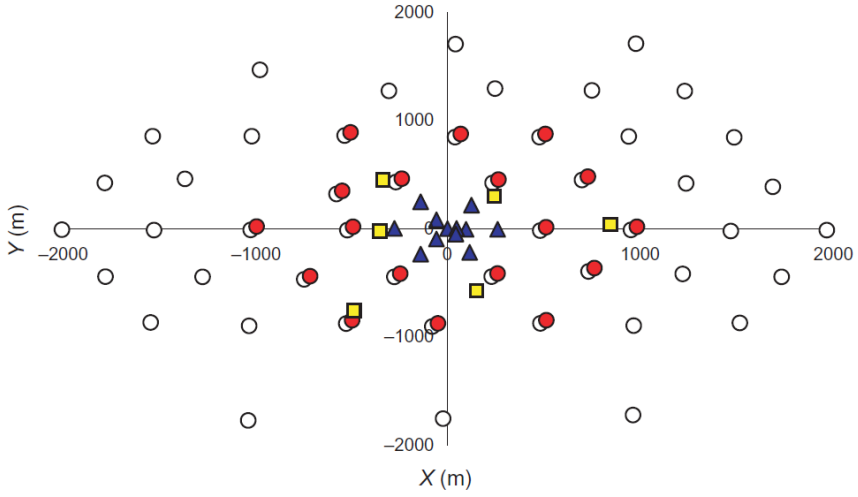


Figure 3.1: The configuration of the Yakutsk array. The scintillation detectors and muon detectors are represented respectively by open circles and filled squares. The C_1 and C_2 Čerenkov detector sub-arrays are represented respectively by filled circles and filled triangles. From [193].

The configuration of Yakutsk changed multiple times. Beginning operation in 1970, an expansion in 1974 resulted in a collecting area of about 18 km^2, however a contraction in 1995 resulted in the detector spanning 10 km^2. Figure 3.1 displays the current configuration of Yakutsk, designed to detect cosmic rays of energy 10^{15} to 10^{20} eV by employing different detector spacings. The Čerenkov array comprises of 48 PMTs divided between two arrays, the first array with a PMT spacing of about 500 m, the second array deployed in 1995 with a PMT spacing between 50 m to 200 m. These arrays are referred to as C_1 and C_2, respectively. There are 58 scintillator stations in the array spanning 10 km^2, in addition to 6 scintillator stations buried underground to detect muons. Each scintillator station consists of two 2 m^2 scintillator detectors. The scin-

tillator detectors trigger the C_1 array, whereas the C_2 array is triggered independently [193]. Imaging atmospheric Čerenkov telescopes will be added to the array, enabling the measurement of the longitudinal development profile of showers, which will significantly improve the ability of Yakutsk to investigate the composition of cosmic rays [194].

3.5 Fly's Eye

The first Fly's Eye detector, known as Fly's Eye I, was built at Dugway Proving Grounds, Utah, and commenced operation in 1981 [18,25,195]. The second detector, Fly's Eye II, was built 3.3 km from the first and came online in 1986 [18, 25, 196]. The array operated with this configuration until 1992 [7, 18]. The Fly's Eye Detector was the first air fluorescence detector to successfully measure large quantities of events.

Both detectors used PMTs to observe atmospheric nitrogen fluorescence. The experiment proved that CRs could be successfully detected independently of ground arrays. Some important findings published by Fly's Eye were the measurement of the CR energy spectrum ankle at 3×10^{18} eV, a change in composition from heavy to light primaries between 0.1 EeV to 15 EeV, and the detection of the highest energy CR recorded to date at an estimated energy of 3×10^{20} eV [197, 198]. The event's shower profile is shown in Figure 3.2.

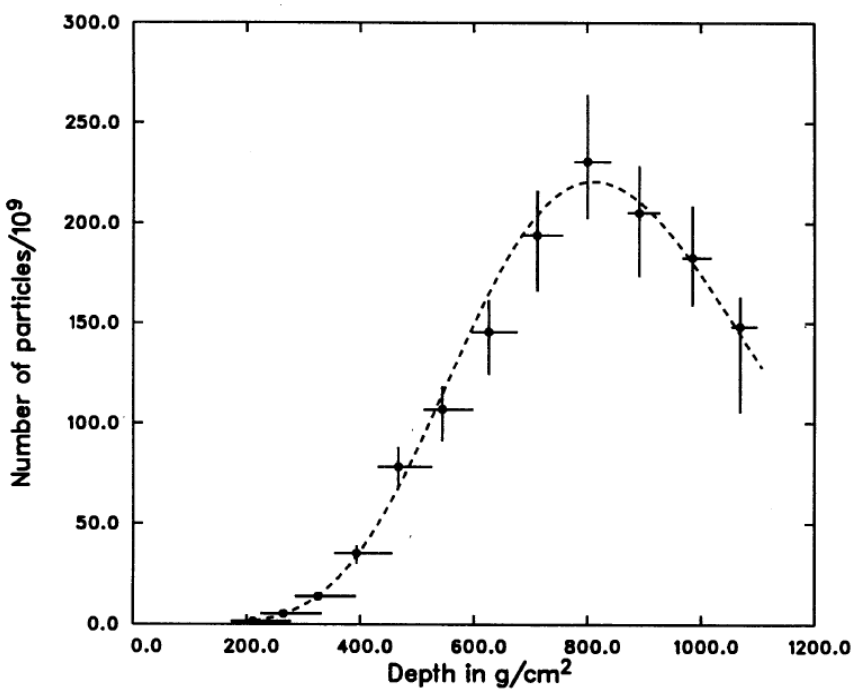


Figure 3.2: The shower profile of the highest energy cosmic ray recorded to date. Measured by Fly's Eye, the estimated energy of the event is 3×10^{20} eV. The dotted line is the best-fit of the Gaisser-Hillas function. From [198].

3.6 AGASA

Located 120 km west of Tokyo, Japan, AGASA (Akeno Giant Air Shower Array) was the largest air shower array in the world before the Pierre Auger Observatory was constructed [7, 18, 200]. The array spanned 100 km^2, consisting of 111 1.2 m^2 plastic scintillator detectors spaced 1 km apart. At 27 of the scintillator sites, muon detectors were installed consisting of proportional counters underneath an absorber giving a vertical muon energy threshold of 0.5 GeV.

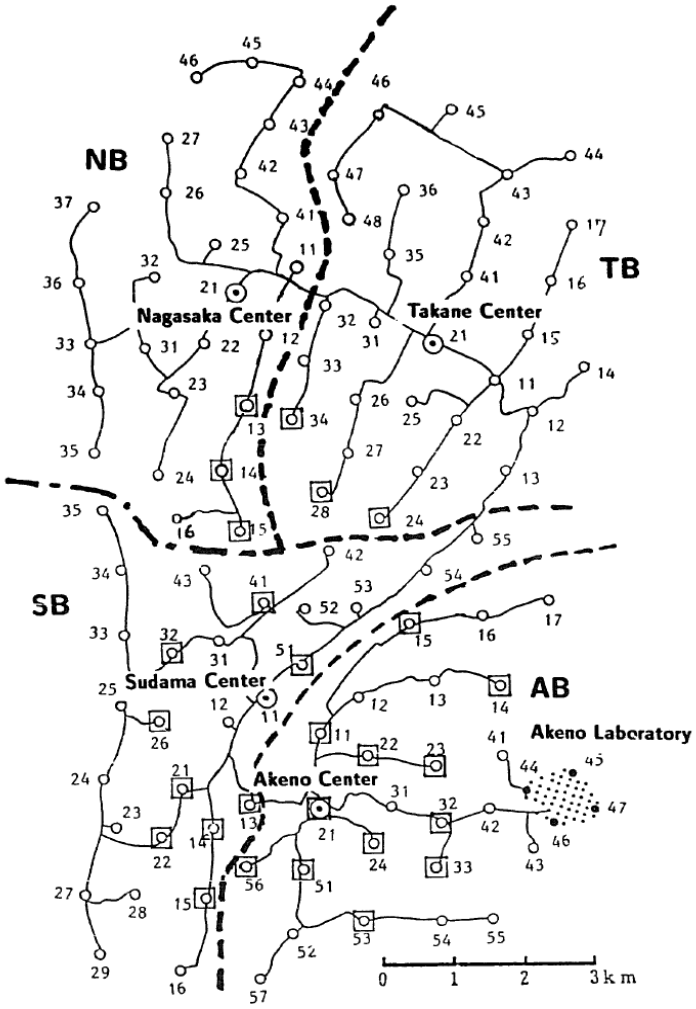


Figure 3.3: The configuration of the AGASA detector. The plastic scintillators and muon detectors are indicated by open circles and squares respectively. The dashed lines show the borders of the four array branches, and the dotted lines show the optical fibre paths connecting the detectors to data processing centres. From [199].

AGASA was originally designed as four separate branches, but in 1995 it was merged to operate as a single detector [7, 199]. Before the consolidation of the array, triggering and shower reconstructions were performed independently by each branch. Upon unification, the effective area of AGASA increased by a factor of ~ 1.7 as showers that triggered detectors across or near branch boundaries were able to be reconstructed. Figure 3.3 shows the configuration of the array. Initial data collection commenced in 1990, continuing until January 2004 [38].

The array measured a significant number of super-GZK events, resulting in the collaboration reporting there was no spectral cut-off consistent with the GZK effect in their UHECR spectrum [200, 201]. Later HiRes (see below) reported the existence of a flux suppression in their measurements, contradicting the AGASA findings. It was eventually discovered that the AGASA experiment had overestimated the energy of measured events, and subsequently revised their findings to state that the existence of a spectral cut-off was now uncertain [202, 203].

The array had a superior angular resolution of approximately 3° at 10^{18} eV and 1.5° at 10^{19} eV which facilitated large and small-scale anisotropy studies [204]. They reported on the discovery of a large scale anisotropy at energies around 10^{18} eV pointing towards the Galactic Centre, the scale of the observed excess being about $\sim 20°$ [204]. They also discovered a small-scale clustering of arrival directions at energies above 4×10^{19} eV, consisting of a cluster of three events and three

clusters of two events in 2.5° regions, with a reported chance of coincidence of less than 1% [205]. However, the true statistical significance of these results is unclear (see Section 2.5).

3.7 HiRes

The High Resolution Fly's Eye observatory succeeded the Fly's Eye experiment. The experiment consisted of two fluorescence detectors named HiRes I and HiRes II which commenced operation in 1997 and 1999 respectively, with the operation of both detectors ending 2006 [7, 46, 206]. The detectors were separated by 12.6 km, consisting of individual modules to cover 360° in azimuth.

The HiRes experiment published findings on a range of topics including composition and anisotropy [46]. A significant discovery was the measurement of a flux suppression above $\approx 6 \times 10^{19}$ eV, which was consistent with the predictions of the GZK effect [112].

3.8 TA

The Telescope Array (TA) observatory is a hybrid detector located in Millard County, Utah which commenced data taking in 2008 [207]. It consists of 507 scintillation surface detectors separated by 1200 m on a square grid, covering an area of approximately 700 km^2. Each surface detector consists of two 1.2 cm thick layers of plastic scintillator of area 3 m^2 [208]. The SD array is overlooked by 3 fluorescence detector sites, consisting of 12 FD telescopes at two of the sites and 14 FD telescopes at the third site. The FD telescopes at the third site were refurbished from the HiRes-1 station of the HiRes experiment. The fluorescence telescopes view an angle range of $3° - 31°$ above the horizon, and an azimuth range of 114° [209].

The Telescope Array and Pierre Auger Observatory have conducted joint studies on the energy spectrum and mass composition of UHECRs (see Chapter 2), and also joint anisotropy studies.

3.9 The Pierre Auger Observatory

In 2004 data collection began at the Pierre Auger Observatory [30], the largest cosmic ray detector in the world. Located in the Pampa Amarilla ($\sim$ 35° S, 69° W), near the town of Malargüe in the Province of Mendoza, Argentina, the site is a semi-arid plateau around 1400 m above sea level, with the detector altitudes ranging from 1340 m to 1610 m. The site was chosen due to its relative flatness allowing easier equipment deployment and maintenance as well as wireless communication, its proximity to the infrastructure of Malargüe, and the clear night skies with low light pollution which are ideal conditions for measuring nitrogen fluorescence emission [41, 43].

In the following sections the components of the Pierre Auger Observatory will be discussed.

3.9.1 Communications and CDAS

A wireless network is used for communications, due to the large size of the array and number of detectors (see Figure 3.4). The surface detector (SD) array is divided into sectors, with each sector assigned a base station to communicate with the SD stations of that sector. At each fluorescence detector (FD) site, four communication towers each hold eight base station units. The array is divided up so the processing load of the array-wide communications is distributed efficiently, whilst preventing data loss from the whole array if a base station fails. The SD stations transmit data to their base stations through the wireless local area network (LAN) which operates in the 902-928 MHz band, while a microwave network in the 7 GHz band trasmits data to the Observatory Campus in Malargüe. Data is then transferred to the Central Data Acquisition System (CDAS) [43]. Data transmission to the CDAS is done directly from Coihueco and Los Leones, whereas Loma Amarilla and Los Morados relay data through the former and latter respectively to the CDAS.

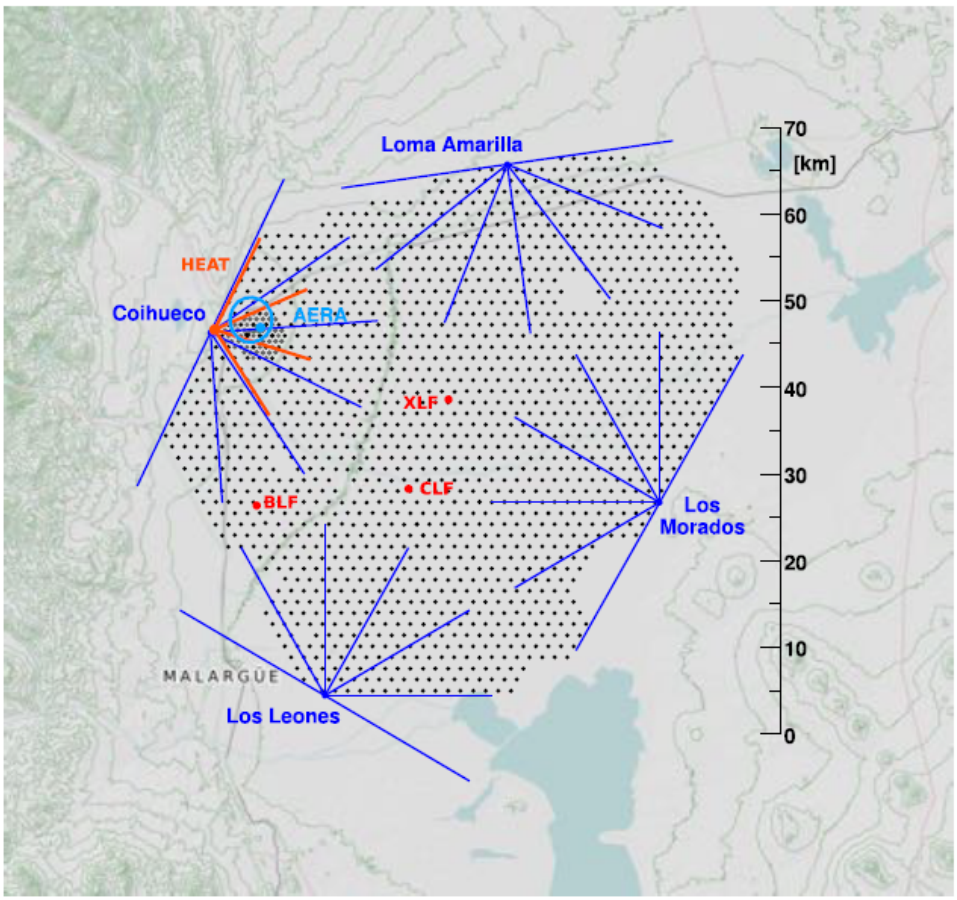


Figure 3.4: Map of the Pierre Auger Observatory. The surface detector stations are represented by black dots. Blue and orange lines show the field of view of the regular and high elevation fluorescence telescopes respectively. From [210].

The Central Data Acquisition System performs a number of roles. Its main job is to process triggers for SD-only and hybrid events, and to initiate data collection. The CDAS controls the SD stations and stores SD data, whilst allowing access to the SD monitoring data and configuration tools. The SD triggering is discussed in further detail in Section 3.9.2. The FD triggering is independent of the CDAS, with a notification transmitted to the CDAS of the detection of an event for the purpose of hybrid triggering, discussed in detail in Section 3.9.4.

3.9.2 Surface Detector

The surface detector array of the Pierre Auger Observatory spans an area of 3000 km^2 and is comprised of 1660 water-Čerenkov detectors spaced 1.5 km apart on a triangular grid. The SD construction began in 1999, stable data collection commenced in January 2004, and construction was finished in 2008 [211]. Water-Čerenkov detectors were chosen due to their low cost, durability in harsh conditions and their almost uniform exposure to primary cosmic rays of incident zenith angles less than 60° [41].

The 1.5 km detector spacing was chosen as a compromise between energy threshold and cost. The lower energy range of the Pierre Auger Observatory measurements was intended to overlap with the energy range of previous cosmic ray experiments, allowing the comparison of results [41]. The detector spacing results in the SD being fully efficient (the trigger probability for the 3ToT trigger reaches 100% - see Section 3.9.2.3) at an energy of 3×10^{18} eV for zenith angles less than 60° [212]. A shower energy of 10^{19} eV will trigger at least five stations [41].

3.9.2.1 Station Design

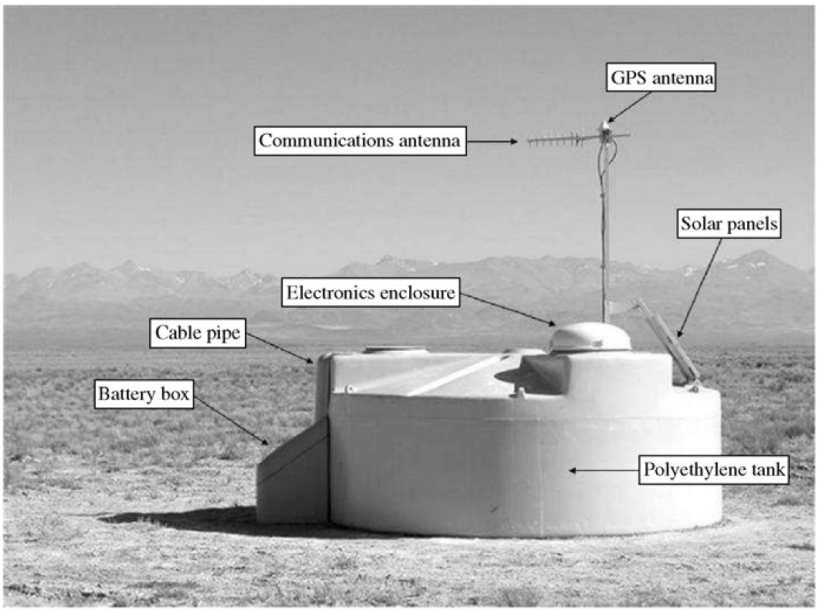


Figure 3.5: An SD station of the Pierre Auger Observatory. The main components are labelled. The flatness of the terrain is typical of the site. From [41].

An SD station is an independently operating, self-contained unit. Figure 3.5 shows an SD station. It consists of a polyethylene tank which protects the detector components from outside elements, except the communications antennae and solar panels which are exposed on the exterior. The tank encloses a cylindrical volume which is 3.6 m in diameter and 1.2 m in depth, containing 12 000 L of ultra-pure water. The ultra-pure water maximises the transmission of Čerenkov light whilst preventing microbe growth that would destabilise detector performance. The depth of the detectors is approximately 3 photon radiation lengths, therefore there is an efficient conversion of photons in an EAS to charged particles, and statistical fluctuations in the sampling of the EAS electromagnetic component is reduced [33].

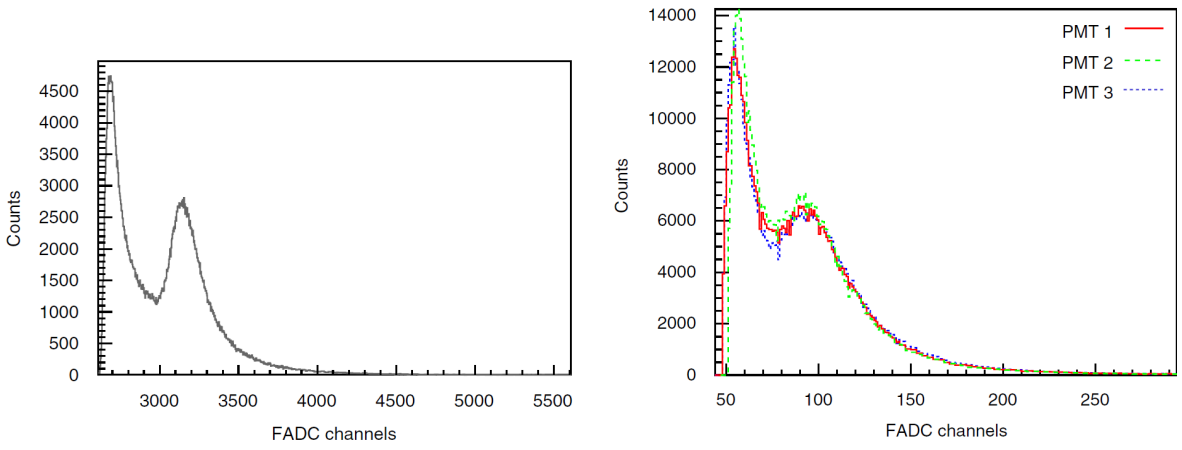
A sturdy, diffusively-reflective plastic liner encloses the water to prevent contamination and block external light from entering. At the top surface of the liner are three windows of UV-transparent polyethylene at a distance of 1.2 m from its centre, with each having a 9 in diameter PMT viewing the water through the window [41, 213].

Enclosed in the top of the station is an electronics package to process signals from the PMTs. A processor in the electronics package allows local data acquisition, detector monitoring, software triggering and storage of data. Measuring signals close to and at large distances from a shower core requires a large dynamic range, which is achieved by reading the output from the anode and the amplified last-stage dynode of the PMTs. Signals ranging from several to 10^5 photoelectrons can be measured due to the large dynamic range. The signals are filtered and digitised for triggering by an FADC operating at 40 MHz (see Section 3.9.2.3). A 100 MHz clock which is synchronised by a GPS receiver at each station is used for the signal timing.

Two solar panels and two lead acid batteries power each station for approximately 99% of the time. An average of 10 W is supplied to the station through the solar power system. Over the lifetime of the SD, possible degradation of the power system will still allow an expected uptime fraction of 98% [41].

3.9.2.2 Calibration

Knowledge of the particle response at each individual station is necessary if accurate energy reconstruction and stable triggering of the SD is to be achieved. This requires constant calibration of each station, which is performed by the continuous autonomous monitoring of signals produced by the atmospheric muon flux [214]. To calibrate a station, data immediately prior to an event trigger is used, in addition to histograms of the charges and pulse heights produced at the PMT outputs by atmospheric muons in the preceding minute (approximately 150,000 muons). Examples of such histograms are displayed in Figure 3.6. All of this information is transmitted to CDAS, allowing the accurate conversion of a measured signal in units of FADC channels into station-independent units known as VEM so that signals can be compared.



(a) (b)

Figure 3.6: Examples of the histograms sent to the CDAS. The histogram in Figure 3.6a displays the sum of the charges for the 3 PMTs over the preceding minute. Figure 3.6b displays the pulse height histogram for each PMT. In both images, the peak on the right is produced by atmospheric muons and is used for calibration, whereas the peak on the left is produced by low-energy particles and therefore not required for calibration calculations. From [214].

A VEM, vertical equivalent muon, is defined as the average measured charge produced at the PMT output by a muon travelling vertically through the centre of a station. If the value of 1 VEM in integrated FADC channels is known, any signal can be converted into VEM. An SD station can not isolate a vertical, central muon, so instead the value of 1 VEM is determined from the charge histogram produced by background muons. An independent experiment showed the position of the second peak in a histogram of the sum of the charges from the 3 PMTs of a station corresponds to 1.09 VEM. When data is analysed a polynomial fit is used to find the position of this peak, and its value used to convert the measured signals to VEM. This conversion has an accuracy of approximately 3%.

The background muon flux is also used for setting the triggering levels of the SD (see Section 3.9.2.3). The thresholds are set with respect to a value derived from the current produced by the PMT as a result of the atmospheric muon flux. Therefore the relationship between this value and FADC channels must be known. The reference unit is I_{VEM}^{peak}, the value of the peak of a histogram of pulse heights measured by the station. I_{VEM}^{peak} is not measured directly as this would require significant downtime for a station, so instead I_{VEM}^{est} is used which is an estimate of I_{VEM}^{peak}. Using I_{VEM}^{est} enables stable triggering of the SD during the continuous calibration. The rate of a calibration trigger is maintained at a pre-defined value using a convergence algorithm. A measured rate of approximately 70 Hz (determined from the reference station) is required for signals above $2.5I_{VEM}^{est}$ in a PMT and above $1.75I_{VEM}^{est}$ in all three PMTs. If this rate is not measured,

the value of I_{VEM}^{est} is adjusted until these conditions are satisfied. The value of I_{VEM}^{peak} is known to an accuracy of 6% when convergence is achieved, providing stable triggering conditions.

3.9.2.3 Trigger System

The SD uses a hierarchical trigger system, starting at 'local' triggers - the low-level triggers independently processed by each station - before leading to high-level triggers that are processed off-line which select the physical events for reconstruction [211, 215].

Local triggering filters out background signals, so only events of interest are transmitted to the CDAS, efficiently using the site's communications bandwidth. The lowest level trigger is T1, requiring either of two conditions to be satisfied. The first is the time over threshold trigger (ToT-T1) requiring 13 FADC bins in a 120 bin window to be above a level of $0.2I_{VEM}^{est}$ for two PMTs in coincidence. The ToT-T1 trigger selects low-amplitude, dispersed signals, typical of those from distant or low energy showers. The rate of this trigger is around 1.6 Hz. The second condition is a threshold trigger (TH-T1) requiring a threshold value of $1.75I_{VEM}^{est}$ to be exceeded in coincidence with all three PMTs. TH-T1 selects the quick signals from the muonic component of highly inclined showers, at the cost of increased background noise. This condition is triggered at a rate of around 100 Hz.

Upon a T1 trigger, the station's processor checks for T2 signal criteria. The objective of these triggers is to reject background noise and select signals that are likely due to showers. The ToT-T1 triggers are all upgraded to a T2, whereas the TH-T1 triggers must pass a threshold of $3.2I_{VEM}^{est}$ in coincidence with the three PMTs in the station. The T2 trigger results are then sent to the CDAS for processing, at a rate of approximately 20 Hz per station.

Array triggering occurs at the T3 level, where the CDAS uses the coincidences of T2 triggers between separate stations to assess the T3 level. If the CDAS recognises a T3 trigger, FADC trace data are collected from every station exhibiting either a T2 or T1 trigger within 30 µs of the T3. Two sets of conditions trigger a T3, illustrated in Figure 3.7. The first set is a coincidence of ToT triggers in 3 stations along with a required level of compactness, referred to as '$ToT2C_1\&3C_2$'. C_n refers to the n^{th} successive hexagon surrounding a given tank. This compactness condition requires that one of the stations has a least one of its nearest neighbours triggered (total of two stations within the smallest hexagon around the station of interest - $2C_1$) in addition to at least one of its second nearest neighbours triggered (total of 3 stations within the second hexagon - $3C_2$). It is also required that the T2 triggers occur within $(6 + 5C_n)$ µs of the first trigger. The rate of this T3 trigger is 1600 times per day, with 90% of these being real events.

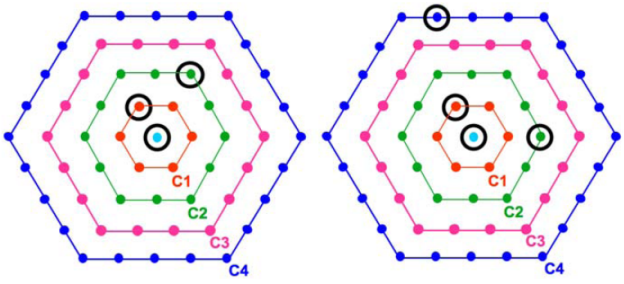


Figure 3.7: Two examples of events satisfying the spatial criteria of the T3 trigger. The left is a $ToT2C_1\&3C_2$, the right is a $2C_1\&3C_2\&4C_4$. From [211].

The second set of T3 trigger is for the detection of highly inclined showers. These showers have a highly attenuated electromagnetic component and the remaining muons produce signals of short duration that are spread over a large area. This T3 trigger is less rigorous and is known as '$2C_1\&3C_2\&4C_4$' and has the same timing criteria as the $ToT2C_1\&3C_2$. This second set of T3 conditions occurs around 1200 times per day, of which 10% are real showers.

The first off-line trigger level is the T4. This trigger is 'physical' and intended to accept real showers and reject noise triggers that passed the T3 level. This trigger has two possible conditions, the first condition (3ToT) being that three stations in a triangular cell pass the ToT trigger requirement with their arrival times consistent with a plane shower front moving at the speed of light. 98% of real showers at fully-efficient energies with zenith angles below 60° pass this condition. The second possible T4 condition (4C1), requiring 4 stations in a hexagon to trigger any type of T2 with the same timing requirements as the 3ToT, accounts for the remaining 2% of real showers. After the selection of T4 triggers, stations which have triggered accidentally are removed by applying timing criteria to stations outside of the triangle of stations which measured the highest event signal. Of the events reaching this stage, 99.9% are fully reconstructed, culminating in their arrival direction, core position and value of S(1000) (see Section 3.9.2.4) being determined.

The final trigger, T5, selects events with well-understood accuracies in energy and geometry. The aim of this trigger is to avoid potentially incorrect shower reconstructions which have emerged from shower cores outside the array boundaries. The T5 trigger checks the shower core is within a triangle of working stations and that the station which measured the highest signal (the 'hottest' station) has 6 working nearest neighbours.

The full trigger chain reduces the single station signal rate of 3 kHz, attributed mainly to atmospheric muons, down to approximately 3 triggers per day from real shower detections. The SD aperture at any time for energies above 3×10^{18} eV is calculated by summing the number of hexagons containing only working stations, multiplied by the effective aperture of a single hexagon. The T5 trigger results in a straightforward calculation of the aperture. In studies were the calculation of the aperture is not important, such as anisotropy investigations, a relaxed T5 condition (T5 ICRC2005) may be used which requires the hottest station be surrounded by 5 instead of 6 working neighbours. This less rigorous trigger means more statistics are available, with an accurate reconstruction of event properties still permitted [215].

3.9.2.4 Event Reconstruction and Performance

Once a cosmic ray event is identified by the CDAS, reconstruction of the event properties is performed. The following discussion relates to the reconstruction of showers with zenith angles smaller than 60°, referred to as 'vertical' showers. Showers that are incident upon the array at larger angles, referred to as 'horizontal' air showers, have a significantly attenuated electromagnetic component, therefore the standard analysis is not used. The reconstruction of 'horizontal' air showers (HAS) is discussed in [36].

To determine the arrival direction of an event measured by the SD, fits are applied to the time of arrival of particles at each station [30]. The shower front is represented by a spherical shower geometry with variable curvature for the purpose of fitting, however the particular shower front functional form chosen does not contribute significant uncertainty to the reconstruction. The precision of the clocks at each station and natural arrival time fluctuations contribute to the uncertainty of the arrival direction. These factors are merged into a time variance model giving the probability distribution of arrival times at each station. Pairs of tanks separated by 11m in the array, called 'doublets', allows the model used to be experimentally validated by repeating the reconstruction of events measured by a doublet with one of the doublet tanks excluded. Consequently, the angular resolution can be calculated on an event-by-event basis. The angular resolution improves with increasing station multiplicity, with a vertical shower involving 3 stations typically having an angular resolution of around 2.2°. For energies above 10 EeV the angular resolution for any zenith angle is better than 1°. Hybrid events can be fully reconstructed with only the SD, therefore independent checks of these resolutions can be performed. The resolution estimated using this method agrees reasonably well with the time variance model [32].

The energies of events detected by the SD are estimated by fitting a lateral distribution function to the signals measured by each station that observed the event. By applying the LDF,

$S(1000)$ - the signal that would have been produced in a station at a core distance of 1000 m - is calculated. $S(1000)$ is then converted into S_{38} - the $S(1000)$ the shower would have produced if it arrived with the median observed zenith angle of 38° - thereby adjusting for the attenuation of the signal due to the atmosphere. The Constant Intensity Cut (CIC) method is used to perform this conversion [216]. Lastly, S_{38} (in units of VEM) is converted to the event's primary energy by using the calibration curve derived from hybrid events. The statistical uncertainty of the SD energy reconstruction is $< 17\%$. Using the FD energy for calibration results in an overall systematic uncertainty of 14% (see Section 3.9.3.4) [217].

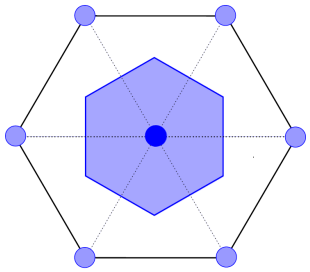


Figure 3.8: The basic hexagonal cell used for exposure calculations. The shaded area illustrates the hexagon corresponding to the central station, surrounded by 6 working stations. From [211].

The state of each SD station at any time is known, allowing the determination of the instantaneous exposure of the array. Above 3 EeV, the array is fully efficient, therefore the aperture calculation is one of geometry. A hexagon defined by the central detector surrounded by 6 working neighbours defines the individual element considered (illustrated in Figure 3.8). The hexagon has an area of 1.95 km^2, resulting in an aperture for zenith angles less than 60° of $a_{cell} \sim 4.6\,\mathrm{km}^2\,\mathrm{sr}$. The total array aperture is $N_{cell}(t) \times a_{cell}$, where N_{cell} is the number of cells, monitored on a second-by-second basis with an uncertainty of $\sim 1.5\%$ due to communication problems between the stations and the CDAS. Uncertainty in the live time of the array is $\sim 1\%$ due to communication problems across the array and local data storage issues. The exposure of the total array is calculated by integrating the total array aperture over time, and the uncertainty in this calculation is less than 3%. Above 3×10^{18} eV the SD exposure increases by approximately $500\,\mathrm{km}^2\,\mathrm{sr\,yr}$ every month with the full array deployed [211].

3.9.3 Fluorescence Detector

The fluorescence detector of the Pierre Auger Observatory consists of 24 telescopes (plus those of HEAT - see Section 3.9.5.1), arranged in 4 groups of 6 telescopes each, located on elevated terrain and looking inwards from the edge of the SD array. Each FD telescope views 28.1° in elevation and 30° in azimuth, therefore each FD site has a 180° azimuthal field of view. The four sites are named Los Leones, Coihueco, Loma Amarilla and Los Morados, located on the southern, western, northern and eastern edges of the SD, respectively. The four FD buildings are climate-controlled to lessen the effect of temperature changes on the detector, and for instrument protection retractable shutters cover the windows of each telescope bay, opened only when the FD is operated. The average duty cycle of the FD is 15%, as the FD only operates on clear nights with minimal moonlight [30]. For energies above 10^{19} eV, over the ground array the FD configuration has a triggering efficiency of 100%.

The FD is capable of measuring stereo and higher multiplicity events. As hybrid detections provide shower geometry reconstructions which are more accurate than those using only the FD, this FD capability is not as important for the Pierre Auger Observatory compared to experiments comprising only fluorescence detectors [24].

3.9.3.1 Telescope Design

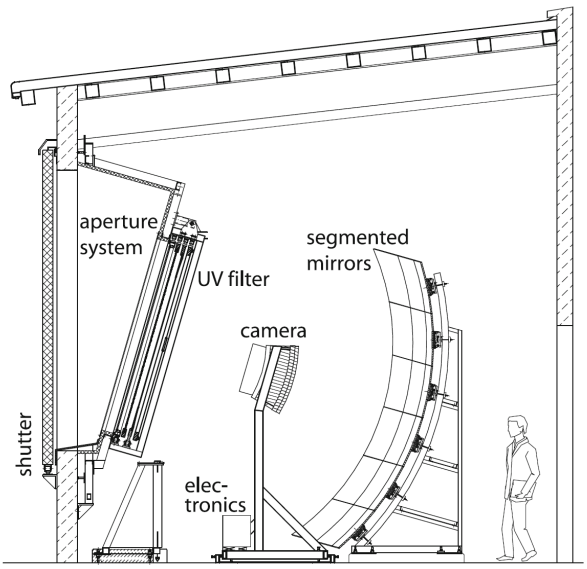


Figure 3.9: A schematic showing the main components of an FD telescope. Image from [24].

Figure 3.9 illustrates the design of a Pierre Auger Observatory FD telescope. Schmidt optics are used by the FD telescopes to image extensive air showers. Each camera is placed in the focal surface of a $\sim 13\,\mathrm{m}^2$ spherical mirror. A camera consists of 440 hexagonal PMT pixels, with each pixel viewing a 1.5° diameter region of sky. The spaces between pixels are covered with a light collector made of reflective 'Mercedes stars', maximising light collection while minimising light falling between pixels. The average light collection efficiency over the camera is 93%, in contrast with an efficiency of 70% without the Mercedes stars [24]. Laser shots and the tracking of bright stars moving across the field of view allows each pixel's absolute pointing direction to be determined to within 0.1° [24, 37].

The camera views a mirror made from segmented pieces. Los Leones and Los Morados have anodised aluminium mirrors comprising 36 rectangular pieces from one manufacturer, whereas from another manufacturer Coihueco and Loma Amarilla have mirrors made of 60 hexagonal glass segments lined with a reflecting aluminium coating, resulting in reflectivities greater than 90% at a wavelength of 370 nm. Both mirror types have irregularities of a scale less than 10 nm.

An annular corrector ring reduces the spherical aberration of the mirror, resulting in a doubling of the optical aperture of the FD telescopes whilst maintaining a small optical spot. The ring consists of 24 glass segments, with the inner and outer radii of the ring being 0.85 and 1.1 m respectively.

An optical filter covers the aperture of each telescope, reducing background light whilst allowing through photons of wavelengths around 290 nm to 410 nm, which is most of the nitrogen fluorescence spectrum range. Without the filter, background photons and the low flux of photons from the EAS would be indistinguishable from each other.

3.9.3.2 Electronics and Triggering

Analogous to the SD, the FD triggering consists of hardware triggers applied online by the telescope electronics to filter out background events, and software triggers applied offline to select real events.

The FD electronics must be able to detect low and high energy events over a large volume, subject to a range of background light conditions. A compression technique detailed in [24] is utilised to achieve a large dynamic range. The First Level Trigger (FLT) boards ensure the pixel

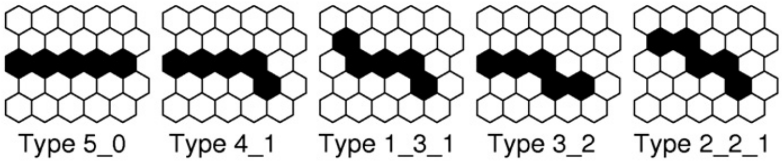


Figure 3.10: The five main trigger patterns used to identify tracks. From [24].

thresholds are adjusted so that a trigger rate of 100 Hz is achieved throughout changing light conditions. Additionally, the FLT generates pixel triggers when the threshold is exceeded.

Output from the FLT boards is passed to the Second Level Trigger (SLT) boards. The SLT boards process data and identify triggered pixel combinations that are consistent with straight tracks. Figure 3.10 displays the five main patterns the SLT boards look for. In case there are untriggered pixels, only four of the five pixels in each combination need to trigger. A straight track is identified at a rate of $0.1 - 10$ Hz per telescope.

Upon satisfying the SLT, the FD data is processed by the Third Level Trigger (TLT) algorithm, which is the first software trigger. The core responsibility of the TLT is to reject triggers caused by lightning. Lightning can trigger hundreds of pixels in a camera, which would cause significant dead time in the FD. The TLT trigger rejects 99% of lightning events at a cost of 0.7% of real events. The TLT conditions are satisfied at a rate of 0.01 Hz per telescope.

When an event satisfies the TLT, an algorithm collects data from all the telescopes of the site and initiates a hybrid trigger, T3. The T3 trigger initiates SD data collection so that a hybrid event is measured. A basic reconstruction of the FD data is performed, providing a location and ground impact time estimate. The FD data and T3 trigger are sent to the CDAS, which then requests signals from the SD stations which are near the event and coincide with the ground impact time. This ensures SD data for events below 3×10^{18} eV are saved, in-case the event triggered one or two SD stations but did not produce a T3. The FD and SD data is stored, and an off-line analysis can then combine these data into a hybrid event (see Section 3.9.4).

3.9.3.3 Calibration and Atmospheric Monitoring

To calibrate the FD, the response of each pixel to an incident photon flux is determined, allowing a conversion of a value in ADC counts to a number of photons, which can then be used to estimate the shower energy. The change in any telescope's response is monitored by performing absolute and relative calibrations.

Absolute calibrations are performed periodically and are time consuming, with each site taking about 3 days to calibrate [24]. To calculate the detector response from light incident on the aperture, an end-to-end calibration is used thereby taking into account the effects of optics and electronics, and eliminating the need to investigate each telescope component separately. The calibration is performed at each telescope by mounting a drum over the aperture, which provides uniform illumination of a known intensity. The pulses of light allow a pixel-by-pixel calculation of the camera response. Measurements are performed at several wavelengths to evaluate each camera's spectral response curve. The pixel response uncertainty from absolute calibration is 9%.

An independent check of the absolute calibration detailed above is provided by a portable 337 nm nitrogen laser. The intensity of light emission from the laser is known, and so by positioning the laser relative to the FD such that the effects of aerosol scattering are minimised, a known fraction of light is scattered to the telescope. The calibration constants from the drum calibration are compared, allowing the response of the telescope to be cross-checked between the two absolute calibration methods.

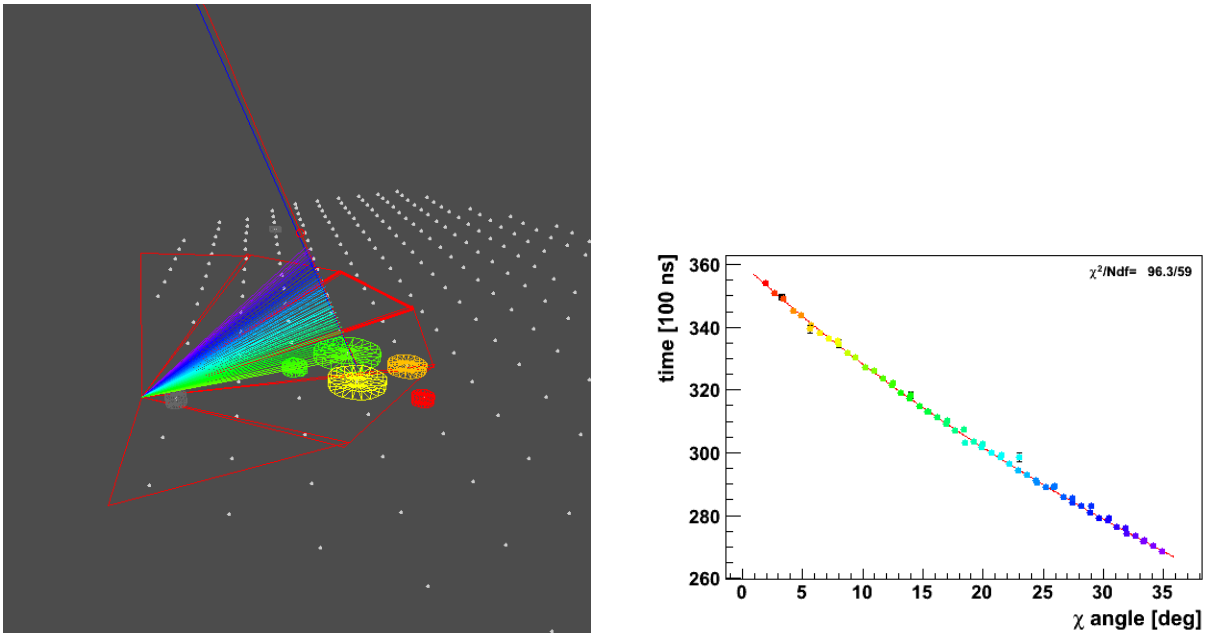
As absolute calibration is only performed periodically, relative calibration is performed on a nightly basis during operation to track the detector response. Common light sources are connected

by fibre optics to one of three positions - A, B or C - in each telescope to illuminate different optical components. Position A results in diffuse illumination at a wavelength of 470 nm on the camera face. Position B illuminates the mirror through a filter which approximates the wavelength acceptance of the telescopes. Position C illuminates diffuse reflectors mounted on the interior of the telescope bay shutters from which light reflects into the telescope aperture. Before and after each night, the relative calibration updates the calibration constants of each pixel in the FD. The FD calibration processes enable an accurate calculation of the number of photons reaching the telescope aperture.

The propagation of photons through the atmosphere to the camera must also be considered. The atmospheric conditions are monitored extensively by weather stations, GDAS (Global Data Assimilation System), infra-red cameras and lasers, providing frequent measurements on atmospheric conditions such as pressure, temperature, humidity, aerosol scattering properties and cloud coverage [44]. These ingredients are taken into account during FD event reconstruction to minimise the effect of atmospheric fluctuations.

3.9.3.4 Performance

The reconstruction of events seen by the FD begins with a reconstruction of their geometry [24]. Figure 3.11 displays an event reconstruction illustration provided by analysis software. The 'shower-detector plane' (SDP) is the plane containing the shower track and the fluorescence telescope, and is determined by a fit of the integrated pulse and pointing direction of each triggered pixel. The accuracy of the fit, which depends on the observed track length, can reach an order of 0.1°. Pixel timing information is then fitted to determine the shower orientation. When using only FD data, there may be large uncertainty in the shower geometry due to degeneracy between the fitted parameters (see Section 1.3.2). The SD operates almost 100% of the time, therefore many events which trigger the FD are observed as hybrids, facilitating a reconstructed shower geometry of excellent resolution (see Section 3.9.4). These uncertainties can be verified by using laser shots from the Central Laser Facility (CLF), located near the centre of the SD array, to compare known geometries with the reconstructed geometry.



(a) A 3D view of the shower.

(b) A timing fit of the shower.

Figure 3.11: A fluorescence detector event reconstruction.

Upon determining the shower geometry, the primary energy of the event can be estimated.

By considering direct and scattered fluorescence light in addition to Čerenkov light, the flux of photons at the aperture is converted to energy deposited in the atmosphere as a function of depth. The primary energy of the shower is calculated by integrating the profile of energy deposited as a function of slant depth. Events with a well-determined shower geometry have a statistical energy uncertainty of $\sim 8\%$ [217]. The systematic uncertainty on the energy scale is $\sim 14\%$, contributed predominantly by an FD calibration uncertainty of $\sim 10\%$ and an FD profile reconstruction uncertainty of $\sim 6\%$ [217]. A fit of the energy deposit profile provides the depth of shower maximum, $X_{\max}$, with a resolution of $< 20\,\mathrm{g\,cm^{-2}}$ for well measured events [30].

Unlike the SD, the FD aperture does not saturate at high energies. As the shower energy increases, so does the amount of light produced, therefore the viewing distance depends on shower energy. The viewing distance also depends on background light and atmospheric conditions. Considering average values of these environmental factors, simulations determine an aperture of $\sim 900\,\mathrm{km^2\,sr}$ at $10^{17.5}\,\mathrm{eV}$, to $\sim 7400\,\mathrm{km^2\,sr}$ at $10^{19}\,\mathrm{eV}$. However, for energy spectrum calculations, the hybrid exposure is more important and thus it is calculated very accurately (see below).

3.9.4 Hybrid Operation

Hybrid detections provide the highest quality data measured by the Pierre Auger Observatory, resulting in excellent shower reconstructions. The hybrid data is used in CR energy spectrum studies [42] and composition studies [21] as these investigations require precise and accurate data.

To achieve accurate hybrid reconstructions, the trigger times between the FD and SD need to be well known so that offsets can be compensated for. Two procedures are applied to calculate the value of timing offsets to within 100 ns [37, 51]. The first procedure uses events that can be reconstructed as hybrids, or with the SD alone, and are referred to as 'golden hybrid' events. As the hybrid reconstructed core location is sensitive to the time offsets between the FD and SD, a systematic shift in the core location of a hybrid reconstruction compared to the SD only reconstruction betrays a timing offset. The second procedure involves using the CLF to fire an inclined laser (oriented towards the FD) at the same time a pulse from this laser is sent through an optical fibre to a nearby tank. Comparing the reconstructed laser firing times between the FD and SD detectors can reveal timing offsets.

Reconstruction of a hybrid event occurs upon the collection of a hybrid trigger (see Section 3.9.3.2). This involves using the pulse time from the SD station measuring the highest signal, and the FD data [42]. Combining the time of arrival at ground level given by the SD, with the longitudinal timing information from the FD, provides a high quality shower geometry reconstruction of the hybrid event. The resolution of the core location and arrival direction is 50 m and 0.6° respectively. These uncertainties are verified with CLF laser shots of known geometries [24].

The energy reconstruction technique is the same for hybrid events and FD-only events. Importantly, events which can be reconstructed as hybrid or SD-only events enable the SD energy scale to be derived from the FD energy scale [218].

Some studies require the hybrid exposure to be well known [50], such as the those of the energy spectrum. Using Monte Carlo simulations that take into account the status of every detector, weather conditions and other applicable factors over time, the aperture of the hybrid detector is calculated. The uncertainty of the calculated hybrid exposure is within 10% at $10^{18}\,\mathrm{eV}$ and 6% above $10^{19}\,\mathrm{eV}$.

3.9.5 Enhancements

3.9.5.1 HEAT

HEAT (High Elevation Auger Telescopes) consists of three fluorescence telescopes, 180 m north-east of the Coihueco site, pointing towards AMIGA (see Section 3.9.5.2) [213, 219]. These telescopes are almost identical to the standard FD telescopes, except they can be titled upwards by

29° and their electronics digitises at 20 MHz instead of 10 MHz. Identical to the standard FD telescopes, HEAT operates as an independent detector, with event data being merged with other detectors offline. When HEAT is combined with the standard FD telescopes of Coihueco, a field of view between 3° and 58° in elevation is observed. The stable operation of HEAT commenced in September 2009.

The purpose of HEAT is to facilitate hybrid measurements of events with energies as little as 10^{17} eV. Showers at these energies develop high in the atmosphere, therefore the shower maximum is normally outside the field of view of the standard FD telescopes and thus the standard telescopes are biased towards seeing deeper showers of particular geometries. Lower energy showers also produce a lower flux of fluorescence photons, consequently these showers are more likely to be detected close to a fluorescence telescope. Extending the viewing area to higher elevations allows an unbiased measurement of low energy showers.

3.9.5.2 AMIGA

AMIGA - Auger Muons and Infill for the Ground Array [219] - was designed to lower the energy threshold at which the SD array becomes fully efficient. This was achieved by building a sub-array with a detector spacing less than the 1.5 km spacing of the standard SD array. The sub-array is fully efficient above 3×10^{17} eV for showers of zenith angles less than 55°. The detectors are positioned on a triangular grid spanning an area of 23.5 km^2, with a spacing of 750 m. As the cosmic ray flux increases rapidly with decreasing energy, this small sub-array can collect sufficient statistics.

Within the 750 m sub-array, is a hexagon region comprised of 7 SD stations (one SD in the centre). These SD stations are paired with 30 m^2 plastic scintillator muon detectors that are buried underground at a depth of $\sim$ 2.3 m. There are also an extra 30 m^2 of plastic scintillators at two station locations so that the accuracy of the muon counting technique can be determined experimentally, and an extra 20 m^2 of plastic scintillators buried at $\sim$ 1.3 m at one station position to analyse shielding features [220]. The SD 750 m array consists of standard water-Čerenkov tanks that operate like the other SDs of the 1500 m array, but those with a corresponding scintillator have the additional task of providing a triggering signal to commence scintillator data recording. The scintillator detector array is designed to measure cosmic ray composition in a way that is compatible with the measurement of the depth of shower maximum by the FD. The scintillator output consists of a logical bit reporting when the signal measured exceeds an adjustable threshold. The scintillator output is sampled at a rate of 320 MHz.

The water-Čerenkov array of AMIGA was completed in September 2011, whereas the muon detectors have been operating since Febuary 2015 [221].

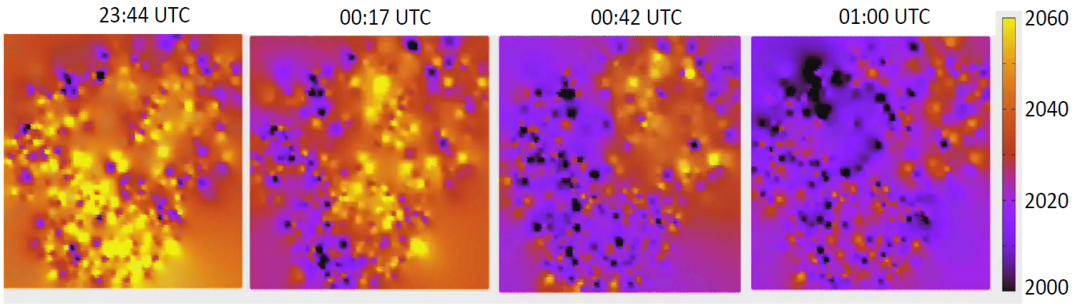
3.9.5.3 Additional Capabilities

The Pierre Auger Observatory has been useful for other studies not related to UHECRs [222].

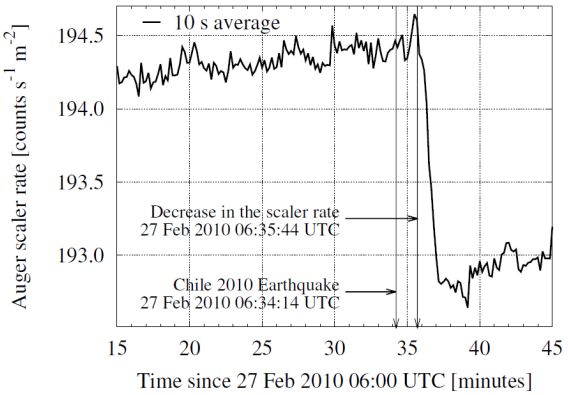
The scaler rate is the frequency at which an individual SD station is triggered. The scaler rate is about 2000 Hz per station, particles of low energies showers (around 1 GeV to 1 TeV) being the main cause of a trigger. Due to the large sample size, a change of the order of 0.1% is statistically significant.

Following a solar coronal mass ejection, the decrease in the low-energy cosmic ray flux (referred to as a Forbush decrease) is observed by the SD array, and these observations correlate with neutron monitor observations. These scaler rate measurements have enabled the Pierre Auger Observatory to contribute to the Space Weather program [223].

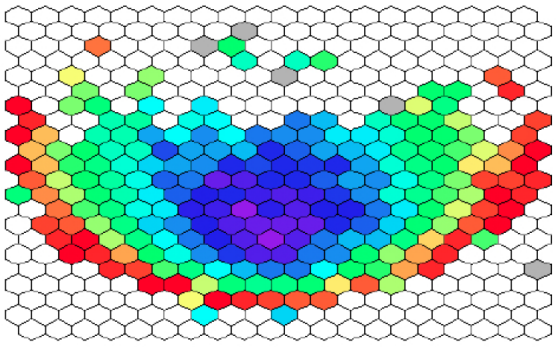
The scaler rates have also been observed to decrease during strong wind conditions (see Figure 3.12a), potentially due to static charge being deposited on the tanks. Following the 2010 Chile earthquake, a 0.8% decrease was observed in the average scaler rate lasting several hours



(a) Effect on the Auger SD scaler rate (units of Hz) of a terrestrial storm on 11 – 12 March 2009. From [222].



(b) Effect on the average Auger SD scaler rate of an 8.8-magnitude earthquake off the coast of central Chile. From [222].



(c) A standard ELVES event measured by the FD. The colours correspond to trigger times, with blue being the earliest time and red the latest. From [30].

Figure 3.12

(see Figure 3.12b). The cause of this phenomena may be understood through a collaboration with the proposed Malargüe seismic array [224].

The FD array is able to detect ELVES (Emissions of Light and Very low frequency perturbations due to Electromagnetic pulse Sources). These are transient luminous events, originating from lightning discharges causing intense electromagnetic pulses at altitudes of $80-95$ km. These events are observable as rapidly expanding quasi-circular fronts (see Figure 3.12c). In 2014, 581 events were measured, and 305 events were observed in the last nine months of 2013, by the FD array. These events were caused by lightning at distances of $300-1000$ km [30].

3.9.6 Future Direction

3.9.6.1 Auger Engineering Radio Array

The Auger Engineering Radio Array (AERA) [225] is designed for the purpose of detecting radio emissions of $30-80$ MHz from extensive air showers of energy greater than 10^{17} eV. AERA consists of 153 radio stations of varying spacing, spanning an area of 17 km^2 within the AMIGA region. The aim is to observe events with four types of detector simultaneously (water-Čerenkov SD, buried scintillation detector, FD, radio detector). By March 2015, 509 SD+AMIGA+AERA events were detected, and 51 SD+AMIGA+FD+AERA events were detected.

3.9.6.2 AugerPrime

AugerPrime (Auger Primary cosmic Ray Identification with Muons and Electrons) [220] is a planned upgrade to the Pierre Auger Observatory. AugerPrime will involve the installation of a $3.8\,\mathrm{m}^2$ plastic scintillator plane above each SD station [226]. This detector is named the SSD (surface scintillator detector), with each SSD unit consisting of a $3.8\,\mathrm{m} \times 1.3\,\mathrm{m}$ box comprised of two scintillator sub-modules. The differing responses of the water-Čerenkov detectors and plastic scintillators to the muonic and electromagnetic components of an EAS will allow these components to be easily distinguished. The muon number will be used to estimate the primary mass composition on an event-by-event basis.

The main objective of the upgrade is to enable the determination of the mass composition of cosmic rays of the highest energy with good resolution, even with only the SD array. If a fraction of the highest-energy cosmic rays are determined to be protons, AugerPrime will be able to provide proton-enriched samples that can be used in anisotropy studies. The expected magnetic deflections of the highest energy protons are very small. Therefore if 10% of the highest-energy cosmic rays are protons, half of which are deflected by less than 3°, then within the anticipated lifetime of the experiment a correlation between the arrival directions and UHECR sources is predicted with a significance $\gg 3\sigma$ (see Figure 3.13).

In March 2016, an engineering array of scintillator detectors was deployed, with the remainder of the array being filled from $2016 - 2018$. Full data taking will commence later in 2018 and end in 2024, providing a $40\,000\,\mathrm{km}^2\,\mathrm{sr\,yr}$ total exposure. The muon number estimation from the Čerenkov-scintillator measurements will be fine tuned by using the buried muon detectors of AMIGA (see Section 3.9.5.2). Furthermore, the duty cycle of the FD will be increased by $\sim 50\%$ by operating the FD during periods of higher night sky background.

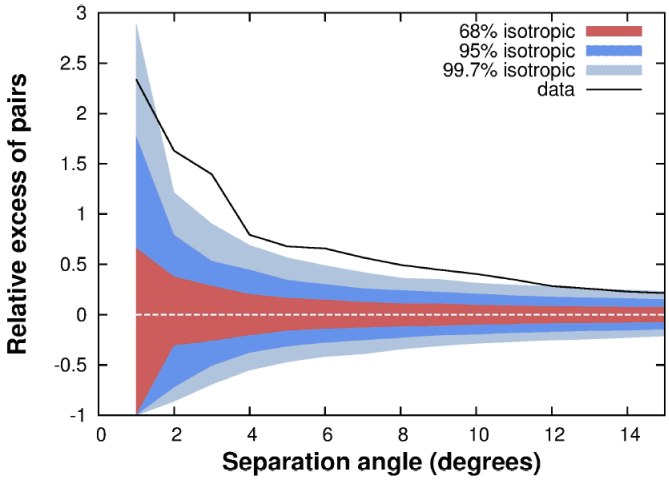


Figure 3.13: The predicted angular correlation between simulated AugerPrime events above 40 EeV from a proton-enriched sample, and AGNs from the Swift-BAT catalogue. This prediction assumes 10% of all events above 40 EeV are protons, of which 50% come within 3% of an AGN in the catalogue. From [220].

Chapter 4

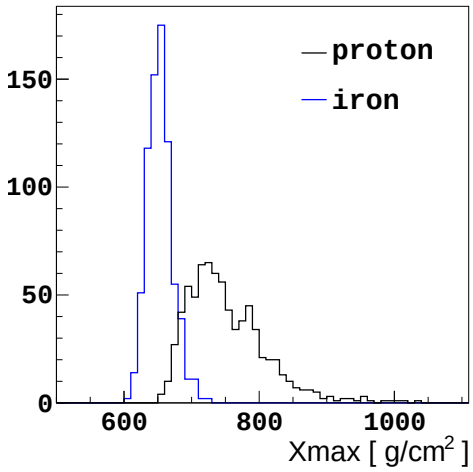
Using Xmax to infer the primary mass composition

4.1 Analysing $X_{\max}$ distributions

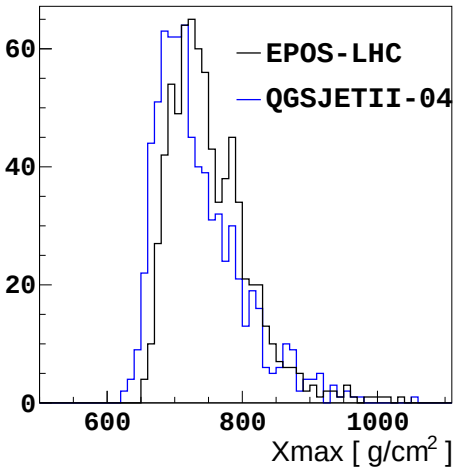
Knowledge of the composition of cosmic rays, and its evolution with energy, is an important ingredient in our understanding of cosmic rays. Investigating the primary mass composition of UHECRs can further our understanding of the origin of these highest energy cosmic rays, their acceleration and propagation, and the nature of the UHECR spectral ankle and suppression. A measured observable commonly used to extract information on the mass composition of UHECRs reaching Earth is $X_{\max}$, the atmospheric depth in g/cm^2 from the top of the atmosphere where the longitudinal development of a cosmic ray extensive air shower reaches the maximum number of particles. Different cosmic ray primaries propagate through the atmosphere differently, therefore the $X_{\max}$ distribution from extensive air showers initiated by a particular primary mass will be unique [227]. Differences in the mode, width and tail of the $X_{\max}$ distributions of different primaries allows information on the mass composition distribution of the events, in addition to hadronic interaction information, to be extracted from an examination of the observed $X_{\max}$ distributions at different energies. [228, 229].

Due to statistical variability in the interaction between cosmic rays of a specific primary mass and the atmosphere, a cosmic ray's primary mass cannot be determined on an event-by-event basis by examining $X_{\max}$. Figure 4.1a illustrates this, showing the $X_{\max}$ distribution resulting from the CONEX v4r37 [230,231] simulation of 750 proton events, and separately 750 iron events, of 10^{18} eV according to the EPOS-LHC hadronic interaction model. We can see that even for cosmic rays of vastly contrasting mass, there is considerable overlap in their $X_{\max}$ distributions. However, the predicted $X_{\max}$ distribution of protons has distinct properties compared to the iron distribution, with the proton distribution having a larger mode, larger spread and larger tail. These distinguishable shape characteristics mean the mass composition distribution of the cosmic rays forming an $X_{\max}$ distribution can be inferred. By utilising the predictions of a hadronic interaction model to parameterise the expected $X_{\max}$ distributions of specific primaries, and by fitting this $X_{\max}$ distribution parameterisation to the observed data, mass composition information is obtained that is based upon the predictions of that hadronic interaction model.

This work will focus on the predicted $X_{\max}$ distributions from the EPOS-LHC [232] , QGSJetII-04 [233] and Sibyll2.3 [234] hadronic interaction models. The Sibyll2.3 hadronic interaction model was only available towards the end of my research, thus this model has been implemented less throughout this work. The particle interaction predictions of these models determine the predicted $X_{\max}$ distributions. These predictions are based on accelerator measurements of particles of energies much lower than the cosmic ray energies measured by Auger. By extrapolating from those lower energy measurements, predictions of the particle interactions at higher energies are formed. The extrapolations towards higher energies differ between hadronic interaction models,



(a) An Xmax distribution of 750 proton events (black), and separately 750 iron events (blue), of energy 10^{18} eV simulated according to the EPOS-LHC hadronic interaction model.



(b) An Xmax distribution of 750 EPOS-LHC simulated proton events (black), and separately 750 QGSJetII-04 simulated proton events (blue), of energy 10^{18} eV.

Figure 4.1

leading to differences in the predicted $X_{\max}$ distributions of UHECRs. The predicted $\langle X_{\max} \rangle$ is particularly sensitive to the model predictions of the following particle interaction properties: total cross section (probability of interaction), multiplicity (number of produced particles), elasticity (the fraction of energy carried away by leading particles) and the ratio of neutral to charged pions created in the interaction [235].

Measurements by the Large Hadron Collider (LHC) have contributed to reducing the separation between the high energy extrapolations of the different hadronic models. For example, Figure 4.2 illustrates the reduced divergence in the proton-air inelastic cross section predictions of different hadronic models, before and after LHC constraints. Figure 4.3 illustrates the proton and iron $\langle X_{\max} \rangle$ predictions of various hadronic models, demonstrating that although the LHC measurements have contributed to reducing the difference between the model predictions of some particle interaction properties, the separation between the $\langle X_{\max} \rangle$ predictions of the hadronic models is still considerable.

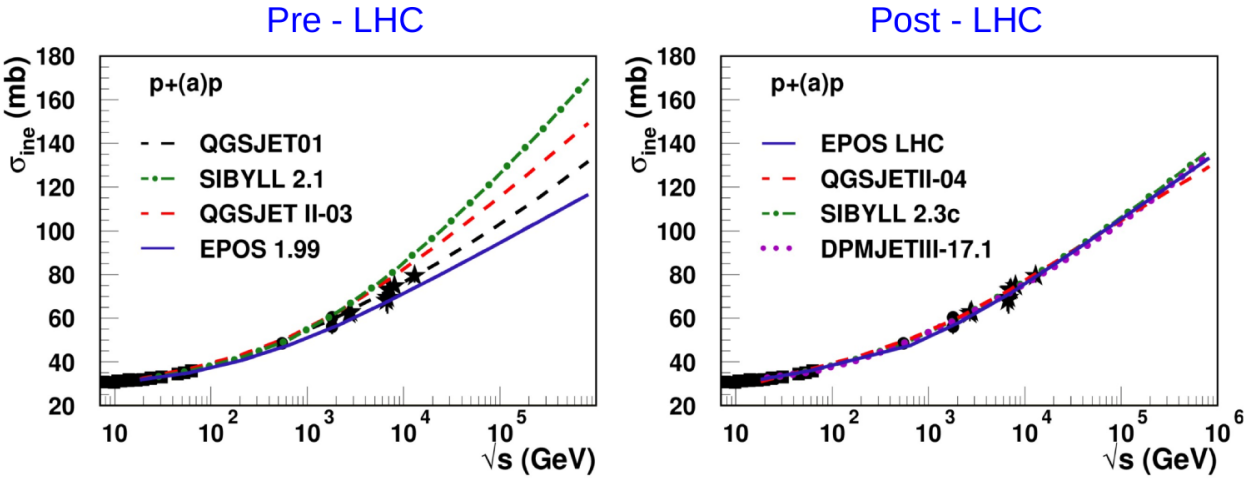


Figure 4.2: Proton-air inelastic cross section predictions of different hadronic models, pre and post LHC measurements. From [236].

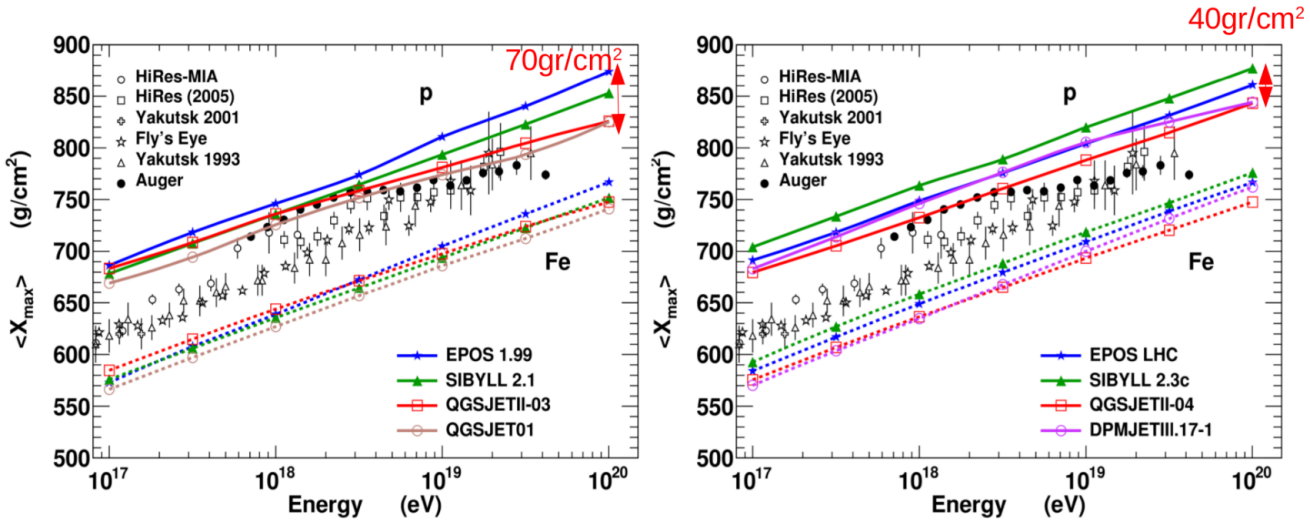


Figure 4.3: Proton and iron $\langle X_{\max} \rangle$ predictions of different hadronic models, pre (left) and post (right) LHC measurements. The spread between the model predictions is denoted by the red arrow and accompanying text at the top right of each plot. From [236].

Figure 4.1b shows the $X_{\max}$ distribution resulting from the simulation of 750 proton events according to the EPOS-LHC model, and separately 750 proton events according to the QGSJetII-04 model, of energy 10^{18} eV. This figure illustrates the differences in the $X_{\max}$ distribution predicted by different post-LHC hadronic interaction models. Most noticeable is the difference in the modes of the distributions, but there are also marginal differences in the width and tails of the distributions. These differences between the hadronic interaction models change with energy to some degree. Although the dissimilarity between these predicted distributions may appear minor, applying a parameterisation based on these different predictions to measured data can have a considerable impact on the mass composition inferred. Consequently, typical mass composition studies of $X_{\max}$ are strongly dependent on the hadronic interaction model assumed.

This work attempts to extract mass composition information that is partially independent of the hadronic model assumed, by fitting some of the coefficients which describe the shape properties of the predicted $X_{\max}$ distributions of specific primaries. Therefore, in addition to a semi-model independent mass composition result, information on the $X_{\max}$ distribution shape properties will be obtained, which relates back to the cosmic ray hadronic interaction predictions.

4.2 Proposed four component composition $X_{\max}$ parameterisation

Given the shape characteristics of the predicted $X_{\max}$ distributions for different primaries, the $X_{\max}$ distributions measured by the Pierre Auger Observatory can be well reproduced by assuming a composition of at least four components consisting of proton, helium, nitrogen and iron [145, 147, 237]. Therefore, in this work we will fit $X_{\max}$ distributions with the relative amounts of p, He, N and Fe required to reproduce the total distribution from these four separate component distributions, thereby retrieving the primary mass composition distribution.

The algorithm CONEX v4r37, along with the hadronic interaction packages EPOS-LHC, QGSJetII-04 and Sibyll2.3, were used to simulate air showers for the purpose of obtaining the predicted $X_{\max}$ distributions of cosmic rays of some primary mass and energy according to these hadronic interaction models. By fitting this Monte Carlo (MC) simulated $X_{\max}$ data, we have developed a parameterisation which describes the expected $X_{\max}$ distribution according to either the EPOS-LHC, QGSJetII-04 or Sibyll2.3 models for cosmic rays of some energy and mass. Our

parameterisation of the X_{max} distributions can then be fitted to observed X_{max} distributions to extract primary mass composition information (mass fractions) from each energy bin. The CONEX v4r37 X_{max} utilised throughout this work is the X-position of maximum of the quadratic fit of the dE/dX(X) profile (XmxdEdX in CONEX v4r37).

An X_{max} distribution of some primary energy and mass can be modelled as the convolution of a Gaussian with an exponential [238]. Three shape parameters (t_0, σ, λ) define the X_{max} distribution:

$$\frac{dN}{dX_{\mathrm{max}}} = \frac{1}{2\lambda} \exp\left(\frac{t_0 - t}{\lambda} + \frac{\sigma^2}{2\lambda^2}\right) Erfc\left(\frac{t_0 - t + \frac{\sigma^2}{\lambda}}{\sigma\sqrt{2}}\right) , \tag{4.1}$$

where t_0 defines the mode of the Gaussian component, σ defines the width of the Gaussian component and λ defines the exponential tail of the X_{max} distribution, and t is the X_{max} bin. The mode and spread of the distribution defined in Equation (4.1) is sensitive to t_0 and σ respectively.

We fit Equation (4.1) to simulated X_{max} distributions of a particular primary energy and mass (either proton, helium, nitrogen or iron primaries) according to either the EPOS-LHC, QGSJetII-04 or Sibyll2.3 hadronic interaction model (these fits are displayed in Appendix A), obtaining the values of t_0, σ and λ for that distribution. These shape parameter (t_0, σ and λ) results are displayed in Figure 4.4, with the solid lines displaying the fits to the shape parameters as a function of energy. The functions fitted are defined as follows:

$$\begin{aligned} t_0(E) &= t_{0_{\mathrm{norm}}} + B \cdot \log_{10}\left(\frac{\log_{10} E}{\log_{10} E_0}\right) , \\ \sigma(E) &= \sigma_{\mathrm{norm}} + C \cdot \log_{10}\left(\frac{E}{E_0}\right) , \\ \lambda(E) &= \lambda_{\mathrm{norm}} - K + K \cdot \left(\frac{\log_{10} E}{\log_{10} E_0}\right)^{\frac{L}{\ln 10}} , \end{aligned} \tag{4.2}$$

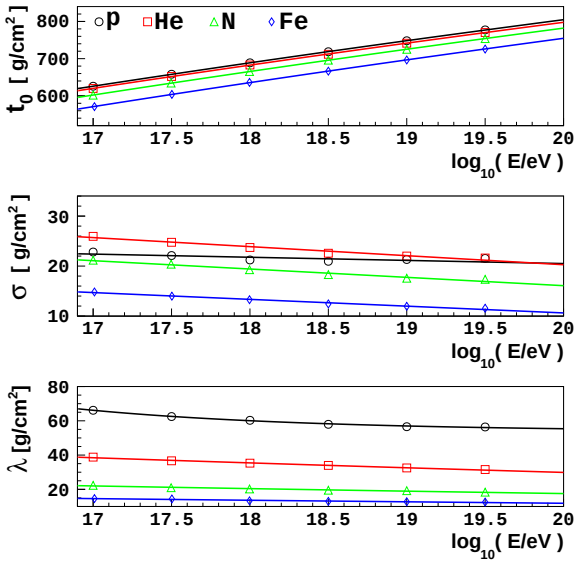
where E is the energy in eV and $E_0 = 10^{18.24}$ eV, the energy at which we choose to normalise the equations. The coefficients in Equation (4.2) are constants, and are specified in Table 4.1 for each mass component according to a particular hadronic model.

The functions of Equation (4.2) consist of two parts, the first part defining the value of a shape parameter at the normalisation energy, and the second part defining the change in the shape parameter as a function of energy. For example, $t_{0_{\mathrm{norm}}}$ for protons would be the value of t_0 for protons at $10^{18.24}$ eV, and similarly σ_{norm} would be the value of σ at $10^{18.24}$ eV. The functions in Equation (4.2) parameterise t_0, σ and λ in Equation (4.1).

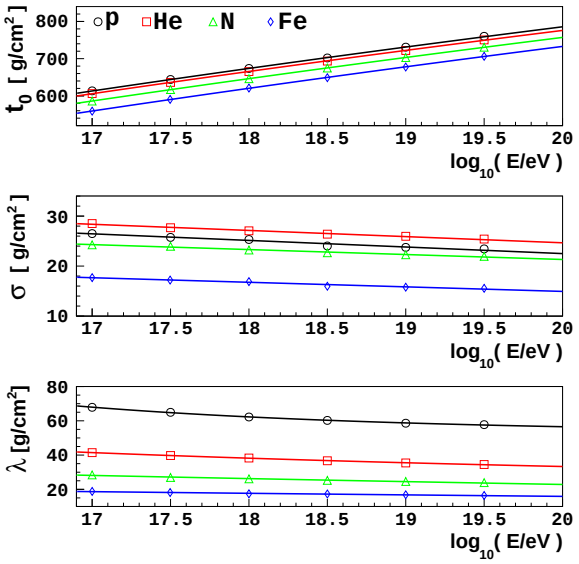
The normalisation energy of $10^{18.24}$ eV corresponds to the energy at which Auger has measured λ for a proton dominated composition [229]. This means that λ_{norm} for protons can be directly compared with Λ_η, the exponential tail measured by Auger, which is shown in Equation (4.3). We even considered adopting Λ_η as the value for λ_{norm}, but this could potentially break self consistency in the models.

$$\Lambda_\eta = [55.8 \pm 2.3(stat) \pm 1.6(sys)] \ \mathrm{g/cm^2} \tag{4.3}$$

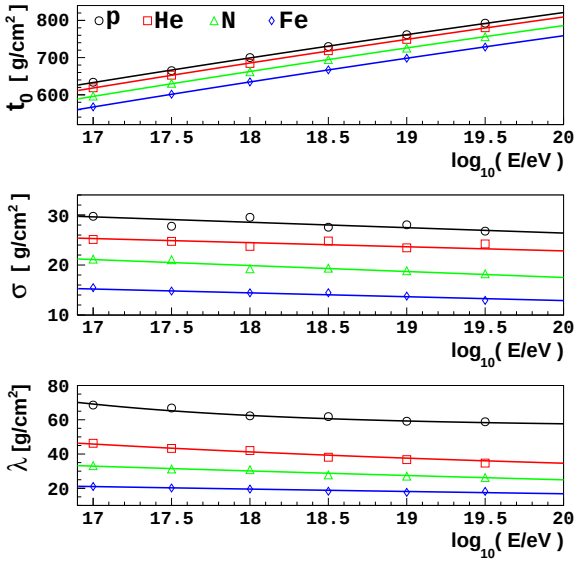
Table 4.1 defines the shape parameter functions for each mass. The shape parameter functions for the four masses could be parameterised by a single shape parameter function containing E and A terms to describe the shape parameter value expected for some energy and mass according to a model. A single function describing the four masses however does not describe the shape parameter values for each mass as adequately as the separate equations for each mass, and therefore this approach is discarded.



(a) EPOS-LHC



(b) QGSJetII-04



(c) Sibyll2.3

Figure 4.4: The shape parameter results (markers) from fits to MC $X_{\max}$ data according to the EPOS-LHC (left), QGSJetII-04 (right) and Sibyll2.3 (bottom) hadronic interaction models. Fitted to the shape parameter results for each primary mass and hadronic model are the functions in Equation (4.2), displayed by coloured lines.

EPOS-LHC	Proton	Helium	Nitrogen	Iron
$t_{0_{\text{norm}}}$	703.41	697.07	680.13	650.56
B	2533.29	2515.46	2548.31	2603.31
σ_{norm}	21.61	23.46	19.01	13.01
C	-0.63	-1.81	-1.67	-1.36
λ_{norm}	59.12	34.74	20.06	13.41
K	5.80	-1913.99	-1828.11	-1406.72
L	-25.93	0.063	0.035	0.027
QGSJetII-04	Proton	Helium	Nitrogen	Iron
$t_{0_{\text{norm}}}$	687.87	679.44	660.20	634.47
B	2444.88	2410.38	2422.37	2460.32
σ_{norm}	24.82	26.83	23.07	16.54
C	-1.32	-1.24	-0.99	-0.91
λ_{norm}	61.29	37.5	25.84	17.46
K	9.35	19.32	-1818.36	-986.08
L	-17.63	-6.08	0.041	0.040
Sibyll2.3	Proton	Helium	Nitrogen	Iron
$t_{0_{\text{norm}}}$	714.46	701.03	678.18	650.19
B	2666.31	2705.43	2695.22	2714.41
σ_{norm}	28.30	24.28	19.61	14.24
C	-1.08	-0.82	-1.20	-0.77
λ_{norm}	61.52	40.31	29.48	19.20
K	5.81	23.70	-1362.17	-1349.93
L	-27.47	-6.84	0.083	0.044

Table 4.1: Parameters adopted in our EPOS-LHC, QGSJetII-04 and Sibyll2.3 X_{max} distribution parameterisations, assuming a normalisation energy of $E_0 = 10^{18.24}$ eV.

4.2.1 Accounting for the detector resolution and acceptance

The measured X_{max} distributions are affected by the detector resolution and the detector acceptance. The X_{max} resolution quantifies the broadening of the original distribution by statistical fluctuations of the reconstructed X_{max} around the true X_{max}. The X_{max} acceptance defines the relative probability of the detector to detect and reconstruct an event of some measured X_{max}. The Pierre Auger X_{max} publication [145] provides parametrisations for the average detector X_{max} resolution as a function of energy, $Res(E)$, and the detector acceptance as a function of X_{max} for each energy bin, $Acc(E,t)$, where t is the X_{max} bin as in Equation (4.1).

The detector X_{max} resolution is accounted for in the expected X_{max} parameterisation by adding the detector resolution in quadrature with the corresponding $\sigma(E)$ (from Equation (4.2)), to provide the total expected value of $\sigma(E)_{tot}$ for some primary (used in Equation (4.1)):

$$\sigma(E)_{tot} = \sqrt{\sigma(E)^2 + Res(E)^2} \tag{4.4}$$

The acceptance profile of the measured X_{max} distributions is also accounted for by adjusting the predicted X_{max} distribution. In the fit procedure, the expected total X_{max} distribution (from the sum of the four mass components, and normalised to an area of 1) is multiplied by the X_{max} acceptance function of that energy bin, and the resulting distribution is then renormalised to 1. The expected X_{max} distribution is then multiplied by the number of events in the energy bin before being compared with the measured X_{max} distribution. Therefore the fit procedure leaves the measured distribution unchanged.

We can combine Equations (4.1), (4.2), (4.4) and the detector acceptance $Acc(E,t)$ to obtain the expected $X_{\max}$ distribution for cosmic rays of a mixture of primary masses in a particular energy bin according to a hadronic interaction model:

$$\left.\frac{dN}{dX_{\max}}(E,t)\right|_{\text{total}} = N(E)Acc(E,t)\sum_{i=p,He,N,Fe} c_i(E)f_i(E)\left.\frac{dN}{dX_{\max}}(E,t)\right|_i \tag{4.5}$$

where $f_p(E)$, $f_{He}(E)$, $f_N(E)$ and $f_{Fe}(E)$ are the fractions of proton, helium, nitrogen and iron events respectively, $c_i(E)$ are the normalisation factors for each mass component, and $N(E)$ is the total number of events. The Poissonian log-likelihood function minimised is:

$$P(E) = -2\sum_{t=0}^{1300}\ln\left(\frac{M(E,t)^{D(E,t)}e^{-M(E,t)}}{D(E,t)!}\right) \tag{4.6}$$

where E is the energy bin, t is the $X_{\max}$ bin, $D(E,t)$ is the data value of an $X_{\max}$ bin and $M(E,t) = \left.\frac{dN}{dX_{\max}}(E,t)\right|_{\text{total}}$ (from Equation (4.5)) is the parameterisation prediction for an $X_{\max}$ bin. The fractions f_p, f_{He}, f_N and f_{Fe} are all correlated. Thus, the range of allowed fraction values is not always $[0,1]$. This range changes depending on the values of the other fractions. For example, if f_p were 0.9, the allowed range for any of the other fractions would be $[0,0.1]$. In order to avoid changing the fraction limits in an iterative way, we have expressed the fractions f_p, f_{He}, f_N and f_{Fe} in terms of η_1, η_2 and η_3 as follows:

$$\begin{aligned} f_p(E) &= \eta_1 \\ f_{He}(E) &= (1-\eta_1)\eta_2 \\ f_N(E) &= (1-\eta_1)(1-\eta_2)\eta_3 \\ f_{Fe}(E) &= 1 - f_p(E) - f_{He}(E) - f_N(E) \end{aligned} \tag{4.7}$$

Therefore, each energy bin has a set of η_1, η_2 and η_3 which defines the mass fractions of that energy bin. The allowed range for η_1, η_2 and η_3 is always $[0,1]$, consequently the mass fractions are constrained to values between 0 and 1 whilst the sum of the mass fractions equals 1. Therefore, in practice we fit η_1, η_2 and η_3 to determine the corresponding fractions (f_p, f_{He}, f_N, f_{Fe}).

These four primary mass components best represent the range of cosmic ray masses. At opposite ends of the mass range, protons and iron nuclei are the most stable cosmic ray particles, while helium and nitrogen represent the intermediate range of nuclear masses.

For an individual $X_{\max}$ distribution described by Equation (4.1), the first two moments of the $X_{\max}$ distribution are defined as follows:

$$\langle X_{\max}\rangle = t_0 + \lambda \tag{4.8}$$

$$\sigma(X_{\max}) = \sqrt{\sigma^2 + \lambda^2} \tag{4.9}$$

The fits of Equation (4.1) to the MC $X_{\max}$ data used to define our $X_{\max}$ parameterisations are shown in Figures A.1, A.2 and A.3 of Appendix A. The difference in the $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ of the MC data versus the equation fitted to the data (Equation (4.1) with Table 4.1) is shown in Figure 4.5. Although Equation (4.1) does not always describe the fitted data precisely, $\langle \mathrm{X}_{\max}\rangle_{\text{fit}}$ is always within 0.1 g/cm^2 of $\langle \mathrm{X}_{\max}\rangle_{\text{data}}$. The values of $\sigma(\mathrm{X}_{\max})_{\text{fit}}$ are always within 3 g/cm^2 of $\sigma(\mathrm{X}_{\max})_{\text{data}}$, which is acceptable. The $\langle \mathrm{X}_{\max}\rangle$ of the distribution is the main characteristic we endeavour to define accurately, as the fitted mass composition is quite sensitive to the predicted $\langle X_{\max}\rangle$ of our parameterisations.

Figure 4.6 displays the $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ predictions of the three parameterisations for each primary. The predicted $\langle X_{\max}\rangle$ separation of each adjacent mass component (eg. proton vs.

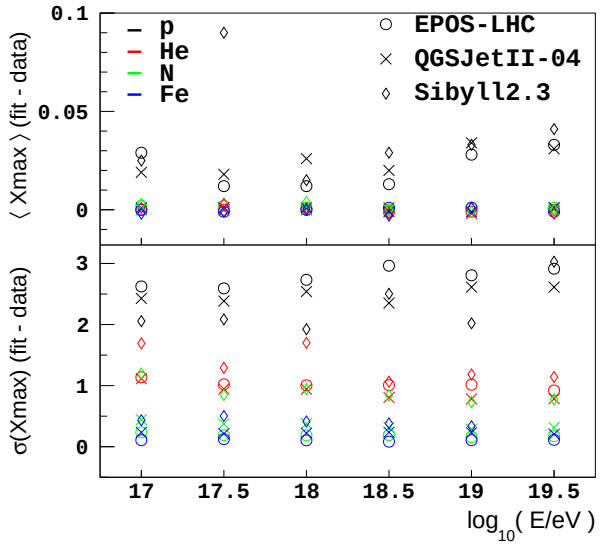


Figure 4.5: Difference in the $\langle X_{max} \rangle$ and $\sigma(X_{max})$ of the MC X_{max} data compared to those derived from the fitted equation. The y-axes have units of g/cm^2.

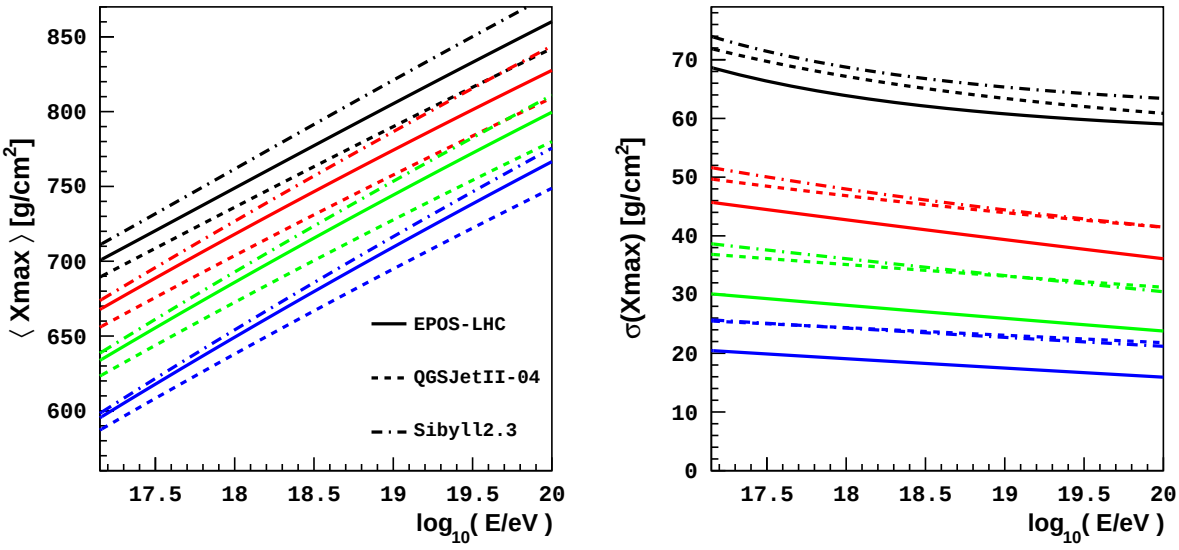


Figure 4.6: The $\langle X_{max} \rangle$ and $\sigma(X_{max})$ predictions of the EPOS-LHC, QGSJetII-04 and Sibyll2.3 X_{max} parameterisations for proton (black), helium (red), nitrogen (green) and iron (blue).

helium, helium vs. nitrogen) within a parameterisation is approximately 30 g/cm^2 to 40 g/cm^2. The predicted $\sigma(X_{max})$ of the primaries is much larger for the QGSJetII-04 and Sibyll2.3 parameterisations compared with the EPOS-LHC parameterisation.

The parameters of Equation (4.5) are fitted to energy binned X_{max} distributions by applying a Poisson log likelihood minimisation. The X_{max} distributions consist of 1 g/cm^2 bins, and the energy binning of the X_{max} distributions is typically 0.1 in $\log_{10}(E/\text{eV})$. The importance of the bin sizes and the reasoning for this binning is discussed in Section 4.5.

4.3 Validation of the parameterisation

Figure 4.7 displays the mass composition results from fitting our EPOS-LHC, QGSJetII-04 or Sibyll2.3 $X_{\max}$ parameterisations to $X_{\max}$ data measured by the Pierre Auger Observatory fluorescence detector (FD) [145]. The error bars represent the standard deviation in the reconstructed mass composition of fitted mock $X_{\max}$ data sets; mock data that was generated from the parameterisation fitted to the initial $X_{\max}$ data (the Auger $X_{\max}$ distributions). The $X_{\max}$ bin size used was $1\,\mathrm{g/cm^2}$, and the detector resolution and acceptance were taken into account. The mass composition obtained using our $X_{\max}$ parameterisations is consistent with the Auger analysis of the same 2014 FD $X_{\max}$ data set (see [147] and Figure 2.13), where $X_{\max}$ distribution templates from hadronic interaction models were compared to the data. The compatibility of our results with the 2014 Auger analysis validates the accuracy of our $X_{\max}$ parameterisations.

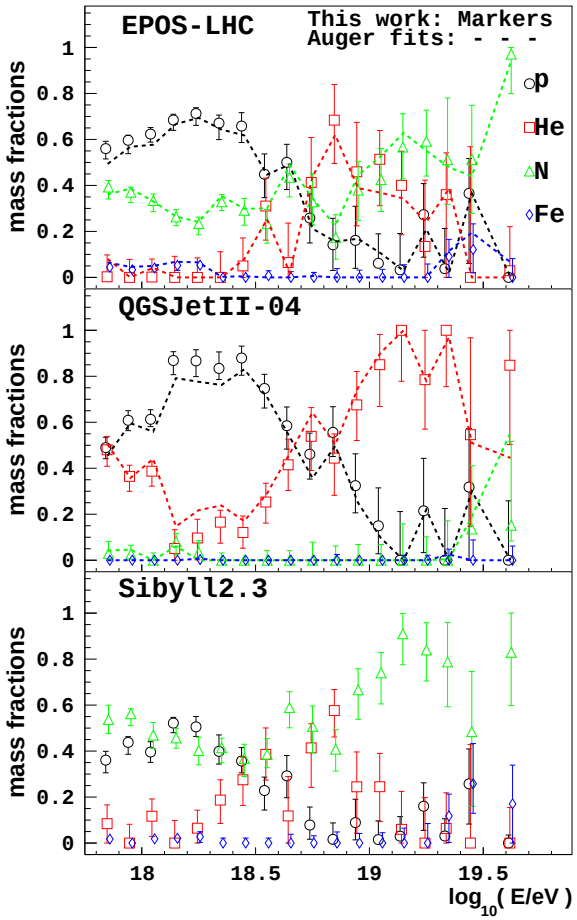
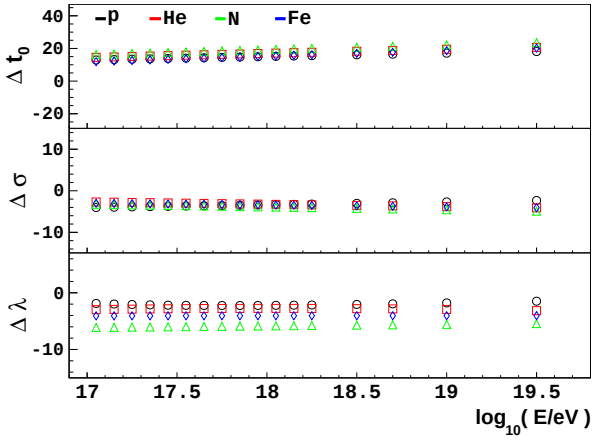


Figure 4.7: Fitting only the mass fractions of our parameterisations to FD $X_{\max}$ data measured by the Pierre Auger Observatory [145]. Included is the mass composition results from the Pierre Auger Observatory analysis using the EPOS-LHC and QGSJetII-04 hadronic interaction models (labelled 'Auger fits'). [147].

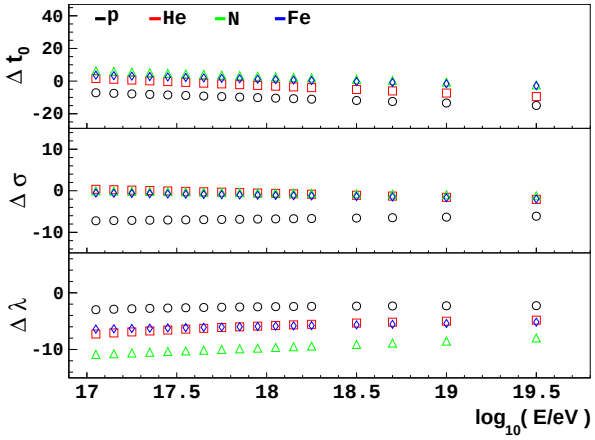
4.4 Method of fitting the shape parameter properties

The coefficients of Equation (4.2), shown in Table 4.1, were obtained from a global fit which included all energy bins. When fitting the $X_{\max}$ distribution data with the mass fraction parameters using our EPOS-LHC, QGSJetII-04 or Sibyll2.3 parameterisation, with the shape coefficients

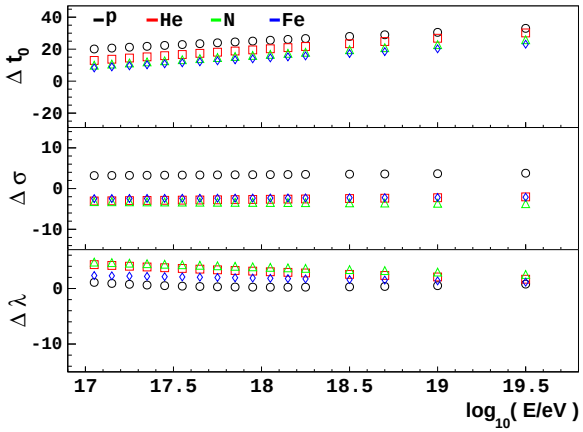
fixed (as in Figure 4.7), the resulting mass composition reflects the characteristics of the hadronic model utilised. Additionally, the mass composition fitted to each energy bin is independent of the mass composition fitted to other energy bins. However, by allowing some of the coefficients shown in Table 4.1 to be fitted, in addition to fitting the mass composition fractions, the mass composition obtained has a reduced dependence on the hadronic interaction model assumed. In this alternative case the mass composition fitted at each energy bin has some dependence with the fits at other energy bins. This is because the fitted shape coefficients of the $X_{\max}$ parameterisation are fitted using all energy bins, while in the first case these coefficients were fixed.



(a) EPOS-LHC shape parameter value minus QGSJetII-04 shape parameter value.



(b) EPOS-LHC shape parameter value minus Sibyll2.3 shape parameter value.



(c) Sibyll2.3 shape parameter value minus QGSJetII-04 shape parameter value.

Figure 4.8: Comparison of the shape parameter values of different $X_{\max}$ parameterisations for some energy and mass. The y-axes have units of g/cm^2.

In principle, if we were able to use the Auger $X_{\max}$ data to perform a global fit of the mass composition and all of the coefficients from Equation 4.2, the resulting composition would be almost independent of the hadronic models, only depending on the assumed functional form of the shape parameter equations. However, the degeneracy between the fitted mass fractions and the coefficients makes it impossible to unambiguously constrain all of these parameters (i.e. the solution would be degenerate). Therefore, we need to identify which coefficients are most relevant for interpreting the mass composition, and evaluate whether we can unambiguously fit these coefficients and the mass composition. One way to identify which coefficients to include in a global fit is to compare the values of t_0, σ and λ for different models. This comparison will identify the parameters that are better or poorly constrained by our current knowledge of the

high energy hadronic interaction physics.

Figure 4.8 illustrates the t_0, σ and λ differences between the EPOS-LHC, QGSJetII-04 and Sibyll2.3 parameterisations at some energy and mass. The differences as a function of energy are relatively small. For example, the slope of Δt_0 as a function of energy is less than $5\,\mathrm{g/cm^2}$ /energy-decade in Figure 4.8a, which is small compared to a t_0 elongation rate of $60\,\mathrm{g/cm^2}$/energy-decade. We have also verified that the separation between different primaries in the t_0, σ and λ space is similar for the three tested models. It remains that the main differences between our EPOS-LHC, QGSJetII-04 and Sibyll2.3 X_{max} parameterisations are the normalisation of t_0 and σ. Therefore, when including $t_{0_{\mathrm{norm}}}$ and σ_{norm} in the global fit, we should obtain a similar interpretation of the mass composition when either the EPOS-LHC, QGSJetII-04 or Sibyll2.3 X_{max} distribution parameterisation is fitted to data. We choose to fit $t_{0_{\mathrm{norm}}}$ and σ_{norm} in the following way:

- $t_{0_{\mathrm{norm}}}$ is fitted such that the absolute values of $t_{0_{\mathrm{norm}}}$ for each primary change by the same amount. Therefore, the difference in $t_{0_{\mathrm{norm}}}$ (and t_0) between primaries is conserved.
- σ_{norm} is fitted such that the ratio of σ (and σ_{norm}) between primaries remains similar to the initial ratio over the energy range (differences in C between primaries prevents the exact conservation of the initial ratio). Therefore, if σ_{norm} for protons changes by Δ, σ_{norm} for other primaries will change by Δ multiplied by the initial average ratio of σ between that primary and protons.

Fitting $t_{0_{\mathrm{norm}}}$ and σ_{norm} with this approach assumes the hadronic models are correctly predicting the separation in t_0 between different elements, and the ratio of σ between different elements, over the fitted energy range.

In Equation (4.2), the values of the shape parameters for helium, nitrogen and iron can be expressed in terms of the corresponding values for protons, therefore fitting $t_{0_{\mathrm{norm}}}$ and σ_{norm} in the way described above can be implemented by simply fitting $t_{0_{\mathrm{norm}}}$ and σ_{norm} for protons.

Figures 4.8b and 4.8c illustrate the difference between the shape parameter predictions of EPOS-LHC compared to Sibyll2.3, and Sibyll2.3 compared to QGSJetII-04, respectively. Across primaries, the differences between the EPOS-LHC and QGSJetII-04 shape parameter predictions (shown in Figure 4.8a) are more consistent than the differences between the Sibyll2.3 shape parameter predictions compared to the EPOS-LHC or QGSJetII-04 predictions. In particular, the separation of t_0 between different primaries according to Sibyll2.3 is not as similar to the EPOS-LHC and QGSJetII-04 predictions, and this is the shape parameter that most influences the fitted composition. Therefore, when fitting $t_{0_{\mathrm{norm}}}$ and σ_{norm} of the Sibyll2.3 parameterisation to either EPOS-LHC or QGSJetII-04 simulated X_{max} data, or vice versa, we do not expect to obtain mass composition results which are as accurate as the fits of our EPOS-LHC or QGSJetII-04 parameterisation to QGSJetII-04 or EPOS-LHC X_{max} data respectively.

In order to avoid unphysical fit results, we constrain the possible fitted values for $t_{0_{\mathrm{norm}}}$ and σ_{norm}. These constraints are significantly wider than the range of the EPOS-LHC, QGSJetII-04 and Sibyll2.3 X_{max} parameterisation predictions for these coefficients. The predicted value of $t_{0_{\mathrm{norm}}}$ for protons according to EPOS-LHC is $\sim 703\,\mathrm{g/cm^2}$, according to QGSJetII-04 is $\sim 688\,\mathrm{g/cm^2}$, and according to Sibyll2.3 is $\sim 714\,\mathrm{g/cm^2}$. The minimum and maximum limits of $t_{0_{\mathrm{norm}}}$ for protons are typically set to $670\,\mathrm{g/cm^2}$ and $725\,\mathrm{g/cm^2}$ respectively. The predicted value of σ_{norm} for protons according to EPOS-LHC, QGSJetII-04 and Sibyll2.3 is $\sim 22\,\mathrm{g/cm^2}$, $\sim 25\,\mathrm{g/cm^2}$ and $\sim 28\,\mathrm{g/cm^2}$ respectively. The minimum and maximum limits of σ_{norm} for protons are set to $5\,\mathrm{g/cm^2}$ and $45\,\mathrm{g/cm^2}$ respectively.

With a suitable shift in $t_{0_{\mathrm{norm}}}$, many primary mixtures which produce a fairly smooth total distribution can be fitted well with a single dominant distribution, instead of a sum of distributions. On the other hand, a distribution dominated by a single primary can be well fitted by a balanced mixture of distributions when $t_{0_{\mathrm{norm}}}$ is shifted appropriately. It is common that X_{max} distributions can be fitted with a value of $t_{0_{\mathrm{norm}}}$ for protons larger than $725\,\mathrm{g/cm^2}$, which results

in the primary mass of the events being overestimated (i.e. biased towards heavier masses). This was noticed in some of the fits of simulated $X_{\max}$ data in Chapter 5. Therefore, we limit the maximum fitted value of $t_{0_{\text{norm}}}$ for protons to approximately 725 g/cm^2, about 11 g/cm^2 above the Sibyll2.3 prediction, which is a significant difference considering the separation between protons and helium in $t_{0_{\text{norm}}}$ according to EPOS-LHC is only ~ 6 g/cm^2.

We have evaluated the performance of fitting $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions using simulated $X_{\max}$ distributions of a known composition (presented in more detail in Chapter 5). Provided there is enough dispersion of masses in the data and enough statistics, it is possible to accurately fit $t_{0_{\text{norm}}}$, σ_{norm} and the corresponding abundance (fractions) of p, He, N and Fe. An important achievement from including $t_{0_{\text{norm}}}$ and σ_{norm} in the fit is that the mass composition interpretation becomes consistent whether using the predicted EPOS-LHC, QGSJetII-04 or Sibyll2.3 parameterisation.

The requirement of a large dispersion of masses is considered over the entire energy range. For example, a data set consisting of a pure proton composition at higher energies can be fitted, provided that at lower energies we have populations consisting of other primaries. If the statistics or mass dispersion were not large enough, there would be significant degeneracy in the fit between the mass fractions and $t_{0_{\text{norm}}}$ and σ_{norm}.

Apart from the dispersion of masses in the data and statistics, the performance of the fit depends on the intrinsic values for σ of the data. This is nature's width for the $X_{\max}$ distributions of the different primaries. Given that the separation of the distribution modes between primaries remains unchanged in the fit, a larger primary $X_{\max}$ distribution width will increase the $X_{\max}$ distribution overlap of adjacent primaries, resulting in the fit of $t_{0_{\text{norm}}}$, σ_{norm} and the mass composition becoming more uncertain.

We have also evaluated the performance of fitting additional shape parameter coefficients in conjunction with $t_{0_{\text{norm}}}$ and σ_{norm} (see Chapter 5). We have tested fitting $t_{0_{\text{norm}}}$, B, and σ_{norm} in addition to the mass fractions, where B defined in Equation (4.2) describes the change in t_0 with energy. As the predicted mass composition is particularly sensitive to the predicted values of t_0, B is a powerful coefficient which can significantly effect the fitted mass composition. Fitting B allows the predicted $\langle X_{\max} \rangle$ elongation rate of the primaries to change. If B is fitted, the method is as follows:

- B is fitted such that the absolute values of B for each primary change by the same amount. The difference in B among primaries is conserved (identical to how $t_{0_{\text{norm}}}$ is fitted).

The predicted value of B for protons according to EPOS-LHC, QGSJetII-04 and Sibyll2.3 is $\sim$ 2533 g/cm^2, $\sim$ 2445 g/cm^2 and $\sim$ 2666 g/cm^2 respectively. With $t_{0_{\text{norm}}}$ normalised at $10^{18.24}$ eV, a change in B of 350 g/cm^2 corresponds to a change in t_0 at $10^{19.5}$ eV of $\sim$10 g/cm^2. The fitting range limits of B for protons are 1000 g/cm^2 and 4000 g/cm^2. Our EPOS-LHC, QGSJetII-04 and Sibyll2.3 $X_{\max}$ parameterisations have similar values for B, therefore we do not expect fits of B to yield results significantly different from the initial prediction of B when we are fitting EPOS-LHC, QGSJetII-04 or Sibyll2.3 simulated $X_{\max}$ data. However, if the values of B predicted by our parameterisations are considerably incorrect for the data being fitted, significant systematics would be introduced to the reconstructed mass composition if B remains fixed.

Data sets that can be fitted with $t_{0_{\text{norm}}}$ and σ_{norm} may not be accurately fitted when B is included in the fit, as fitting extra coefficients increases the degeneracy between the fitted parameters. Accurately fitting these three coefficients requires a greater spread of primaries and/or statistics than fitting just $t_{0_{\text{norm}}}$ and σ_{norm}.

We have also considered constraining t_0 at 10^{14} eV, where the hadronic models are more reliable, and fitting only B, σ_{norm} and the mass fractions. As the energy of the anchor point is well below the energy of the fitted data, fitting B in this way can provide a consistent mass fraction result between the EPOS-LHC, QGSJetII-04 and Sibyll2.3 parameterisation fits of simulated $X_{\max}$ data, as the t_0 prediction in the fitted energy range adjusts in a way that is similar to the

$t_{0_{\text{norm}}}$ fit, with the added advantage that unlike the $t_{0_{\text{norm}}}$ fit, the resulting fitted parameterisation of t_0 is consistent with the hadronic model predictions at lower energies. However, we have found that over the energy range of interest (10^{17} eV to 10^{20} eV), fitting $t_{0_{\text{norm}}}$ and σ_{norm} results in a more accurate mass composition reconstruction compared to fitting B and σ_{norm}. This is because there is less degeneracy between the fitted mass fractions and shape parameters when fitting $t_{0_{\text{norm}}}$ and σ_{norm}. Additionally, a t_0 parameterisation constrained at $10^{18.24}$ eV describes the energy range of interest better than a t_0 parameterisation extrapolated from 10^{14} eV. If a wider energy range was being fitted, then a $t_{0_{\text{norm}}}$ and σ_{norm} fit would be less accurate, because the t_0 and σ parameterisations of different models do not adequately align over a wider energy range by only adjusting their normalisations.

When constraining at 10^{14} eV, the fitting range limits of B for protons are set to 2290 g/cm^2 to 2724 g/cm^2. This corresponds to a separation of ~ 50 g/cm^2 between the minimum and maximum limits of t_0 at $10^{18.24}$ eV. However, the t_0 range across models at 10^{17} eV is ~ 37 g/cm^2, and at $10^{19.5}$ eV it is ~ 62 g/cm^2. When $t_{0_{\text{norm}}}$ and σ_{norm} are fitted and the normalisation energy is set to $10^{18.24}$ eV, the separation between the minimum and maximum limits of t_0 at any energy is 55 g/cm^2.

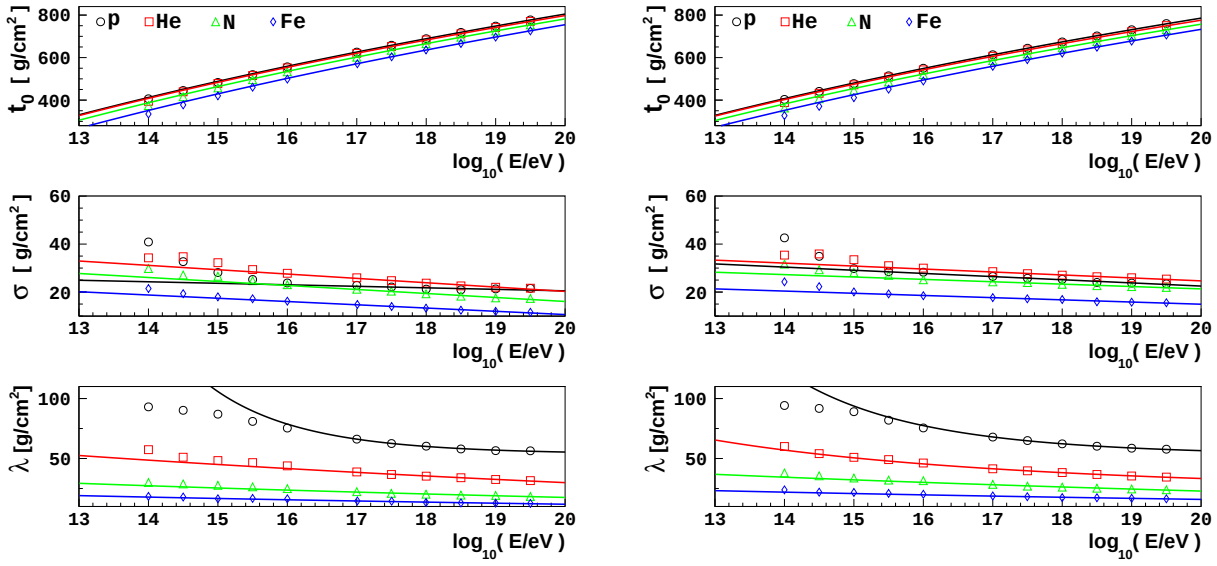
EPOS-LHC	Proton	Helium	Nitrogen	Iron
$t_{0_{\text{norm}}}$	407.63	400.70	379.90	341.81
B	2575.61	2580.66	2613.85	2688.64
QGSJetII-04	Proton	Helium	Nitrogen	Iron
$t_{0_{\text{norm}}}$	404.82	395.60	370.90	336.77
B	2464.1	2472.49	2521.61	2592.9
Sibyll2.3	Proton	Helium	Nitrogen	Iron
$t_{0_{\text{norm}}}$	408.11	390.19	368.51	338.31
B	2666.31	2705.43	2695.22	2714.41

Table 4.2: Parameters adopted in our EPOS-LHC, QGSJetII-04 and Sibyll2.3 t_0 parameterisations, assuming a normalisation energy of $E_0 = 10^{14}$ eV.

The coefficients for the t_0 parameterisations constrained at 10^{14} eV are shown in Table 4.2. These t_0 parameterisations were obtained by fitting the shape parameter results displayed in Figure 4.4, in addition to the shape parameter results from fits of Equation 4.1 to CONEX v4r37 X_{max} distributions with fixed energies between 10^{14} eV to 10^{16} eV. The shape parameter values of t_0, σ and λ from 10^{14} eV to $10^{19.5}$ eV are displayed in Figure 4.9, with the t_0 parameterisation constrained at 10^{14} eV displayed. Data below 10^{17} eV was not obtained for Sibyll2.3, therefore the standard Sibyll2.3 t_0 parameterisation (obtained from fits of energies above 10^{17} eV) was extrapolated to 10^{14} eV to determine an anchor value at this energy.

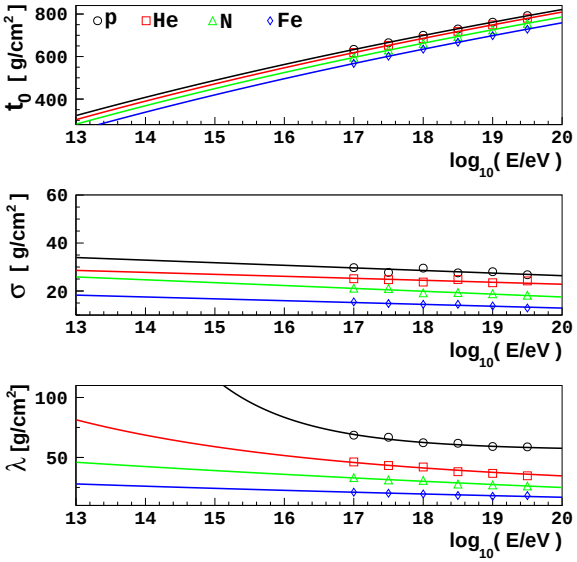
In the σ and λ panels, the fitted functions displayed are those normalised at $10^{18.24}$ eV (from Table 4.1), parameterisations which were determined by fitting only the shape parameter values above 10^{17} eV, therefore these parameterisations do not necessarily align with the shape parameter results below 10^{17} eV. It is possible to fit functions which describe σ and λ over the extended energy range. However, as we are interested in X_{max} data of energies above 10^{17} eV, and as we intend to fit σ_{norm} and possibly λ_{norm}, it is advantageous to apply a parameterisation which describes the energy range of interest as accurately as possible, instead of a parameterisation that is partially burdened by accommodating a wider energy range. When normalising t_0 at 10^{14} eV to fit B, we apply the same parameterisations of σ and λ used previously (normalised at $10^{18.24}$ eV).

It is important to recognise that the fits of how t_0 changes with energy, and therefore the $\langle X_{\text{max}} \rangle$ elongation rate, are restricted. When fitting B, we are assuming a functional form for how t_0 changes with energy, and only fitting the rate of change of the $\log_{10}\left(\frac{\log_{10} E}{\log_{10} E_0}\right)$ factor. To



(a) EPOS-LHC

(b) QGSJetII-04



(c) Sibyll2.3

Figure 4.9: The shape parameter results (markers) from fits to simulated data according to the EPOS-LHC (left), QGSJetII-04 (right) and Sibyll2.3 (bottom) hadronic interaction models. Fitted to the shape parameter results for each primary mass and hadronic model are the functions in Equation (4.2), displayed by coloured lines. Only the t_0 function is fitted to the full energy range results.

properly fit the slope of t_0 with energy would require the fit of a third t_0 parameter, for example, fitting B and x in $B \cdot \log_{10}\left(\frac{\log_{10} E}{\log_{10} E_0}\right)^x$, where x currently equals 1.

The normalisation of λ, denoted λ_{norm}, is another shape coefficient of interest. We have evaluated the performance of fitting $t_{0_{\text{norm}}}$, σ_{norm} and λ_{norm}, in conjunction with the mass fractions. We fit λ_{norm} in the same way σ_{norm} is fitted:

- λ_{norm} is fitted such that the ratio of λ (and λ_{norm}) between primaries remains similar to the initial ratio over the energy range. Differences in K and L (λ coefficients) between primaries prevents the exact conservation of the initial ratio. If λ_{norm} for protons changes

by Δ, λ_{norm} for other primaries will change by Δ multiplied by the initial average ratio of λ between that primary and protons.

This approach assumes that over the fitted energy range, the hadronic models are correctly predicting the ratio of λ between different elements.

The predicted value of λ_{norm} for protons is $\sim 59\,\text{g/cm}^2$ for EPOS-LHC, $\sim 61\,\text{g/cm}^2$ for QGSJetII-04, and $\sim 62\,\text{g/cm}^2$ for Sibyll2.3. The values of λ for protons are quite similar between the models. However, for other masses there is greater dissimilarity in the λ predictions between models. As mentioned earlier regarding fits of B, even though the X_{max} parameterisations from different models may agree with each other on many aspects of the shape parameter predictions, these consistent predictions may incorrectly describe nature.

We limit λ_{norm} for protons to a value between $35\,\text{g/cm}^2$ and $85\,\text{g/cm}^2$, a considerably wide range. Including λ_{norm} in the fit of data sets of a realistic mass composition can often result in an incorrect reconstructed mass, as breaking the degeneracy between the mass fractions, λ_{norm} and the other fitted shape parameter coefficients is difficult. This is partly because both $t_{0_{\text{norm}}}$ and λ_{norm} define the predicted $\langle X_{\text{max}} \rangle$ of a primary, a property which the fits are quite sensitive to.

The four components of proton, He, N and Fe present an even spread across the X_{max} range. The fits of the shape parameter coefficients are dependent on choosing four components that are evenly spread over X_{max}, which are able to match the details of the fitted X_{max} distribution. If for example, we removed the nitrogen component, the fit of $t_{0_{\text{norm}}}$ and σ_{norm} would break down, as the distributions of p, He and Fe would need to be shifted significantly in $t_{0_{\text{norm}}}$ and σ_{norm} to compensate for the gap left by the nitrogen distribution. Therefore, the fitted values of $t_{0_{\text{norm}}}$ and σ_{norm} are unable to be compared to the initial predictions of the models, as these coefficients are not shifting to correct the initial model assumptions, but are instead shifting to correct a gap in the distribution. However, if we instead replaced the nitrogen component with a component of similar mass, the effect on the fit would be minimal. For example, if oxygen replaced nitrogen, in terms of the mass composition, a small fraction of events that were previously interpreted to be nitrogen would be assigned helium, and some events that were previously interpreted as iron would be assigned oxygen. For most mass distributions, this would be the general form of the change in the mass fractions fitted to the data.

Given the current statistics of the measured data, even if only the mass fractions are fitted, including more than four individual mass components in the fit is problematic due to the degeneracy between the mass fraction coefficients alone. The degeneracy of the mass fraction coefficients is increased further if shape parameter coefficients are fitted.

4.5 Importance of bin sizes in the fit

We have evaluated the effect of different X_{max} bin sizes and energy bin sizes on the performance of the fit. When fitting only the mass fractions, $1\,\text{g/cm}^2$ X_{max} binning gives marginally more accurate results than $20\,\text{g/cm}^2$ X_{max} binning ($20\,\text{g/cm}^2$ is the X_{max} bin size of the Auger X_{max} distributions published in [145]). The absolute improvement in the fitted mass fractions of an energy bin is no greater than 3%. However, when fitting shape coefficients (such as $t_{0_{\text{norm}}}$ and σ_{norm}) in addition to the mass fractions, it is important that a small X_{max} binning is used, otherwise the chosen center of the X_{max} bins can significantly affect the fitted results, especially if the statistics are not substantial. The predicted separation between different primaries in $t_{0_{\text{norm}}}$ and σ_{norm} can be very small. For example, our EPOS-LHC parameterisation predicts the difference in $t_{0_{\text{norm}}}$ between protons and helium is only $\sim 6\,\text{g/cm}^2$. Therefore, a $20\,\text{g/cm}^2$ X_{max} binning can be too coarse, and can shift the apparent $\langle X_{\text{max}} \rangle$ of the distribution, which affects the fit of $t_{0_{\text{norm}}}$.

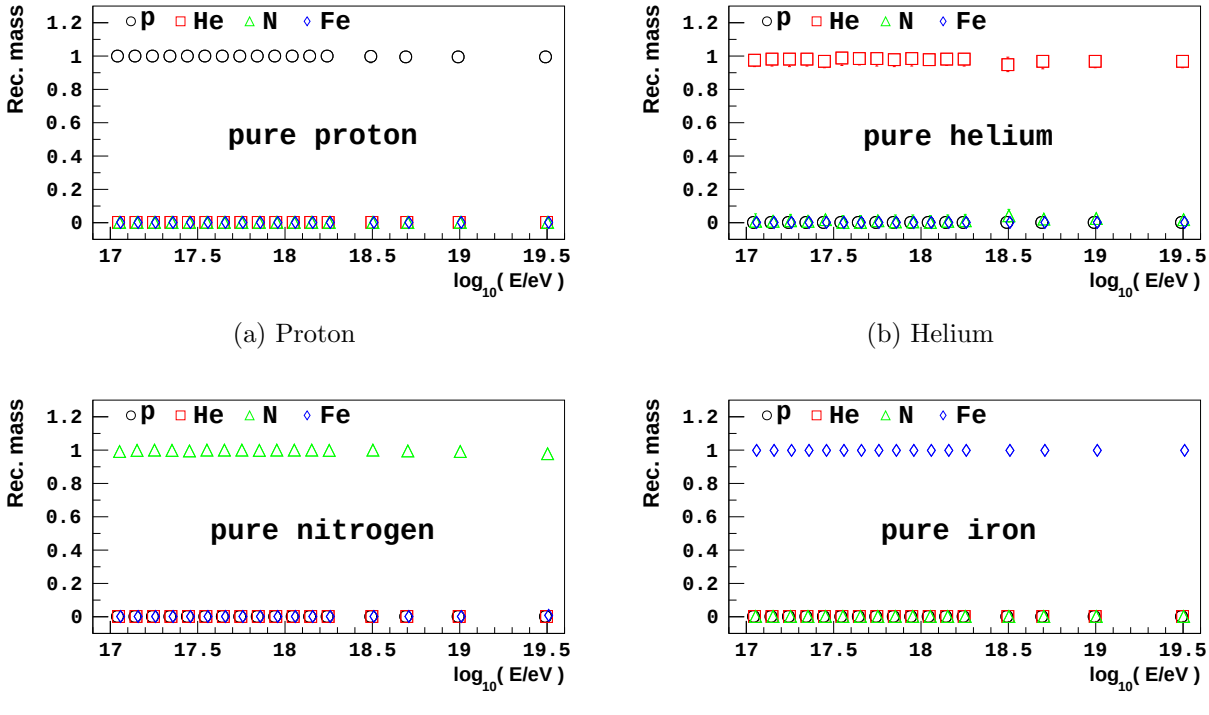
Due to similar reasons, the energy bin size is also important. Energy binning that is too

large can result in data from the same primary mass, but on opposite extremes of the energy bin, being evaluated as data from different primaries. This is a consequence of the separation in the predicted X_{max} distributions between primaries being small compared to the shift in these X_{max} distributions with energy. We find that an energy binning of 0.1 in $\log_{10}(E/\mathrm{eV})$ is sufficient.

Chapter 5

Evaluation of the fit performance using simulated $X_{\max}$ data

5.1 Fit performance on generic mass compositions



(a) Proton

(b) Helium

(c) Nitrogen

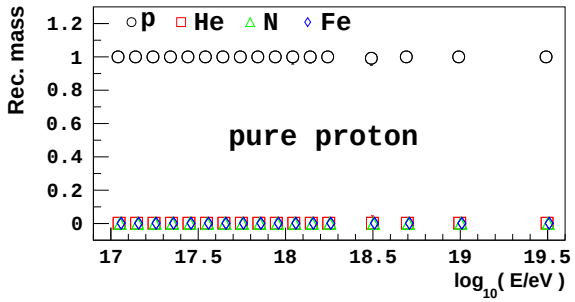
(d) Iron

Figure 5.1: Fitting only the mass fractions to mock $X_{\max}$ data sets. The data sets have been generated using the EPOS-LHC model and assume a pure primary composition over the whole energy range. The composition fits were performed using the $X_{\max}$ parameterisations from the EPOS-LHC model predictions. Throughout this work, 'Rec. mass' refers to the mass fractions fitted to the data.

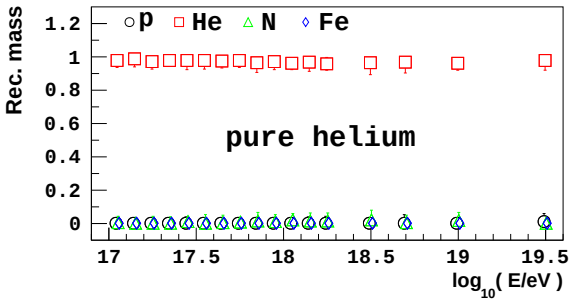
Using CONEX v4r37, 100 $X_{\max}$ data sets were generated according to the EPOS-LHC and QGSJetII-04 hadronic interaction models for a number of different mass compositions. The data consists of 17 energy bins, of which there are 13 energy bins of a width of 0.1 in $\log_{10}(E/\mathrm{eV})$ between 10^{17} eV and $10^{18.3}$ eV, and 4 fixed energy bins at $10^{18.5}$ eV, $10^{18.7}$ eV, 10^{19} eV and $10^{19.5}$ eV. Each energy bin contains approximately 750 events.

We have fitted only the mass fractions (all coefficients from the $X_{\max}$ parameterisation were

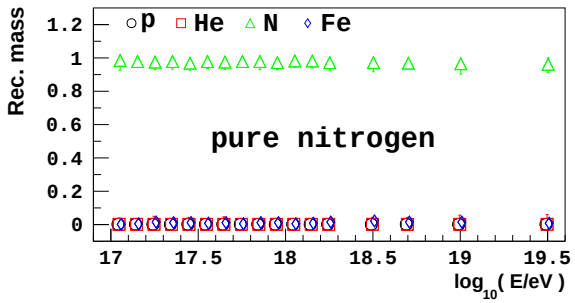
kept fixed) to data of a single primary generated with the same hadronic interaction model that the parameterisation fitted is based on. Figures 5.1a to 5.1d summarises the results (of these 100 fits) for the EPOS-LHC hadronic model and Figures 5.2a to 5.2d for the QGSJetII-04 model. The markers represent the medians of the fitted mass fractions and the error bars represent the standard deviation. The results show that our $X_{\max}$ parameterisations are an accurate description of the expected $X_{\max}$ distribution of a primary according to the EPOS-LHC or QGSJetII-04 hadronic interaction models. Both our EPOS-LHC and QGSJetII-04 $X_{\max}$ parameterisation fits can accurately determine the mass composition of data from their corresponding hadronic model.



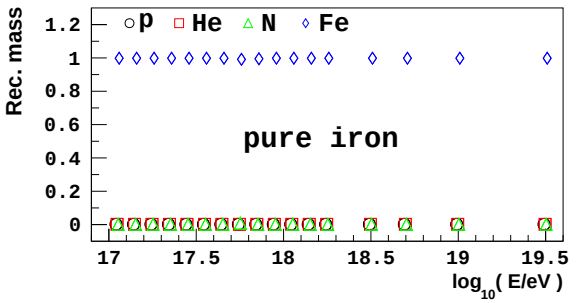
(a) Proton



(b) Helium



(c) Nitrogen



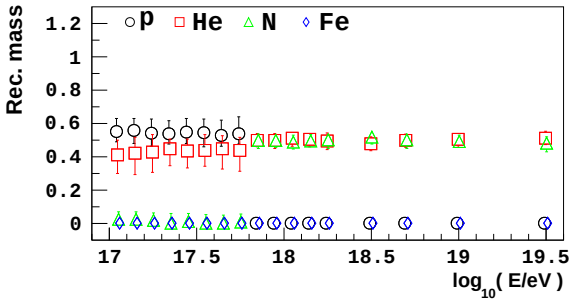
(d) Iron

Figure 5.2: Fitting only the mass fractions to mock $X_{\max}$ data sets. The data sets have been generated using the QGSJetII-04 model and assume a pure primary composition over the whole energy range. The composition fits were performed using the $X_{\max}$ parameterisations from the QGSJetII-04 model predictions.

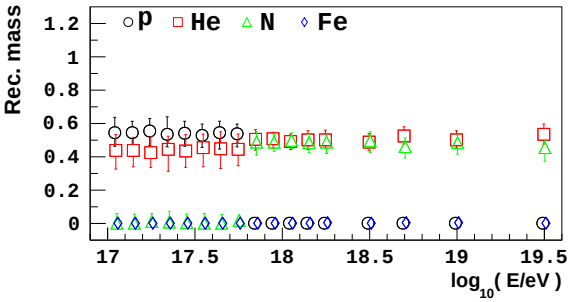
Figure 5.3a to Figure 5.4a summarises the results of fits to 100 $X_{\max}$ data sets with a true mass composition consisting of 50% protons and helium in the first 8 energy bins, and 50% helium and nitrogen in the remaining 9 energy bins. When fitting only the mass fractions of our parameterisations (i.e. keeping fixed the coefficients of the $X_{\max}$ distribution parameterisation) to CONEX v4r37 $X_{\max}$ data based on the same model (as seen in Figures 5.3a and 5.3b), the fits are able to reconstruct the mass composition to within an absolute offset in the median of 10% from the true mass.

Figure 5.4a shows the results of fitting $t_{0_{\text{norm}}}$ and σ_{norm}, in addition to the mass fractions, of the QGSJetII-04 parameterisation to QGSJetII-04 data sets. These QGSJetII-04 $X_{\max}$ distributions do not provide sufficient constraints on our fitted parameterisation, resulting in a mass composition reconstruction that does not resemble the true mass composition. In order to successfully fit $t_{0_{\text{norm}}}$ and σ_{norm} to data of a similar distribution, a greater dispersion of primary masses over the energy range of the data is required (wider than the one in the given example). For example, in Figure 5.4b the range of primary masses has been increased by replacing helium with iron in the last energy bin. The subsequent fit of the mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm} re-

sults in an absolute offset in the median of less than 15% from the true mass fractions, which is comparable to a fit of only the mass fractions to data of a similar composition.

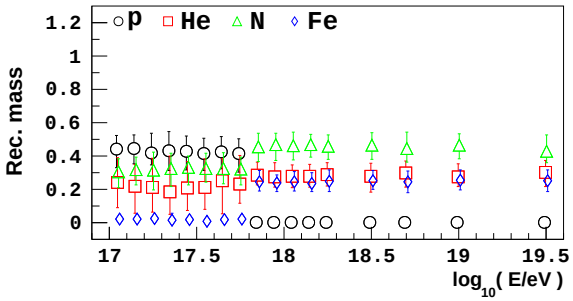


(a) Fitting only the mass fractions of the EPOS-LHC parameterisation to EPOS-LHC $X_{\max}$ data.

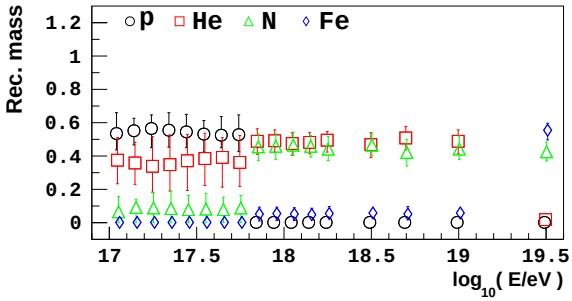


(b) Fitting only the mass fractions of the QGSJetII-04 parameterisation to QGSJetII-04 $X_{\max}$ data.

Figure 5.3: Fitting a particular $X_{\max}$ parameterisation to $X_{\max}$ data of the same hadronic interaction model. The true mass composition of the data is 50% protons and helium in the first 8 energy bins, and 50% helium and nitrogen in the remaining bins.

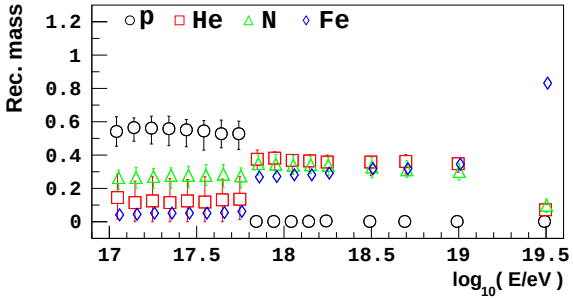


(a) The true mass composition of the data is 50% protons and helium in the first 8 energy bins, and 50% helium and nitrogen in the remaining bins.

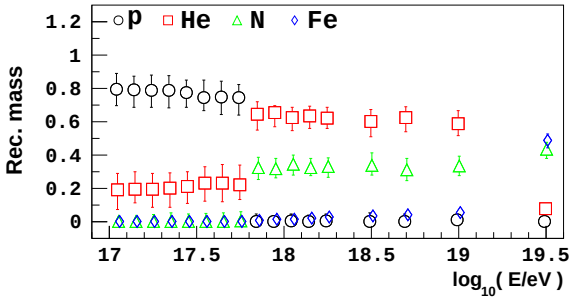


(b) The true mass composition is the same as the plot on the left, but the last energy bin instead contains 50% nitrogen and iron.

Figure 5.4: Fitting the mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm} of the QGSJetII-04 parameterisation to QGSJetII-04 $X_{\max}$ data.

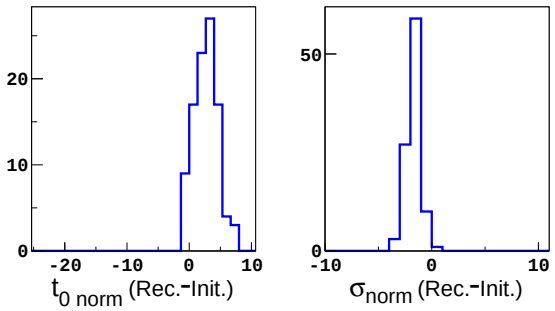


(a) Fitting only the mass fractions (i.e. $t_{0_{\text{norm}}}$ and σ_{norm} are kept fixed) of our EPOS-LHC parameterisation to QGSJetII-04 $X_{\max}$ data.

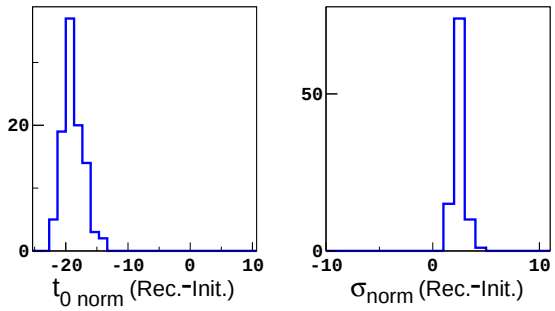


(b) Fitting the mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm} of our EPOS-LHC parameterisation to QGSJetII-04 $X_{\max}$ data.

Figure 5.5: The true mass composition of the data is 50% protons and helium in the first 8 energy bins, 50% helium and nitrogen in the next 8 energy bins, and 50% nitrogen and iron in the last energy bin.



(a) Fitting the QGSJetII-04 $X_{\max}$ parameterisation. See Figure 5.4b.



(b) Fitting the EPOS-LHC $X_{\max}$ parameterisation. See Figure 5.5b.

Figure 5.6: Change in $t_{0_{\text{norm}}}$ and σ_{norm} for protons. The data being fitted was generated with the QGSJetII-04 model in both cases.

Compare Figure 5.5a with Figure 5.5b, which shows the composition fits when using the EPOS-LHC parameterisation to fit QGSJetII-04 data, with $t_{0_{\text{norm}}}$ and σ_{norm} fixed in the former, and $t_{0_{\text{norm}}}$ and σ_{norm} fitted in the latter. Fitting these two coefficients is enough to result in a reconstructed mass which resembles the true mass to a much greater degree, despite the fitted data originating from a different model. By fitting $t_{0_{\text{norm}}}$ and σ_{norm}, there is no longer a significant iron component reconstructed where there should only be 50% helium and nitrogen, and in the bins where the true composition is 50% protons and helium there is no longer a fitted nitrogen component larger than the helium fraction.

Figures 5.6a and 5.6b show the difference between the fitted values and initial values of $t_{0_{\text{norm}}}$ and σ_{norm} for protons in the fitted parameterisation, when fitting the data with iron added in the last energy bin. Figure 5.6a displays the results of fitting QGSJetII-04 data with our QGSJetII-04 parameterisation (the reconstructed mass is shown in Figure 5.4b), and as expected the difference between the reconstructed and initial values of our coefficients is minimal. Figure 5.6b displays the results of fitting the same QGSJetII-04 data with our EPOS-LHC parameterisation (the reconstructed mass is shown in Figure 5.5b), and we see that $t_{0_{\text{norm}}}$ and σ_{norm} are shifted towards the QGSJetII-04 values for these coefficients. The initial EPOS-LHC proton $t_{0_{\text{norm}}}$ and σ_{norm} values are $\sim 703\,\text{g/cm}^2$ and $\sim 22\,\text{g/cm}^2$ respectively, while the initial QGSJetII-04 proton $t_{0_{\text{norm}}}$ and σ_{norm} values (and therefore the approximate values of the QGSJetII-04 MC data) are $\sim 688\,\text{g/cm}^2$ and $\sim 25\,\text{g/cm}^2$ respectively.

As the absolute separation between σ for different primaries is similar in the EPOS-LHC and QGSJetII-04 parameterisations (like t_0), marginally better results would be obtained in Figure 5.5b if instead of fitting σ_{norm} such that the initial ratios of σ among primaries is conserved, σ_{norm} was fitted such that the initial separation between σ_{norm} among primaries was conserved (like $t_{0_{\text{norm}}}$). However, conserving the initial ratios of σ is the more physical approach, because if σ_{norm} for protons changes by $10\,\text{g/cm}^2$, we would not expect σ_{norm} for iron to also change by $10\,\text{g/cm}^2$. Additionally, nature does not necessarily conform to the EPOS-LHC or QGSJetII-04 predictions of the absolute separation of σ_{norm} among primaries.

If the mass dispersion of the data was greater, the $t_{0_{\text{norm}}}$ and σ_{norm} fit of the EPOS-LHC parameterisation to QGSJetII-04 data (Figure 5.5b) would reconstruct the mass composition with even greater accuracy. For example, if instead of large energy bin segments with the same mass composition, there was more variability in the transition of the mass composition over the energy range, this increase in the mass dispersion would result in a more accurate mass composition reconstruction.

A data set consisting of 25% proton, helium, nitrogen and iron across the whole energy range does not constrain a $t_{0_{\text{norm}}}$ and σ_{norm} fit of the corresponding parameterisation particularly well, as for each energy bin the $\langle X_{\max} \rangle$ of the total distribution and of each primary component is

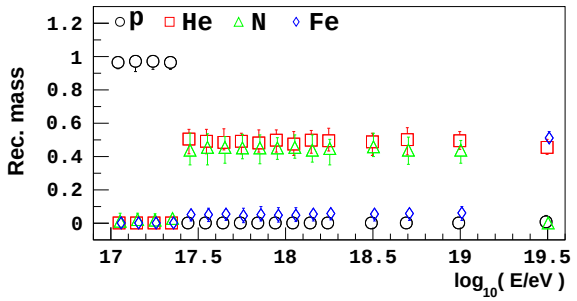
consistently positioned relative to the predicted $\langle X_{\max} \rangle$ given some value of $t_{0_{\text{norm}}}$. Therefore, each energy bin constrains the fitted parameters in the same way. If the correct mass fractions were applied to each energy bin, and a parameter space scan of $t_{0_{\text{norm}}}$ and σ_{norm} was performed, the Poisson log likelihood space of each energy bin across $t_{0_{\text{norm}}}$ and σ_{norm} would be similar. Data spanning a wider energy range will generally constrain the $t_{0_{\text{norm}}}$ and σ_{norm} fit with more power, as more energy bins typically means more mass dispersion, and therefore the Poisson log likelihood space across $t_{0_{\text{norm}}}$ and σ_{norm} is more likely to change between energy bins. This helps to rule out incorrect values of $(t_{0_{\text{norm}}}, \sigma_{\text{norm}})$ that can fit data of a particular composition quite well with the appropriate fitted mass fractions over the energy range, but can not adequately describe data of a different composition.

A pure composition in some energy bins can help constrain the fit of $t_{0_{\text{norm}}}$ and σ_{norm} with more power than energy bins containing a mix of masses. In particular, energy bins containing pure proton or pure iron provide strong constraints on the fitted shape parameter coefficient values, as these masses are on the extreme ends of the possible fitted composition, therefore requiring a significant shift in the predicted $X_{\max}$ distributions of each mass if intermediate masses are to accommodate these data. As the $X_{\max}$ distribution shift is applied over the whole energy range, if some energy bins contain intermediate masses, these bins are unlikely to be fitted adequately when a large shift is applied to the $X_{\max}$ distributions so that pure proton or iron elsewhere can be accommodated by the intermediate masses. Additionally, if a significant proton fraction exists, the fitted proton $X_{\max}$ parameterisation cannot be shifted too far towards smaller $X_{\max}$ values (by shifting $t_{0_{\text{norm}}}$), as the proton $X_{\max}$ distribution is the upper limit. Likewise, if a significant iron fraction exists, the fitted iron $X_{\max}$ parameterisation cannot be shifted too far towards larger $X_{\max}$ values.

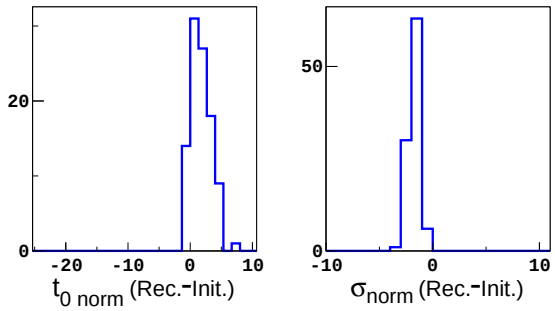
Demonstrating the constraining power of a data set containing pure proton in some energy bins is Figure 5.7, which displays the results of fitting $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions to QGSJetII-04 data consisting of 100% protons in the first 4 energy bins, then 50% helium and nitrogen, until the last bin which contains 50% helium and iron. The fit of the QGSJetII-04 parameterisation accurately reconstructs the mass composition, as the correct $t_{0_{\text{norm}}}$ and σ_{norm} values are able to be fitted. The EPOS-LHC fit of the QGSJetII-04 data results in absolute offsets in the median reconstructed mass fractions from the true mass of less than 10%, a much more accurate fit than in Figure 5.5b.

Notice that in Figures 5.3 to 5.7, the bins containing a helium and nitrogen mix are reconstructed better than those containing a proton and helium mix. Proton and helium distributions are harder to reconstruct due to their wider spread and their larger overlap. A wider spread means that for a given number of events, fewer events will populate individual bins. Therefore, proton and helium fits have larger statistical uncertainties. Additionally, the $X_{\max}$ parameterisations for lighter masses do not describe the CONEX v4r37 EPOS-LHC and QGSJetII-04 MC data as accurately. Figure 4.5 illustrates that as the primary mass of the distribution increases, the $X_{\max}$ parameterisations reproduce the true $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$ of the distributions with better accuracy. Appendix A shows that for proton and helium data especially, the fits of Equation (4.1) to MC data of either hadronic model tend to overestimate the number of events at the mode of the distribution. When fitting mixes of protons and helium, our fits tend to have a reconstruction bias towards protons.

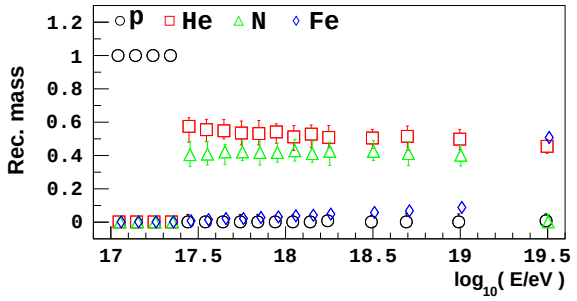
The shape coefficient parameters can be adjusted to increase the accuracy with which a particular mix of masses is reconstructed. For example, if we wanted to increase the accuracy with which a 50% proton and helium mix was fitted, the mass fractions would be appropriately fixed to a large data set of this mass mix, and one at a time each shape parameter coefficient would be fitted to the MC data iteratively, until each coefficient converges to the optimum value for describing this distribution. However, a parameterisation developed from this iteration would not reconstruct alternative mixes with as much accuracy as the original parameterisation.



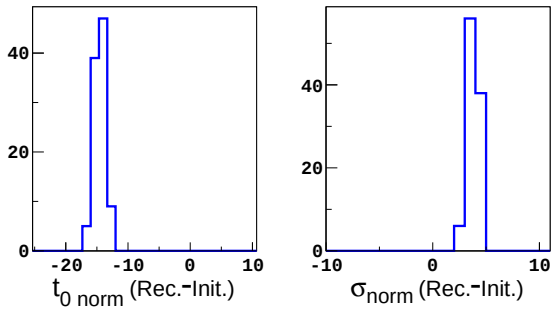
(a) QGSJetII-04 fitted to QGSJetII-04 $X_{\max}$ data.



(b) Fitted coefficients from Figure 5.7a.



(c) EPOS-LHC fitted to QGSJetII-04 $X_{\max}$ data.



(d) Fitted coefficients from Figure 5.7c.

Figure 5.7: Fitting the mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm}. The true mass composition of the data is 100% protons in the first 4 energy bins, 50% helium and nitrogen in the remaining energy bins, except the last bin which contains 50% helium and iron.

5.2 Evaluating the fit performance for a mass composition consistent with the Auger $X_{\max}$ distributions

We fit our EPOS-LHC, QGSJetII-04 and Sibyll2.3 parameterisations to mock $X_{\max}$ data sets that resemble the Auger $X_{\max}$ distributions presented in Chapter 8. These mock data sets were produced using the corresponding mass fractions fitted to the Auger $X_{\max}$ data (where the fits were performed with all coefficients of the $X_{\max}$ parameterisation fixed).

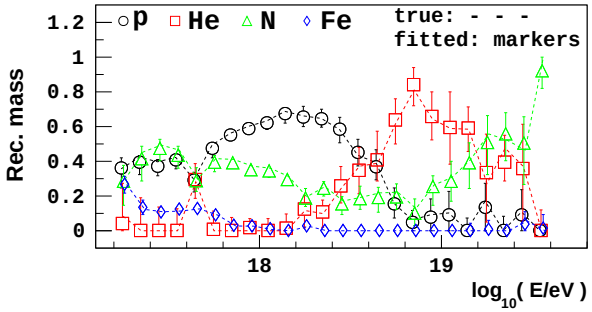
The measured FD and HEAT $X_{\max}$ distributions are broadened due to the $X_{\max}$ resolution of the detector, and are affected by the detector acceptance, therefore the mock $X_{\max}$ data generated from the fitted composition fractions are convolved with the same detector effects. The $X_{\max}$ resolution and acceptance of the Auger data is taken into account when fitting this mock Auger $X_{\max}$ data. Our mock $X_{\max}$ distributions and the $X_{\max}$ distributions measured by Auger are treated with exactly the same approach.

The purpose of this section is to determine the ability of the shape coefficient and mass fraction fits of the three $X_{\max}$ parameterisations to reconstruct the true mass composition and true shape coefficients of the data, when applied to $X_{\max}$ data sets that are similar to those measured by Auger. This section mainly focuses on the combined FD and HEAT Auger $X_{\max}$ data set. Various combinations of $t_{0_{\text{norm}}}$, B, σ_{norm} and λ_{norm} are fitted to the data. As before, the mass composition plots in this section display the median reconstructed mass composition fractions from the fits of 100 mock $X_{\max}$ data sets, and the error bars represent the 1σ standard deviation of these 100 fits.

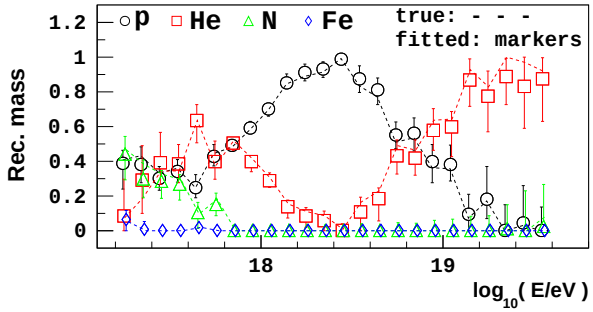
5.2.1 Using FD and HEAT $X_{\max}$ data

We fit the $X_{\max}$ parameterisations to 100 mock data sets of a composition based on the results of fitting only the mass fractions of our $X_{\max}$ parameterisations to the combined Auger FD and HEAT $X_{\max}$ data (see Section 8.1.1). Each mock data set is comprised of two separate data sets, a mock HEAT $X_{\max}$ data set and a mock FD $X_{\max}$ data set, with the same statistics as the Auger data. The HEAT $X_{\max}$ data set spans an energy of $10^{17.2}$ eV to $10^{18.1}$ eV and contains 16778 events, while the FD $X_{\max}$ data set spans an energy of $10^{17.8}$ eV to 10^{20} eV and contains 25884 events (see Appendix B for the statistics of each energy bin). Identical to Section 8.1, the FD and HEAT mock data sets are treated separately when calculating the Poisson log likelihood, that is the two data sets are not summed together to form a single $X_{\max}$ distribution which is then fitted.

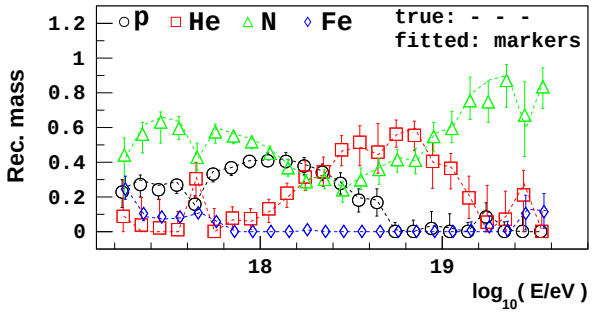
5.2.1.1 Fitting only the mass fractions



(a) Fits of EPOS-LHC based $X_{\max}$ data with the EPOS-LHC parameterisation.



(b) Fits of QGSJetII-04 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



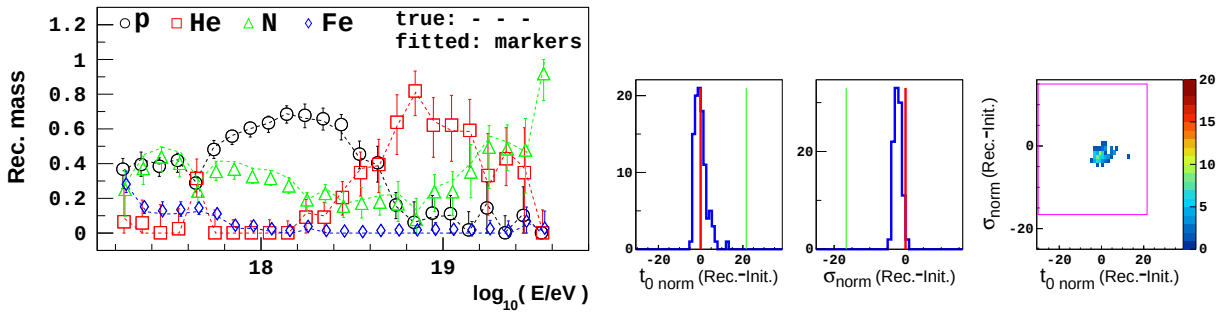
(c) Fits of Sibyll2.3 based $X_{\max}$ data with the Sibyll2.3 parameterisation.

Figure 5.8: Fitting only the mass fractions of an $X_{\max}$ parameterisation to mock data based on the same $X_{\max}$ parameterisation that is fitted.

For each parameterisation, sets of mock data were generated, the respective mock data then fitted with its corresponding parameterisation. Figure 5.8 displays the reconstructed mass composition from fitting only the mass fractions to the mock data, compared to the true mass composition of this data. As the data was fitted with the $X_{\max}$ parameterisation it was generated from, and only the mass fractions are fitted (therefore the applied shape coefficients are correct), the systematic offsets in the mass fraction reconstruction should be small. The absolute offsets in the median reconstructed mass fractions from the true mass fractions are less than 5% in most energy bins, demonstrating that our parameterisation fits have the ability to accurately reconstruct the mass composition if the correct shape parameter values are applied.

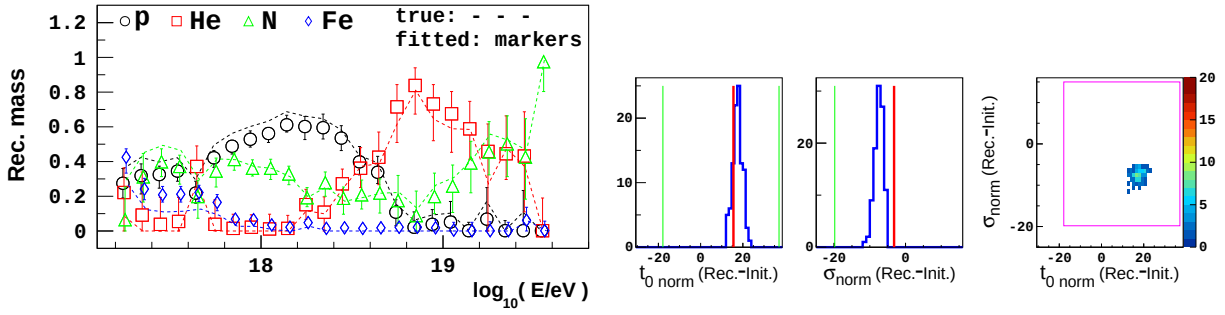
The various data sets consist of a reasonable mass dispersion and statistics, therefore the fits of $t_{0_{\text{norm}}}$ and σ_{norm} should reconstruct the mass composition with good accuracy. The $\langle X_{\max} \rangle$ of the mock data sets generated from different parameterisations is very similar, but the $\sigma(X_{\max})$ of the mock data sets will differ substantially (see Figure 8.2b of Section 8.1.1). The $\sigma(X_{\max})$ of the QGSJetII-04 based mock data is larger than the EPOS-LHC and Sibyll2.3 mock data.

5.2.1.2 Fitting $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions



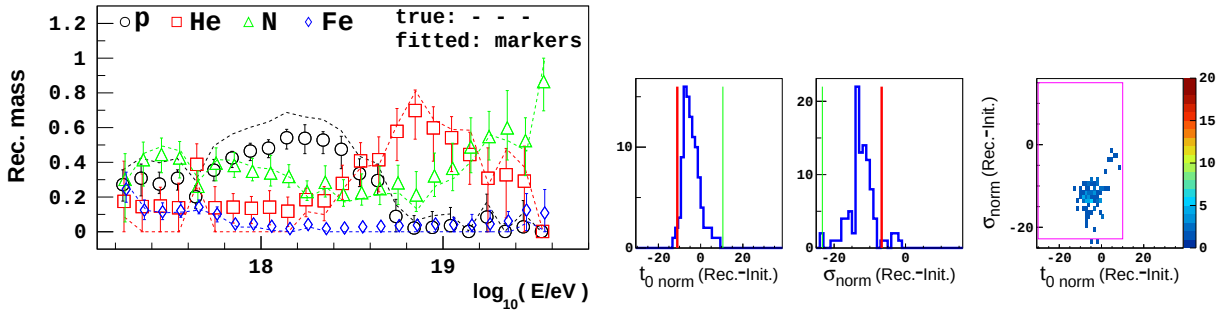
(a) (b)

Figure 5.9: Fits of EPOS-LHC based $X_{\max}$ data with the EPOS-LHC parameterisation. See the text for an explanation of the vertical lines and the magenta boxes.



(a) (b)

Figure 5.10: Fits of EPOS-LHC based $X_{\max}$ data with the QGSJetII-04 parameterisation.



(a) Trimmed. (b)

Figure 5.11: Fits of EPOS-LHC based $X_{\max}$ data with the Sibyll2.3 parameterisation.

Figures 5.9a, 5.10a and 5.11a displays the mass composition results from fitting the mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm} of either the EPOS-LHC, QGSJetII-04 or Sibyll2.3 parameterisations respectively, to 100 $X_{\max}$ data sets generated from the parameterisation which resulted when

only the mass fractions of the **EPOS-LHC** parameterisation were fitted to the combined FD and HEAT Auger $X_{\max}$ data. The true mass composition of the mock data is therefore the mass composition which resulted from the EPOS-LHC fit to the FD and HEAT $X_{\max}$ data (Section 8.1.1). Figures 5.9b, 5.10b and 5.11b displays the fitted values of $t_{0_{\text{norm}}}$ and σ_{norm} for protons, and their correlation, from the respective EPOS-LHC, QGSJetII-04 or Sibyll2.3 fits of the mock data.

The vertical red lines in the 1D histogram plots indicate the values of $t_{0_{\text{norm}}}$ and σ_{norm} for protons that were fixed to the fitted Auger $X_{\max}$ data, relative to the initial $t_{0_{\text{norm}}}$ and σ_{norm} proton values of the parameterisation being fitted to the mock data. Therefore, if the mean $t_{0_{\text{norm}}}$ proton value fitted to the mock data sets by a parameterisation is equal to the true $t_{0_{\text{norm}}}$ proton value of the data, the blue histogram will be centred on the red line. The vertical green lines indicate the limits applied to the coefficient fit. The limits are the same for each parameterisation fit. For $t_{0_{\text{norm}}}$, the lower and upper limits (in relation to protons) are $670\,\text{g/cm}^2$ and $725\,\text{g/cm}^2$ respectively, and for σ_{norm} they are $5\,\text{g/cm}^2$ and $45\,\text{g/cm}^2$.

The magenta box of the 2D histogram indicates the range of fitted shape coefficient values that were used in the accompanying mass fraction plot. If the mass fraction figure is labelled 'Trimmed', fits which resulted in a shape coefficient reaching its upper or lower limit are not included in the results displayed in the mass fraction plot, and accordingly these coefficient values will be outside of the magenta box.

Fits which reach the coefficient limits are excluded as it is unlikely they found the true minimum of the log likelihood space, consequently including these fits would distort the representation of the mass fraction fits.

Sometimes the $t_{0_{\text{norm}}}$ and/or σ_{norm} distributions display bi-modality. The mass fraction reconstructions of substantially separate coefficient solutions will not be consistent. Combining these solutions and displaying the median mass fractions would provide a misleading representation. Throughout this chapter, where the fitted coefficients indicate two separate solutions, if the corresponding mass fraction figure is labelled 'Solution 1', the mass fraction results displayed will correspond to the coefficient solution closest to the true solution (the solution closest to the red lines of the 1D histograms), and accordingly the magenta box will encompass the fits corresponding to 'Solution 1'. It can be assumed that the secondary solution ('Solution 2') will not result in an accurate mass composition reconstruction. The secondary solution is not always the location of the deepest minimum for that data set. Occasionally the fit becomes stuck in a local minimum. The local minima tend to be located at larger values of $t_{0_{\text{norm}}}$.

To help prevent fits becoming stuck in local minima, if the fitted value of $t_{0_{\text{norm}}}$ is above a particular threshold, the data is re-fitted with the upper limit of $t_{0_{\text{norm}}}$ set to the threshold, to check if a deeper minimum can be found at smaller values of $t_{0_{\text{norm}}}$. To ensure that our fits of the measured Auger distributions in Chapter 8 did not become stuck in a local minimum, we perform scans over reasonably large ranges of the shape parameter space, detailed in Chapter 6. If secondary solutions exist, they are revealed by these scans.

In Figure 5.9a, the EPOS-LHC $X_{\max}$ parameterisation fit to mock data based on the EPOS-LHC parameterisation results in a good mass fraction reconstruction, which is expected as the correct parameterisation is allowed to be fitted to the data. The absolute offsets in the medians of the fitted mass fractions are less than 10% from the true mass in most energy bins. The statistical error, and generally the systematic offsets, are greater in the higher energy bins due to the lack of statistics in these bins. We see in Figure 5.9b that the $t_{0_{\text{norm}}}$ distribution from the EPOS-LHC parameterisation fits is centred on the true $t_{0_{\text{norm}}}$ value of the data, while the σ_{norm} distribution is slightly offset. As the EPOS-LHC parameterisation is being fitted to mock data based on the SAME EPOS-LHC parameterisation, the blue distributions are expected to be centred on the red lines, and the median reconstructed mass composition centred around the true mass.

Figures 5.10a and 5.10b show that despite the differences between the EPOS-LHC and QGSJetII-04 parameterisations (which are not limited to different $t_{0_{\text{norm}}}$ and σ_{norm} predictions),

by allowing $t_{0_{\text{norm}}}$ and σ_{norm} to be fitted to the mock data based on the EPOS-LHC parameterisation, the QGSJetII-04 Xmax parameterisation fits to this data reconstructs the true mass fractions with an overall accuracy comparable to the EPOS-LHC fit of this data. The absolute offsets in the medians of the reconstructed mass fractions are less than 10% from the true fractions in most energy bins, demonstrating that fitting $t_{0_{\text{norm}}}$ and σ_{norm} significantly reduces the differences between the EPOS-LHC and QGSJetII-04 $X_{\max}$ parameterisations. As we are fitting the QGSJetII-04 parameterisation to EPOS-LHC mock data, we do not expect the average fitted values of $t_{0_{\text{norm}}}$ and σ_{norm} to be centred on the red lines if no systematic offset was present in the reconstruction. This is because the separation of these coefficients among masses differs between the EPOS-LHC and QGSJetII-04 parameterisations, thus if the fitted QGSJetII-04 value of $t_{0_{\text{norm}}}$ for protons was equal to the EPOS-LHC value of $t_{0_{\text{norm}}}$ for protons, the appropriately adjusted $t_{0_{\text{norm}}}$ values for He, N and Fe of the QGSJetII-04 parameterisation would be different to the EPOS-LHC description. Similarly, if the residuals of the $t_{0_{\text{norm}}}$ and σ_{norm} values of these QGSJetII-04 fits were centred on the red line, there could still be a small bias in the fitted composition.

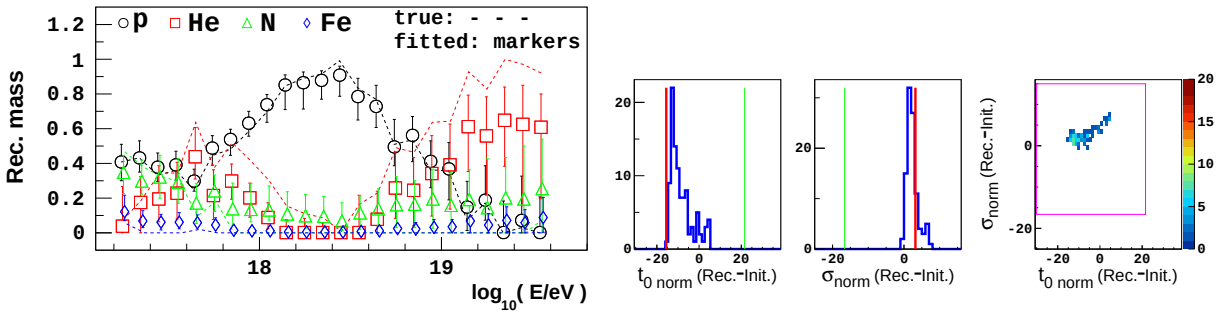
The mass composition reconstruction accuracy (absolute offsets in the medians) of the EPOS-LHC fit to EPOS-LHC based data changes less with energy than the accuracy of the QGSJetII-04 fit to the EPOS-LHC data. The mass reconstruction of the QGSJetII-04 fit is less accurate at lower energies compared to higher energies, because the shape parameter differences between EPOS-LHC and QGSJetII-04 are not constant over the energy range, or the same for all masses (see Figure 4.8a). The QGSJetII-04 parameterisation fitted to EPOS-LHC based data will not be able to fit a proton value of $t_{0_{\text{norm}}}$ and σ_{norm} that results in the correct t_0 and σ at all energies for all primary masses. Even if this were possible, differences in λ between the models would result in a mass reconstruction that is more accurate at particular energies for particular masses.

The fixed λ values of the fitted parameterisations impact the $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$ of the primary distributions fitted (see Equation (4.8)), resulting in a potential discrepancy with the data. To partially compensate for the effect of incorrect λ values on the $\langle X_{\max} \rangle$ of the fitted parameterisation, $t_{0_{\text{norm}}}$ may adjust accordingly. Likewise, σ_{norm} can adjust appropriately to partially compensate for the effect of λ on the fitted $\sigma(X_{\max})$.

Figure 5.11a shows the Sibyll2.3 fit to the same EPOS-LHC data results in a reconstructed mass that is representative of the true mass, but this mass reconstruction is not as accurate as the EPOS-LHC and QGSJetII-04 fits to this data. This is because a $t_{0_{\text{norm}}}$ and σ_{norm} shift of the Sibyll2.3 parameterisation does not align the Sibyll2.3 t_0 and σ parameterisations with the EPOS-LHC (or QGSJetII-04) descriptions as adequately as the EPOS-LHC or QGSJetII-04 descriptions can be aligned with each other (compare Figures 4.8a, 4.8b and 4.8c). Larger differences in the λ Sibyll2.3 parameterisation relative to EPOS-LHC further hinders an accurate mass reconstruction. The shape parameter disagreement between Sibyll2.3 and other models changes more with energy than the disagreement between EPOS-LHC and QGSJetII-04, and the Sibyll2.3 proton shape parameter separation from heavier nuclei is also greater.

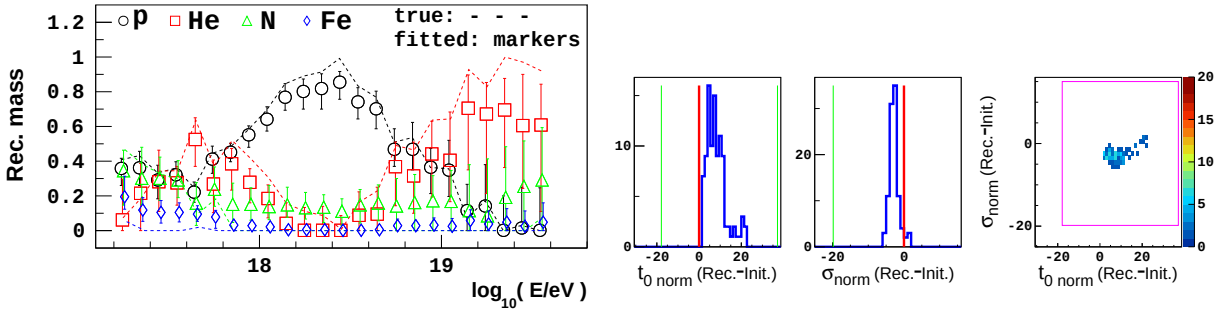
For the Sibyll2.3 fit to EPOS-LHC data, the fitted proton $t_{0_{\text{norm}}}$ is smaller on average than the true proton $t_{0_{\text{norm}}}$ of the data, as less of a shift in $t_{0_{\text{norm}}}$ is needed to adequately align the He, N and Fe Sibyll2.3 t_0 predictions with the EPOS-LHC predictions. As the $t_{0_{\text{norm}}}$ shifts have resulted in the t_0 Sibyll2.3 descriptions of heavier nuclei being better aligned with the data than the proton description, a more accurate mass composition is fitted in the bins which do not contain a significant fraction of EPOS-LHC proton data. Unlike $t_{0_{\text{norm}}}$, the $\langle X_{\max} \rangle$ of a primary's fitted distribution is not directly sensitive to the fit of σ_{norm}, consequently the fitted σ_{norm} has less impact on the reconstructed mass.

The spread of $t_{0_{\text{norm}}}$ and σ_{norm} from the Sibyll2.3 fits is greater than the spread of these coefficients in the EPOS-LHC and QGSJetII-04 fits (compare Figure 5.11b with Figures 5.9b and 5.10b), indicating the fitted $X_{\max}$ data is able to constrain the EPOS-LHC and QGSJetII-04 fits of $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions better than the Sibyll2.3 fit.



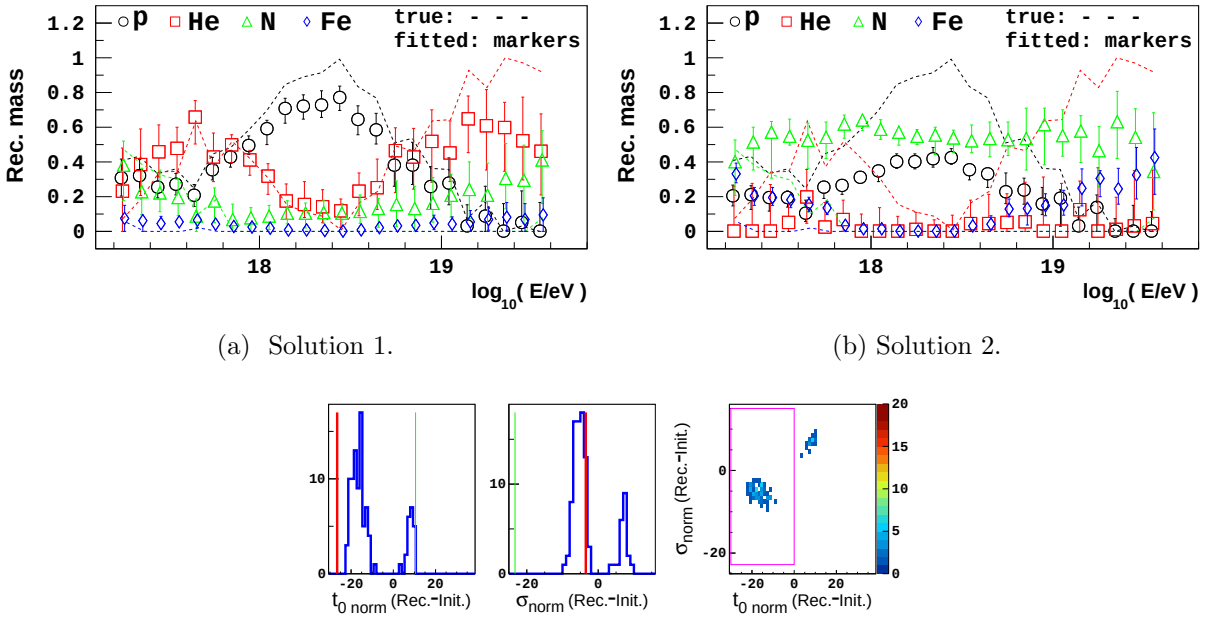
(a) (b)

Figure 5.12: Fits of QGSJetII-04 based $X_{\max}$ data with the EPOS-LHC parameterisation.



(a) (b)

Figure 5.13: Fits of QGSJetII-04 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



(a) Solution 1. (b) Solution 2.

(c) Fits inside the magenta region correspond to Solution 1, those outside correspond to Solution 2.

Figure 5.14: Fits of QGSJetII-04 based $X_{\max}$ data with the Sibyll2.3 parameterisation.

Similar to the earlier figures presented, Figures 5.12a, 5.13a, 5.14a and 5.14b displays the mass composition results from fitting the mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm} of either the EPOS-LHC, QGSJetII-04 or Sibyll2.3 parameterisations respectively, to 100 data sets generated from the parameterisation which resulted when the mass fractions of the **QGSJetII-04** parameterisation were fitted to the FD and HEAT Auger $X_{\max}$ data. The true mass composition of the mock data is the mass composition from this QGSJetII-04 fit to the FD and HEAT $X_{\max}$ data (Section 8.1.1). Figures 5.12b, 5.13b and 5.14c display the fitted values of $t_{0_{\text{norm}}}$ and σ_{norm} from the EPOS-LHC, QGSJetII-04 or Sibyll2.3 fits respectively to the QGSJetII-04 based data.

The mass fraction, $t_{0_{\text{norm}}}$ and σ_{norm} fit of the EPOS-LHC parameterisation to QGSJetII-04 based mock data reconstructs the mass composition fractions with absolute offsets in the median from the true mass that are slightly worse overall than the QGSJetII-04 parameterisation fit to the same data. Both fits reconstruct the mass composition to within absolute offsets in the medians of less than 15% at most energies below $10^{18.8}$ eV, but above this energy the mass composition reconstruction becomes less accurate due to the fewer statistics of these bins.

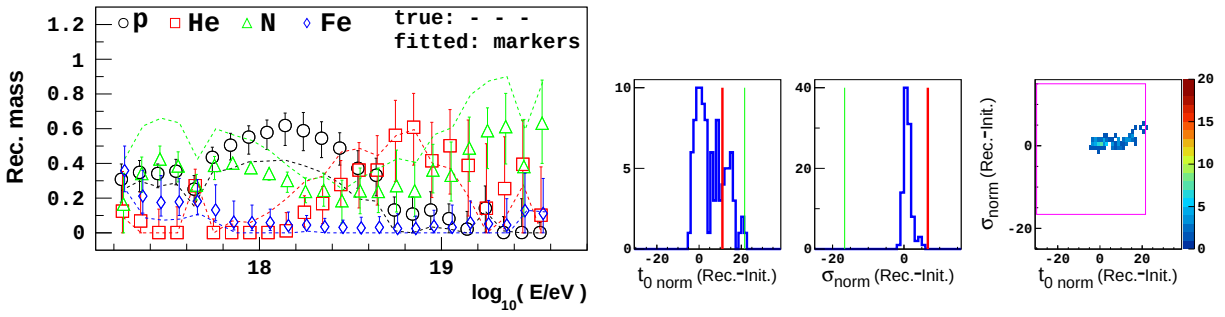
Compared to the fits of the EPOS-LHC based data, the absolute offsets in the median mass fractions from the true mass are worse for the fits to the QGSJetII-04 based data because this data consists of predominantly protons and helium, whereas the EPOS-LHC data consists of a greater mass dispersion over the energy range which is easier to fit accurately. Additionally, protons and helium are harder to fit than heavier masses, particularly if there are few events. The helium distribution is wide and overlaps considerably with adjacent primaries, which means at the highest energies where the statistics are low, the data does not form a clearly defined distribution. Therefore, it is easy for the fit to shift the shape coefficients such that the scattered helium data is partially described by adjacent primaries. The QGSJetII-04 based helium data in particular is harder to fit than EPOS-LHC helium data, as it has a wider spread (see Figure 4.6). Therefore $t_{0_{\text{norm}}}$ and σ_{norm} are not constrained as strongly in the fits of the QGSJetII-04 based data. This is evidenced by the spread in the fitted values of $t_{0_{\text{norm}}}$, which is greater for the fits to QGSJetII-04 data than the fits to EPOS-LHC data (compare Figures 5.12b and 5.13b to Figures 5.9b and 5.10b).

As we are not fitting CONEX v4r37 $X_{\max}$ data, the less accurate representation of proton and helium CONEX v4r37 data by our $X_{\max}$ parameterisations (see Section 5.1) has no effect on the poorer fit of this QGSJetII-04 parameterisation based data.

Comparing Figures 5.12a and 5.13a, we see in the energy range of $10^{18.15}$ eV to $10^{18.65}$ eV, the EPOS-LHC fit of QGSJetII-04 data reconstructions the mass composition more accurately than the QGSJetII-04 fit. This is because the disagreement between the EPOS-LHC and QGSJetII-04 shape parameter predictions are not constant with energy (see Figure 4.8a), therefore at some energies the EPOS-LHC shape parameter values are better aligned with the QGSJetII-04 based data than the QGSJetII-04 parameterisation, despite the QGSJetII-04 fit being better aligned overall.

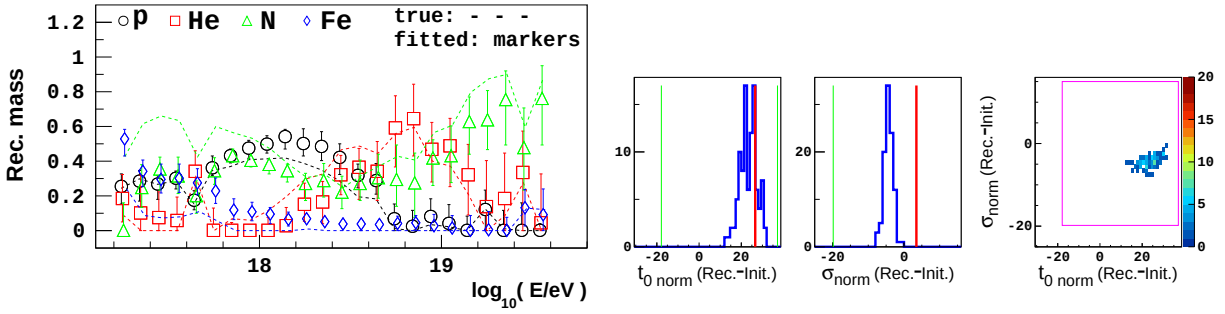
The Sibyll2.3 fit, in this case to QGSJetII-04 based data, again does not reconstruct the mass composition as precisely as the EPOS-LHC or QGSJetII-04 fits. Figure 5.14c shows that the Sibyll2.3 fits consist of two $t_{0_{\text{norm}}}$ and σ_{norm} populations, a consequence of the degeneracy between the fitted coefficients and the mass fractions. If the mass fractions are taken only from the $t_{0_{\text{norm}}}$ and σ_{norm} distributions closest to the true values of the data (Solution 1 [enclosed by the magenta box], Figure 5.14a), the reconstructed mass composition is representative of the true mass, but the absolute offsets in the medians are not as good as those of the EPOS-LHC fit to this QGSJetII-04 data. If the alternate $t_{0_{\text{norm}}}$ and σ_{norm} solution is taken (Solution 2, Figure 5.14b), the reconstructed mass composition does not reflect the true mass at all.

Figures 5.15, 5.16 and 5.17 displays the results from fitting the mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm} of either the EPOS-LHC, QGSJetII-04 and Sibyll2.3 parameterisations respectively, to 100 data sets generated from the parameterisation which resulted when the mass fractions of the **Sibyll2.3** parameterisation were fitted to the FD and HEAT Auger $X_{\max}$ data. For the Sibyll2.3



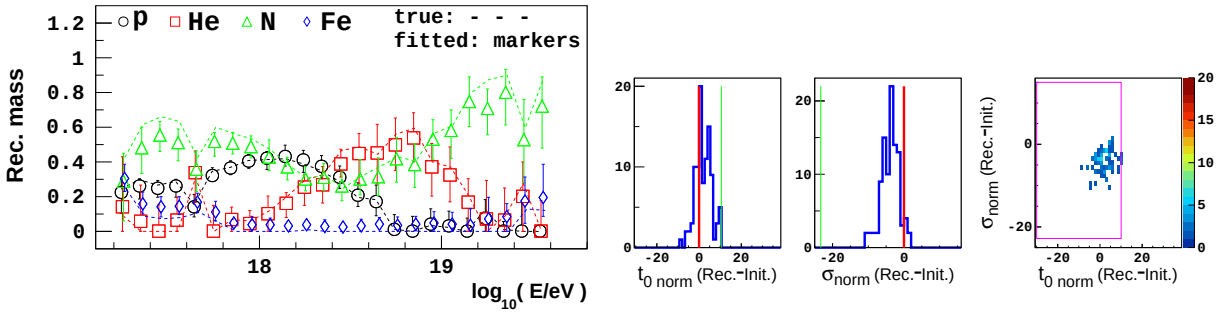
(a) (b)

Figure 5.15: Fits of Sibyll2.3 based $X_{\max}$ data with the EPOS-LHC parameterisation.



(a) (b)

Figure 5.16: Fits of Sibyll2.3 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



(a) Trimmed. (b)

Figure 5.17: Fits of Sibyll2.3 based $X_{\max}$ data with the Sibyll2.3 parameterisation.

fit to Sibyll2.3 based data, the absolute offsets in the medians of the fitted mass fractions are less than 10% from the true mass at most energies.

As the Sibyll2.3 $X_{\max}$ parameterisation is not as similar to the EPOS-LHC and QGSJetII-04 parameterisations as the latter are to each other, the EPOS-LHC and QGSJetII-04 fits to the Sibyll2.3 based data struggle to reconstruct the true mass composition, but the general mass composition transition with energy indicated by the fits is correct. The QGSJetII-04 parameterisation is more similar to the Sibyll2.3 parameterisation than EPOS-LHC, consequently the QGSJetII-04 fit results in a better mass reconstruction. The shape coefficients have shifted such that the EPOS-LHC and QGSJetII-04 parameterisations align with the Sibyll2.3 parameterisation (and therefore with the data) fairly well at higher energies, but are poorly aligned at lower energies, resulting in a very inaccurate mass composition reconstruction in the lower energy bins.

Conclusions

The ability of the $t_{0_{\mathrm{norm}}}$ and σ_{norm} fits of the EPOS-LHC, QGSJetII-04 and Sibyll2.3 X_{max} parameterisations to reconstruct the general mass composition trend of data based on any of these three parameterisations, shows that the normalisations of t_0 and σ are the most relevant differences between these parameterisations in regards to reconstructing the mass composition. Fitting the mass fractions, $t_{0_{\mathrm{norm}}}$ and σ_{norm} of any of the three X_{max} parameterisations to mock X_{max} data similar to the Auger distributions results in a reconstruction of the mass composition that represents the true mass distribution trend of the data, regardless of the model the mock data is based upon. Therefore, the EPOS-LHC, QGSJetII-04 and Sibyll2.3 fits of the mass fractions, $t_{0_{\mathrm{norm}}}$ and σ_{norm} to the combined FD and HEAT Auger data are expected to result in a consistent and accurate description of the mass composition transition, provided the shape coefficients which are fixed to the data are a reasonably correct description of nature. We apply this mass fraction, $t_{0_{\mathrm{norm}}}$ and σ_{norm} fit to the Auger data in Section 8.1.2.

5.2.1.3 Fitting $t_{0_{\mathrm{norm}}}$, σ_{norm}, λ_{norm} and the mass fractions

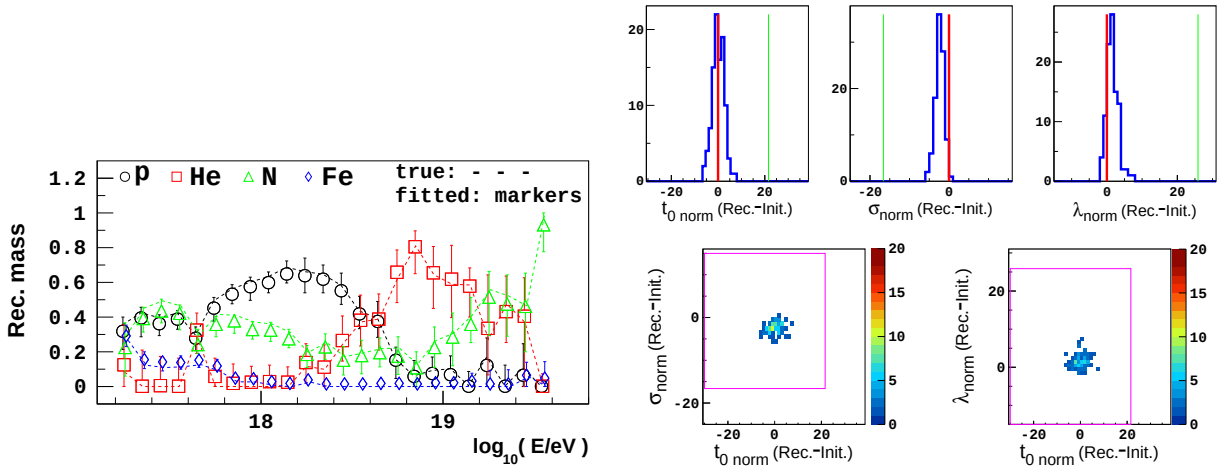
Figures 5.18 to 5.26 displays the mass composition and coefficient results from fitting the mass fractions, $t_{0_{\mathrm{norm}}}$, σ_{norm} and λ_{norm} of either the EPOS-LHC, QGSJetII-04 or Sibyll2.3 parameterisations to 100 X_{max} data sets generated from the parameterisations which resulted when only the mass fractions of a parameterisation were fitted to the combined FD and HEAT Auger X_{max} data.

Comparing the figures in this section to those of the previous section where only $t_{0_{\mathrm{norm}}}$, σ_{norm} and the mass fractions were fitted, we see that adding λ_{norm} to the fit of these distributions will not usually introduce significant systematic offsets to the mass composition reconstruction. When data is fitted with its corresponding parameterisation, there is very little change in the reconstruction. The most noticeable example of an increase in the absolute offsets in the median reconstructed mass from the true mass is the fit of QGSJetII-04 data with the Sibyll2.3 parameterisation. Comparing Figure 5.23a to Figure 5.14a, the absolute offsets in the median are about 10% greater in the fit of $t_{0_{\mathrm{norm}}}$, σ_{norm} and λ_{norm}.

Fitting three coefficients compared to just $t_{0_{\mathrm{norm}}}$ and σ_{norm} can sometimes result in additional bi-modality, or less. For example, the Sibyll2.3 parameterisation fit of these three coefficients to the EPOS-LHC data results in some additional bi-modality in $t_{0_{\mathrm{norm}}}$ and σ_{norm} (compare Figure 5.20b to Figure 5.11b). However, there is significantly reduced bi-modality in the three-coefficient fit of the Sibyll2.3 parameterisation to QGSJetII-04 data, compared to the fit of only $t_{0_{\mathrm{norm}}}$ and σ_{norm} (compare Figure 5.23b to Figure 5.14c).

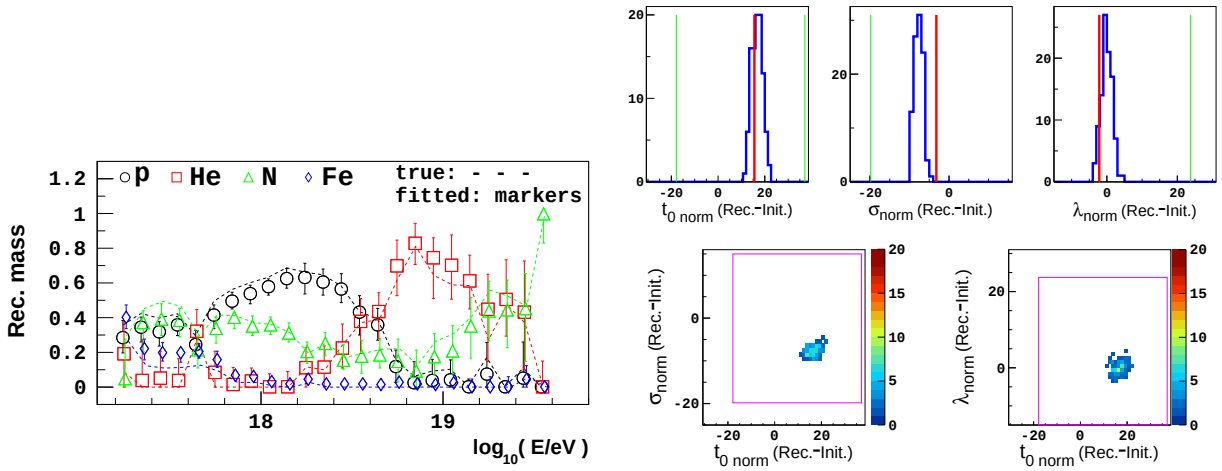
The figures show that accurate values of λ_{norm} are fitted, which is not a huge feat as none of the fitted parameterisations required significant shifts of λ_{norm} to align with the data. The X_{max} parameterisation predictions of λ are fairly consistent between the three hadronic models, the greatest discrepancy is between EPOS-LHC and Sibyll2.3. In Figure 4.8, we see that the scale of the t_0 difference between models is greater than the scale of the λ difference ($\sim 0 - 30\,\mathrm{g/cm^2}$ in t_0 compared to $\sim 0 - 10\,\mathrm{g/cm^2}$ in λ). Additionally, the fitted t_0 parameterisation exerts much greater influence on the fitted mass composition than λ. As the data fitted in this section was generated from values of λ_{norm} and λ which are somewhat consistent between the parameterisations, in terms of the reconstructed mass composition, when fitting a parameterisation to data of a different model, the three-coefficient fit does not result in a drastic improvement. The λ_{norm} applied in the two-coefficient fits was already close to the true values of the data.

Adding a third coefficient to the fit does increase the degeneracy between the fitted parameters, which is the cause of the marginal decrease in the accuracy of the reconstructed mass compared to the two-coefficient fit. The increased degeneracy also results in increased statistical error, which can be seen clearly in the fit of QGSJetII-04 data with the QGSJetII-04 parameterisation (compare Figures 5.22a and 5.13a).



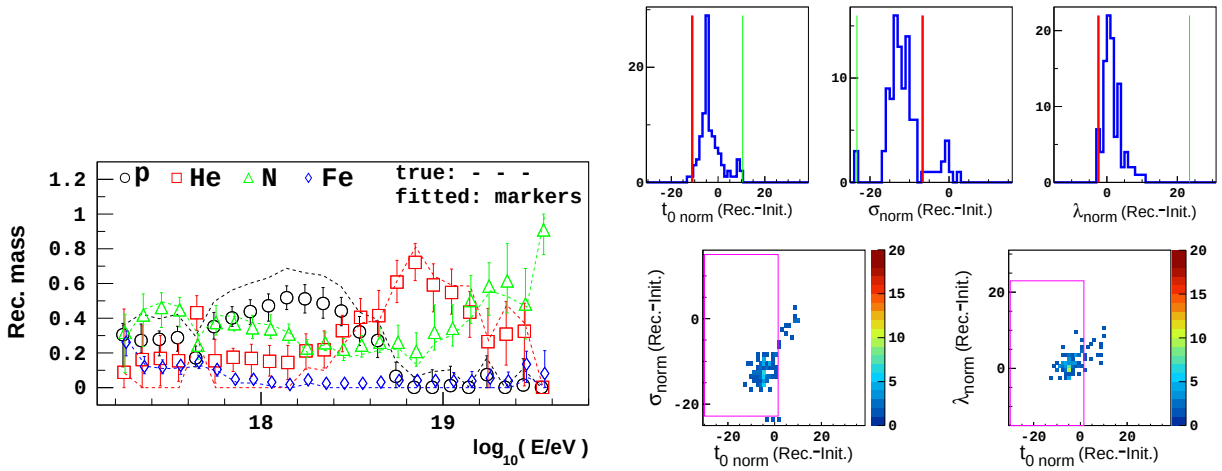
(a) (b)

Figure 5.18: Fits of EPOS-LHC based $X_{\max}$ data with the EPOS-LHC parameterisation.



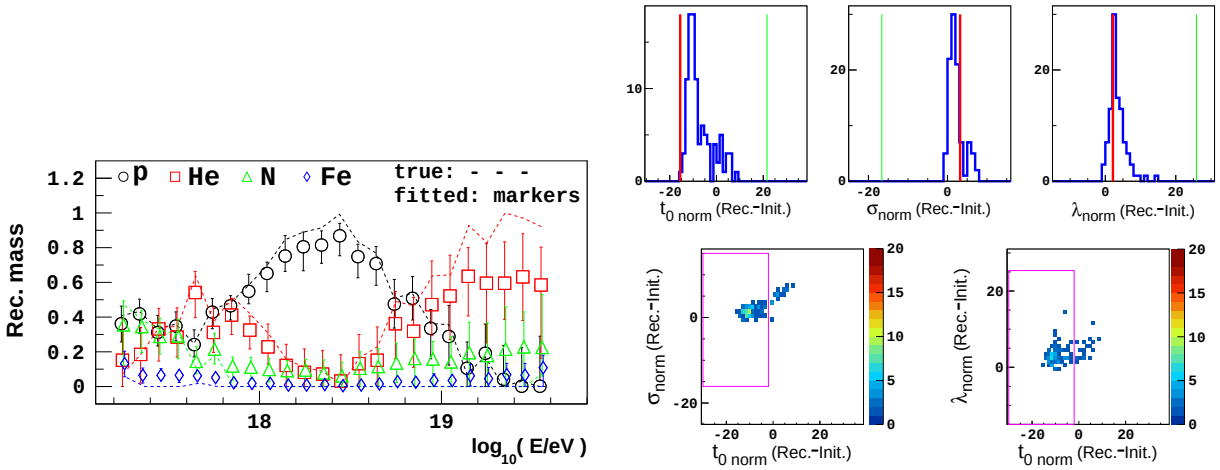
(a) (b)

Figure 5.19: Fits of EPOS-LHC based $X_{\max}$ data with the QGSJetII-04 parameterisation.



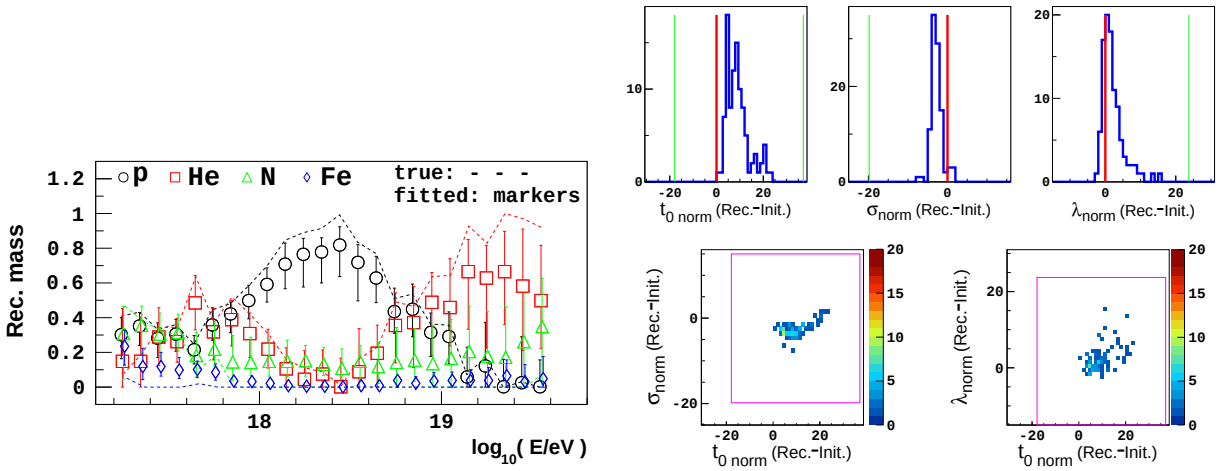
(a) Solution 1. (b)

Figure 5.20: Fits of EPOS-LHC based $X_{\max}$ data with the Sibyll2.3 parameterisation.



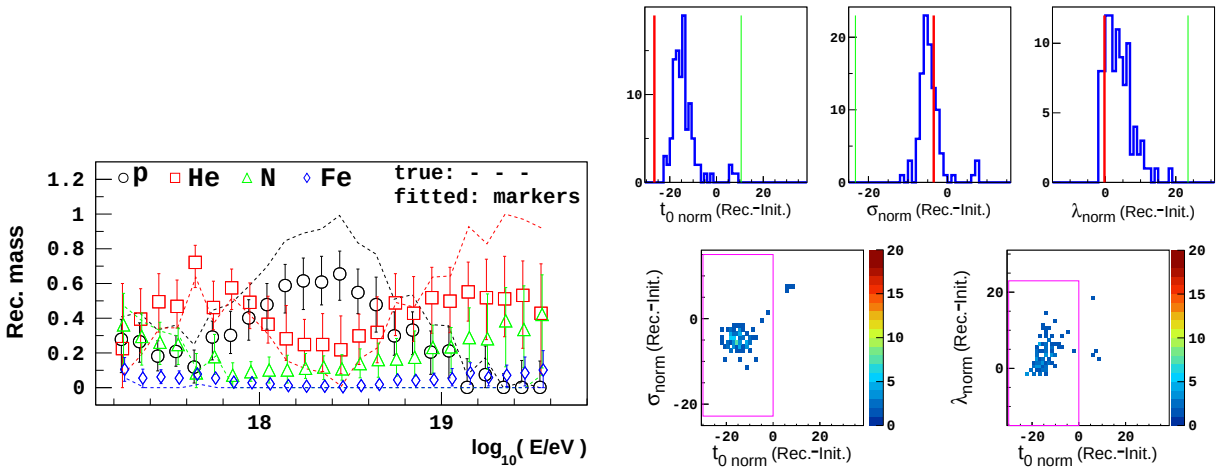
(a) Solution 1. (b)

Figure 5.21: Fits of QGSJetII-04 based $X_{\max}$ data with the EPOS-LHC parameterisation.



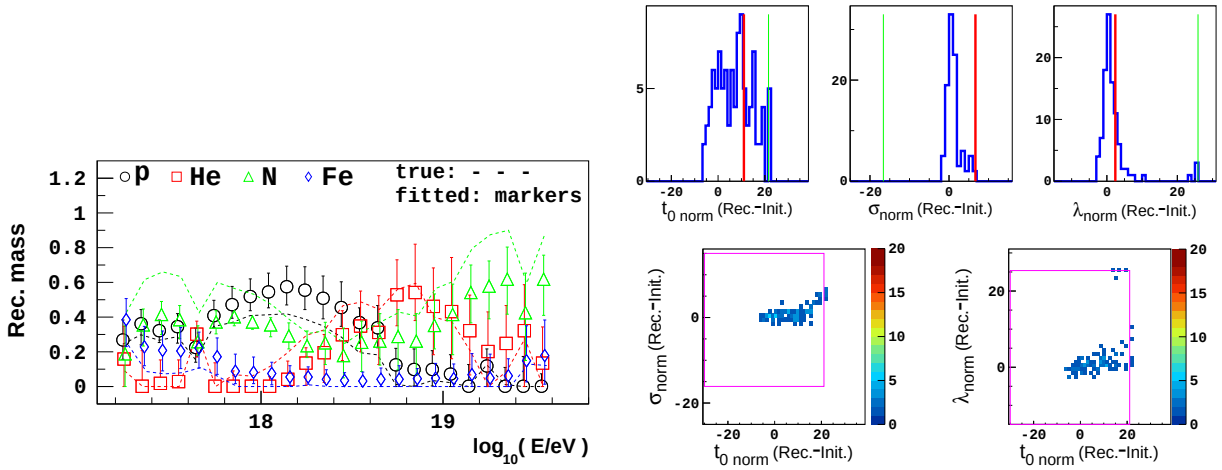
(a) (b)

Figure 5.22: Fits of QGSJetII-04 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



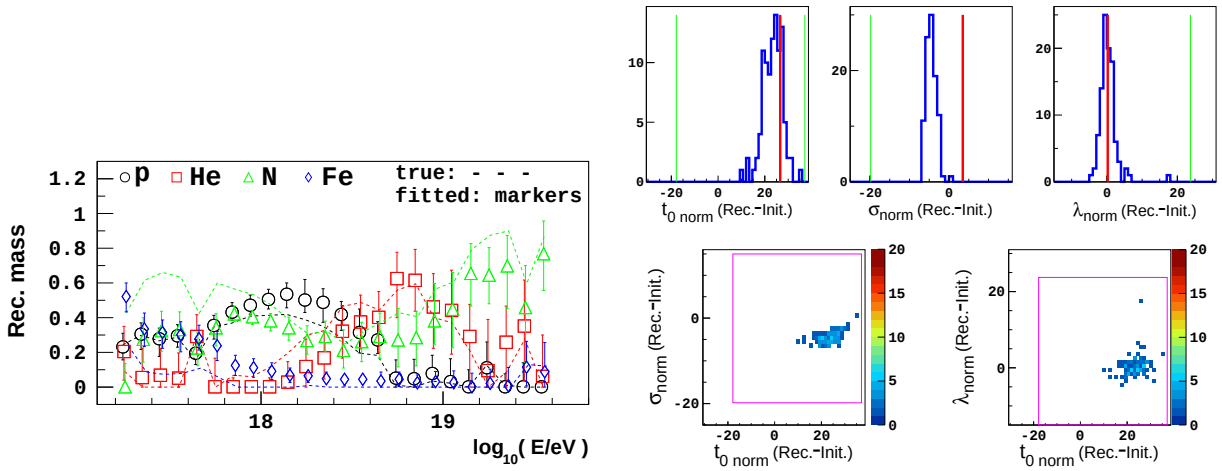
(a) Solution 1. (b)

Figure 5.23: Fits of QGSJetII-04 based $X_{\max}$ data with the Sibyll2.3 parameterisation.



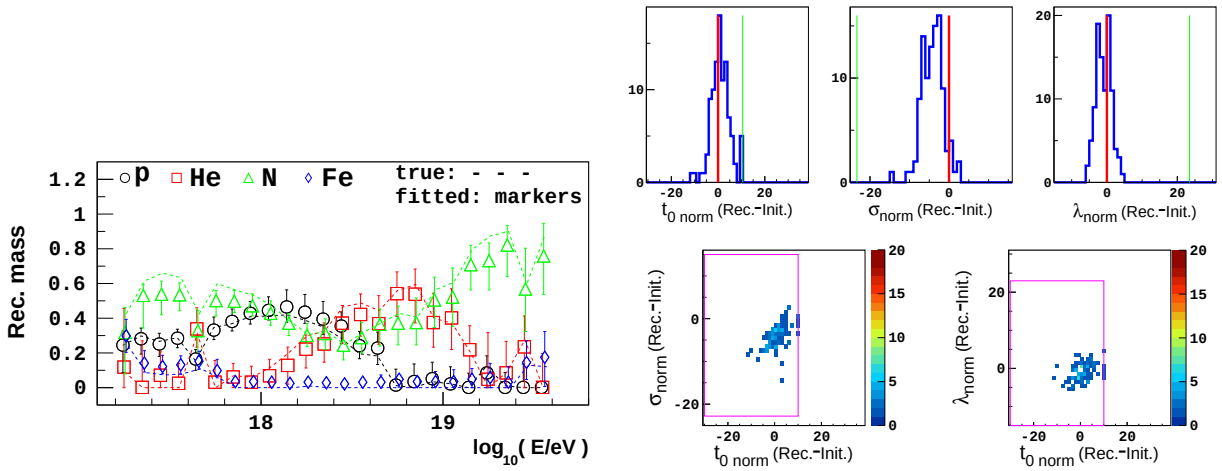
(a) Trimmed. (b)

Figure 5.24: Fits of Sibyll2.3 based $X_{\max}$ data with the EPOS-LHC parameterisation.



(a) (b)

Figure 5.25: Fits of Sibyll2.3 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



(a) Trimmed. (b)

Figure 5.26: Fits of Sibyll2.3 based $X_{\max}$ data with the Sibyll2.3 parameterisation.

Fitting mock data with a value of λ_{norm} significantly different to the initial predictions of the $X_{\max}$ parameterisations

In the following figures, we apply our three-coefficient fit to data that was generated from a value of λ_{norm} that is shifted from the prediction of our EPOS-LHC parameterisation. The purpose of this section is to see if the three-coefficient fit is able to fit the correct value of λ_{norm}, as well as the correct $t_{0_{\text{norm}}}$, σ_{norm} and mass fractions, when the true value of λ_{norm} is significantly different to the initial predictions of our three parameterisations.

The value of λ_{norm} in our EPOS-LHC parameterisation was shifted by $+10\text{g/cm}^2$, and the value of $t_{0_{\text{norm}}}$ was shifted by -10g/cm^2, and then only the mass fractions of this modified EPOS-LHC parameterisation were fitted to the FD and HEAT Auger data. This fitted parameterisation was then used to generate mock data sets as before. As the shifts in $t_{0_{\text{norm}}}$ and λ_{norm} were applied before fitting the Auger data, the mock distributions generated are similar to the measured Auger distributions, but the true mass composition will be considerably different from the standard EPOS-LHC fit. The mass fractions, $t_{0_{\text{norm}}}$, σ_{norm} and λ_{norm} of the standard EPOS-LHC, QGSJetII-04 and Sibyll2.3 parameterisations were then fitted to this modified EPOS-LHC based mock data. Figures 5.27 to 5.29 displays the results of these fits.

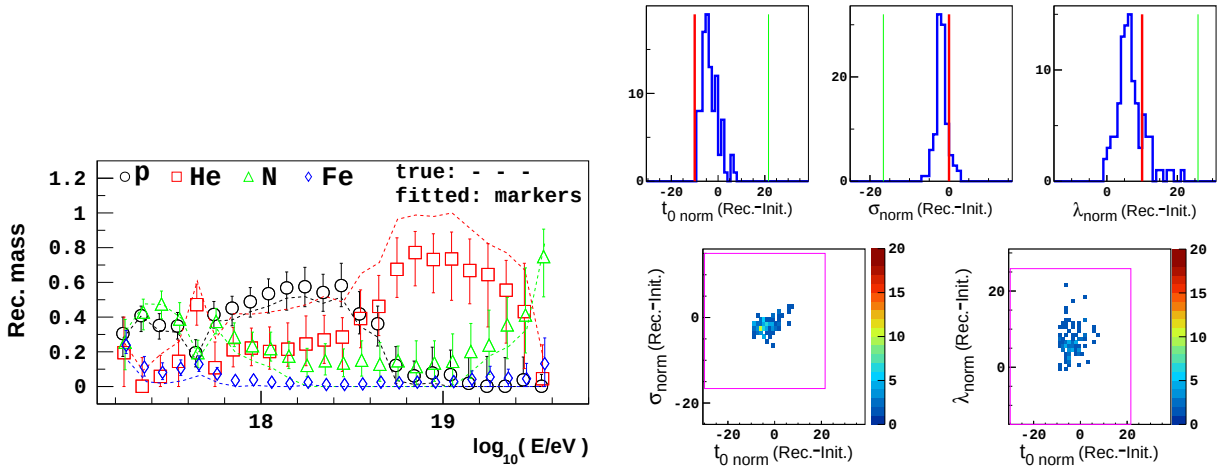
We see that the fitted values of $t_{0_{\text{norm}}}$, σ_{norm} and λ_{norm} are shifted in the appropriate directions, but they do not match the data well enough to produce a reconstruction as accurate as the previous fits. The general trend of the mass composition is reconstructed, but there are substantial systematic offsets. As seen in the fits of QGSJetII-04 based data, this modified EPOS-LHC data is difficult to fit as it is dominated by protons and helium. Additionally, as the true λ_{norm} of the data is larger than the standard EPOS-LHC value, the distribution tails of the primaries are greater, resulting in a more difficult fit as the $\sigma(X_{\max})$ of each primary distribution is larger.

The fit of the modified EPOS-LHC data with the EPOS-LHC parameterisation is capable of reconstructing the correct mass fractions, if the correct coefficients are fitted. Unfortunately, the fitted $t_{0_{\text{norm}}}$ values are overestimated, while the fitted λ_{norm} values are underestimated. However, these $t_{0_{\text{norm}}}$ and λ_{norm} shifts allow the $\langle X_{\max}\rangle$ of the fitted primary distributions to be similar to the $\langle X_{\max}\rangle$ of the true primary distributions, as $\langle X_{\max}\rangle = t_0 + \lambda$. The fits will consistently reproduce the true $\langle X_{\max}\rangle$ of the data provided the primary distributions are capable of spanning the smallest and largest $X_{\max}$ values (i.e. there is not any detector selection bias).

In Figures 5.30 to 5.32, the mass fractions, $t_{0_{\text{norm}}}$, σ_{norm} and λ_{norm} of the three parameterisations were fitted to mock data based on a modified QGSJetII-04 mass fraction fit of the FD and HEAT Auger data, where the QGSJetII-04 λ_{norm} value was shifted by -10g/cm^2 and $t_{0_{\text{norm}}}$ shifted by $+10\text{g/cm}^2$ before fitting the parameterisation to the Auger data. The absolute offsets in the median fitted mass fractions from the true mass are much smaller for these fits, as shape coefficient values were fitted that are closer to the true values of the data. The EPOS-LHC and QGSJetII-04 fits reproduce the true mass with very good accuracy considering the distribution consists of mainly protons and helium. The Sibyll2.3 fit also manages to reproduce the true mass composition trend. The true primary distributions have less spread as the tails are smaller, which helps the mass composition reconstruction.

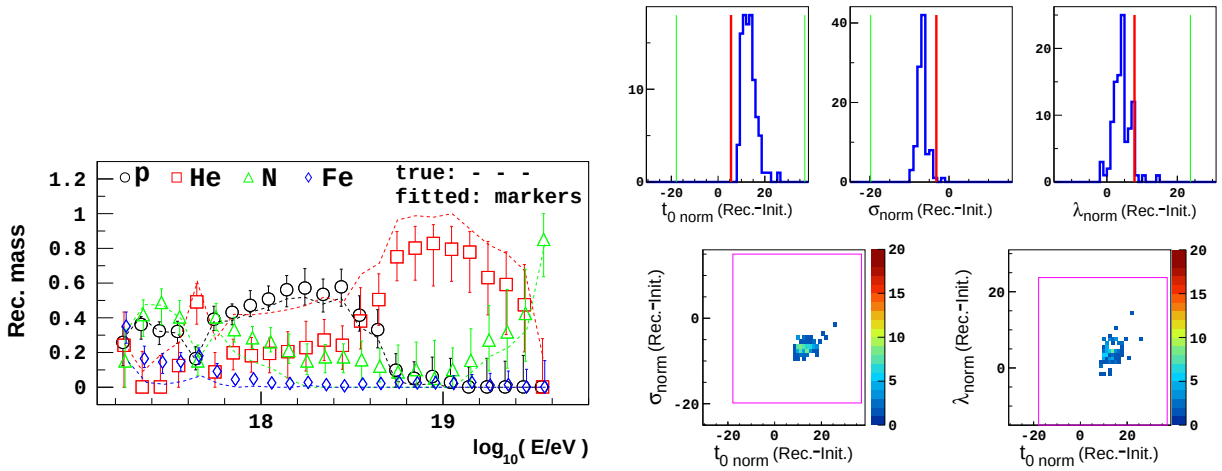
Conclusions

Although the $X_{\max}$ parameterisations have similar predictions for λ_{norm} and λ, these predictions may not align with nature. The fits of $t_{0_{\text{norm}}}$, σ_{norm}, λ_{norm} and the mass fractions are able to fit a value of λ_{norm} that is representative of the true value of the data, even if the true value is considerably different from the initial predictions. Compared to the $t_{0_{\text{norm}}}$ and σ_{norm} fit, adding λ_{norm} does not introduce considerable systematics to the reconstructed mass or fitted shape parameter coefficients. This three-coefficient fit is applied to the combined FD and HEAT Auger data in Section 8.1.3.



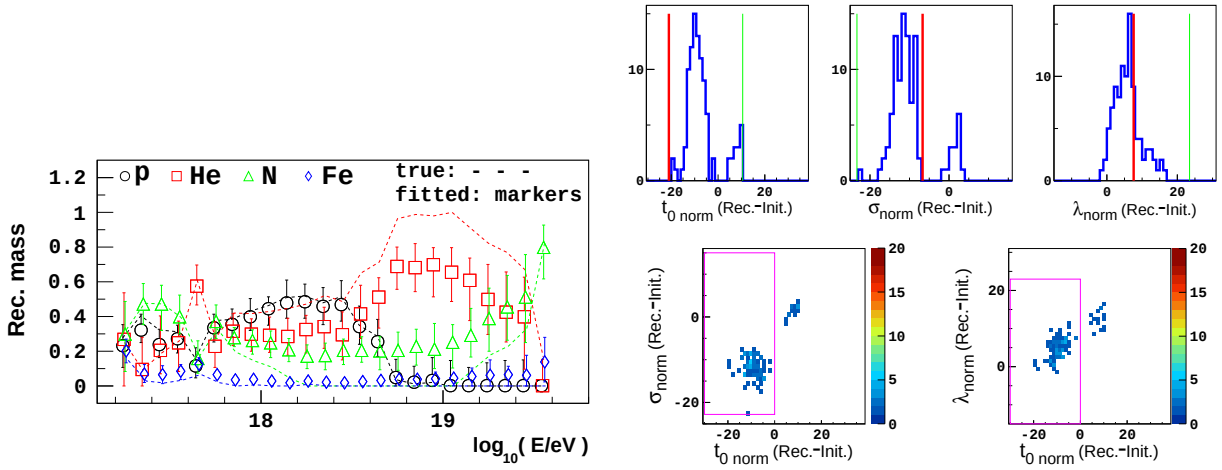
(a) (b)

Figure 5.27: Fitting the EPOS-LHC parameterisation to EPOS-LHC based $X_{\max}$ data with a $-10\mathrm{g/cm}^2$ shift in $t_{0_{\mathrm{norm}}}$ and $+10\mathrm{g/cm}^2$ shift in λ_{norm}.



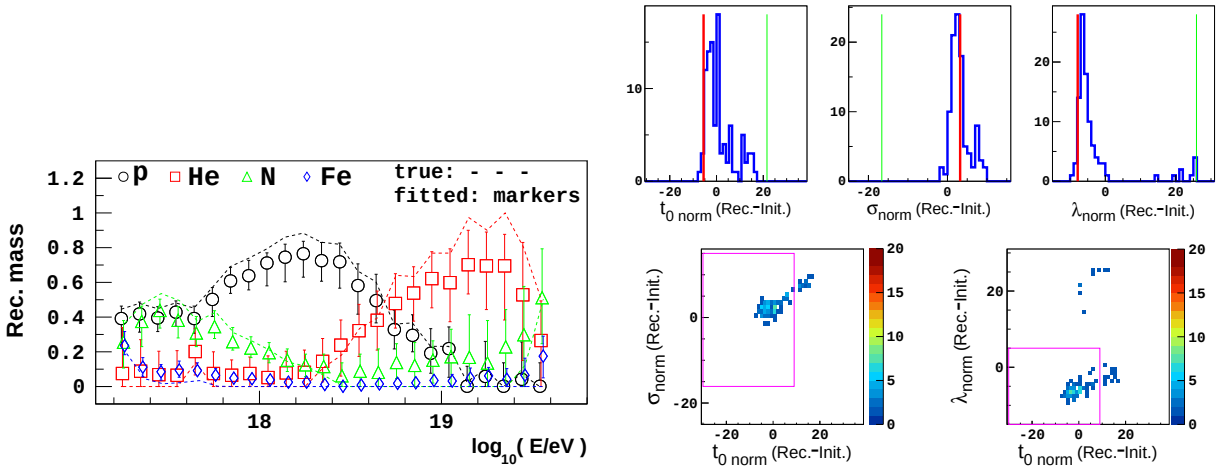
(a) (b)

Figure 5.28: Fitting the QGSJetII-04 parameterisation to EPOS-LHC based $X_{\max}$ data with a $-10\mathrm{g/cm}^2$ shift in $t_{0_{\mathrm{norm}}}$ and $+10\mathrm{g/cm}^2$ shift in λ_{norm}.



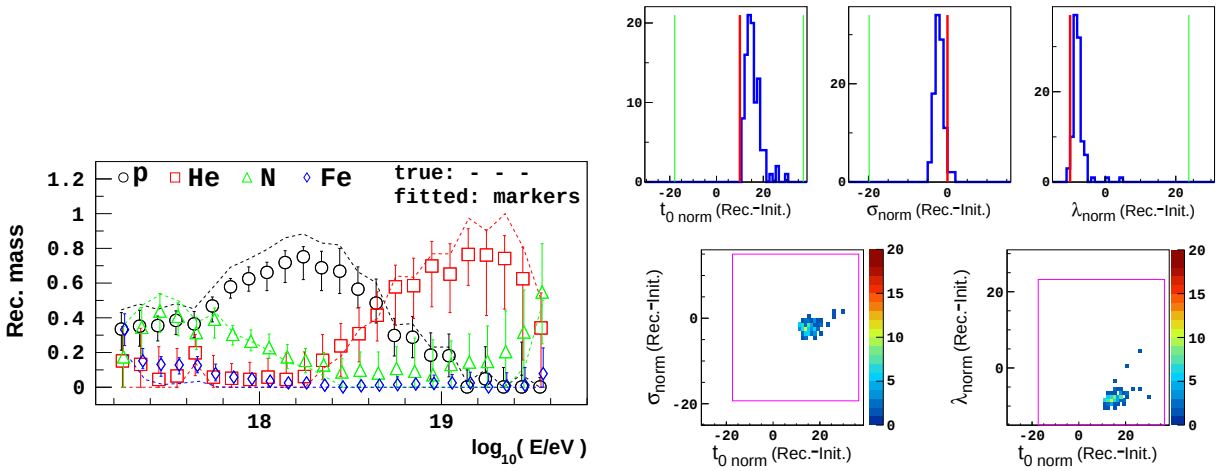
(a) Solution 1. (b)

Figure 5.29: Fitting the Sibyll2.3 parameterisation to EPOS-LHC based $X_{\max}$ data with a $-10\mathrm{g/cm}^2$ shift in $t_{0_{\mathrm{norm}}}$ and $+10\mathrm{g/cm}^2$ shift in λ_{norm}.



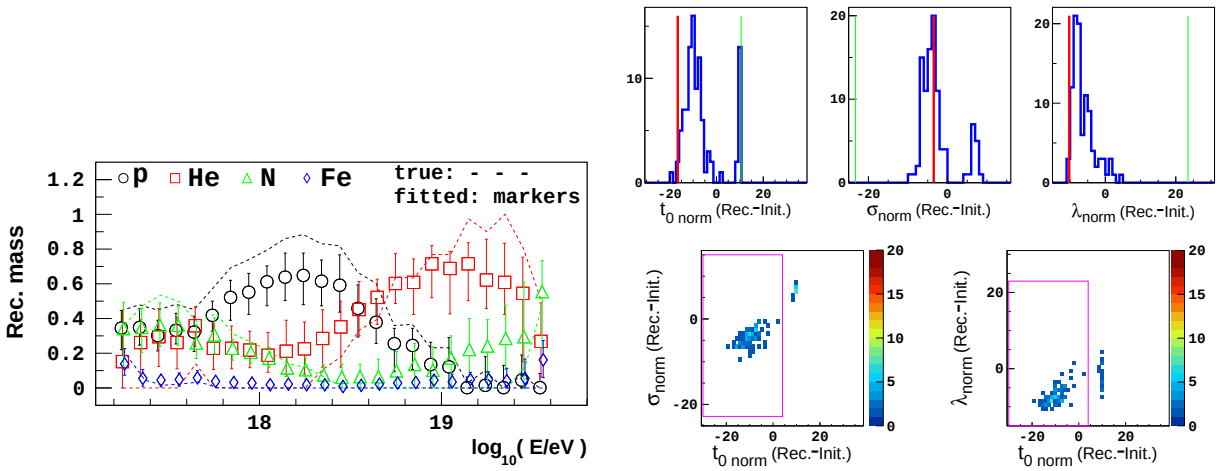
(a) Solution 1. (b)

Figure 5.30: Fitting the EPOS-LHC parameterisation to QGSJetII-04 based $X_{\max}$ data with a $+10\text{g/cm}^2$ shift in $t_{0_{\text{norm}}}$ and -10g/cm^2 shift in λ_{norm}.



(a) (b)

Figure 5.31: Fitting the QGSJetII-04 parameterisation to QGSJetII-04 based $X_{\max}$ data with a $+10\text{g/cm}^2$ shift in $t_{0_{\text{norm}}}$ and -10g/cm^2 shift in λ_{norm}.



(a) Solution 1. (b)

Figure 5.32: Fitting the Sibyll2.3 parameterisation to QGSJetII-04 based $X_{\max}$ data with a $+10\text{g/cm}^2$ shift in $t_{0_{\text{norm}}}$ and -10g/cm^2 shift in λ_{norm}.

5.2.1.4 Fitting $t_{0_{\text{norm}}}$, B, σ_{norm} and the mass fractions

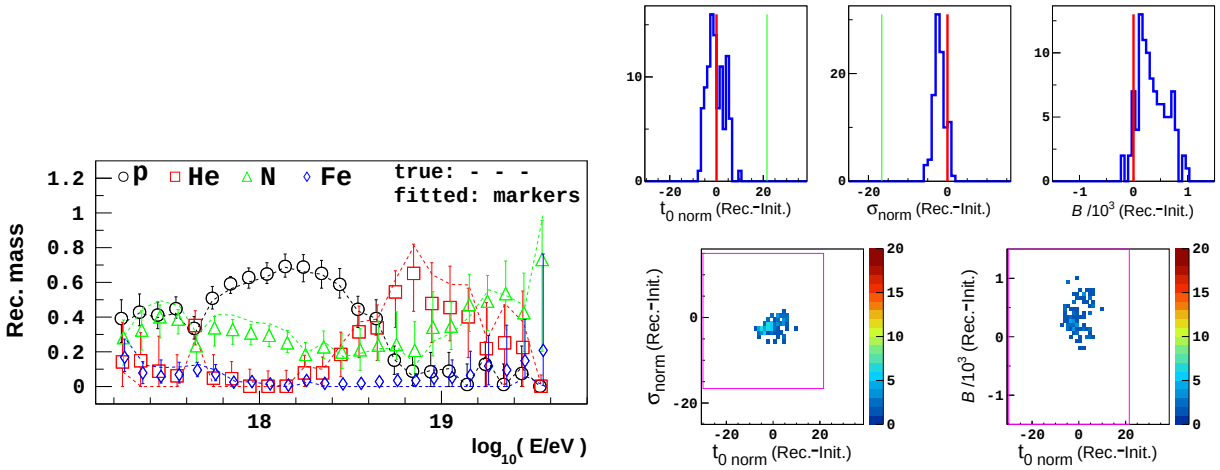
The values of B are reasonably consistent between the three parameterisations (see Table 4.1), therefore in the previous coefficient fits, the true mass composition trend of the data was able to be fitted without requiring a fit of B. If the data can constrain a fit of $t_{0_{\text{norm}}}$, B and σ_{norm}, the accuracy of the mass reconstruction will be improved when a parameterisation is fitted to data that is based on a different model.

The $\langle X_{\max} \rangle$ elongation rate applied to the data impacts the mass composition interpretation considerably. Fitting both $t_{0_{\text{norm}}}$ and B substantially increases the possible t_0 parameterisations that can be fitted to the data. The mass dispersion of the data must be greater if a fit of $t_{0_{\text{norm}}}$, B and σ_{norm} is to result in an accurate reconstruction. The added degeneracy introduced by including B will result in larger statistical errors compared to a fit of only $t_{0_{\text{norm}}}$ and σ_{norm}.

Figures 5.33, 5.34 and 5.35 displays the mass composition and coefficient results from fitting the mass fractions, $t_{0_{\text{norm}}}$, B and σ_{norm} of either the EPOS-LHC, QGSJetII-04 or Sibyll2.3 parameterisations respectively, to 100 $X_{\max}$ data sets generated from the parameterisation which resulted when only the mass fractions of the **EPOS-LHC** parameterisation were fitted to the combined FD and HEAT Auger $X_{\max}$ data.

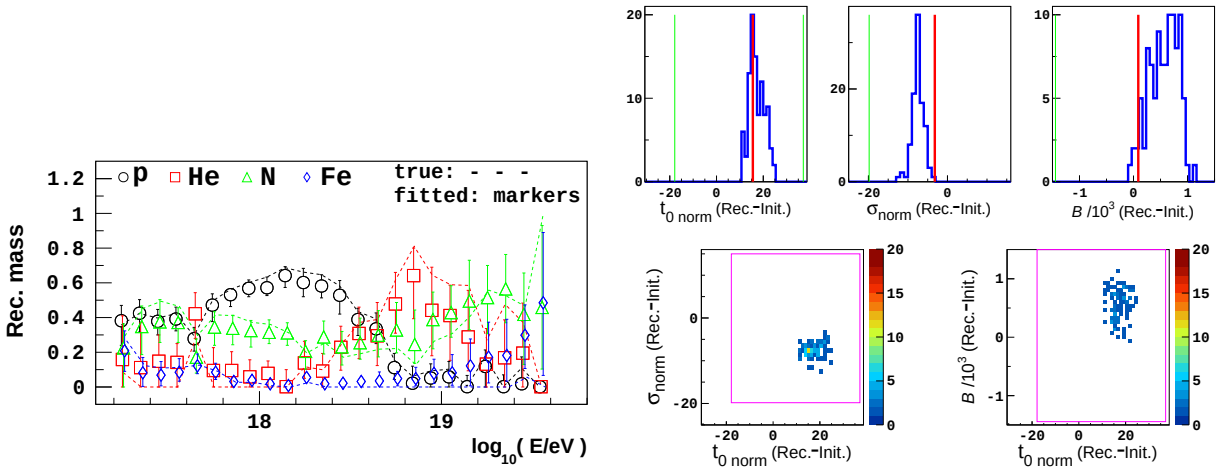
Comparing the $t_{0_{\text{norm}}}$, B and σ_{norm} coefficient fits of Figures 5.33a, 5.34a and 5.35a to the respective fits of only $t_{0_{\text{norm}}}$ and σ_{norm} in Figures 5.9a, 5.10a and 5.11a, we find that at lower energies the mass composition reconstruction is of comparable accuracy, but above $10^{18.8}$ eV the accuracy of the three-coefficient fit noticeably declines. This is due to the three-coefficient fit not constraining B adequately. As the fit gets further away from the normalisation energy of $10^{18.24}$ eV, the mass composition reconstruction accuracy will be worse, because the incorrectly fitted t_0 elongation rate increases the discrepancy between the fitted t_0 parameterisation and the true t_0 of the data. Comparing Figures 5.33b, 5.34b and 5.35b to the respective fits of only $t_{0_{\text{norm}}}$ and σ_{norm} in Figures 5.9b, 5.10b and 5.11b, we see that adding B to the fit adds some spread in the fitted values of $t_{0_{\text{norm}}}$ and σ_{norm}, but the systematic offsets in the mean fitted values from the true values are similar. In Figure 5.33b, where the EPOS-LHC parameterisation is fitted to EPOS-LHC data, we would expect the distribution of B to be centred on the true value if no systematic offset was present.

The $t_{0_{\text{norm}}}$, B and σ_{norm} fits of QGSJetII-04 based data (Figures 5.36 to 5.38) and Sibyll2.3 based data (Figures 5.39 to Figure 5.41) result in the same story. When B is added to the fit, the absolute offsets in the median reconstructed mass fractions from the true mass do not noticeably increase at lower energies, but at energies above $10^{18.8}$ eV the fit becomes less accurate, predominantly due to an incorrect value of B being fitted. However, the $t_{0_{\text{norm}}}$, B and σ_{norm} fits of the QGSJetII-04 and Sibyll2.3 data generally resulted in less of a decrease in the mass reconstruction accuracy than the EPOS-LHC fits, when compared to the $t_{0_{\text{norm}}}$ and σ_{norm} fits. This is because in the fits of QGSJetII-04 and Sibyll2.3 based data, the fitted values of B were closer to the true value of B. In the case of the Sibyll2.3 fit to QGSJetII-04 data, the three-coefficient fit improved the mass reconstruction at higher energies (compare Figures 5.38a and 5.14a).



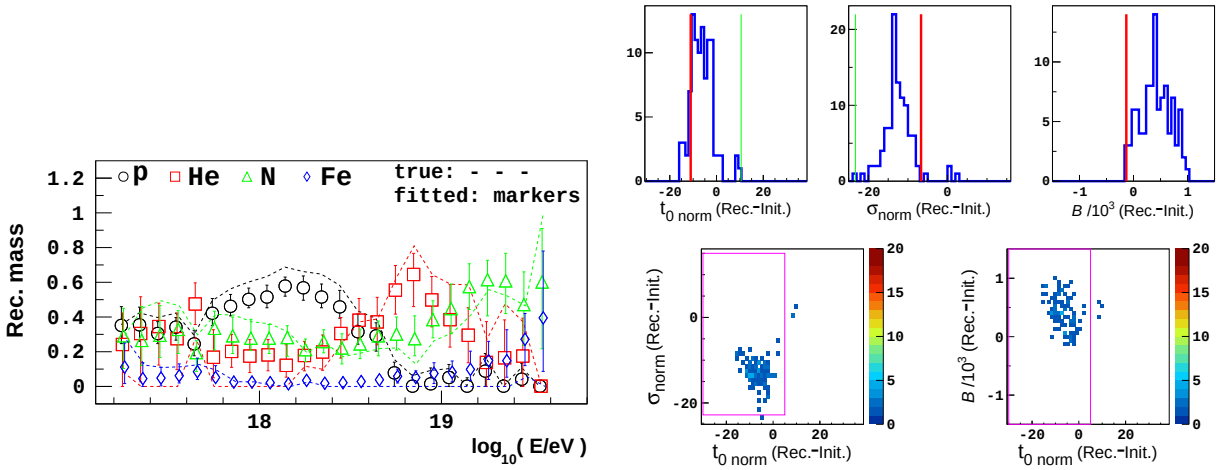
(a) (b)

Figure 5.33: Fits of EPOS-LHC based $X_{\max}$ data with the EPOS-LHC parameterisation.



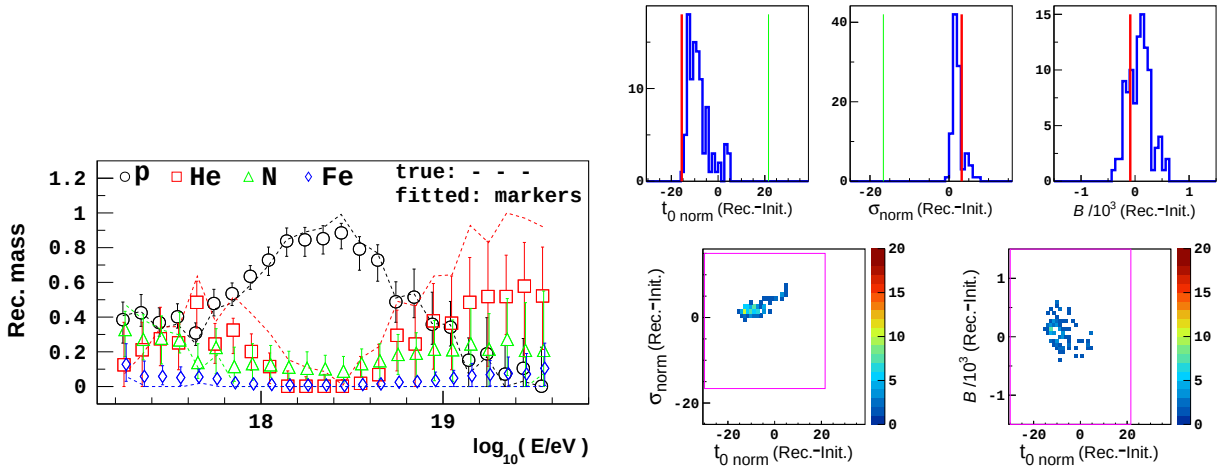
(a) (b)

Figure 5.34: Fits of EPOS-LHC based $X_{\max}$ data with the QGSJetII-04 parameterisation.



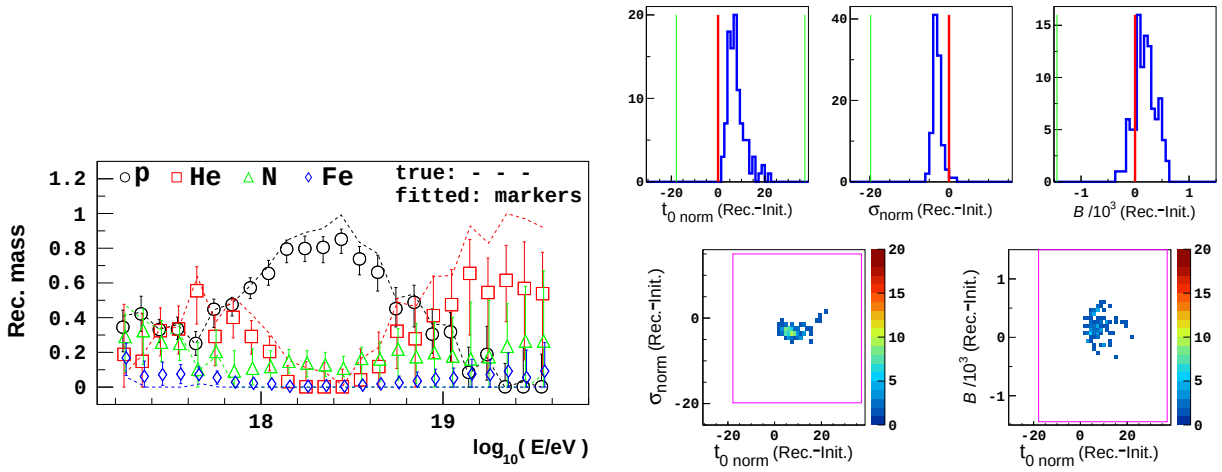
(a) Solution 1. (b)

Figure 5.35: Fits of EPOS-LHC based $X_{\max}$ data with the Sibyll2.3 parameterisation.



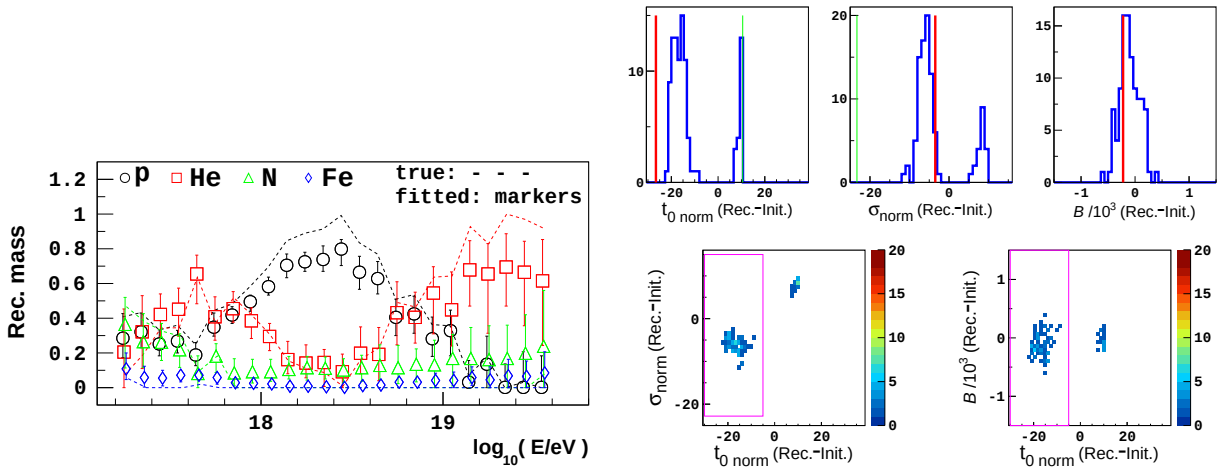
(a) (b)

Figure 5.36: Fits of QGSJetII-04 based $X_{\max}$ data with the EPOS-LHC parameterisation.



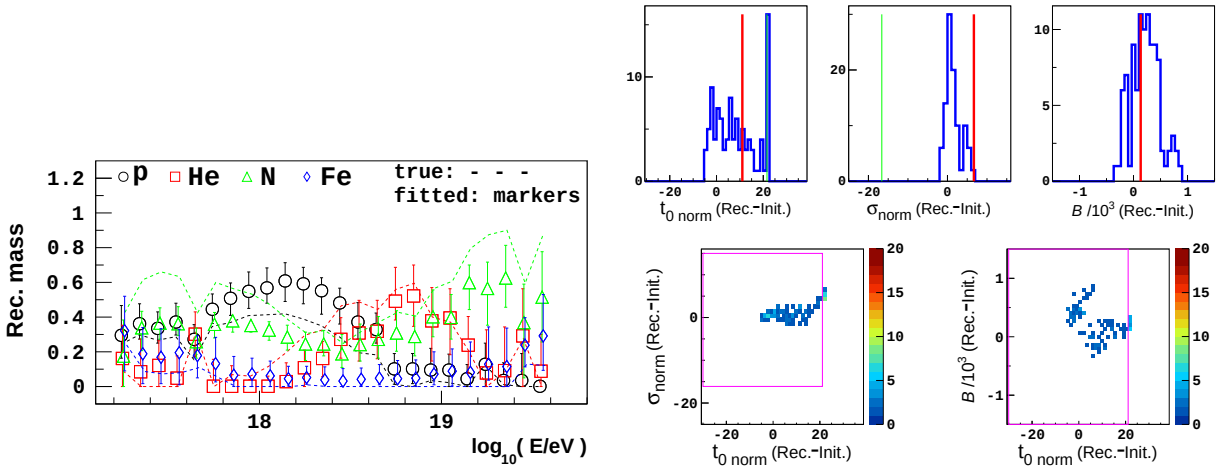
(a) (b)

Figure 5.37: Fits of QGSJetII-04 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



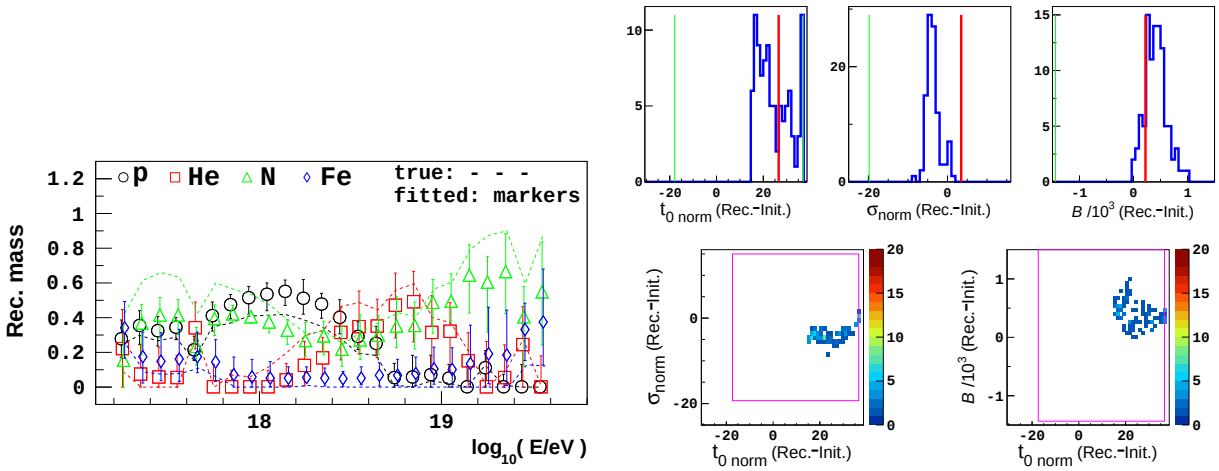
(a) Solution 1. (b)

Figure 5.38: Fits of QGSJetII-04 based $X_{\max}$ data with the Sibyll2.3 parameterisation.



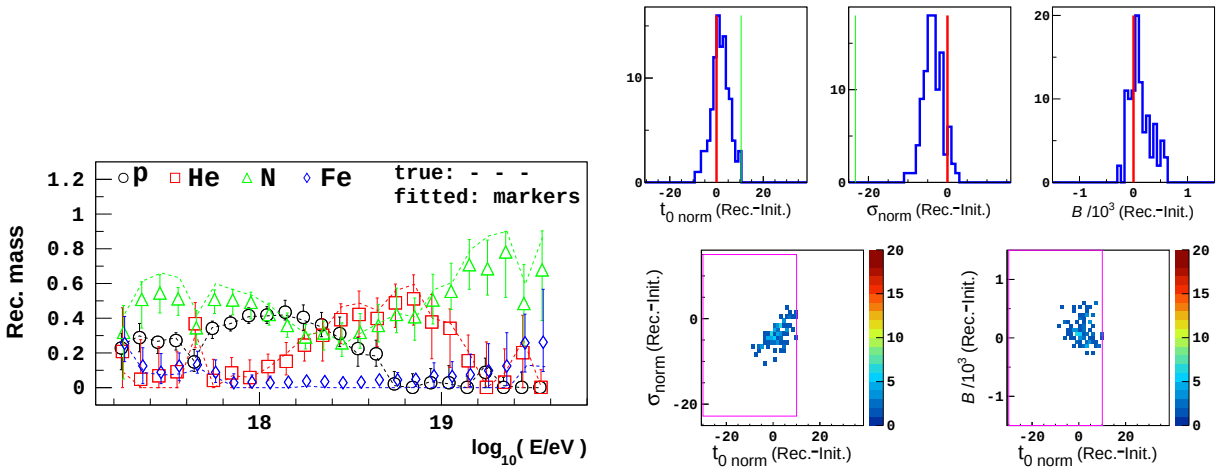
(a) Trimmed. (b)

Figure 5.39: Fits of Sibyll2.3 based $X_{\max}$ data with the EPOS-LHC parameterisation.



(a) Trimmed. (b)

Figure 5.40: Fits of Sibyll2.3 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



(a) Trimmed. (b)

Figure 5.41: Fits of Sibyll2.3 based $X_{\max}$ data with the Sibyll2.3 parameterisation.

Fitting mock data with a value of B significantly different to the initial predictions of the $X_{\max}$ parameterisations

The purpose of this section is to determine if our parameterisations can fit the correct value of B when the true B value of the data is significantly different from the initial predictions of the models. In the following figures, we apply our three-coefficient fit to data that was generated from a value of B that is shifted from the prediction of our EPOS-LHC parameterisation. The mass fractions of a modified EPOS-LHC parameterisation were fitted to the FD and HEAT Auger data, applying a value of B that was shifted by $+700\text{g/cm}^2$ (B is now $\sim 3200\text{g/cm}^2$ instead of the standard value of $\sim 2500\text{g/cm}^2$, which corresponds to a change in the t_0 elongation between $10^{18}\,\text{eV}$ and $10^{19}\,\text{eV}$ of $\sim 16.4\,\text{g/cm}^2$). This fitted parameterisation was then used to generate mock data sets. Again, the mock distributions generated are similar to the measured Auger distributions, but the true mass composition will be considerably different from the standard EPOS-LHC fit.

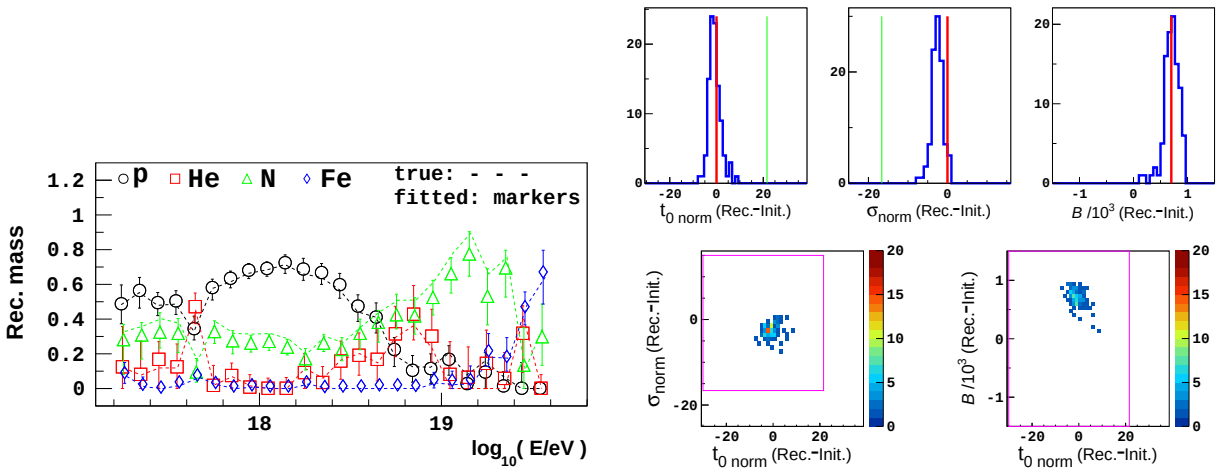
We have fitted $t_{0_{\text{norm}}}$, B, σ_{norm} and the mass fractions of our three parameterisations to this mock data based on the modified EPOS-LHC parameterisation. Figures 5.42 to 5.44 displays the results of these fits. We see that the fitted values of $t_{0_{\text{norm}}}$, B and σ_{norm} are shifted appropriately to match the data, and that B is well constrained unlike previous fits. In the case of the EPOS-LHC and QGSJetII-04 fits, this results in a very accurate mass reconstruction. The Sibyll2.3 fit also reconstructs the mass adequately, with absolute offsets in the median mass fractions of less than 15% in most energy bins. The true mass of the data consists of a significant dispersion of masses, with a greater abundance of heavier masses at the highest energies than the previous fits. This strongly constrains the fitted coefficients.

In Figures 5.45 to 5.47, $t_{0_{\text{norm}}}$, B, σ_{norm} and the mass fractions of the three parameterisations were fitted to mock data based on a different modified EPOS-LHC parameterisation. For this modified EPOS-LHC data, the value of B was shifted by -700g/cm^2 (B is now $\sim 1800\text{g/cm}^2$ instead of the standard value of $\sim 2500\text{g/cm}^2$, which corresponds to a change in the t_0 elongation between $10^{18}\,\text{eV}$ and $10^{19}\,\text{eV}$ of $\sim -16.4\,\text{g/cm}^2$), before the mass fractions were fitted to the FD and HEAT Auger data. The three-coefficient fits to this mock data do not reconstruct the highest energy bins well, as we have seen in previous data sets that contain predominantly proton-like data at lower energies and then helium-like data at higher energies. The mass composition trend indicated at the highest energies is misleading, with nitrogen mixing in equal proportion with helium above $10^{19.2}\,\text{eV}$ for the QGSJetII-04 and Sibyll2.3 fits, despite a true helium fraction of at least 85%. This mix does not constrain the fitted parameters adequately, however the coefficients are shifted in the correct direction.

Conclusions

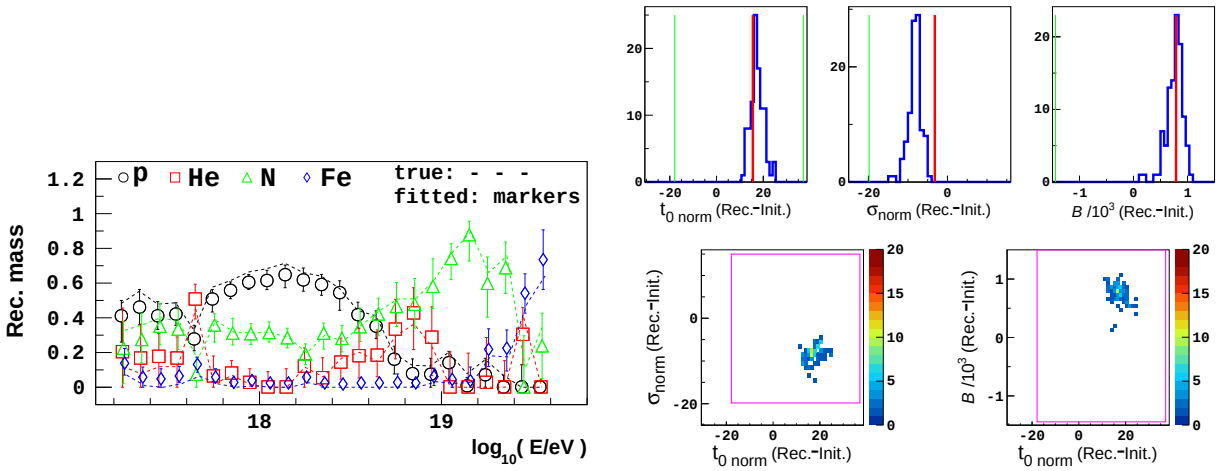
The three-coefficient fits performed in this subsection to mock Auger data sets indicate that a $t_{0_{\text{norm}}}$, B and σ_{norm} fit applied to Auger data will not reconstruct the mass fractions as accurately as a fit of $t_{0_{\text{norm}}}$ and σ_{norm}, if the shape coefficients fixed to the data are reasonably correct. This is particularly true at higher energies where the statistics are lower. If the initial parameterisation prediction for B is reasonably close to the true value of B, the three-coefficient fit is unlikely to predict a mass composition trend that is misleading, assuming the other fixed parameters are reasonable approximations of the data.

The EPOS-LHC, QGSJetII-04 and Sibyll2.3 parameterisations predict similar values for B. If this prediction of B was incorrect to a significant degree, it is possible that a fit of $t_{0_{\text{norm}}}$, B and σ_{norm} would be able to fit the correct value of B (and $t_{0_{\text{norm}}}$ and σ_{norm}) and therefore accurately reconstruct the mass composition. We apply this three-coefficient fit to the Auger data in Section 8.1.4.



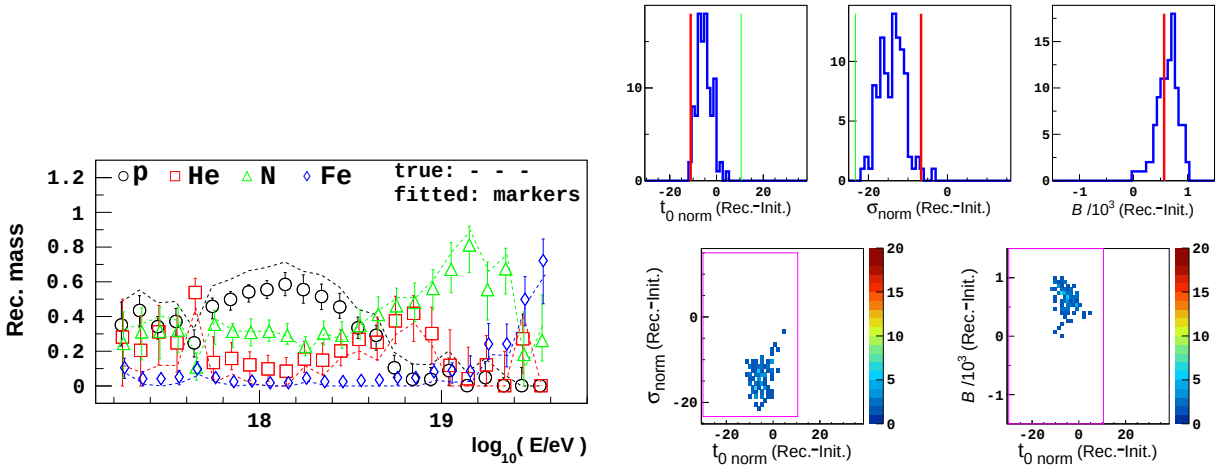
(a) (b)

Figure 5.42: Fitting the EPOS-LHC parameterisation to EPOS-LHC based $X_{\max}$ data with a $+700\mathrm{g/cm}^2$ shift in B.



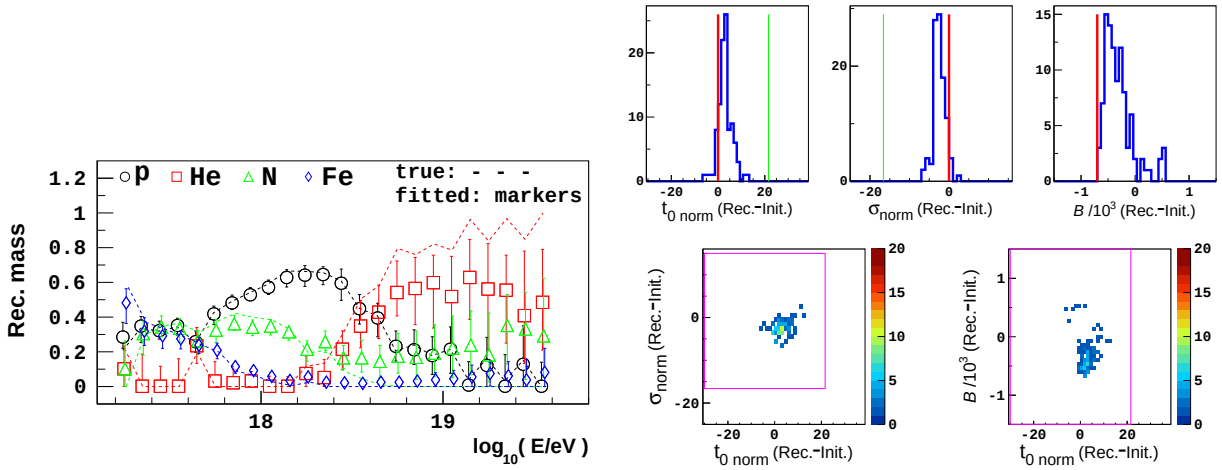
(a) (b)

Figure 5.43: Fitting the QGSJetII-04 parameterisation to EPOS-LHC based $X_{\max}$ data with a $+700\mathrm{g/cm}^2$ shift in B.



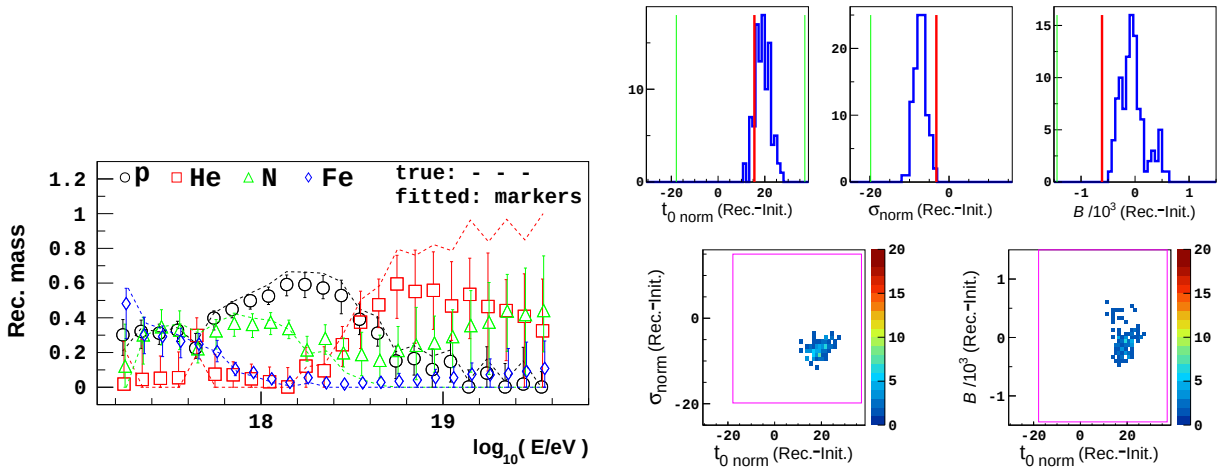
(a) (b)

Figure 5.44: Fitting the Sibyll2.3 parameterisation to EPOS-LHC based $X_{\max}$ data with a $+700\mathrm{g/cm}^2$ shift in B.



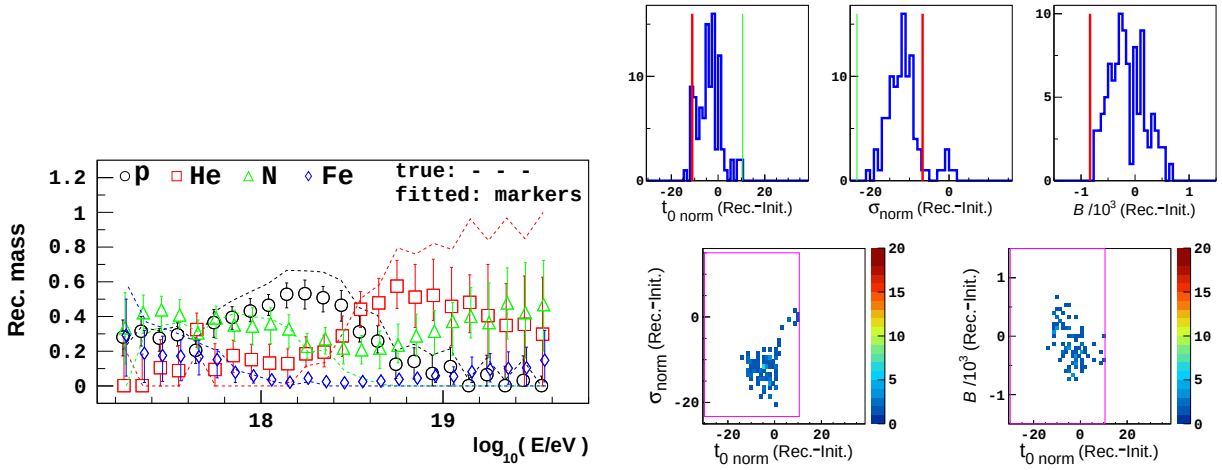
(a) (b)

Figure 5.45: Fitting the EPOS-LHC parameterisation to EPOS-LHC based $X_{\max}$ data with a $-700\mathrm{g/cm}^2$ shift in B.



(a) (b)

Figure 5.46: Fitting the QGSJetII-04 parameterisation to EPOS-LHC based $X_{\max}$ data with a $-700\mathrm{g/cm}^2$ shift in B.

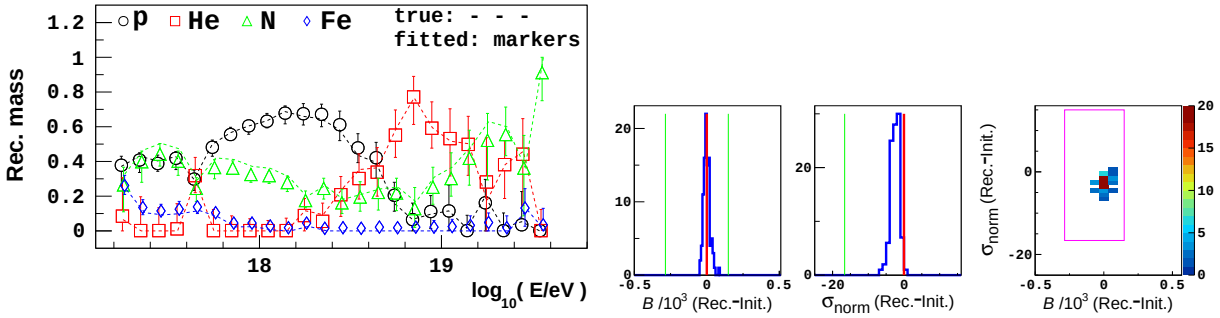


(a) Solution 1. (b)

Figure 5.47: Fitting the Sibyll2.3 parameterisation to EPOS-LHC based $X_{\max}$ data with a $-700\mathrm{g/cm}^2$ shift in B.

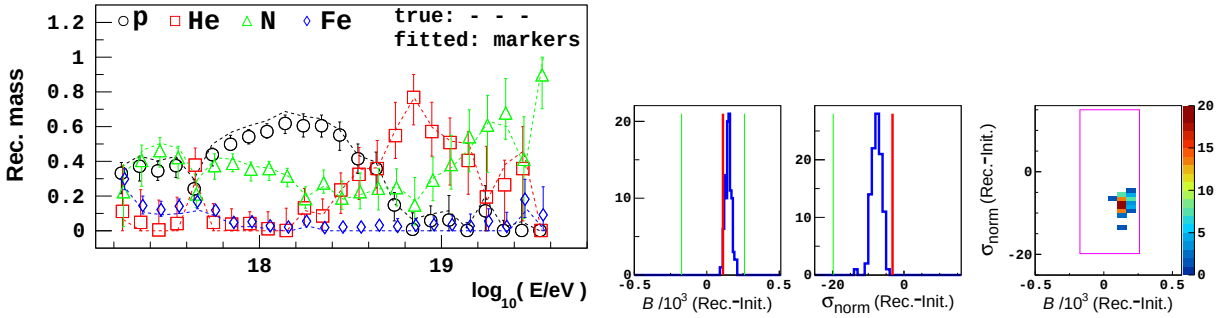
5.2.1.5 Fitting B, σ_{norm} and the mass fractions

Figures 5.48 to 5.56 displays the mass composition and coefficient results from fitting the mass fractions, B and σ_{norm} of either the EPOS-LHC, QGSJetII-04 or Sibyll2.3 parameterisations to 100 $X_{\max}$ data sets generated from the parameterisations which resulted when only the mass fractions of a parameterisation were fitted to the combined FD and HEAT Auger $X_{\max}$ data. The parameterisations used in this section apply a t_0 parameterisation constrained at 10^{14} eV. For the EPOS-LHC and QGSJetII-04 parameterisations, the t_0 parameterisation normalised at 10^{14} eV is slightly different to the one normalised at $10^{18.24}$ eV. As these parameterisations are initially fitted to the FD and HEAT Auger data to obtain the function used to generate mock data, the true mass composition of the EPOS-LHC and QGSJetII-04 mock data in this section will be slightly different to that of the previous sections.



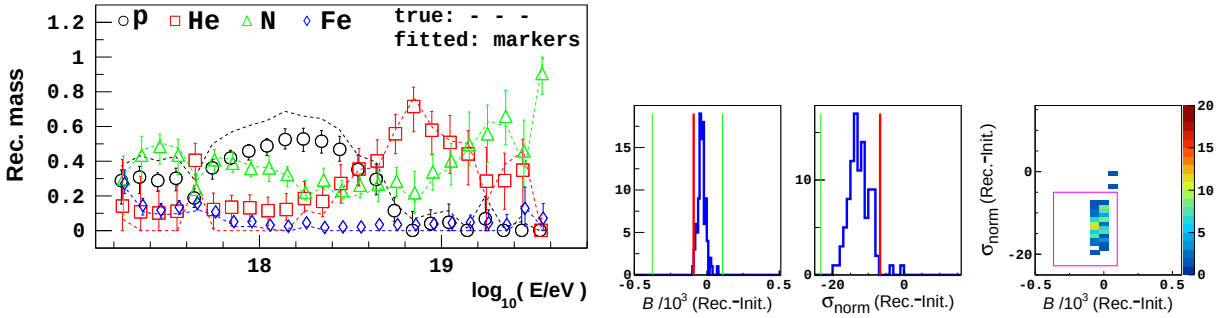
(a) (b)

Figure 5.48: Fits of EPOS-LHC based $X_{\max}$ data with the EPOS-LHC parameterisation.



(a) (b)

Figure 5.49: Fits of EPOS-LHC based $X_{\max}$ data with the QGSJetII-04 parameterisation.



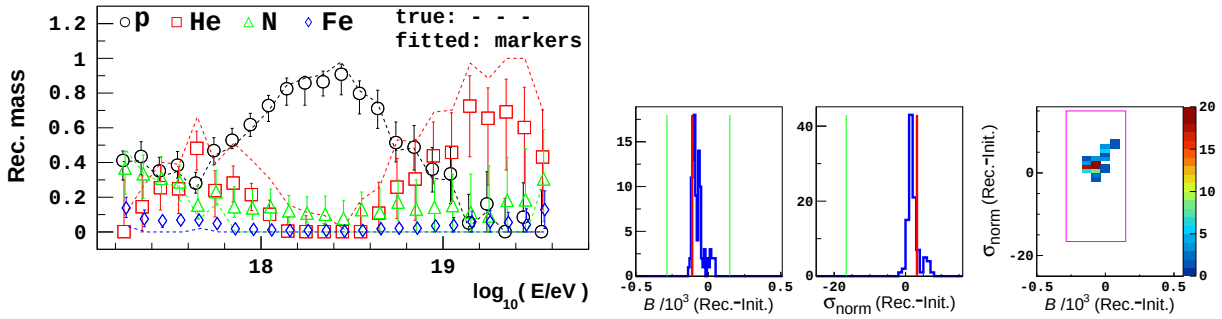
(a) Solution 1. (b)

Figure 5.50: Fits of EPOS-LHC based $X_{\max}$ data with the Sibyll2.3 parameterisation.

For these fits, the limits of B are usually 2290-2724 g/cm^2. However, when the Sibyll2.3 parameterisation is being fitted to data, or Sibyll2.3 based data is fitted, the upper limit of B has been increased to 2774 g/cm^2 in this section, to prevent fewer fits becoming stuck at the upper limit.

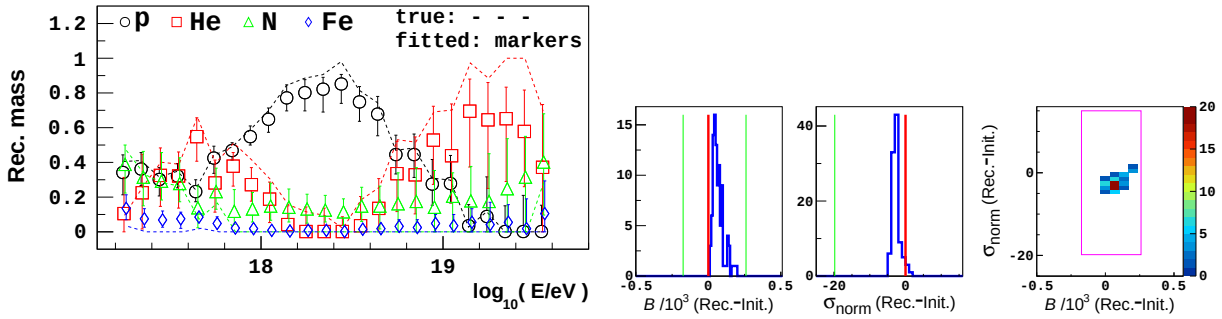
Comparing the fits in this section to those where only $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions were fitted, fitting B and σ_{norm} results in similar systematic offsets in the reconstructed mass fractions and fitted σ_{norm}. The statistical error is marginally larger in the fits of B and σ_{norm}.

We see in the fits of Sibyll2.3 data that the fitted value of B often becomes stuck at the upper limit. This could be due to the fit becoming stuck in a local minimum, or the fit attempting to reach a deeper minimum above the upper limit.



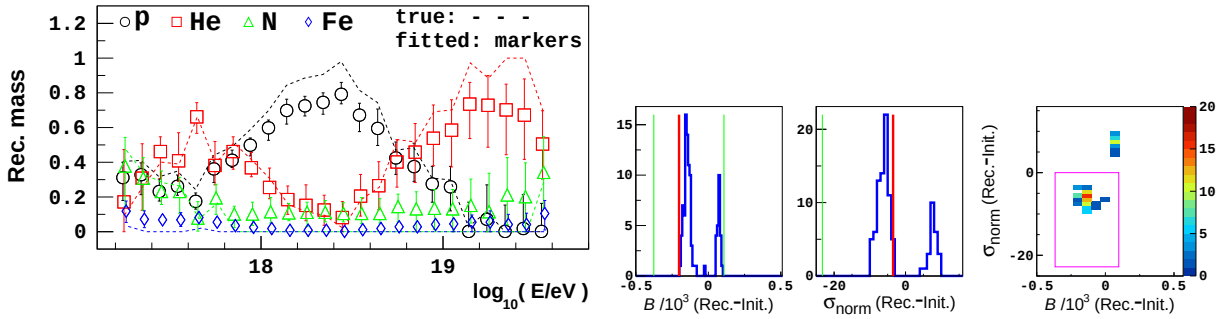
(a) (b)

Figure 5.51: Fits of QGSJetII-04 based $X_{\max}$ data with the EPOS-LHC parameterisation.



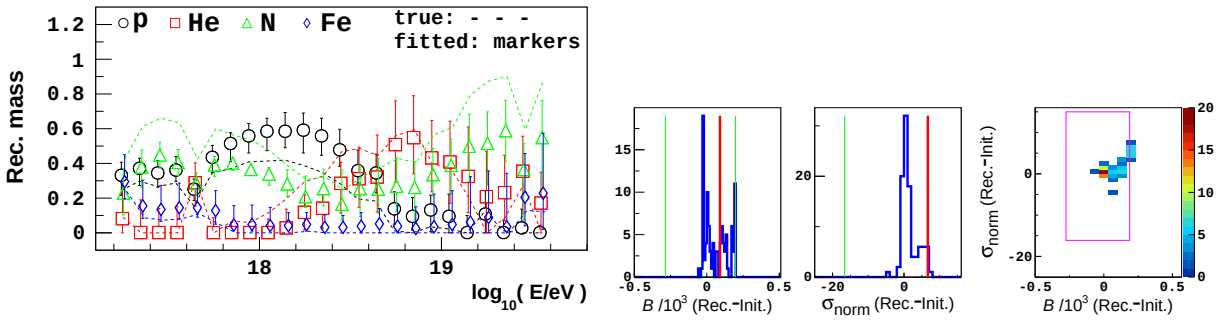
(a) (b)

Figure 5.52: Fits of QGSJetII-04 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



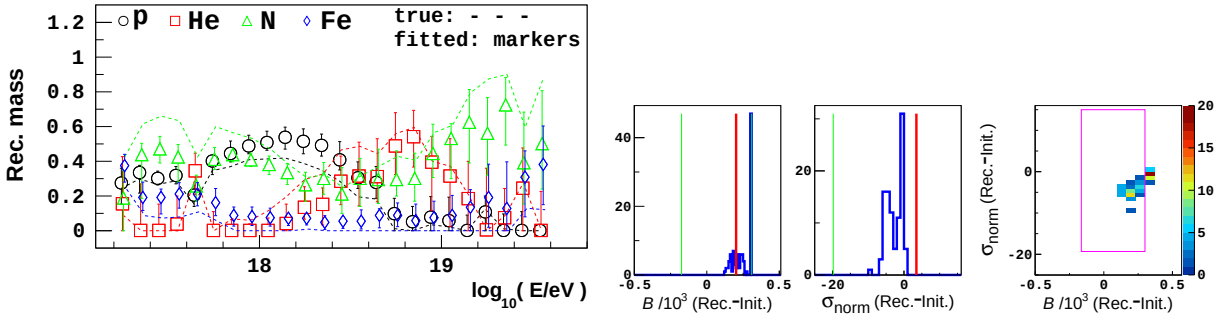
(a) Solution 1. (b)

Figure 5.53: Fits of QGSJetII-04 based $X_{\max}$ data with the Sibyll2.3 parameterisation.



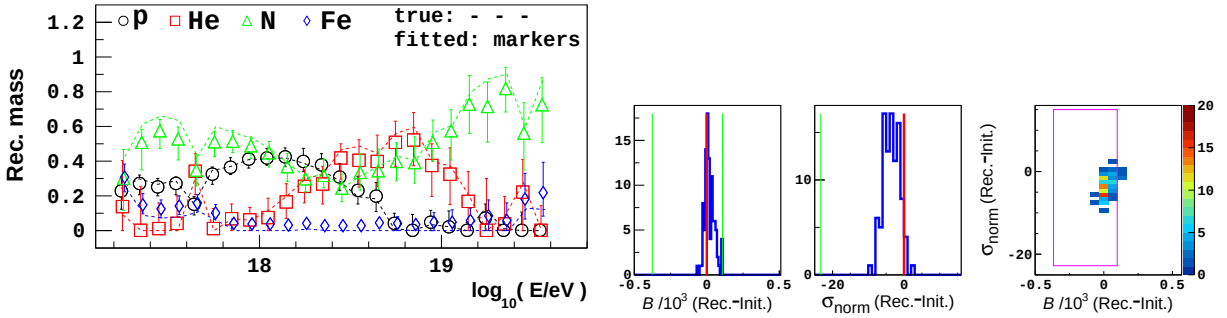
(a) Trimmed. (b)

Figure 5.54: Fits of Sibyll2.3 based $X_{\max}$ data with the EPOS-LHC parameterisation.



(a) Trimmed. (b)

Figure 5.55: Fits of Sibyll2.3 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



(a) Trimmed. (b)

Figure 5.56: Fits of Sibyll2.3 based $X_{\max}$ data with the Sibyll2.3 parameterisation.

Fitting mock data with a value of B significantly different to the initial predictions of the $X_{\max}$ parameterisations

In the following figures, we apply the B and σ_{norm} fit to mock data that was generated using a value of B that is shifted from the initial prediction of our EPOS-LHC parameterisation. The value of B in this modified EPOS-LHC parameterisation was shifted by $+300\text{g/cm}^2$ (B is now $\sim 2800\text{g/cm}^2$ instead of the standard value of $\sim 2500\text{g/cm}^2$, which corresponds to a change in the t_0 elongation between 10^{18} eV and 10^{19} eV of $\sim 7\,\text{g/cm}^2$), before the mass fractions of the EPOS-LHC parameterisation were fitted to the FD and HEAT Auger data. The fitted parameterisation was then used to generate mock data sets as before.

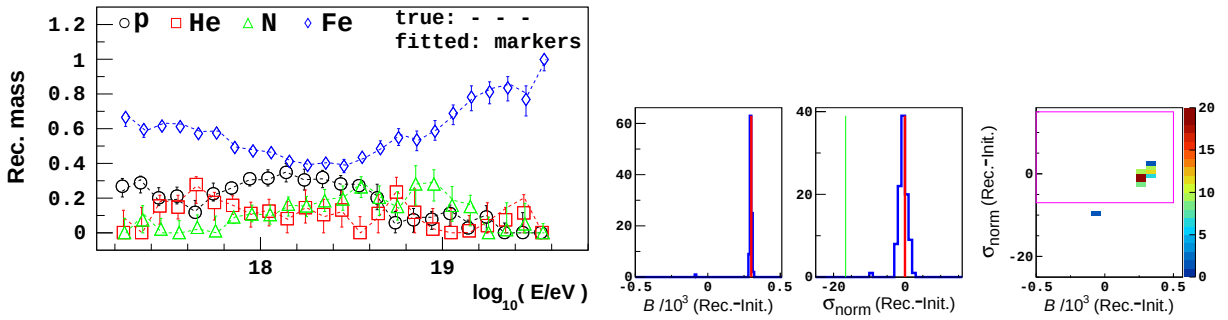
In fitting B, σ_{norm} and the mass fractions of our three parameterisations to this modified EPOS-LHC based mock data, the minimum and maximum limits of B were increased to 1800g/cm^2 and 3200g/cm^2 respectively. Setting reasonable limits reduces the time required to perform a fit, prevents the fit from becoming stuck in a local minimum, and prevents the fit from reaching an un-physical result. Therefore, B is normally limited to 2290-2724 g/cm^2 when fitting Auger data, as parameter space scans are used to ensure the region of the true minimum is found (see Chapter 6).

Figures 5.57 to 5.59 displays the results of the modified mock data fits. We see that the fitted values of B and σ_{norm} are shifted appropriately to match the data. The mass reconstruction is very accurate as the data set is dominated by iron, with a sufficient mix of lighter masses to constrain the fit.

In Figures 5.60 to 5.62, B and σ_{norm} are fitted to mock data based on a different modified EPOS-LHC parameterisation, in this case with a value of B that was shifted by -150g/cm^2 from the standard EPOS-LHC prediction (B is now $\sim 2350\text{g/cm}^2$ instead of the standard value of $\sim 2500\text{g/cm}^2$, which corresponds to a change in the t_0 elongation between 10^{18} eV and 10^{19} eV of $\sim -3.5\,\text{g/cm}^2$). The fitted shape coefficients do not align with the true values, consequently the mass reconstructions of these data sets are not as accurate. The data is dominated by protons and helium, which is much more difficult to fit than the iron dominated data set fitted earlier.

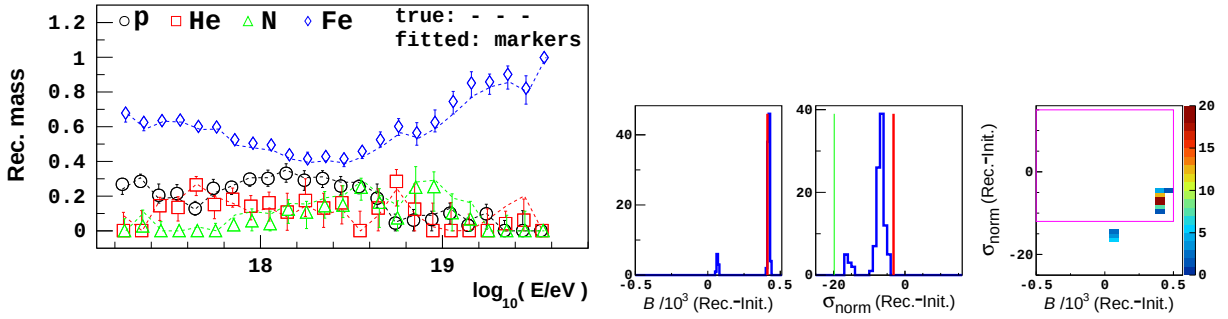
Conclusions

The B, σ_{norm} and mass fraction fits applied to mock Auger data sets indicate that this fit, applied to Auger data, will be able to reconstruct the mass fractions accurately, provided the fixed shape coefficients and the shape parameter functional forms describe the data reasonably correctly. If the initial $X_{\max}$ parameterisation prediction of B was incorrect to a significant degree, it is possible that a fit of B and σ_{norm} would be able to fit the correct values for these coefficients, and therefore accurately reconstruct the mass composition. We apply this fit to Auger data in Section 8.1.5.



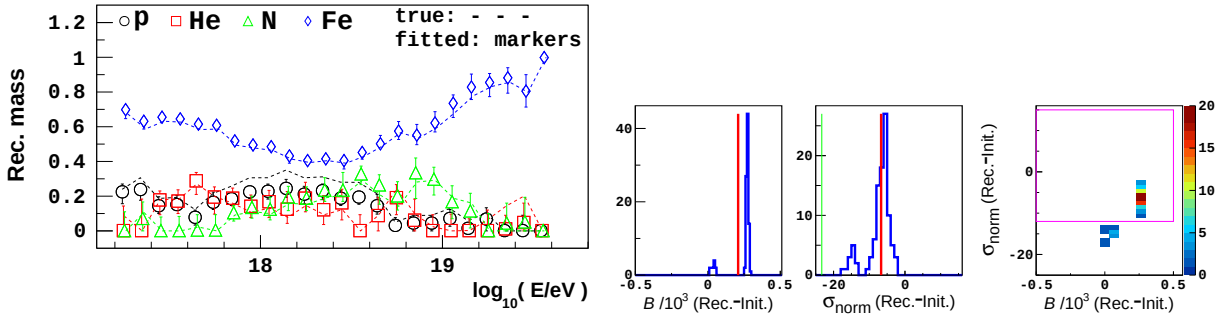
(a) Solution 1. (b)

Figure 5.57: Fitting the EPOS-LHC parameterisation to EPOS-LHC based $X_{\max}$ data with a $+300\mathrm{g/cm}^2$ shift in B.



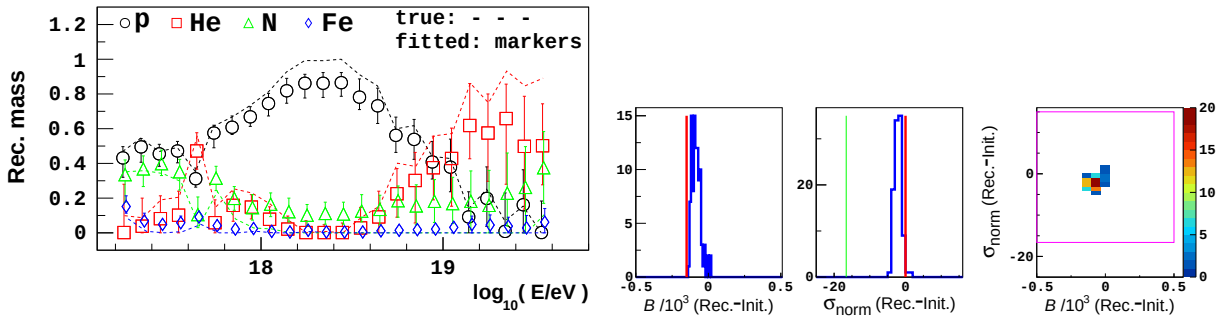
(a) Solution 1. (b)

Figure 5.58: Fitting the QGSJetII-04 parameterisation to EPOS-LHC based $X_{\max}$ data with a $+300\mathrm{g/cm}^2$ shift in B.



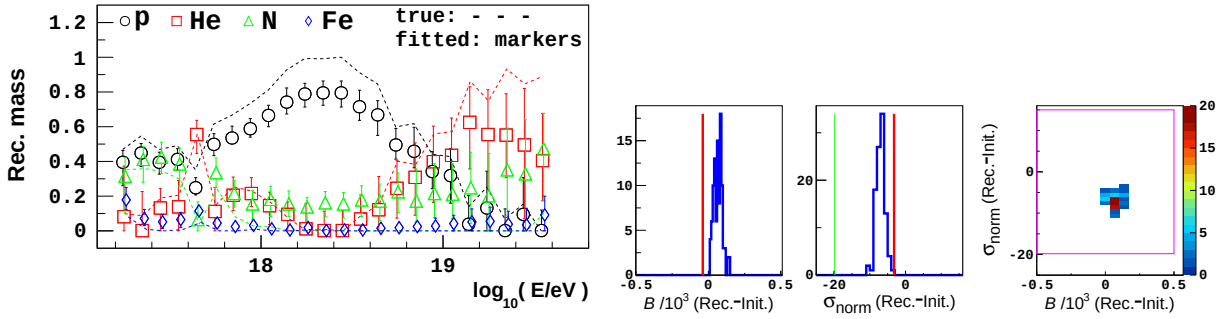
(a) Solution 1. (b)

Figure 5.59: Fitting the Sibyll2.3 parameterisation to EPOS-LHC based $X_{\max}$ data with a $+300\mathrm{g/cm}^2$ shift in B.



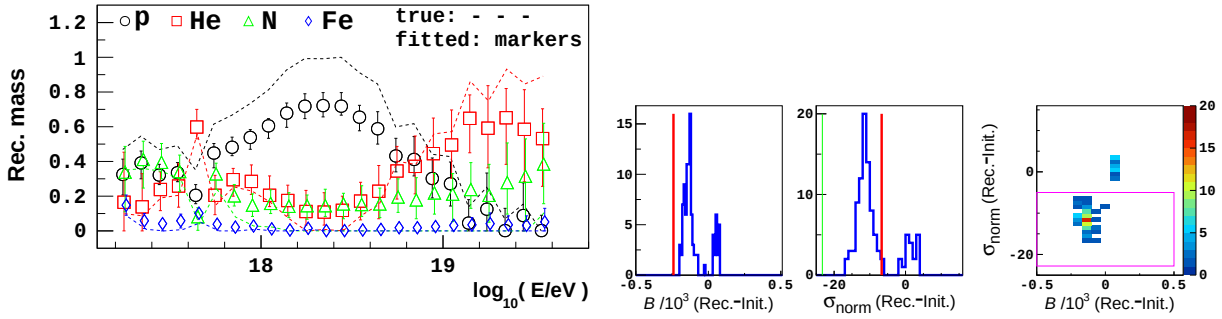
(a) (b)

Figure 5.60: Fitting the EPOS-LHC parameterisation to EPOS-LHC based $X_{\max}$ data with a $-150\mathrm{g/cm}^2$ shift in B.



(a) (b)

Figure 5.61: Fitting the QGSJetII-04 parameterisation to EPOS-LHC based $X_{\max}$ data with a $-150\mathrm{g/cm}^2$ shift in B.

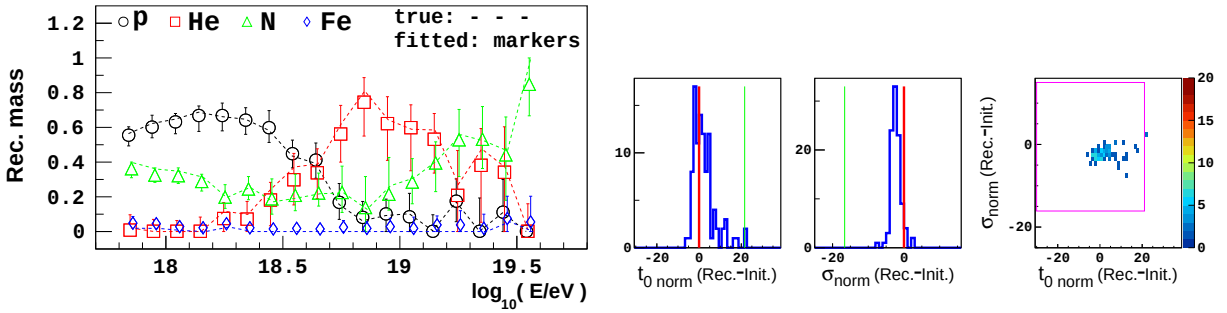


(a) Solution 1. (b)

Figure 5.62: Fitting the Sibyll2.3 parameterisation to EPOS-LHC based $X_{\max}$ data with a $-150\mathrm{g/cm}^2$ shift in B.

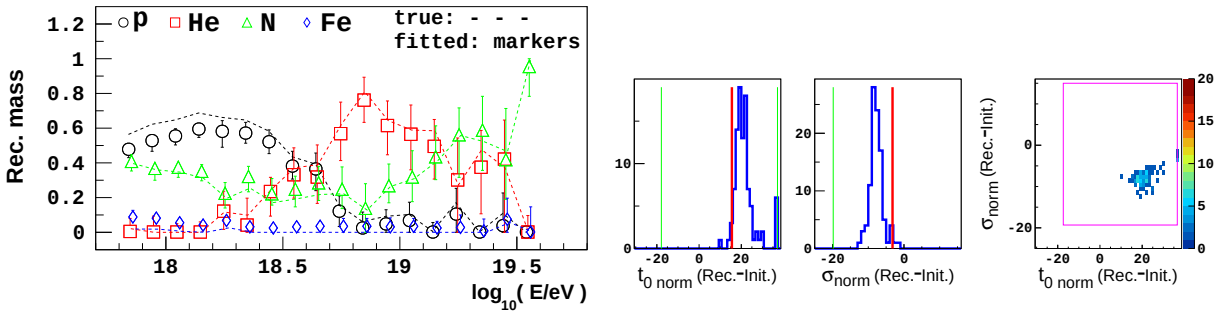
5.2.2 Using only FD $X_{\max}$ data

5.2.2.1 Fitting $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions



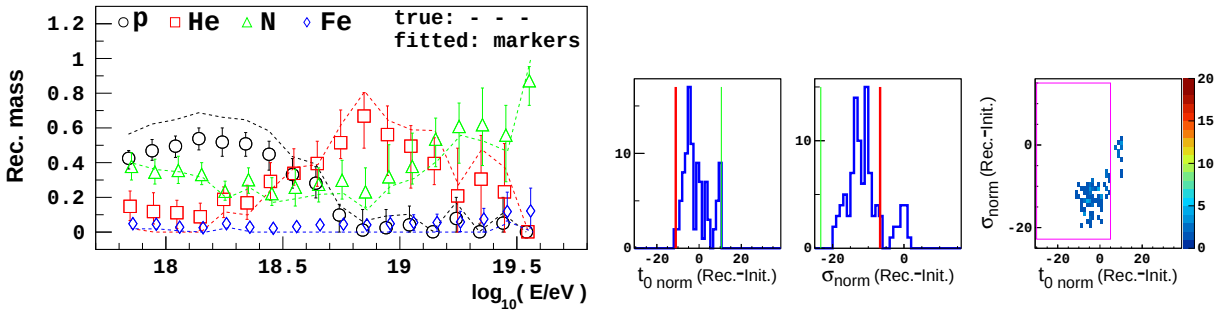
(a) Trimmed. (b)

Figure 5.63: Fits of EPOS-LHC based $X_{\max}$ data with the EPOS-LHC parameterisation.



(a) Trimmed. (b)

Figure 5.64: Fits of EPOS-LHC based $X_{\max}$ data with the QGSJetII-04 parameterisation.



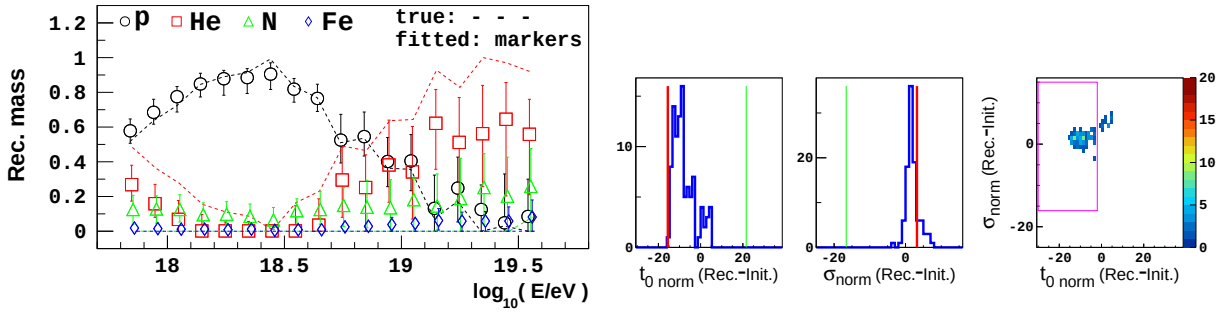
(a) Solution 1. (b)

Figure 5.65: Fits of EPOS-LHC based $X_{\max}$ data with the Sibyll2.3 parameterisation.

Figures 5.63 to 5.71 displays the mass composition and coefficient results from fitting the mass fractions, $t_{0_{\text{norm}}}$, σ_{norm} of either the EPOS-LHC, QGSJetII-04 or Sibyll2.3 parameterisations to 100 $X_{\max}$ data sets generated from the parameterisations which resulted from fitting only the mass fractions to the FD Auger $X_{\max}$ data. Comparing the figures in this section to those in Section 5.2.1.2, where FD and HEAT mock data was fitted with $t_{0_{\text{norm}}}$ and σ_{norm}, we find that over the energy range of the FD data, the absolute offsets in the median reconstructed mass fractions from the true mass are slightly larger when fitting only the FD data.

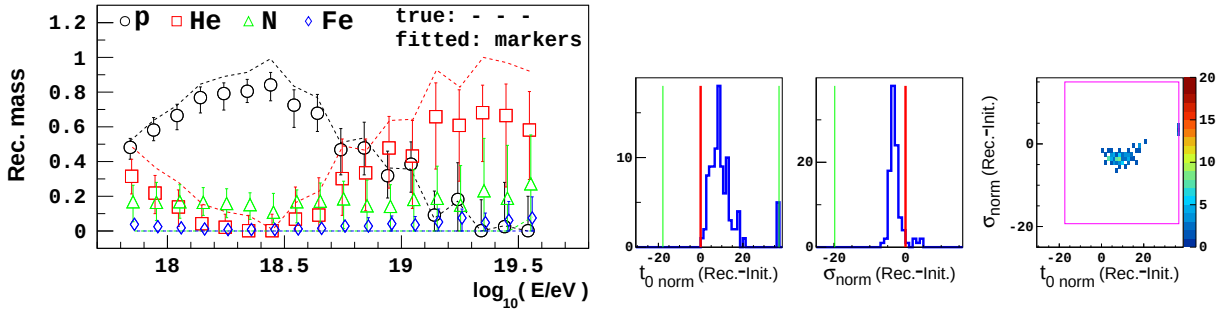
The FD and HEAT data consisted of 42662 events, while the FD data consists of 25884 events. The FD data does not constrain the fit as strongly as the combined FD and HEAT data. This

is expected due to the reduced statistics and mass dispersion of this data, resulting in greater degeneracy between the fitted fractions and coefficients, and consequently multiple solutions exist that well describe the data. Compared to the combined FD and HEAT mock data fits, the fits of only the FD mock data become stuck at the upper limit of $t_{0_{\text{norm}}}$ with more frequency, particularly for the fit of the Sibyll2.3 based data.



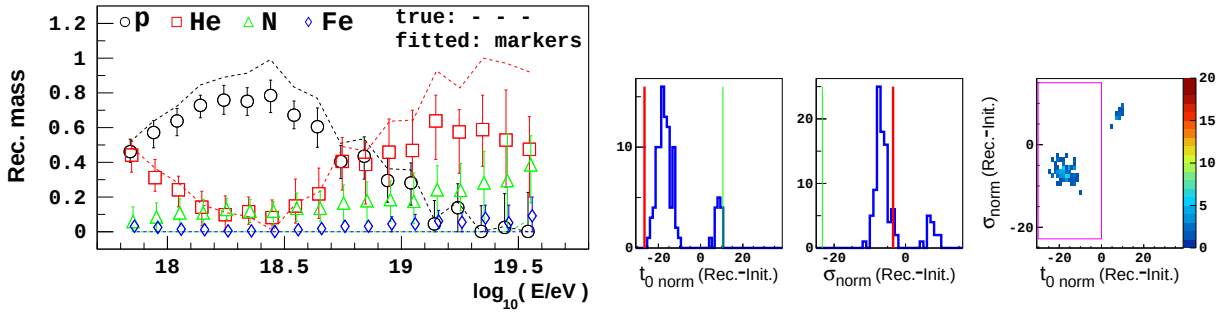
(a) Solution 1. (b)

Figure 5.66: Fits of QGSJetII-04 based $X_{\max}$ data with the EPOS-LHC parameterisation.



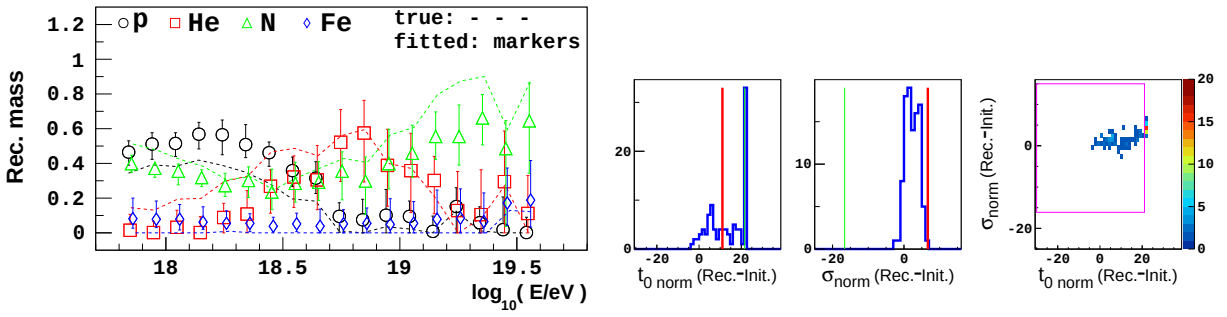
(a) Trimmed. (b)

Figure 5.67: Fits of QGSJetII-04 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



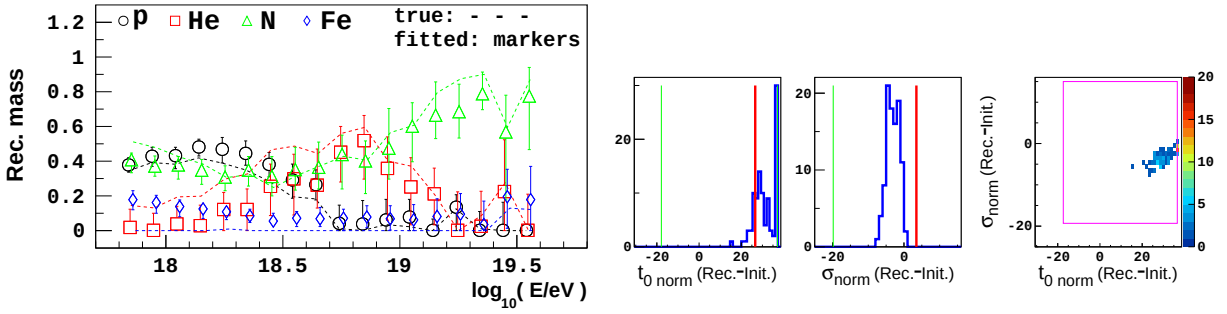
(a) Solution 1. (b)

Figure 5.68: Fits of QGSJetII-04 based $X_{\max}$ data with the Sibyll2.3 parameterisation.



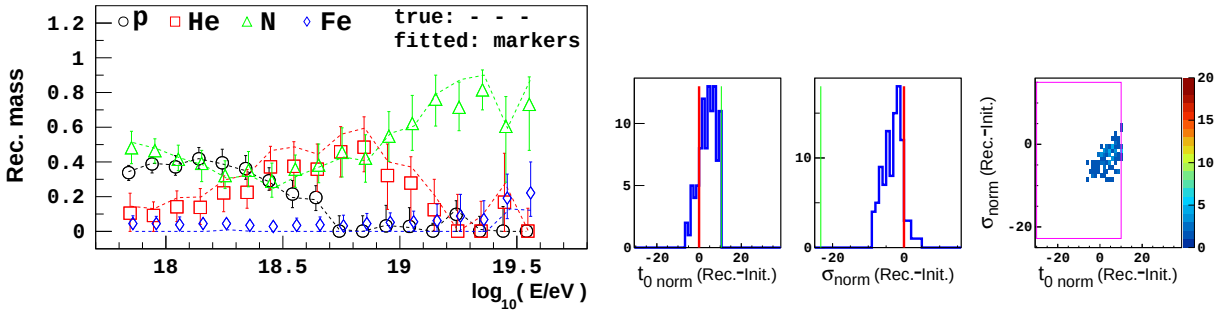
(a) Trimmed. (b)

Figure 5.69: Fits of Sibyll2.3 based $X_{\max}$ data with the EPOS-LHC parameterisation.



(a) Trimmed. (b)

Figure 5.70: Fits of Sibyll2.3 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



(a) Trimmed. (b)

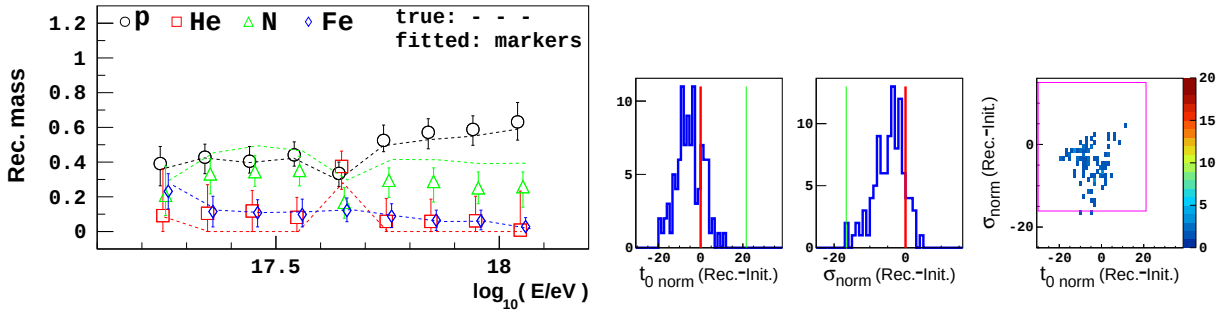
Figure 5.71: Fits of Sibyll2.3 based $X_{\max}$ data with the Sibyll2.3 parameterisation.

Conclusions

The $t_{0_{\text{norm}}}$, σ_{norm} and mass fraction fit of the Auger FD data is able to accurately reconstruct the mass composition if the correct region of $t_{0_{\text{norm}}}$ and σ_{norm} is considered. Applying the $t_{0_{\text{norm}}}$ and σ_{norm} fit to only the FD data, and comparing the shape coefficient results to those obtained from the FD and HEAT fit, can indicate if the shape coefficients which are fixed to the data are correctly describing the measured data. For example, if the predicted $\langle X_{\max} \rangle$ elongation rate was incorrect, the value of $t_{0_{\text{norm}}}$ fitted to only the FD data may differ significantly from the $t_{0_{\text{norm}}}$ value fitted to the FD and HEAT data which spans a larger energy range. If the fixed shape parameters are correctly describing the data over the full energy range, and the FD data consists of a mass dispersion that can constrain the fitted parameters, the values of $t_{0_{\text{norm}}}$ and σ_{norm} fitted to only the FD data, and to the combined FD and HEAT data, should be consistent.

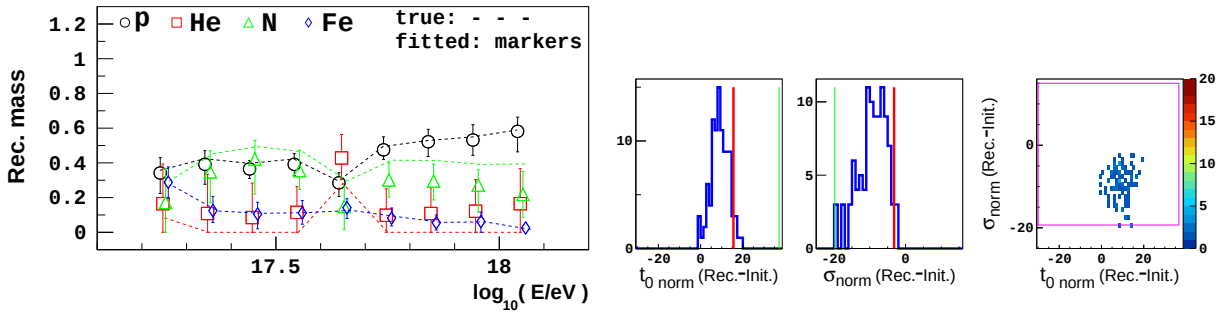
5.2.3 Using only HEAT $X_{\max}$ data

5.2.3.1 Fitting $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions



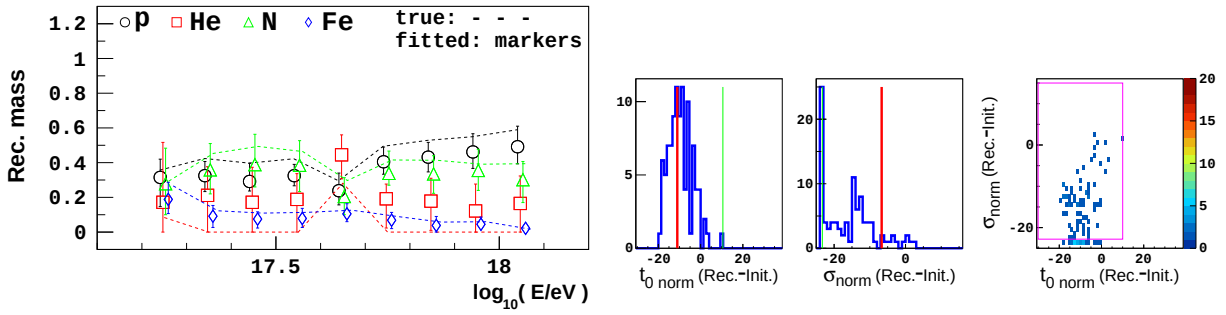
(a) Trimmed. (b)

Figure 5.72: Fits of EPOS-LHC based $X_{\max}$ data with the EPOS-LHC parameterisation.



(a) Trimmed. (b)

Figure 5.73: Fits of EPOS-LHC based $X_{\max}$ data with the QGSJetII-04 parameterisation.

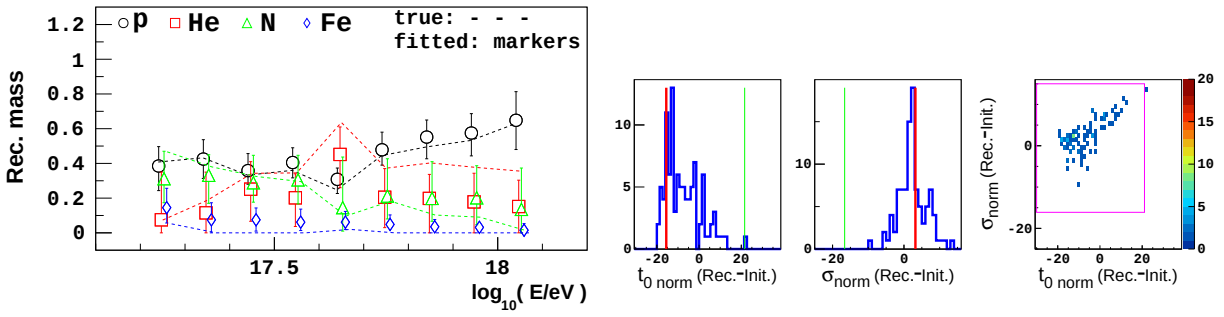


(a) Trimmed. (b)

Figure 5.74: Fits of EPOS-LHC based $X_{\max}$ data with the Sibyll2.3 parameterisation.

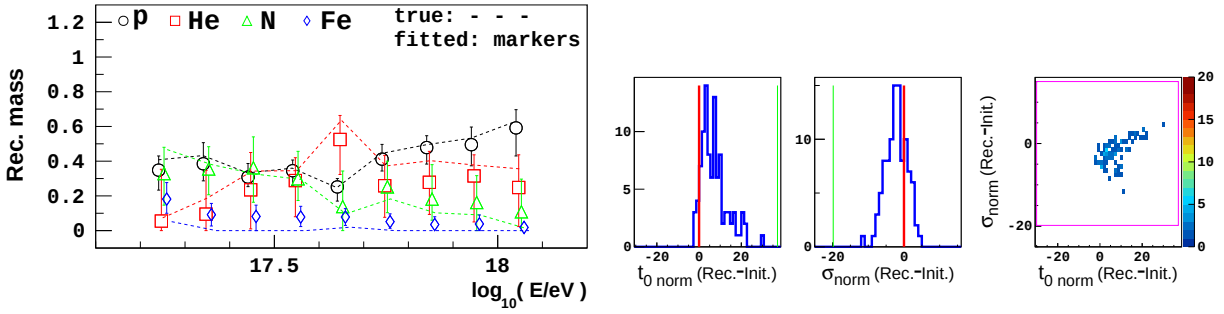
Figures 5.72 to 5.80 displays the results from fitting the mass fractions, $t_{0_{\text{norm}}}$, σ_{norm} of either the EPOS-LHC, QGSJetII-04 or Sibyll2.3 parameterisations to 100 $X_{\max}$ data sets generated from the parameterisations which resulted when only the mass fractions were fitted to the HEAT Auger $X_{\max}$ data.

Comparing the figures in this section to those in Section 5.2.1.2, where FD and HEAT mock data was fitted, we see that over the energy range of the HEAT data, the statistical error and the absolute offsets in the median reconstructed mass fractions from the true mass are greater in the fits of only the HEAT mock data, as each mock HEAT data set contains only 16778 events and spans a much smaller energy range.



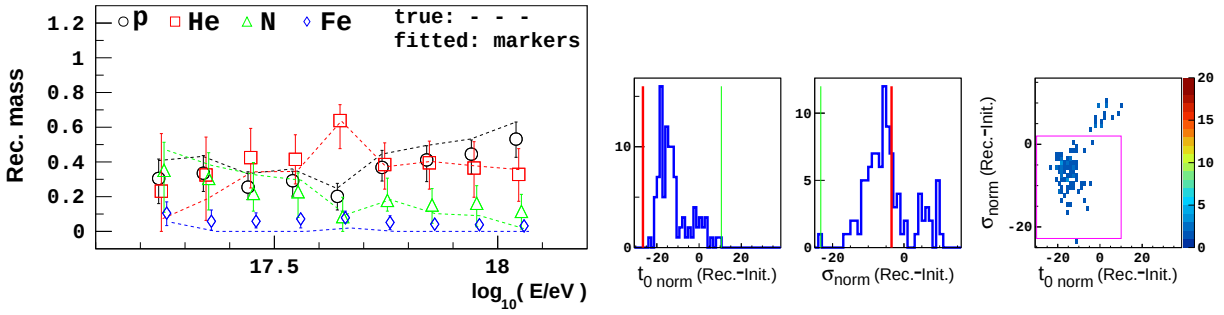
(a) Trimmed. (b)

Figure 5.75: Fits of QGSJetII-04 based $X_{\max}$ data with the EPOS-LHC parameterisation.



(a) (b)

Figure 5.76: Fits of QGSJetII-04 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



(a) Solution 1. (b)

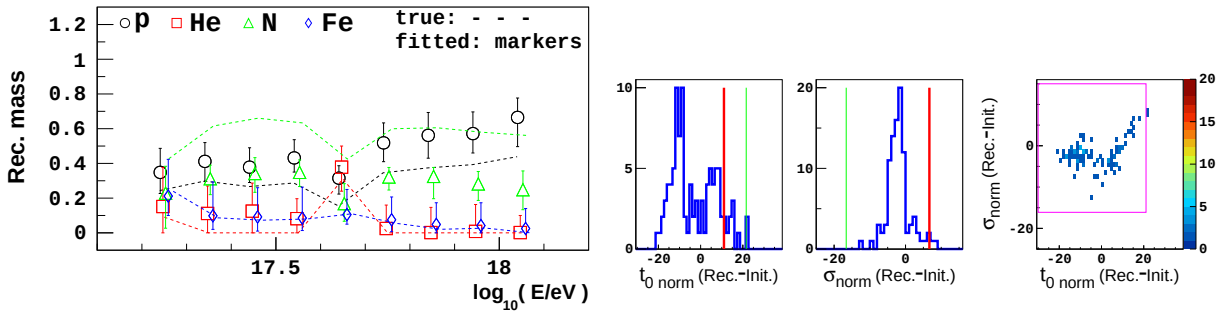
Figure 5.77: Fits of QGSJetII-04 based $X_{\max}$ data with the Sibyll2.3 parameterisation.

The fits of the EPOS-LHC mock HEAT data with the EPOS-LHC and QGSJetII-04 parameterisations results in absolute offsets in the median reconstructed mass fractions of less than 15% in most energy bins. The Sibyll2.3 fit is slightly less accurate.

The QGSJetII-04 and Sibyll2.3 fits of QGSJetII-04 mock data also result in absolute offsets of less than 15% in most energy bins. However, the EPOS-LHC fit to QGSJetII-04 data struggles, which is surprising as the fit of $t_{0_{\text{norm}}}$ and σ_{norm} significantly reduces the differences between these two parameterisations. The problem is that the EPOS-LHC fit is poorly constrained, demonstrated by the large spread in the distributions of the fitted $t_{0_{\text{norm}}}$ and σ_{norm} values.

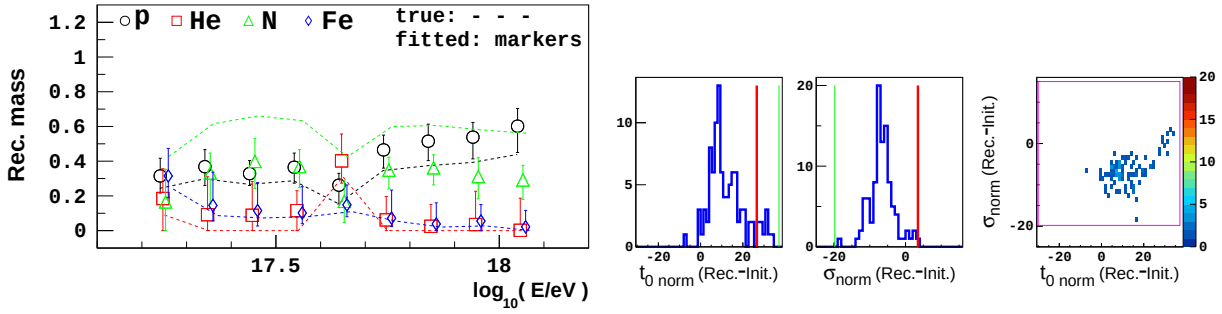
As seen before, the EPOS-LHC and QGSJetII-04 fits to Sibyll2.3 data struggle. The Sibyll2.3 fit of Sibyll2.3 data consists of absolute offsets in the median or around $15-20\%$, which is more than this same fit of the FD and HEAT data.

The coefficient distributions of this section illustrate the large spread in the fitted values of



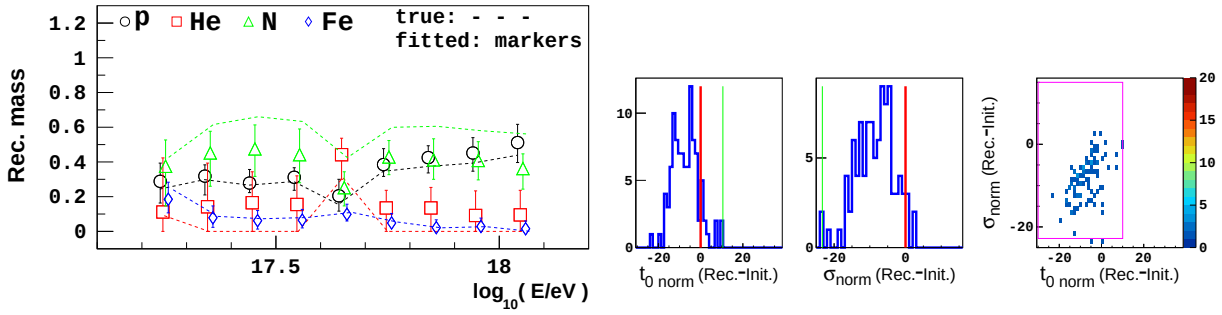
(a) Trimmed. (b)

Figure 5.78: Fits of Sibyll2.3 based $X_{\max}$ data with the EPOS-LHC parameterisation.



(a) (b)

Figure 5.79: Fits of Sibyll2.3 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



(a) Trimmed. (b)

Figure 5.80: Fits of Sibyll2.3 based $X_{\max}$ data with the Sibyll2.3 parameterisation.

$t_{0_{\text{norm}}}$ and σ_{norm}. The $t_{0_{\text{norm}}}$ and σ_{norm} fits of the combined FD and HEAT data resulted in significantly less spread. The HEAT mock data does not constrain the fitted coefficients with as much power, consequently the fitted parameters are very degenerate. The result is increased bi-modality in the fitted coefficient values, and increased systematics in the fitted mass fractions.

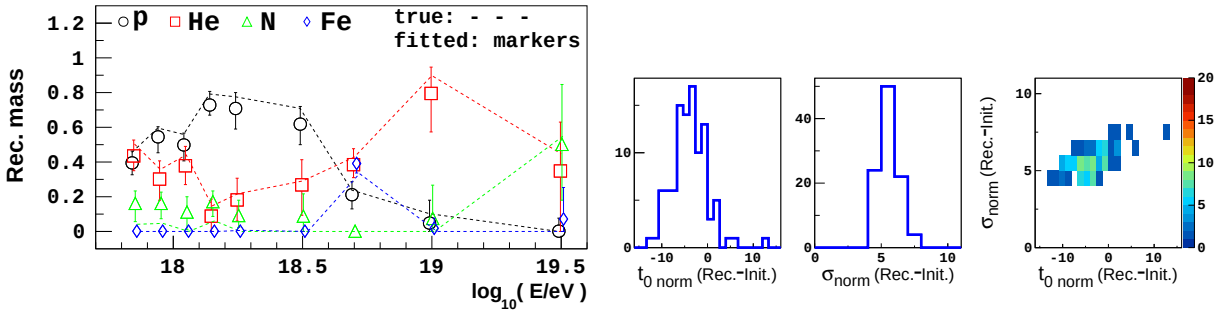
With enough mass dispersion and statistics, it is possible that the $t_{0_{\text{norm}}}$ and σ_{norm} fit of data over a narrower energy range can produce results that are more accurate across the three models, than fits of a wider energy range with more mass dispersion and statistics. Fitting $t_{0_{\text{norm}}}$ and σ_{norm} allows the t_0 and σ parameterisations to be shifted by a constant value over the energy range, but the differences in the shape parameter predictions of the models varies to some degree with energy (see Figure 4.8). Over a narrower energy range, the t_0 and σ parameterisations of different models are able to align with each other better.

Conclusions

The spread of the fitted coefficients in this section indicates that the $t_{0_{\text{norm}}}$ and σ_{norm} fits of the Auger HEAT $X_{\max}$ data will consist of significant uncertainty, which translates to the reconstructed mass composition. The mass composition reconstructions of the HEAT fits are of acceptable accuracy, but this accuracy is much worse than the fits of the combined FD and HEAT data.

Similar to the fits of the only the FD data, if the shape coefficients fixed to the data are not a reasonable description of the data, the $t_{0_{\text{norm}}}$ and σ_{norm} values fitted to the HEAT data may be considerably different to those fitted to the combined FD and HEAT data, or only the FD data. However, due to the large uncertainty in the HEAT fits, a $t_{0_{\text{norm}}}$ and σ_{norm} fit to HEAT data which is inconsistent with the values fitted to other data sets will not be conclusive.

5.3 Effect of $X_{\max}$ systematic uncertainties on the fit performance



(a) (b)

Figure 5.81: Fits of $t_{0_{\text{norm}}}$ and σ_{norm} to $X_{\max}$ data consisting of a $-10\,\text{g/cm}^2$ systematic offset in $X_{\max}$. The $X_{\max}$ data was also smeared by a Gaussian distributed random variable of $\sigma = 20\,\text{g/cm}^2$, of which only $10\,\text{g/cm}^2$ was accounted for in the initial fitted $X_{\max}$ parameterisation.

Systematics in the measured $X_{\max}$ distributions are expressed in the mean and spread of the distribution. Fitting $t_{0_{\text{norm}}}$ can compensate for systematic offsets in $X_{\max}$, while fitting σ_{norm} can compensate for systematic errors in the estimation of the detector resolution of $X_{\max}$. Figure 5.81 shows the results of fitting the mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm} of our QGSJetII-04 parameterisation to 100 CONEX v4r37 QGSJetII-04 $X_{\max}$ data sets of a mass composition similar to the QGSJetII-04 2014 Auger FD $X_{\max}$ analysis [145]. Across the whole energy range, the data was shifted by a systematic offset of -10 g/cm^2. The data was also smeared by a Gaussian distributed random variable of $\sigma = 20\,\text{g/cm}^2$, of which only $10\,\text{g/cm}^2$ was accounted for in the resolution of the applied $X_{\max}$ parameterisation. The CONEX v4r37 $X_{\max}$ distribution data does not consist of all the energy bins contained in the Auger data (the statistics of 9 energy bins are not present), consequently the mass dispersion and statistics of these missing energy bins are not present in the data. Therefore, some iron was added in the third to last energy bin, allowing the data to adequately constrain the $t_{0_{\text{norm}}}$ and σ_{norm} fit.

The fitted $t_{0_{\text{norm}}}$ values are shifted by approximately -5 g/cm^2 on average, to compensate mainly for the -10 g/cm^2 $X_{\max}$ systematic (Figure 5.81b). As t_0 changes by the same amount for each primary when $t_{0_{\text{norm}}}$ is fitted, and the $X_{\max}$ systematic was applied consistently to all data, the $t_{0_{\text{norm}}}$ fit is capable of completely accounting for the $X_{\max}$ systematic offset. However, σ_{norm} for each primary is changed by different absolute amounts when fitting this coefficient, but all of the data is smeared (all masses are consistently smeared), consequently the correct σ_{norm} cannot be fitted for each primary, which partially affects the fit of $t_{0_{\text{norm}}}$. Despite the fit

of σ_{norm} being unable to thoroughly account for the $10\,\text{g/cm}^2$ systematic in the resolution, the absolute offsets in the median reconstructed mass fractions are less than 15% from the true mass in most energy bins, due to a combined shift of $t_{0_{\text{norm}}}$ and σ_{norm} in the appropriate directions. Reasonable detector resolution systematics and systematic offsets in $X_{\max}$ will not significantly affect the accuracy of the reconstructed mass composition.

If the data was not smeared by a Guassian random variable, and only shifted by a constant $X_{\max}$ offset, the $t_{0_{\text{norm}}}$ and σ_{norm} fit of this shifted data would result in a change in the fitted $t_{0_{\text{norm}}}$ (compared to the $t_{0_{\text{norm}}}$ fitted to the un-shifted data) which is equal to the value of the $X_{\max}$ offset. Shifting the $X_{\max}$ data by a constant value has essentially the same effect on the fit as shifting the parameterisations by a constant value, with a very minuscule difference arising if the detector acceptance is not equal to 1 for all $X_{\max}$ and was not shifted by the same offset to account for the applied $X_{\max}$ offset.

5.4 Conclusions

The main differences between the $X_{\max}$ distribution predictions of EPOS-LHC, QGSJetII-04 and Sibyll2.3 are the normalisation values of t_0 and σ for each primary. By fitting parameters which alter the normalisations of t_0 and σ for each primary in an appropriate way, a mass composition can be fitted that is less dependent on the hadronic model assumed. Provided that the $X_{\max}$ distributions consist of adequate statistics and a sufficiently changing composition over the energy range, the mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm} of an $X_{\max}$ distribution parameterisation can be fitted to the data to obtain a mass composition reconstruction that is representative of the true mass distribution of the data, and one that is consistent regardless of the hadronic interaction model assumed.

Mock $X_{\max}$ distributions similar to the measured Auger FD and HEAT $X_{\max}$ distributions have been fitted. The $t_{0_{\text{norm}}}$, σ_{norm} and mass fraction fits of the mock Auger data successfully reconstruct the true mass composition trend of the data and approach the true shape parameter coefficients, indicating that these fits applied to the measured Auger distributions will produce results that can be relied upon. Furthermore, fitting $t_{0_{\text{norm}}}$ and σ_{norm} can account for reasonable detector resolution systematics and systematic offsets in $X_{\max}$, reducing the systematic uncertainty in the mass composition reconstruction.

Additional shape parameter coefficients can be added to the $t_{0_{\text{norm}}}$, σ_{norm} and mass fraction fit of the Auger FD and HEAT data. Fitting $t_{0_{\text{norm}}}$, σ_{norm} and λ_{norm} to mock Auger data sets did not add any significant systematics to the reconstruction. Fitting $t_{0_{\text{norm}}}$, B and σ_{norm} added additional systematics to the reconstructed mass composition at the highest energies, but this three-coefficient fit is still worth applying to the measured Auger distributions. A fit of B and σ_{norm}, with t_0 constrained at $10^{14}\,\text{eV}$, can also be applied to the Auger data with reasonable confidence.

Fitting $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions to only mock FD data, or only mock HEAT data, resulted in reconstructions with acceptable systematics. In the case of the FD fits, the systematics in the reconstructed mass and shape parameter coefficients were marginally worse than the fits of the FD and HEAT mock Auger data, whereas the systematics of the HEAT fit were considerably worse, due to the reduced statistics, energy range and mass dispersion of this data.

The fits of EPOS-LHC or QGSJetII-04 based data with the Sibyll2.3 parameterisation, and the fits of Sibyll2.3 based data with the EPOS-LHC or QGSJetII-04 parameterisations, demonstrated that the Sibyll2.3 $X_{\max}$ parameterisation compared to the EPOS-LHC and QGSJetII-04 parameterisations consists of differences which a fit of $t_{0_{\text{norm}}}$ and σ_{norm} (and any other shape parameter coefficients) can not reconcile. The main cause of this discrepancy is that the Sibyll2.3 proton shape parameter predictions relative to the Sibyll2.3 predictions for other primaries are not consistent with the EPOS-LHC and QGSJetII-04 predictions. The fit of Sibyll2.3 to the Auger data will likely result in a reconstructed mass composition that is not as consistent with

the EPOS-LHC and QGSJetII-04 results.

The fits performed to mock Auger data in this section resulted in mass composition reconstructions and shape parameter coefficient fits of acceptable systematics, therefore the fit approaches implemented here are applied to the Auger distributions in Chapter 8. Due to the statistics, at energies above 10^{19} eV, the systematics in the reconstructed mass increased considerably when a parameterisation was fitted to mock data based on another parameterisation. However, the three parameterisations fitted to a particular data set reconstructed a mass composition trend that is consistent. If the fits of each parameterisation to the Auger data predict a mass composition above 10^{19} eV that is broadly consistent, we can be confident these fits are accurately representing the general mass composition trend at these very high energies, given the coefficients that are fixed when fitting the data. If the predictions of the air shower properties, according to the fixed coefficients, do not adequately describe nature, the fitted mass composition will not represent the true mass composition of nature.

Chapter 6

Scans of the shape coefficient parameter space using the Auger X_{max} data

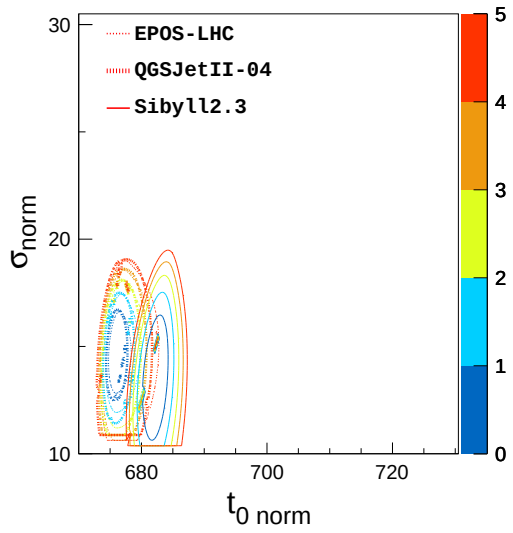
This section presents the parameter space scans using the Auger X_{max} data sets [148]. These scans involve fitting only the mass fractions to the data, while scanning over a range of shape parameter coefficient values. The minimised Poisson log likelihood of each mass fraction fit for some value of the shape parameter coefficients is shown, illustrating the coefficient values which would correspond to the true minimum of the fit when these coefficients and the mass fractions are fitted to the data. The purpose of these scans is to ensure that the results presented in Chapter 8, where the shape parameter coefficients and mass fractions are fitted to the Auger data, correspond to the true minimum of the fit, and not a local minimum. The parameter space scans allow us to see not only the solution that corresponds to the deepest minimum, but additional viable solutions that are not as deep, which given a small change in the fixed coefficients of the X_{max} parameterisation could become the deepest minimum. It is possible the false solution better fits the observed data. Additionally, nature could inherently result in two plausible solutions given the assumptions of an X_{max} parameterisation.

In the following figures, the scanned shape coefficient values for proton are shown. When values of the shape coefficients are mentioned in the text, the proton value is being referred to unless stated otherwise. The coefficient values of the heavier nuclei change (relative to protons) in the way the shape coefficient would be fitted, outlined in Section 4.4. Therefore, if the three parameterisations are fixed to the data with the same $t_{0_{\mathrm{norm}}}$ proton value, the change in the $t_{0_{\mathrm{norm}}}$ values of the heavier nuclei will vary between the parameterisations, depending on the initial predicted proton $t_{0_{\mathrm{norm}}}$ value of the model.

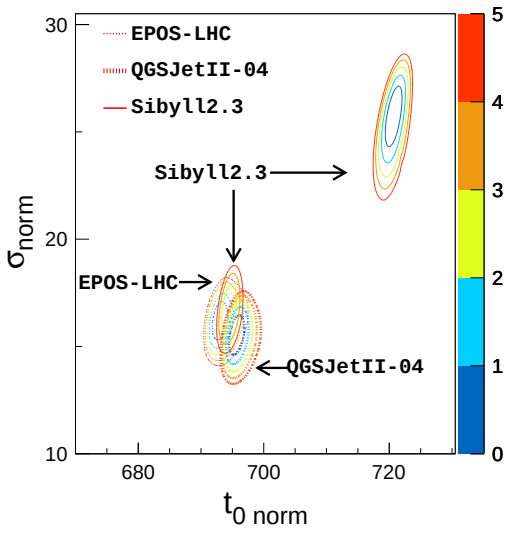
6.1 $t_{0_{\mathrm{norm}}}$ and σ_{norm} parameter space scans

Figure 6.1 shows the minimised Poisson log likelihood space from the mass fraction fits of the parameterisations to the Auger X_{max} data, where $t_{0_{\mathrm{norm}}}$ and σ_{norm} have been fixed to some particular value (indicated by the x and y axes). The z-axis shows the difference between the minimised probability for some value of $t_{0_{\mathrm{norm}}}$ and σ_{norm}, and the deepest minimised probability obtained from the $t_{0_{\mathrm{norm}}}$ and σ_{norm} values which best fitted the data for a particular model. A difference of 1 in the minimised Poisson log likelihood corresponds to 1σ. $t_{0_{\mathrm{norm}}}$ and σ_{norm} were scanned in $0.25\,\mathrm{g/cm^2}$ intervals.

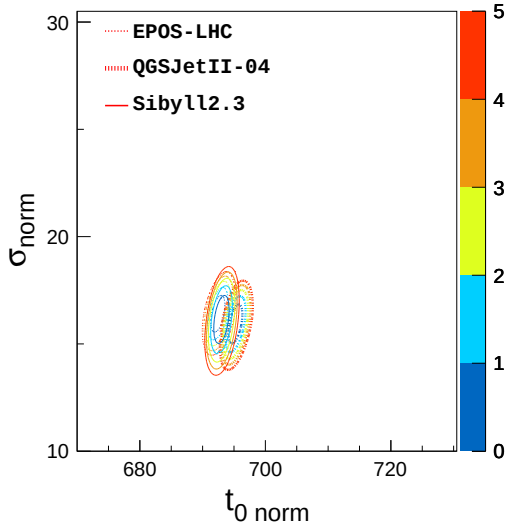
In Figure 6.1b, the absolute minimum of the EPOS-LHC and QGSJetII-04 fits to the Auger FD data correspond to a similar value of $t_{0_{\mathrm{norm}}}$ and σ_{norm}, whereas the absolute minimum of the Sibyll2.3 fit is located at a significantly larger value of $t_{0_{\mathrm{norm}}}$ and σ_{norm}. However, the Sibyll2.3 fit



(a) Parameter space scan of the Auger HEAT $X_{\max}$ data.



(b) Parameter space scan of the Auger FD $X_{\max}$ data.



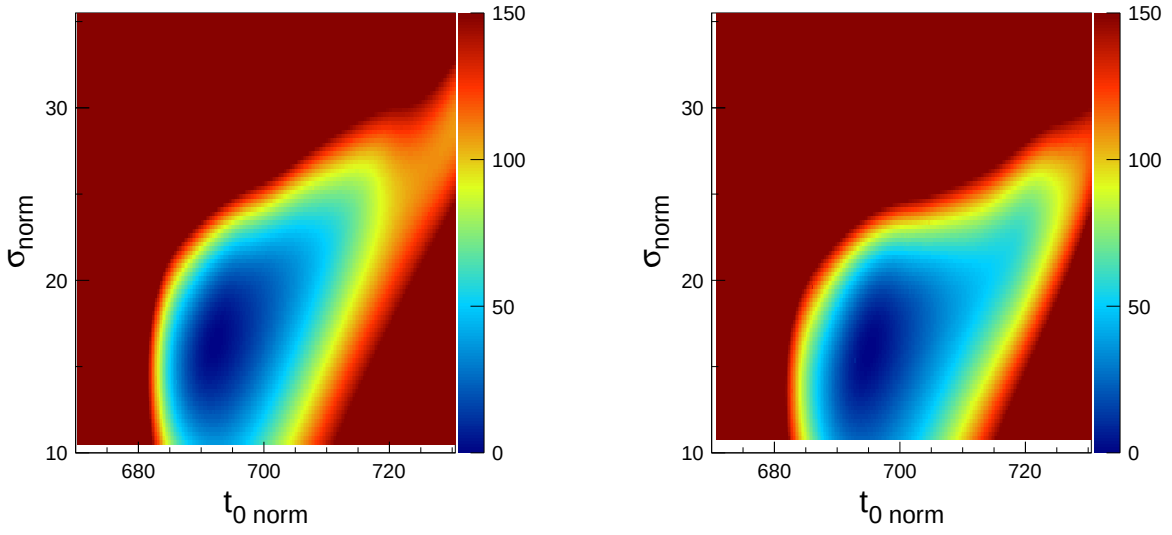
(c) Parameter space scan of the combined Auger FD and HEAT $X_{\max}$ data.

Figure 6.1: For specific values of $t_{0_{\text{norm}}}$ and σ_{norm}, the mass fractions of each parameterisation are fitted to the data, with the resulting minimised Poisson log likelihood relative to the deepest minimum for this fit displayed. The first 5σ contours of the minimised Poisson log likelihood are shown.

also has a local minimum within 4σ of the absolute minimum of the EPOS-LHC and QGSJetII-04 fits. The Auger FD $X_{\max}$ data is not able to constrain the Sibyll2.3 fit with as much certainty as the EPOS-LHC and QGSJetII-04 fits. If the mass dispersion, statistics and/or energy range of the data is increased, it is possible that the Sibyll2.3 fit can be better constrained, resulting in a deeper absolute minimum. By combining the Auger FD and HEAT $X_{\max}$ data and performing the parameter space scan, the results of which are shown in Figure 6.1c, we see the absolute minimum of the Sibyll2.3 fit is now located in the vicinity of the EPOS-LHC and QGSJetII-04 absolute minima. The added information of the HEAT $X_{\max}$ data has resulted in the best fit of the Sibyll2.3 parameterisation shifting to a value of $\sim 693\,\text{g/cm}^2$ and $\sim 16\,\text{g/cm}^2$ for $t_{0_{\text{norm}}}$ and σ_{norm} respectively, consistent with the other parameterisation fits.

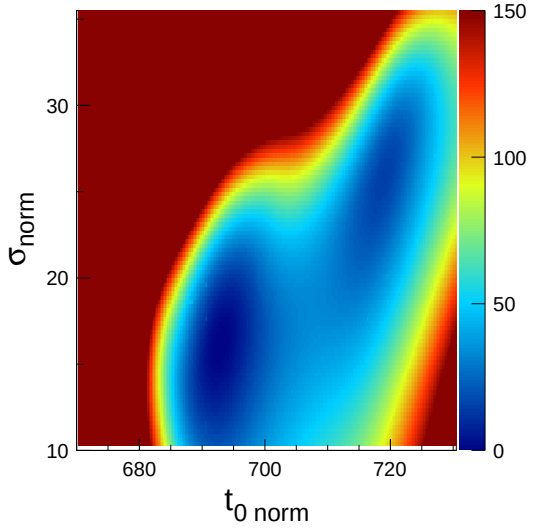
In Figure 6.1a, the absolute minima of the EPOS-LHC and QGSJetII-04 parameterisations coincide. The Sibyll2.3 absolute minimum is located at a value of $t_{0_{\text{norm}}}$ which is $\sim 8\,\text{g/cm}^2$ greater. The HEAT data contains less statistics, therefore the span of the first 5σ contours in $t_{0_{\text{norm}}}$ and σ_{norm} is larger. The results of the HEAT data scan confirm the results of Section 5.2.3, which illustrated the fits of $t_{0_{\text{norm}}}$ and σ_{norm} to HEAT data were more uncertain than the fits to FD data or the combined FD and HEAT data.

In Figure 6.1b, the two separate solutions of the Sibyll2.3 FD fit within 5σ of each other will not necessarily disappear if more statistics were present in the FD data. The nature of the hadronic interactions and mass composition in this energy range may inherently result in two separate solutions of similar minima given the Sibyll2.3 hadronic interaction predictions.



(a) EPOS-LHC

(b) QGSJetII-04



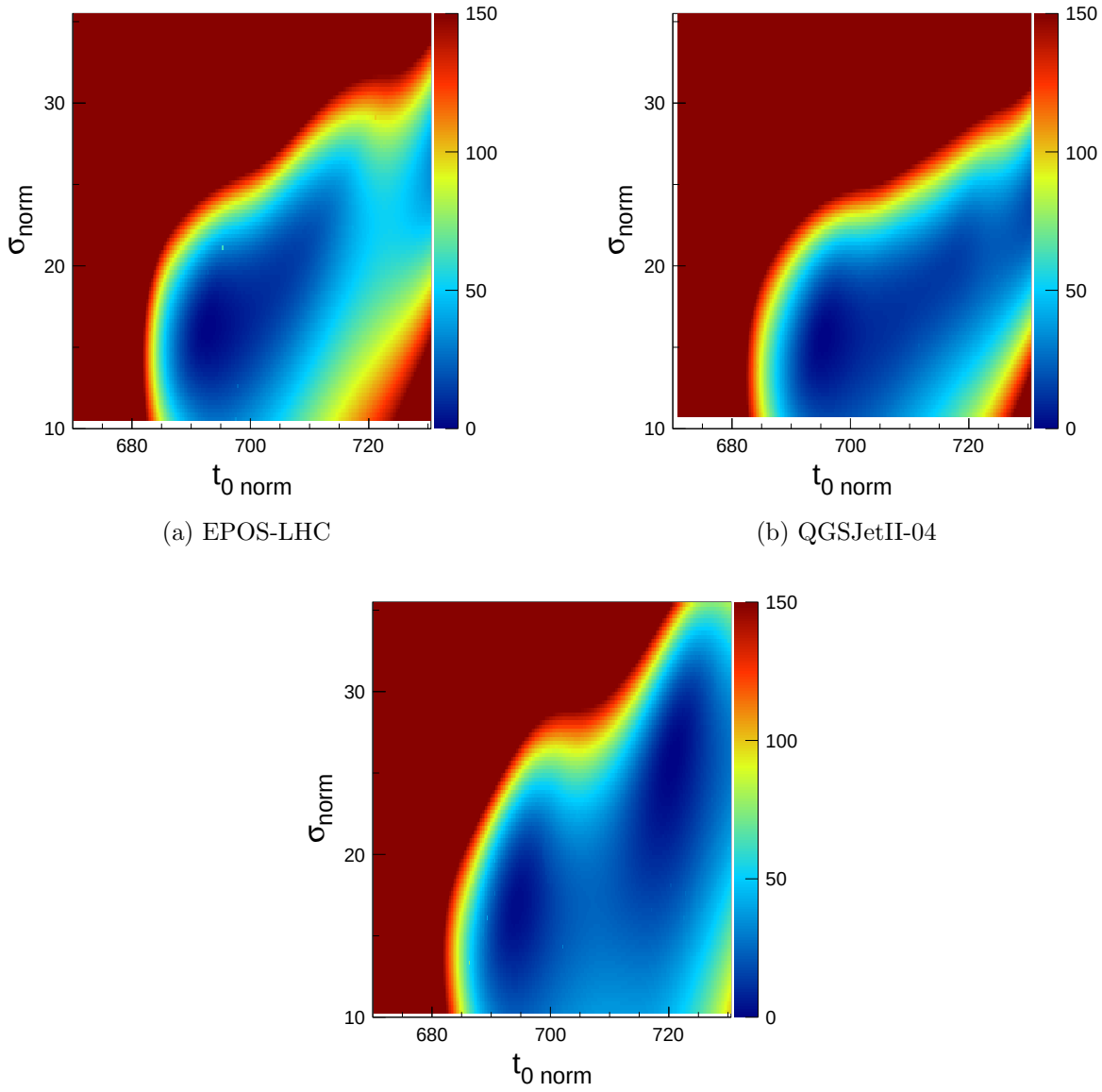
(c) Sibyll2.3

Figure 6.2: Fitting the mass fractions to the combined FD and HEAT $X_{\max}$ data, and scanning over $t_{0_{\text{norm}}}$ and σ_{norm} for each model.

Typically, the EPOS-LHC and QGSJetII-04 solutions from fitting $t_{0_{\text{norm}}}$ and σ_{norm} closely match because the differences in the other coefficients of these parameterisations is minimal. The Sibyll2.3 parameterisation is not as similar to these other two parameterisations. The separation

between the proton shape parameter predictions and heavier nuclei is larger in the Sibyll2.3 parameterisation, therefore the shape parameters for all nuclei of the Sibyll2.3 parameterisation can not align with the EPOS-LHC or QGSJetII-04 parameterisations as adequately as the EPOS-LHC and QGSJetII-04 predictions can align with each other.

Figure 6.2 shows the $t_{0_{\text{norm}}}$ and σ_{norm} scans of the combined FD and HEAT data for each parameterisation, within 150σ of the deepest minimum. The Sibyll2.3 fit of the FD and HEAT data contains a local minimum at a $t_{0_{\text{norm}}}$ value of $\sim 720\,\text{g/cm}^2$, a similar position to the true minimum of the Sibyll2.3 scan of only the FD data. However, this local minimum is not particularly deep compared to the true minimum. The minimised log likelihood space of all three parameterisations is skewed towards larger values of $t_{0_{\text{norm}}}$ and σ_{norm}.



(a) EPOS-LHC

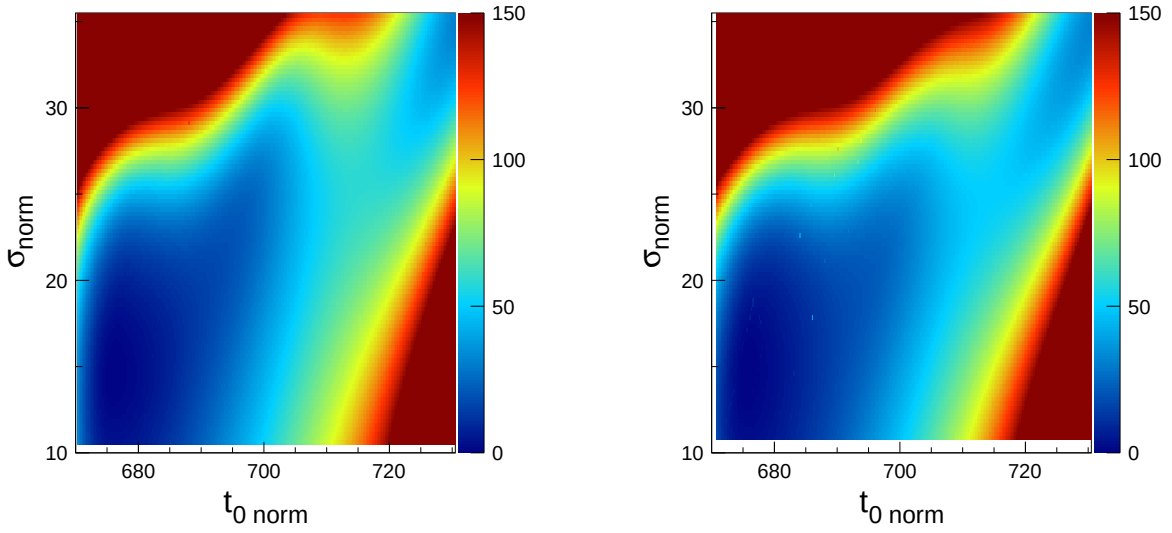
(b) QGSJetII-04

(c) Sibyll2.3

Figure 6.3: Fitting the mass fractions to the FD $X_{\max}$ data, and scanning over $t_{0_{\text{norm}}}$ and σ_{norm} for each model.

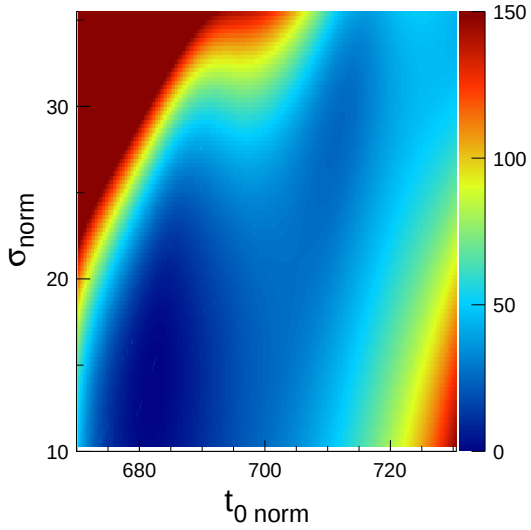
The full $t_{0_{\text{norm}}}$ and σ_{norm} scans of each model to only the FD data, and only the HEAT data, are shown in Figures 6.3 and 6.4 respectively. The FD scan shows the QGSJetII-04 fit contains a local minimum at a $t_{0_{\text{norm}}}$ of $\sim 720\,\text{g/cm}^2$, while the Sibyll2.3 local minimum is around $693\,\text{g/cm}^2$.

The EPOS-LHC and QGSJetII-04 scans of the FD data indicate there is a local minimum outside of the scan range (above 730 g/cm^2 in $t_{0_{\text{norm}}}$). Similarly, the scans of the HEAT data consist of local minima at larger values of $t_{0_{\text{norm}}}$.



(a) EPOS-LHC

(b) QGSJetII-04



(c) Sibyll2.3

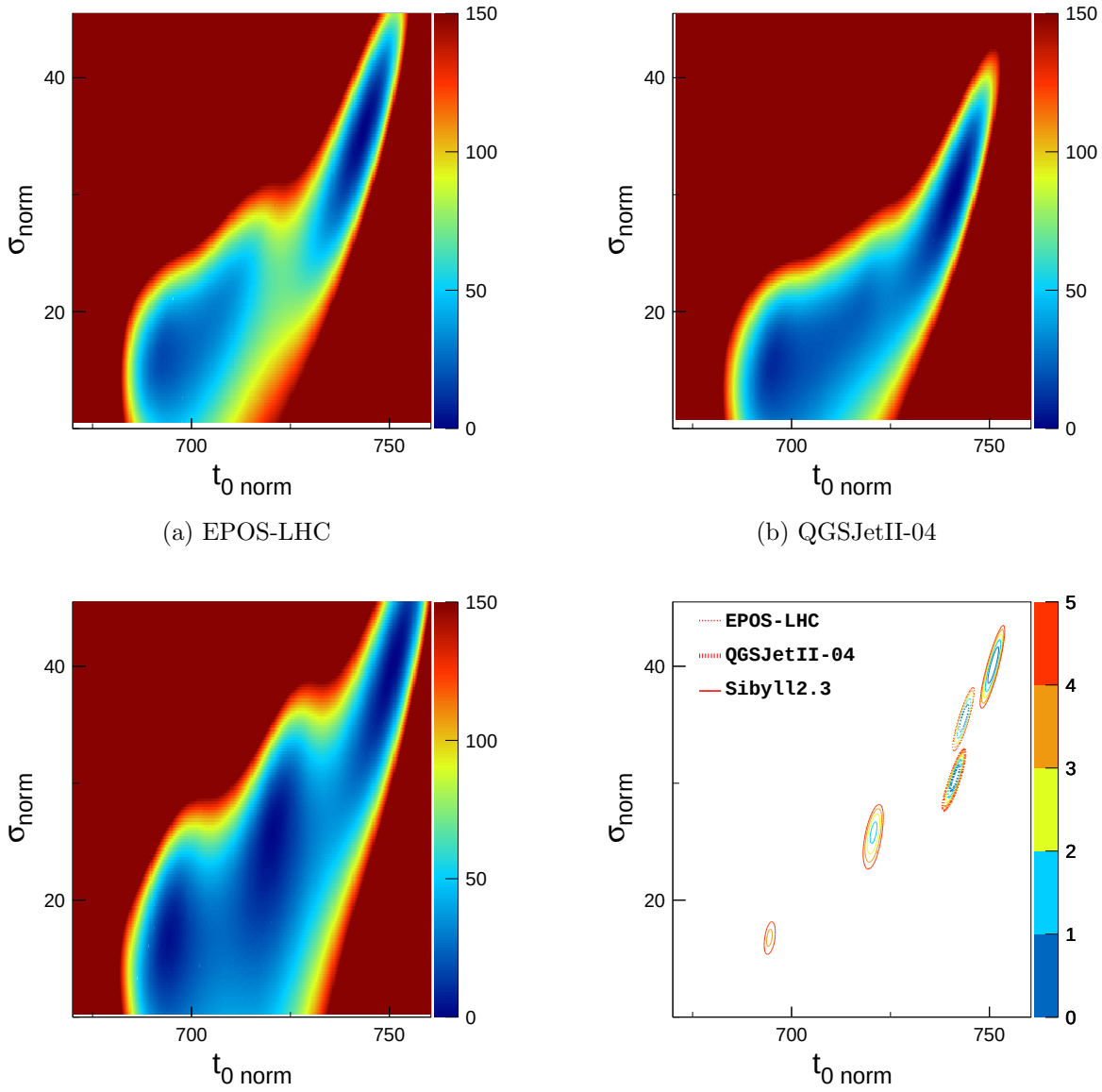
Figure 6.4: Fitting the mass fractions to the HEAT $X_{\max}$ data, and scanning over $t_{0_{\text{norm}}}$ and σ_{norm} for each model.

6.1.1 Extended $t_{0_{\text{norm}}}$ and σ_{norm} parameter space scans

To identify possible solutions at much larger values of $t_{0_{\text{norm}}}$ and σ_{norm}, the $t_{0_{\text{norm}}}$ scan range was extended to 670 g/cm^2 to 760.5 g/cm^2, and the σ_{norm} scan range was extended to 10 g/cm^2 to 45.5 g/cm^2, with a scanning interval of 0.25 g/cm^2. Figure 6.5 illustrates the results of this extended $t_{0_{\text{norm}}}$ and σ_{norm} scan on the FD $X_{\max}$ data with the EPOS-LHC, QGSJetII-04 and Sibyll2.3 $X_{\max}$ parameterisations. The result is absolute minima at a $t_{0_{\text{norm}}}$ value of between 740 g/cm^2 to 752 g/cm^2 and a σ_{norm} value of between 30 g/cm^2 to 40 g/cm^2. The mass composition, $\ln A$ moment and $X_{\max}$ moment results which correspond to the absolute minima are shown in Appendix C. The Sibyll2.3 scan results in two local minima within 5σ of the absolute

minimum, at values of $t_{0_{\mathrm{norm}}}$ of $\sim 694\,\mathrm{g/cm^2}$ and $\sim 721\,\mathrm{g/cm^2}$.

Extended $t_{0_{\mathrm{norm}}}$ and σ_{norm} scans of the combined FD and HEAT data, and of only the HEAT data, do not provide results which are consistent with the extended scan of only the FD data. The extended scans of the combined FD and HEAT data, and of only the HEAT data, are consistent with the standard scans displayed earlier in this chapter.



(a) EPOS-LHC

(b) QGSJetII-04

(c) Sibyll2.3

(d) 5σ contour summary of each model.

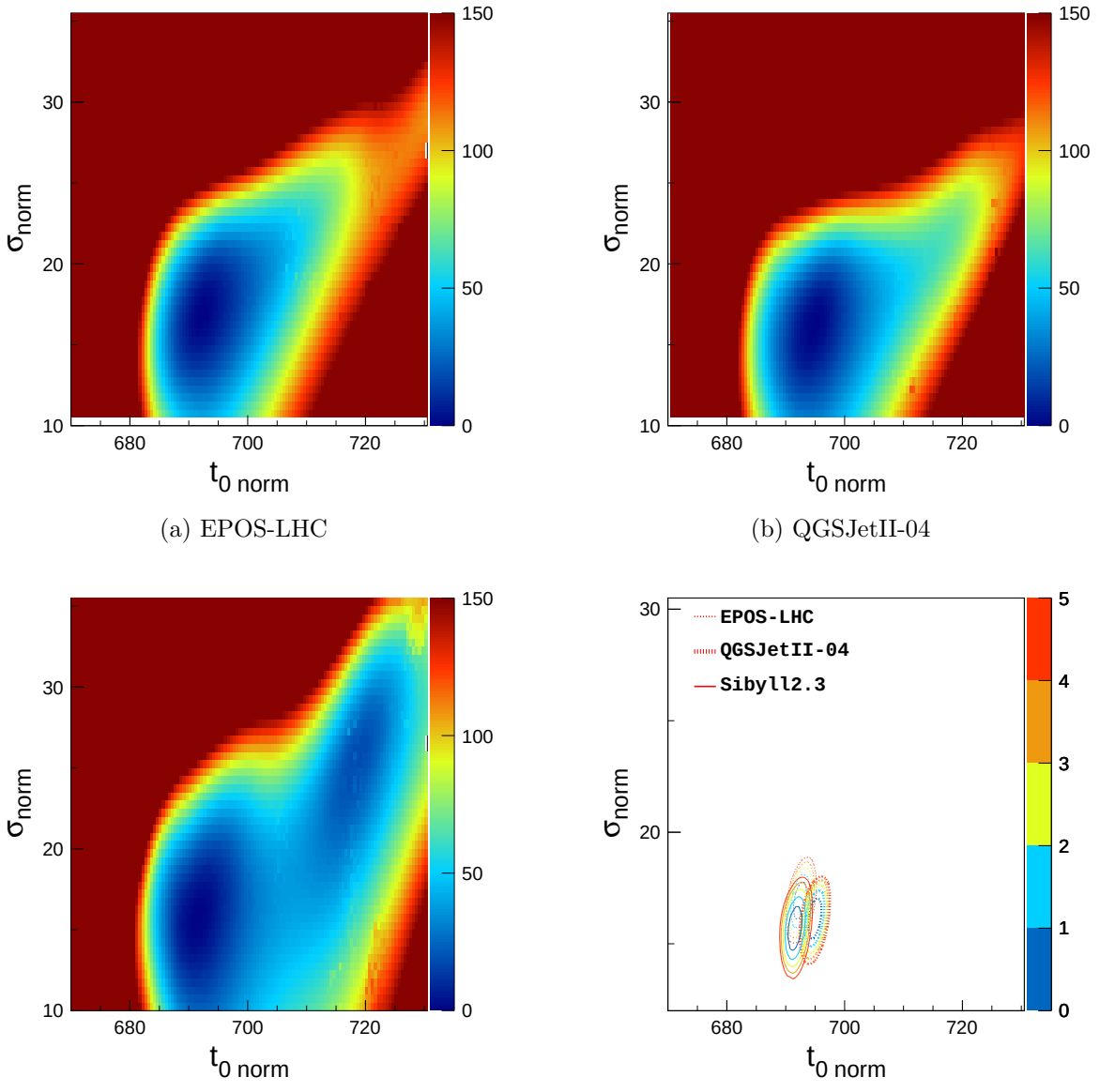
Figure 6.5: Fitting the mass fractions to the FD X_{max} data, and scanning over an extended range of $t_{0_{\mathrm{norm}}}$ and σ_{norm} for each model.

The initial $t_{0_{\mathrm{norm}}}$ predictions of EPOS-LHC, QGSJetII-04 and Sibyll2.3 are approximately $703\,\mathrm{g/cm^2}$, $688\,\mathrm{g/cm^2}$ and $714\,\mathrm{g/cm^2}$ respectively, much smaller than the $t_{0_{\mathrm{norm}}}$ values of the absolute minima in the extended scans of the FD data. Correspondence with Sergey Ostapchenko and Tanguy Pierog (developers of the QGSJetII-04 and EPOS-LHC models) has indicated that the hadronic interaction models are not expected to underestimate the fluctuations in X_{max}, therefore the σ_{norm} fitted to the Auger data is not expected to be smaller than the initial σ_{norm} of the models, unless the estimated X_{max} detector resolution is overestimated. The smallest proton σ_{norm} prediction of the three initial parameterisations is $21.61\,\mathrm{g/cm^2}$, from the EPOS-

LHC parameterisation. In the previous standard scans of $t_{0_{\text{norm}}}$ and σ_{norm} (not over the extended range) with the EPOS-LHC, QGSJetII-04 and Sibyll2.3 parameterisations, the absolute minima have corresponded with a proton σ_{norm} that is smaller than $21.61\,\text{g/cm}^2$, with the exception of the Sibyll2.3 scan of the FD data.

In Section 5.2, where the mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm} were fitted to mock Auger $X_{\max}$ distributions corresponding to a $t_{0_{\text{norm}}}$ solution below $725\,\text{g/cm}^2$, increasing the upper limit of the $t_{0_{\text{norm}}}$ fit to $760.5\,\text{g/cm}^2$ does not cause fits which previously found a best fit of $t_{0_{\text{norm}}}$ below $725\,\text{g/cm}^2$ to be fitted to a larger $t_{0_{\text{norm}}}$ value.

6.2 $t_{0_{\text{norm}}}$, σ_{norm} and λ_{norm} parameter space scans



(a) EPOS-LHC

(b) QGSJetII-04

(c) Sibyll2.3

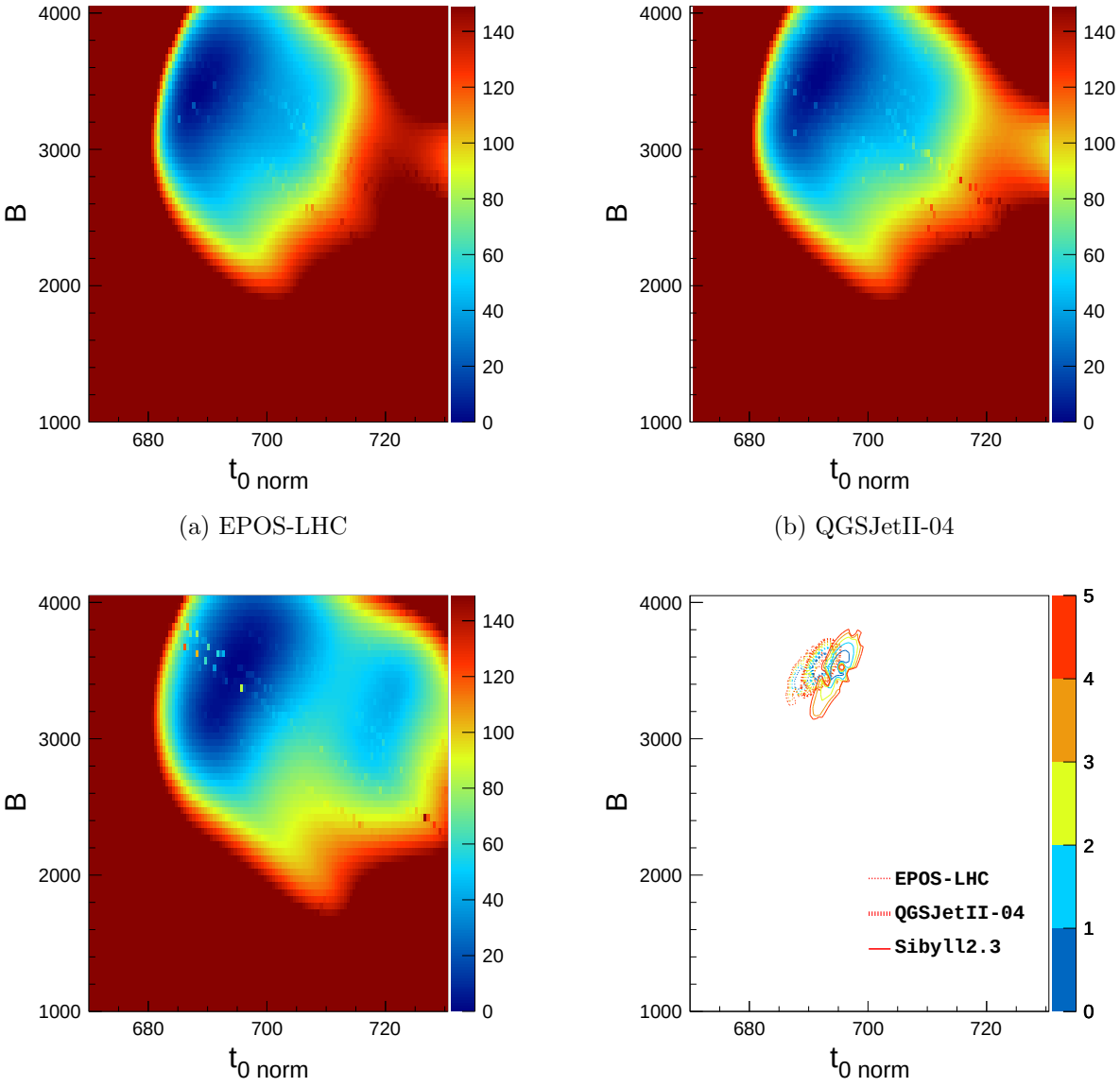
(d) 5σ contour summary of each model.

Figure 6.6: Fitting the mass fractions to the combined FD and HEAT $X_{\max}$ data, and scanning over $t_{0_{\text{norm}}}$, σ_{norm} and λ_{norm} for each model. The relationship between the relative minimised log likelihood and λ_{norm} is not shown.

Figure 6.6 illustrates the results of the $t_{0_{\text{norm}}}$, σ_{norm} and λ_{norm} scans of each model to the

combined FD and HEAT data, with the relative minimised log likelihood shown in terms of the fixed values of $t_{0_{\text{norm}}}$ and σ_{norm}. $t_{0_{\text{norm}}}$, σ_{norm} and λ_{norm} were scanned in 0.5 g/cm^2 intervals. The absolute minima are located at a proton λ_{norm} value of 57.6 g/cm^2, 58.3 g/cm^2 and 59.0 g/cm^2 for the EPOS-LHC, QGSJetII-04 and Sibyll2.3 models respectively. The overall likelihood space and the position of the absolute minima in $t_{0_{\text{norm}}}$ and σ_{norm} are consistent with the scans in the previous section (see Figure 6.2), where λ_{norm} was fixed to the initial predictions of the models.

6.3 $t_{0_{\text{norm}}}$, B and σ_{norm} parameter space scans



(a) EPOS-LHC

(b) QGSJetII-04

(c) Sibyll2.3

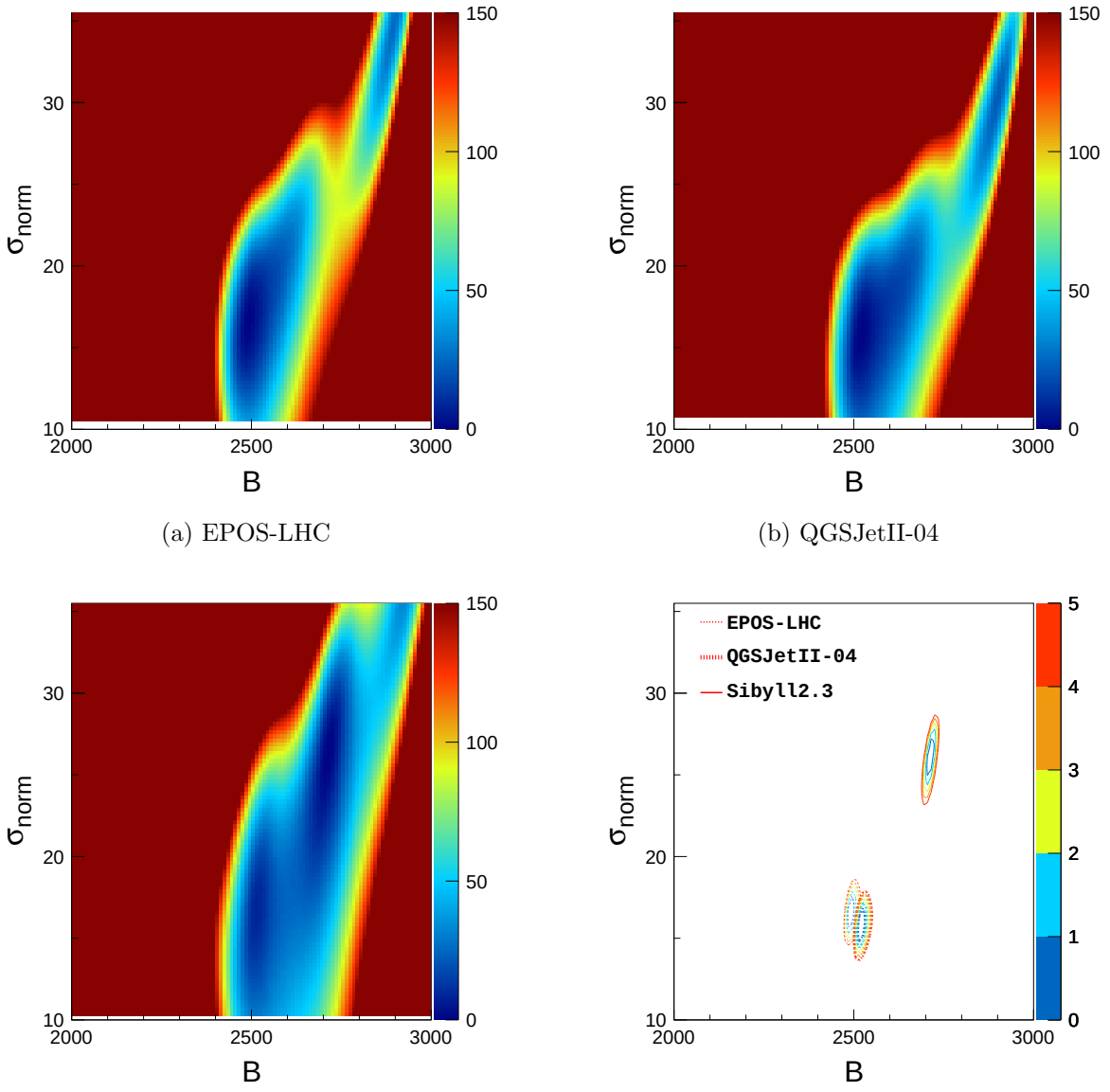
(d) 5σ contour summary of each model.

Figure 6.7: Fitting the mass fractions to the combined FD and HEAT $X_{\max}$ data, and scanning over $t_{0_{\text{norm}}}$, B and σ_{norm} for each model. The relationship between the relative minimised log likelihood and σ_{norm} is not shown.

Figure 6.7 illustrates the results of the $t_{0_{\text{norm}}}$, B and σ_{norm} scans of the combined FD and HEAT data, with the relative minimised log likelihood shown in terms of the fixed values of $t_{0_{\text{norm}}}$ and B. $t_{0_{\text{norm}}}$, B and σ_{norm} were scanned in intervals of 0.5 g/cm^2, 50 g/cm^2 and 0.5 g/cm^2

respectively. The absolute minimum corresponds to a proton σ_{norm} value of 17.1, 15.8 and 15.8 g/cm^2 for the EPOS-LHC, QGSJetII-04 and Sibyll2.3 models respectively. These scans indicate that for the three models, a B of around 3400 g/cm^2 best fits the Auger data (the initial model values of B are $\sim$ 2500 g/cm^2), a value which significantly increases the predicted $\langle X_{\max} \rangle$ elongation rate.

6.4 B and σ_{norm} parameter space scans



(a) EPOS-LHC

(b) QGSJetII-04

(c) Sibyll2.3

(d) 5σ contour summary of each model.

Figure 6.8: Fitting the mass fractions to the combined FD and HEAT $X_{\max}$ data, and scanning over B and σ_{norm} for each model.

Figure 6.8 illustrates the results of the B and σ_{norm} scans of each model to the combined FD and HEAT data. The t_0 parameterisations were constrained at 10^{14} eV. B and σ_{norm} were scanned in intervals of 10 g/cm^2 and 0.25 g/cm^2 respectively. The true minima of the EPOS-LHC and QGSJetII-04 scans are at a B of $\sim$ 2500 g/cm^2, with the Sibyll2.3 scan consisting of a local minima at this location. A B of 2500 g/cm^2 corresponds to a t_0 elongation rate of

$\sim 59\,\mathrm{g/cm^2}$ between $10^{18}\,\mathrm{eV}$ and $10^{19}\,\mathrm{eV}$. The deepest Sibyll2.3 minimum coincides with a B of $\sim 2710\,\mathrm{g/cm^2}$, which corresponds to a t_0 elongation rate of $\sim 64\,\mathrm{g/cm^2}$ between $10^{18}\,\mathrm{eV}$ and $10^{19}\,\mathrm{eV}$. The EPOS-LHC and QGSJetII-04 scans contain local minima at a B of $\sim 2850\,\mathrm{g/cm^2}$. For the three parameterisations, the predicted $\langle X_{\mathrm{max}} \rangle$ of the true minima from the B and σ_{norm} scans of the FD and HEAT data are consistent with the FD data $t_{0_{\mathrm{norm}}}$ and σ_{norm} scans of the corresponding parameterisations (see Figure 6.3).

6.5 Conclusions

Parameter space scans of the Auger X_{max} data reveal the shape coefficient values corresponding to the absolute minima and local minima of the fits. The scans of the combined FD and HEAT data, compared to the scans of only the FD data or only the HEAT data, demonstrate that data of greater statistics, energy range and/or mass dispersion are better constrained. The standard scans of the combined FD and HEAT data with EPOS-LHC, QGSJetII-04 and Sibyll2.3 over $t_{0_{\mathrm{norm}}}$ and σ_{norm}, and over $t_{0_{\mathrm{norm}}}$, σ_{norm} and λ_{norm}, reveal a consistent 5σ absolute minimum region. The standard $t_{0_{\mathrm{norm}}}$ and σ_{norm} scans of the FD data with the EPOS-LHC and QGSJetII-04 parameterisations result in a 5σ absolute minimum region which is consistent with the FD and HEAT scan. The Sibyll2.3 fit consists of a local minimum in this region within 4σ of the absolute minimum of this Sibyll2.3 fit, the absolute minimum located at a larger value of $t_{0_{\mathrm{norm}}}$ and σ_{norm}.

Extending the $t_{0_{\mathrm{norm}}}$ and σ_{norm} scan range results in a drastically different solution for the FD data, at a $t_{0_{\mathrm{norm}}}$ and σ_{norm} much larger than the initial model predictions. This solution corresponds to a much heavier mass composition. Given the current knowledge of the hadronic physics occurring at the highest energies, we are uncertain of the range at which to restrict the fitted shape coefficients. Additionally, the shape coefficients which are fixed to the data, which have varying levels of uncertainty, can strongly impact the mass composition solution. For example, the standard scans of $t_{0_{\mathrm{norm}}}$, B and σ_{norm} reveal absolute minima which correspond to a larger X_{max} elongation rate than predicted by the models, which will result in a heavier mass composition at the highest energies. A greater understanding of the hadronic physics at the highest energies will help us to determine which shape coefficients to fit, and reduce the limits. Additionally, X_{max} data of greater statistics and spanning a large energy range will better constrain the fit. Regardless of the combination of shape coefficients scanned, the scan range of the coefficients, or the data fitted, we find that a particular fit of the EPOS-LHC, QGSJetII-04 and Sibyll2.3 X_{max} parameterisations will tend to result in a solution that is consistent between these parameterisations.

In Chapter 8, using the standard limits for the shape coefficients, the shape coefficient fit combinations presented in this chapter have been applied to the Auger X_{max} data.

Chapter 7

Statistical errors, systematic errors and p-values of the fits

Described in this chapter are the methods used to determine the statistical errors and systematic errors of the fitted mass fractions, fitted shape coefficients and any other properties sensitive to these fitted parameters. The method for determining the goodness of fit, which is quantified by the calculated p-values, is also described in this chapter. The methods outlined in this chapter are applied to the fits of the Auger $X_{\max}$ distributions in Chapter 8.

7.1 Statistical errors

The statistical errors were determined by using the parameterisation fitted to the $X_{\max}$ data to generate 100 mock data sets of equal statistics. Each mock data set is then fitted in the same way, the 100 fits resulting in distributions of the fitted mass fractions, coefficients and other determinable quantities such as $\langle X_{\max} \rangle$. The 1σ standard deviation of a fitted property was then calculated by determining the middle 68% range of values in a distribution, the lower or upper statistical error equal to the difference between the median of the distribution and the 16^{th} or 84^{th} percentile respectively.

7.2 Systematic errors

The systematic errors were determined by shifting the $X_{\max}$ data and the predicted $X_{\max}$ detector resolution by relative amounts within the defined lower and upper systematic errors. Distributions of fitted fractions, coefficients and other quantifiable observables are obtained, each element corresponding to a specific systematic shift. The maximum and minimum values of a quantity, relative to the initial fitted value, defines the upper and lower systematic error respectively.

For example, in one instance, the FD and HEAT $X_{\max}$ data is shifted down over the whole range by 80% of the lower systematic error in $X_{\max}$ for these respective data sets, while the predicted FD and HEAT $X_{\max}$ resolution is shifted up by 40% of the upper systematic error in the resolution of these respective data sets. The fit, using the adjusted $X_{\max}$ detector resolution, is then applied to the adjusted data. The relative shifts are randomly determined. The process is performed 100 times, resulting in a set of fitted mass fractions (and if the coefficients are fitted, a set of coefficient values), each result corresponding to some shifted combination of the $X_{\max}$ data and detector resolution. We also perform the fit process on data subject to the maximum shifts of the lower or upper systematics in $X_{\max}$, or the lower or upper systematics in the resolution, as well as for combinations of these maximum limits.

From the sets of fitted fractions, coefficients and other $X_{\max}$ property results, we determine the minimum and maximum values, defining our lower and upper systematic error bars for each

quantity. In the case of fitting at least $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$, the systematic shifts in the $X_{\rm max}$ data and the resolution are accounted for by these coefficients, resulting in small systematics in the fitted mass composition. This was demonstrated in Section 5.3. In some instances, a $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$ fit becomes stuck at the defined limits for these coefficients, or two solutions exist for the $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$ fit within the set of results, solutions which may or may not be the true minimum for the particular data set fitted. In this case, only the fits which result in a solution that corresponds to the initial fitted solution are considered ($t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$ shifted appropriately considering the systematic offsets), otherwise if a single fit happens to find a different solution for $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$, the systematic errors for the mass fractions will be very large. Consequently, the systematic errors displayed in this work represent the systematic errors of the accepted solution.

In the fits to shifted $X_{\rm max}$ data, or mock $X_{\rm max}$ data (for example, to determine the statistical errors), it is sometimes necessary to increase the allowed limits of a fitted coefficient, if the initial fit of the data resulted in a coefficient fitted near the limit.

7.3 p-values

The fit quality is measured by calculating the p-value, defined as the probability of obtaining a worse fit (larger likelihood ratio $\mathcal{L}$) than that obtained with the data, assuming the distribution predicted by the fit of the data is correct. The likelihood ratio for an energy bin is:

$$\mathcal{L} = \sum_i -2 * (d_i \cdot log(f_i) - d_i \cdot log(d_i) - f_i + d_i) \ , \tag{7.1}$$

where i is the $X_{\rm max}$ bin, d_i the measured data point, and f_i the model prediction.

The likelihood ratio of an energy bin from the fit of the data can be compared to a distribution of likelihood ratios obtained by performing the same fits to the mock data, mock data which was generated from the parameterisation fitted to the initial $X_{\rm max}$ data. Figure 7.1 shows an example of the likelihood ratio distribution obtained from the fits of the mock data, compared to the likelihood ratio of the measured data. The p-value for an energy bin is the proportion of likelihood ratios from the mock fits which are greater than the likelihood ratio of the fit of the initial data. A fit that does not describe the data well will have a smaller p-value.

The $X_{\rm max}$ binning used to calculate the likelihood ratio significantly affects the p-value obtained. In the results presented in this work, the parameterisation fits of the data are always performed using $1\,{\rm g/cm^2}$ bins. This small binning means we are essentially performing an unbinned likelihood fit and consequently the goodness of fit is difficult to determine [239]. Using a binning of $1\,{\rm g/cm^2}$ in the calculation of the likelihood ratio resulted in many p-values close to 1 for the fits of Auger data, which is unrealistic given the fit does not perfectly describe the data. Therefore, different bin sizes were used to calculate the likelihood ratio (the fits were still performed with $1\,{\rm g/cm^2}$ binning however). The 15 bins used in the calculation of the likelihood ratios are $[0.25, 1, 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60]$ in ${\rm g/cm^2}$. Applying different bin sizes to calculate the likelihood ratios did not add considerable computational time to the fit procedure.

A mock FD and HEAT $X_{\rm max}$ data set was generated, which was treated as the measured data, and the mass fractions of the EPOS-LHC parameterisation were fitted to this data. The fitted parameterisation was then used to generate 100 mock data sets of the 'measured data', each of which was then fitted in the same way as the initial data, to obtain a distribution of likelihood ratios for determining the p-values.

The resulting p-values from different likelihood ratio bin sizes (calculating the likelihood ratio using different binning) are shown in Figure 7.2 for each energy bin. We see that p-value obtained depends significantly on the bin size used to calculate the likelihood ratios of the initial fit and the fits of the mock data.

The general approach applied to obtain p-values that are unbiased by the choice of likelihood binning is as follows. The likelihood ratios for a set of bin sizes are computed, for both the

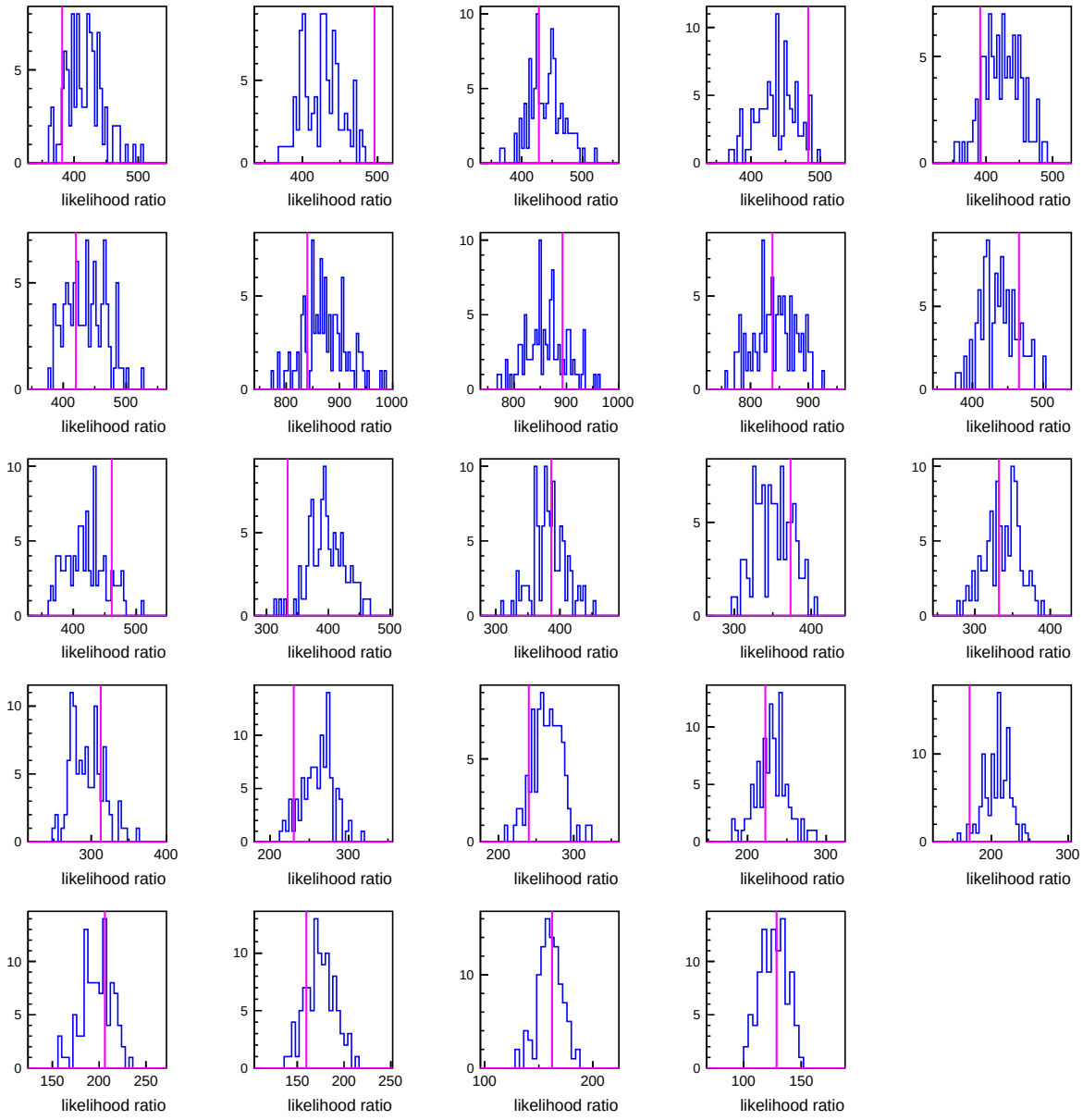


Figure 7.1: The likelihood ratio distributions for each energy bin, using an $X_{\max}$ bin size of $1\,\mathrm{g/cm^2}$. From left to right, top to bottom, each plot is a different energy bin, starting from $10^{17.2-17.3}$ eV, and ending with the final energy bin of $10^{19.5-20}$ eV. The blue distribution illustrates the 100 likelihood ratios obtained for the fits of the mock data, and the magenta line represents the likelihood ratio for the fit of the measured data.

measured data and the 100 mock data sets, which are then used to calculate the measured data p-value for each bin size. The smallest p-value of each energy bin is then chosen as the measured data p-value. The same procedure is then repeated, with the mock data sets now treated as the measured data (the real measured data are no longer used). Therefore additional mock data sets are generated for each initial mock data set, to obtain a distribution of mock data p-values. Finally, by comparing the distribution of mock data p-values to the measured data p-value (of the corresponding energy bin), a final p-value is calculated.

Described below in more detail are the steps used to determine p-values which are not biased by our choice of the likelihood binning:

1. The initial data is fitted, and then using different bin sizes, the likelihood ratio of the fit to

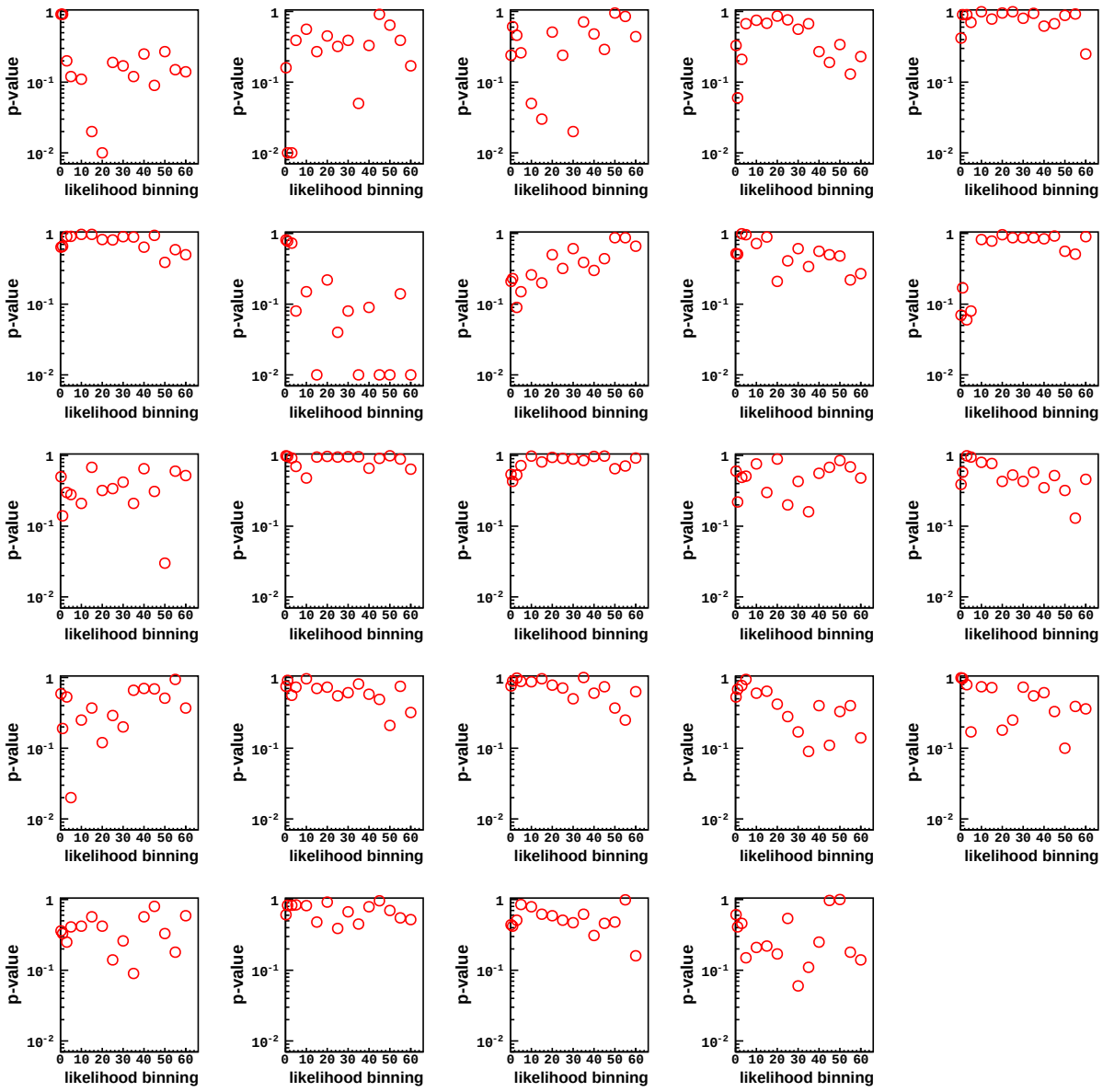


Figure 7.2: The p-values from using a particular bin size in the calculation of the likelihood ratio. From left to right, top to bottom, each plot is a different energy bin, starting from $10^{17.2-17.3}$ eV, and ending with the final energy bin of $10^{19.5-20}$ eV.

the data is determined ($\mathcal{L}_d$).

2. 100 mock data sets are generated from the parameterisation fitted to the initial data. Each mock data set is then fitted in the same way as the initial data, and then using different bin sizes, the likelihood ratios of the mock data fits are determined ($\mathcal{L}_m$) by using the mock data fit as the model prediction. Therefore, for each bin size, we have a distribution of 100 likelihood ratios. The mock data sets fitted in this step will be defined as the 'primary' mock data in the following steps.

3. For each bin size, $\mathcal{L}_d$ is compared to the corresponding $\mathcal{L}_m$ distribution. Therefore, a set of p-values are obtained for each bin size, with the lowest p-value of the set selected (p_{ref}).

4. For each fit of the primary mock data in step 2, we generate 100 mock data sets from the parameterisation fitted to this mock data. We repeat step 2 on this second set of mock

data, obtaining likelihood ratios for this secondary mock data ($\mathcal{L}_s$). Therefore, for each bin size, we have a distribution of 100 likelihood ratios corresponding to each fit of a primary mock data set.

5. For each bin size, $\mathcal{L}_s$ is compared to the corresponding $\mathcal{L}_m$ value. Therefore, a set of p-values are obtained for each bin size, and again the lowest p-value of the set is selected (p_m).

6. p_{ref} (the lowest p-value for the fit of the initial data) is then compared to p_m, the distribution of lowest p-values from the fits of the primary mock data. The proportion of p_m which are less than p_{ref} determines our final p-value (p_f), giving us a measurement of the goodness of the fit to the initial data.

Applying these steps to the mock FD and HEAT $X_{\max}$ data mentioned previously (which we are treating as the measured data), we obtain the following p-values in Figure 7.3. The initial mock data set was generated from our EPOS-LHC parameterisation, therefore the p-values of our fit to this data should be uniformly distributed between 0 and 1 over the energy range. In Figure 7.3, we see the p-values are uniformly distributed. There are 24 energy bins, so we expect 2.4 p-values (10%) on average to be below 0.1.

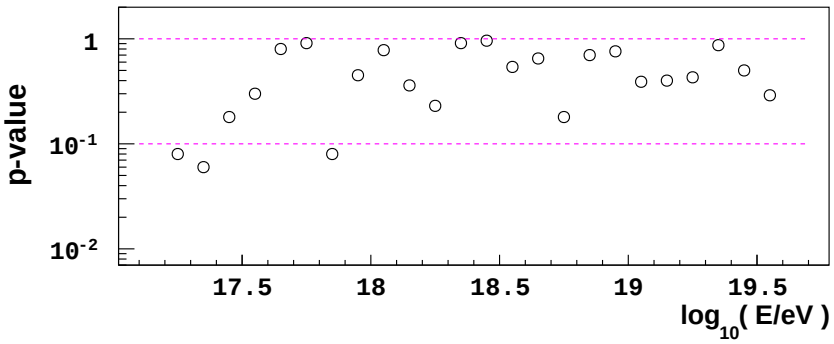


Figure 7.3: Final p-values from the fit of a mock FD and HEAT $X_{\max}$ data set.

The method applied above, which will be defined as the 'proper method', involves fitting many mock data sets, and is therefore computationally intensive if applied for fits of the shape parameter coefficients. We instead apply a 'simplified method', which will be used not only when fitting the shape coefficients and the mass fractions, but also when fitting just the mass fractions, allowing the p-values of these fits to be compared with each other. The simplified method alters steps 2 and 4 of the 'proper method'. In step 2 of the simplified method, 2 different sets of 100 mock data are generated from the parameterisation fitted in step 1. The mock data of the first set is compared to the model prediction, but instead of fitting this mock data and using the fit as the model prediction in the calculation of $\mathcal{L}_m$, the parameterisation used to generate this mock data is used as the model prediction. Identical to the 'proper method', the $\mathcal{L}_m$ obtained here will be used in step 3, but will not be used in step 5. The mock data of the second set is fitted (this fit will be used to generate secondary mock data in step 4 as usual). Identical to the 'proper method', the fit of the mock data of this second set will be used as the model prediction. Each of these mock data $\mathcal{L}_m$ will then be used as before in step 5 (they will not be used in step 3). In step 4 of the simplified method, similar to the process applied to the first set of mock data in step 2, instead of fitting the secondary mock data and using the fit as the model prediction in the calculation of the likelihood ratio, the parameterisation used to generate this secondary mock data is used as the model prediction. Consequently, in step 4, a fit of the secondary mock data is no longer needed to acquire $\mathcal{L}_s$.

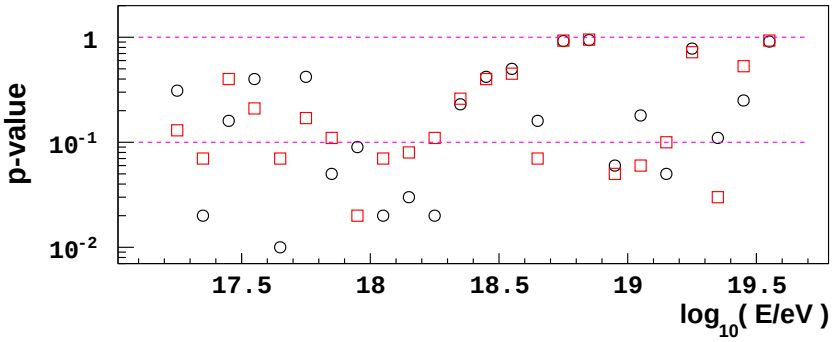


Figure 7.4: p-values for the mass fraction EPOS-LHC parameterisation fit of older FD and HEAT Auger $X_{\max}$ data, using different methods to calculate the p-values. The black points are from the simplified method, the red from the proper method.

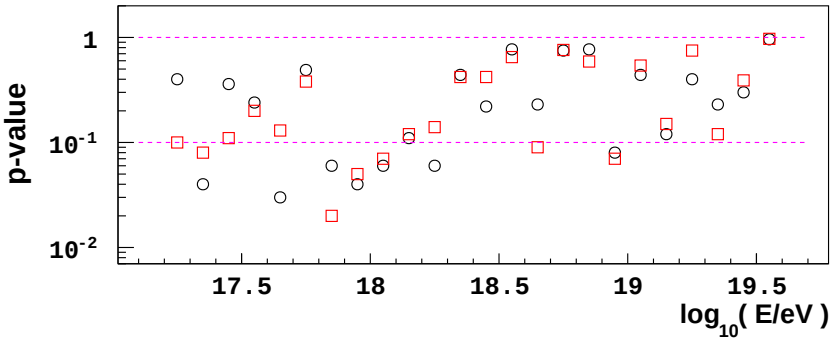


Figure 7.5: p-values for the $t_{0_{\text{norm}}}$, σ_{norm} and mass fraction EPOS-LHC parameterisation fit of older FD and HEAT data, using different methods to calculate the p-values. The black points are from the simplified method, the red from the proper method.

The results of this 'proper method' and 'simplified method' are compared in Figure 7.4, where the mass fractions of the EPOS-LHC parameterisation were fitted to an older Auger FD and HEAT $X_{\max}$ data set. We see the 'simplified method' (black points) results in a fit quality description that is representative of the fit quality description from the 'proper method' (red points). The 'simplified method' tends to underestimate the p-values, implying the fit quality is worse overall.

In Figure 7.5, the results of the different methods are compared, for the fit of $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions of the EPOS-LHC parameterisation to an older Auger FD and HEAT $X_{\max}$ data set. We see that the 'simplified method' again presents a fit quality description representative of the 'proper method'.

As we generate only 100 sets of mock data, the lowest p-value that can be obtained is 10^{-2}. Increasing the number of mock data sets, for example to 1000 (allowing a lower limit of 10^{-3}), would mean we can be more certain of which energy bins are fitted poorly. However, this would increase the computational time significantly. When the p-values are close to 10^{-2}, we can assume the goodness of fit is very poor. To determine the quality of a fit, we are mainly interested in the proportion of p-values less than 0.1.

We do not calculate systematic errors for the p-values (statistical errors are not attainable). Shifting the measured $X_{\max}$ events and the predicted $X_{\max}$ resolution by some combination of the

minimum and maximum limits of the estimated systematic errors, and then performing the fit, does not result in a lower and upper systematic error for the p-values which is centred around the p-value of the fit to the original data. The p-values obtained from systematic shifts are scrambled. For example, shifting the $X_{\max}$ data up by the upper systematic error, and shifting the $X_{\max}$ data down by the lower systematic error, can result in p-values that are in both instances larger than the p-values from the fit of the un-shifted data. Applying systematic shifts in the $X_{\max}$ resolution produce the same outcome.

Chapter 8

Analysis of Pierre Auger Observatory X_{max} data

8.1 FD + HEAT X_{max} data composition results

The EPOS-LHC, QGSJetII-04 and Sibyll2.3 X_{max} parameterisations were fitted to X_{max} data measured by the FD and HEAT detectors of the Pierre Auger Observatory [148]. This hybrid data spans an energy range of $10^{17.2}$ eV to 10^{20} eV. The number of events in each energy bin is detailed in Appendix B. In the energy bins where the FD and HEAT data overlap, the $\langle X_{\mathrm{max}} \rangle$ of the HEAT data is on average 2.3 g/cm^2 larger than that of the FD distributions. Therefore, the X_{max} values of the HEAT data are shifted by -2.3 g/cm^2 to align the FD and HEAT distributions in the three overlapping energy bins.

8.1.1 Mass fraction fit of the combined Auger FD and HEAT X_{max} data

The results of fitting only the mass fractions our EPOS-LHC, QGSJetII-04 or Sibyll2.3 X_{max} distribution parameterisations to the combined FD and HEAT X_{max} data are displayed in Figure 8.1. The top three panels display the fitted mass fractions for each parameterisation, and the bottom panel shows the p-values for these fits. The statistical errors are indicated by the error bar lines, while the systematic error bars are indicated by the square brackets. The fits of these parameterisations to the measured X_{max} distributions are shown in Appendix D.1

Figure 8.1 shows the reconstructed mass composition depends significantly on the hadronic interaction model assumed. The EPOS-LHC fit predicts a mass composition consisting of mainly protons and nitrogen around $10^{17.4}$ eV, which transitions to a larger proton fraction up to $10^{18.4}$ eV, with helium becoming dominant around $10^{18.8}$ eV, and then nitrogen mixing with helium at higher energies. The QGSJetII-04 fit on the other hand predicts a mix of protons, helium and nitrogen which transitions to a composition dominated by protons around $10^{18.3}$ eV, and then helium dominates above $10^{19.2}$ eV, with minimal nitrogen above $10^{17.7}$ eV. The Sibyll2.3 fit predicts a considerable nitrogen component across the whole energy range, with a consistent proton component up to $10^{18.6}$ eV, and a helium component growing and declining between 10^{18} eV to $10^{19.2}$ eV. All three models predict an almost null abundance of iron between $10^{17.8}$ eV and $10^{19.4}$ eV, with EPOS-LHC and Sibyll2.3 predicting a small fraction of iron creeping in above $10^{19.4}$ eV. Above 10^{18} eV, the EPOS-LHC, QGSJetII-04 and Sibyll2.3 fits predict to varying degrees, a composition transitioning to heavier nuclei with increasing energy. This transition reflects a Peters cycle [240], where the maximum acceleration energy of a species is proportional to its charge Z. A significant modification of the hadronic models is required to accommodate a proton dominant composition above 10^{18} eV [241].

The EPOS-LHC and QGSJetII-04 fits both predict a significant proton fraction increasing up to the ankle region (around $10^{18.2}$ eV), and then declining. The upper limits on the large scale

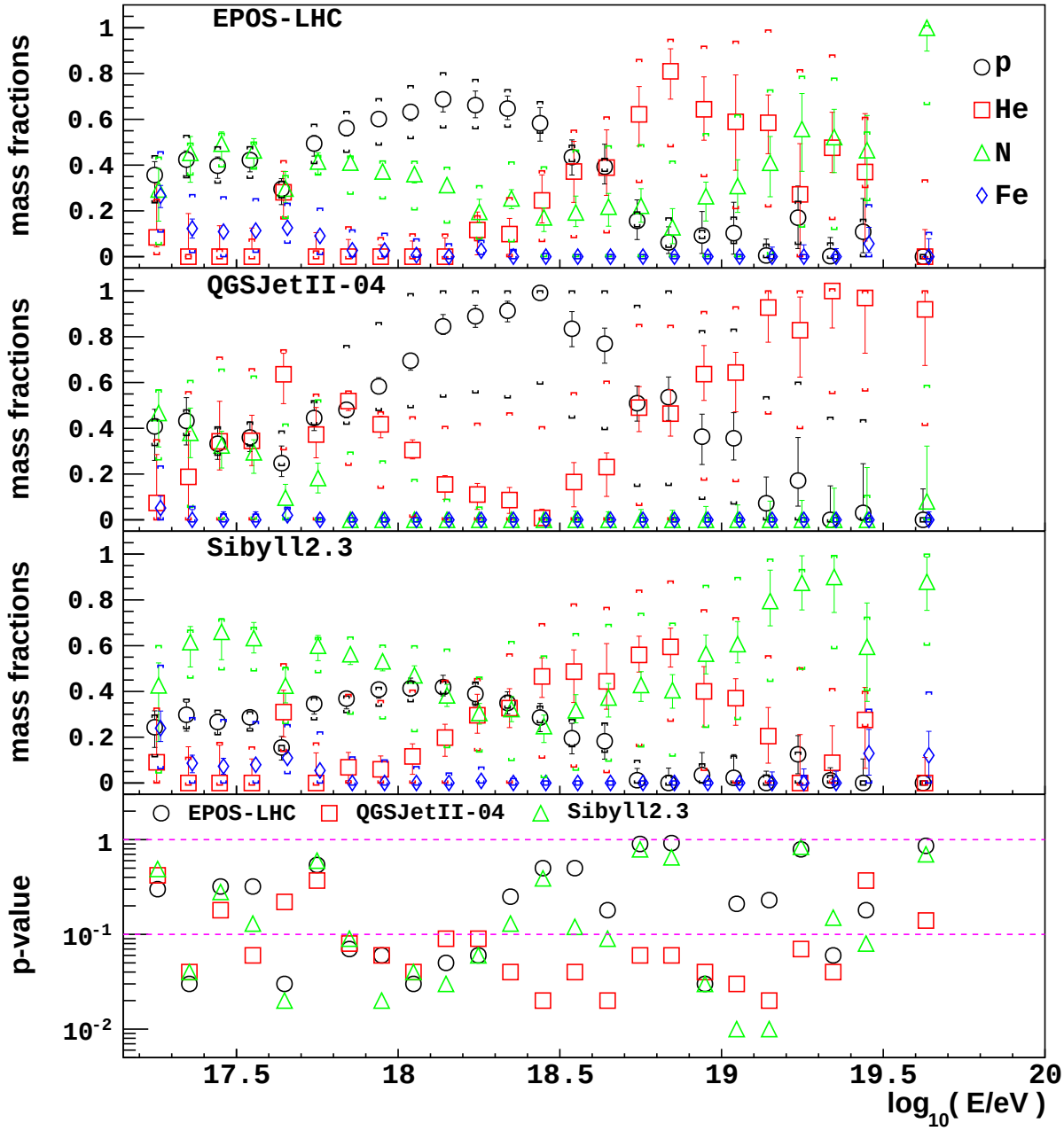
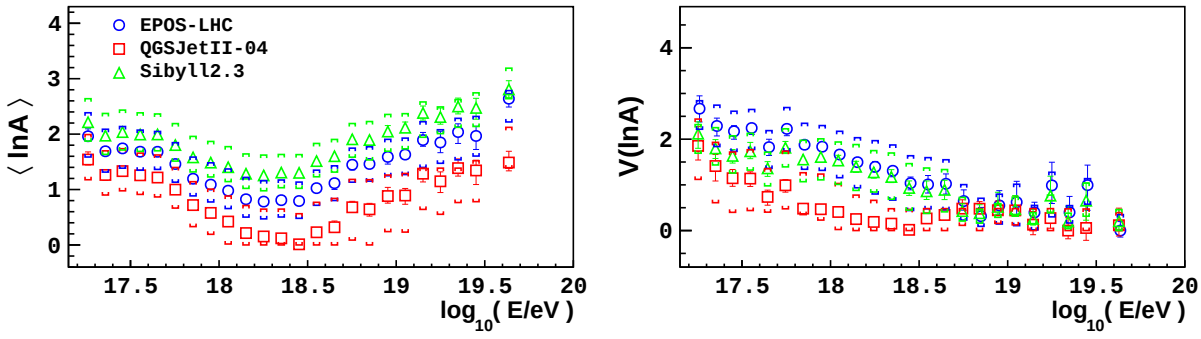


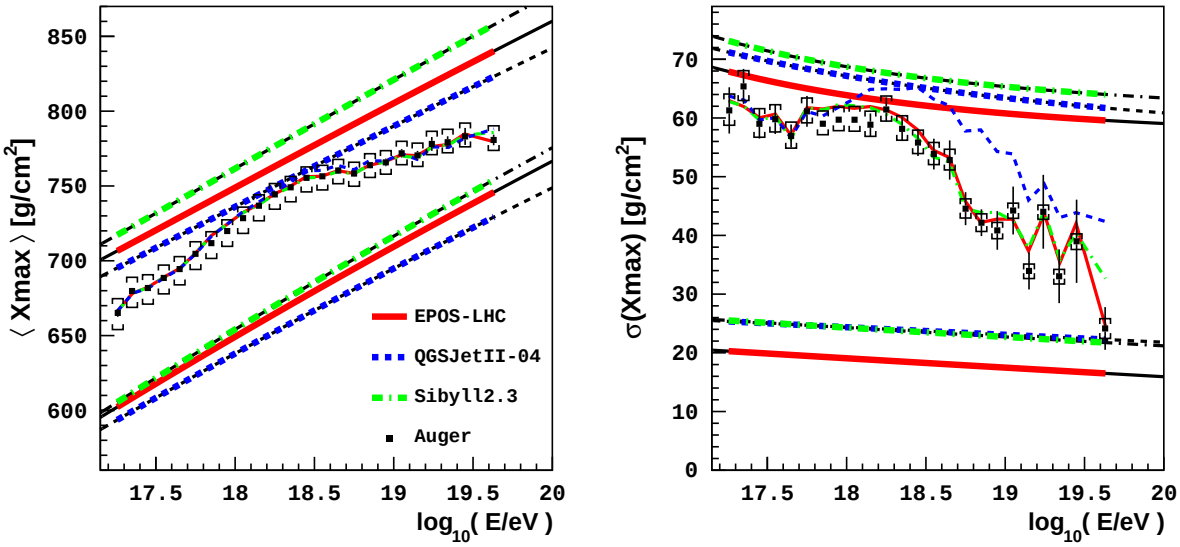
Figure 8.1: Fitting only the mass fractions of our parameterisations to FD and HEAT $X_{\max}$ data measured by the Pierre Auger Observatory. The fitted mass composition and p-values for each fitted model are shown.

anisotropy predict protons below $10^{18.5}$ eV are most likely of extragalactic origin [242]. If cosmic rays are assumed to transition from Galactic to extragalactic around the ankle [243], the proton fraction predicted by these fits is quite large.

For a good fit over the whole energy range, the p-values are expected to be uniformly distributed between 0 and 1 over the energy range. Given 24 energy bins are fitted, in the ideal scenario we would expect 2.4 (10%) of the p-values on average to be below the 0.1 line. Smaller p-values indicate a poorer fit. The p-values in Figure 8.1 indicate the mass fraction fits of the EPOS-LHC and Sibyll2.3 parameterisations are a better fit to the data than the QGSJetII-04



(a) The first two moments of the ln A distribution estimated from the fitted fractions.



(b) The black lines show the $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ initially predicted by the $X_{\max}$ parameterisations for proton and iron. The red, blue and green lines show the $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ from the fits of the $X_{\max}$ parameterisations, as well as the proton and iron $X_{\max}$ moments resulting from the fit (which are unchanged as only the mass fractions were fitted).

Figure 8.2: Fitting only the mass fractions of our parameterisations to FD and HEAT $X_{\max}$ data measured by the Pierre Auger Observatory.

mass fraction fit. The QGSJetII-04 fit consists of 18 p-values below 0.1, many more than the EPOS-LHC and Sibyll2.3 fits with 9 and 12 respectively. The large number of p-values below 0.1 for all of the fitted parameterisations indicates the mass fraction fits were unable to find a combination of p, He, N and Fe that described the details of the $X_{\max}$ distributions over the energy range, given the particle interaction predictions of the hadronic models utilised.

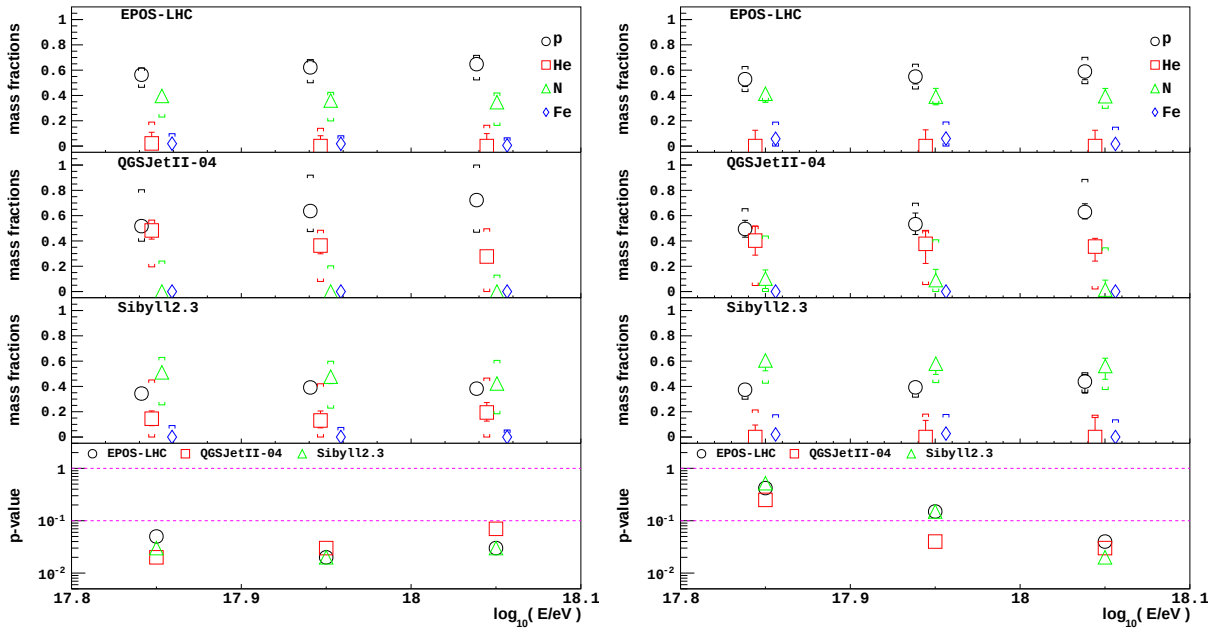
The p-values from the fits of the EPOS-LHC and Sibyll2.3 parameterisations appear to follow each other with energy. As the predicted $X_{\max}$ parameterisations of the EPOS-LHC and Sibyll2.3 models consist of significant differences, it is unexpected that energy bins which appear to be well fitted (large p-value) by the EPOS-LHC model, tend to also be well fitted by the Sibyll2.3 model, and similarly these two models appear to fit the same energy bins poorly.

Figure 8.2 shows the estimated ln A moments from the fits of the mass fractions, and the $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ of the fitted parameterisations compared to the first two $X_{\max}$ moments of the Auger $X_{\max}$ distributions [148]. In terms of the shape parameters, for a particular mass, $\langle X_{\max}\rangle = t_0 + \lambda$, and $\sigma(X_{\max}) = \sqrt{\sigma^2 + \lambda^2}$.

The predicted proton and iron $X_{\max}$ moments from the fit are unchanged from the initial predictions as the shape parameter coefficients were not fitted. The total $\langle X_{\max}\rangle$ of the fitted

parameterisations is consistent with the measured data, whereas the total $\sigma(X_{\max})$ of the fitted parameterisations is not as consistent, especially for the QGSJetII-04 fit. Figure D.2 in Appendix D.1 shows that around the mode of the measured distributions, the fitted QGSJetII-04 $X_{\max}$ parameterisation underestimates the data, particularly at higher energies. As the parameterisation is normalised to the number of events, the fitted parameterisation is therefore overestimating the number of events away from the mode. The poor reproduction of the measured data is reflected in the p-values of the QGSJetII-04 fit.

When fitting only the mass fractions, the mass composition fitted in each energy bin is independent of the fits to other energy bins. Fitting just the mass fractions to only the FD $X_{\max}$ data, or only the HEAT $X_{\max}$ data, results in a reconstructed mass composition that is the same as the mass fraction fit of the combined FD and HEAT $X_{\max}$ data for the energy bins where HEAT data, or FD data, respectively, are not present. Figure 8.3 shows the results of fitting the mass fractions to only the FD data, and only the HEAT data, in the energy range where the FD and HEAT data overlap. As the measured HEAT data is appropriately shifted to align with the FD data set, the fitted mass composition is consistent.



(a) FD fit.

(b) HEAT fit.

Figure 8.3: Fitting the mass fractions of our parameterisations to only the FD $X_{\max}$ data (left), or only the HEAT $X_{\max}$ data (right). The mass composition of the energy bins where there is no overlap between the FD and HEAT $X_{\max}$ data sets are not shown, as the results are identical to Figure 8.1.

8.1.2 $t_{0_{\text{norm}}}$, σ_{norm} and mass fraction fit of the combined Auger FD and HEAT $X_{\max}$ data

Figure 8.4 displays the results from fitting the mass fractions, and the coefficients $t_{0_{\text{norm}}}$ and σ_{norm}, of our EPOS-LHC, QGSJetII-04 and Sibyll2.3 parameterisations to the combined FD and HEAT $X_{\max}$ data. The systematic errors in the reconstructed mass are significantly reduced by fitting $t_{0_{\text{norm}}}$ and σ_{norm}, as systematic offsets in the $X_{\max}$ values and the $X_{\max}$ resolution can be compensated for by $t_{0_{\text{norm}}}$ and σ_{norm} respectively (see Section 5.3). The statistical errors remain similar to the fit of only the mass fractions. The fits of the parameterisations to the $X_{\max}$ distributions are shown in Appendix D.2.

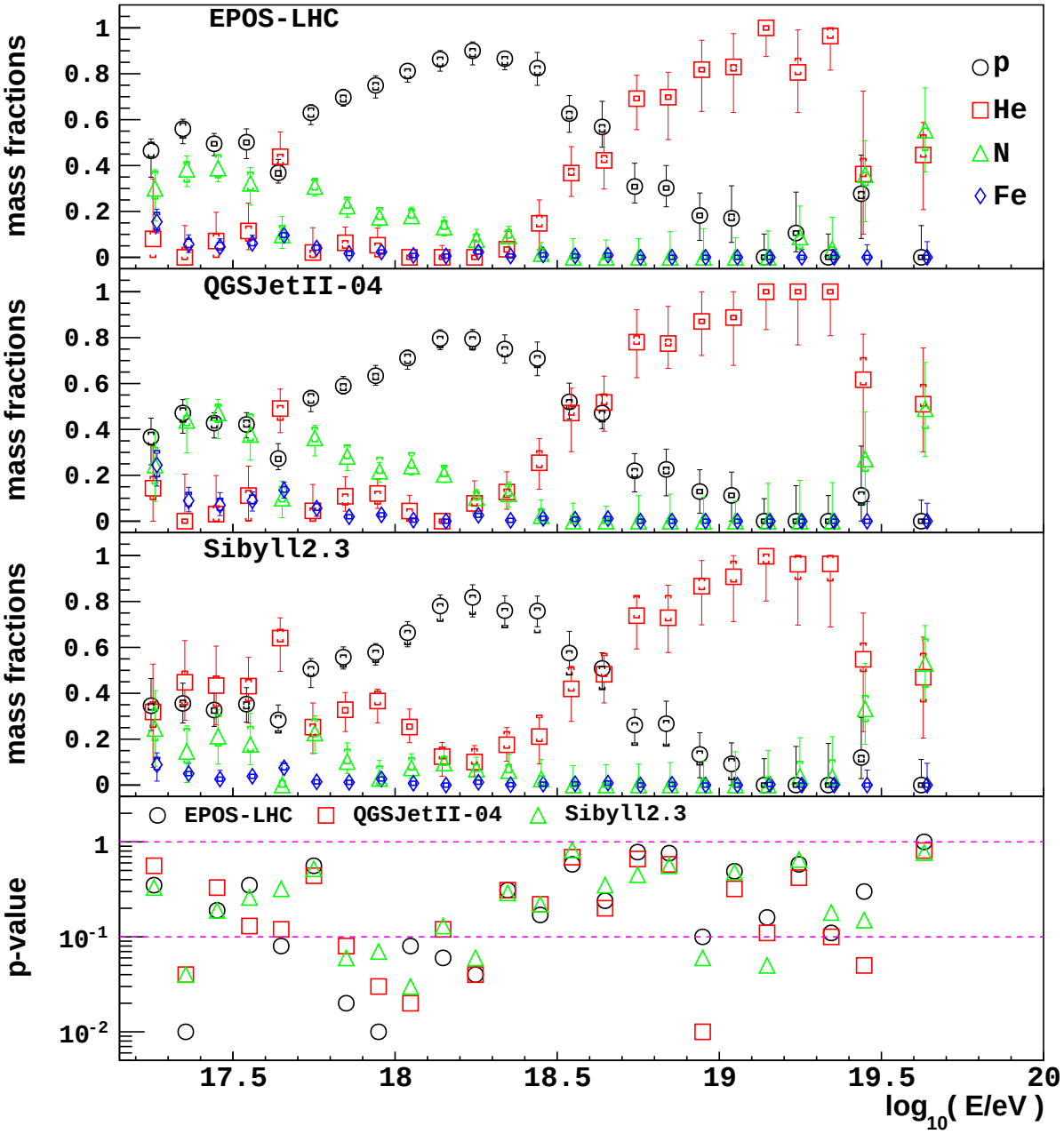
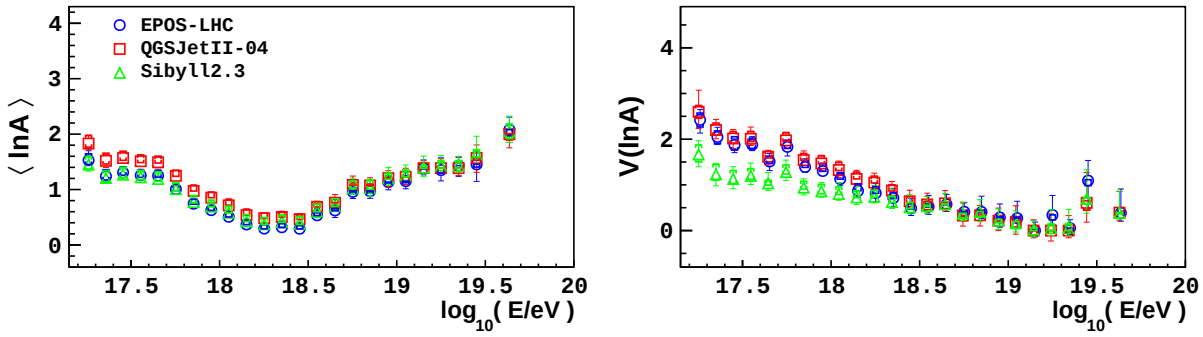
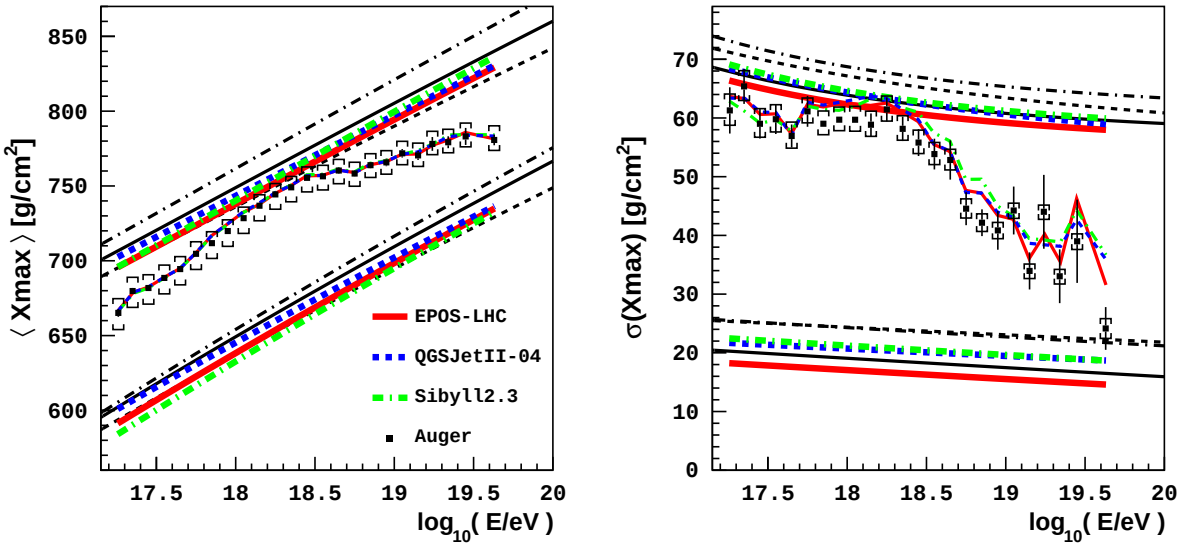


Figure 8.4: Fitting $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions of our parameterisations to the combined FD and HEAT $X_{\max}$ data measured by the Pierre Auger Observatory. The fitted mass fractions and p-values for each fitted model are shown.

Fitting $t_{0_{\text{norm}}}$ and σ_{norm} of the EPOS-LHC, QGSJetII-04 and Sibyll2.3 parameterisations results in a remarkably consistent mass composition, particularly above $10^{18.4}$ eV. The fits of the three parameterisations suggest the mass composition gradually transitions to a proton dominant composition around $10^{18.2}$ eV, which then transitions to helium dominant around $10^{19.2}$ eV. A potentially growing nitrogen composition is hinted around $10^{19.5}$ eV, with more data required at the highest energies to confirm this. Apart from the nitrogen component hinted at the highest energies, the composition above 10^{18} eV resulting from the $t_{0_{\text{norm}}}$ and σ_{norm} coefficient fit is very similar to the composition predicted by the QGSJetII-04 fit of only the mass fractions.



(a) First two moments of the ln A distribution estimated from the fitted fractions.



(b) The black lines show the $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ initially predicted by the $X_{\max}$ parameterisations for proton and iron. The red, blue and green lines show the new predictions for the $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ from the fits of $t_{0_{\text{norm}}}$ and σ_{norm}.

Figure 8.5: Fitting $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions of our parameterisations to FD and HEAT $X_{\max}$ data measured by the Pierre Auger Observatory.

EPOS-LHC	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\text{norm}}}$	692.5 (stat.)$^{+2}_{-1}$ (sys.)$^{+8}_{-8}$	686.1	669.2	639.6
fitted σ_{norm}	16.4 (stat.)$^{+1}_{-1}$ (sys.)$^{+1}_{-2}$	17.9	14.5	9.9
QGSJetII-04	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\text{norm}}}$	695.2 (stat.)$^{+3}_{-1}$ (sys.)$^{+8}_{-8}$	686.7	667.5	641.8
fitted σ_{norm}	15.9 (stat.)$^{+1}_{-2}$ (sys.)$^{+2}_{-2}$	17.1	14.7	10.6
Sibyll2.3	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\text{norm}}}$	693.0 (stat.)$^{+3}_{-1}$ (sys.)$^{+8}_{-8}$	679.6	656.8	628.8
fitted σ_{norm}	16.1 (stat.)$^{+2}_{-2}$ (sys.)$^{+2}_{-2}$	13.8	11.2	8.1

Table 8.1: The $t_{0_{\text{norm}}}$ and σ_{norm} coefficients fitted to the Auger FD and HEAT $X_{\max}$ data, assuming a normalisation energy of $E_0 = 10^{18.24}$ eV.

The main difference between the predicted composition of the three fits is the proportion of nitrogen below $10^{18.4}$ eV. The EPOS-LHC and QGSJetII-04 fits predict a nitrogen component that gradually declines from $10^{17.3}$ eV to $10^{18.4}$ eV, while the Sibyll2.3 fit predicts a minimal nitrogen component and instead a helium component which declines over this energy range. Given the similarity between the EPOS-LHC and QGSJetII-04 parameterisations when $t_{0_{\text{norm}}}$ and σ_{norm} is allowed to shift (see Section 5.2.1.2), the consistency of the EPOS-LHC and QGSJetII-04 results is unsurprising.

As seen in the fit of only the mass fractions, the ankle coincides with the mass composition being dominated by protons, which is interesting as the ankle is commonly assumed to be the point of transition from Galatic to extragalactic cosmic rays [243]. Also consistent with the fit of only the mass fractions, the two-coefficient fit predicts a transition towards heavier cosmic rays with increasing energy.

The data fitted in Section 5.2.1.2 was based on the results of fitting only the mass fractions to the Auger data. Of the three sets of mock data fitted, the QGSJetII-04 mock data was of a composition most similar to the results of the fits in this section. At energies above 10^{19} eV, the fits of the QGSJetII-04 mock data with the QGSJetII-04 parameterisation resulted in absolute systematic offsets as large as 40% in the median reconstructed mass composition from the true mass. These large offsets are due to primaries that are adjacent to the true mass being interpreted as the source of the data. At energies above $10^{18.2}$ eV however, the $\sigma(X_{\max})$ of that QGSJetII-04 mock data is much larger than the $\sigma(X_{\max})$ of the Auger data (see Figure 8.2b). The true primary $X_{\max}$ distributions of that mock data are wider than the distributions predicted from the $t_{0_{\text{norm}}}$ and σ_{norm} QGSJetII-04 fit of the Auger data in this section, because the fitted σ_{norm} is smaller than the initial QGSJetII-04 value (see Table 8.1, and compare with Table 4.1). In particular, the helium data from the initial QGSJetII-04 prediction is significantly wider than the QGSJetII-04 helium distribution resulting from the $t_{0_{\text{norm}}}$ and σ_{norm} fit. Therefore accurately reconstructing the mass of that mock data will be more difficult than fitting mock Auger data based on the $t_{0_{\text{norm}}}$ and σ_{norm} fit results of this section.

Mock data was generated based on the fit results of this section, and then the mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm} of the three parameterisations were fitted to this mock data. The results are displayed in Appendix E. When the correct model was fitted to the data, at energies above 10^{19} eV the absolute offsets in the median mass fractions are less than 20% from the true mass. The mass fraction systematic offsets are smaller than those of the QGSJetII-04 mock data fits in Section 5.2.1.2.

The fits to the Auger data using the three parameterisations (based on EPOS-LHC, QGSJetII-04 and Sibyll2.3) all predict a dominant helium component above 10^{19} eV, with varying amounts of nitrogen above $10^{19.4}$ eV. As the three models are consistent, we can be confident that a substantial helium component exists above 10^{19} eV, assuming the shape coefficients fixed to the data are reasonably correct.

The predicted helium abundance is promising for anisotropy studies. Figure 8.6 displays the predicted length scale of energy losses (propagation length) due to photo-nuclear processes in the CMB and IRB (infrared background) for various nuclei at the source. The figure shows that at energies between 10^{19} eV to 10^{20} eV, helium nuclei lose energy much more rapidly than other nuclei. If a particular source is producing helium cosmic rays, contributing to the significant abundance of helium above 10^{19} eV at Earth, the source is predicted to be nearby. The abundance of helium could also be caused by heavier nuclei disintegrating to lighter nuclei as they propagate to Earth.

Figure 8.5a displays the moments of the $\ln A$ distribution estimated from the fitted mass fractions. The estimated $\ln A$ moments from the three fits are very consistent. The first two moments of the Auger $X_{\max}$ distributions, and the $X_{\max}$ parameterisation predictions of the moments for proton and iron before and after the fit, are shown in Figure 8.5b. Compared to Figure 8.2b where only the mass fractions were fitted, the fits of $t_{0_{\text{norm}}}$ and σ_{norm} result

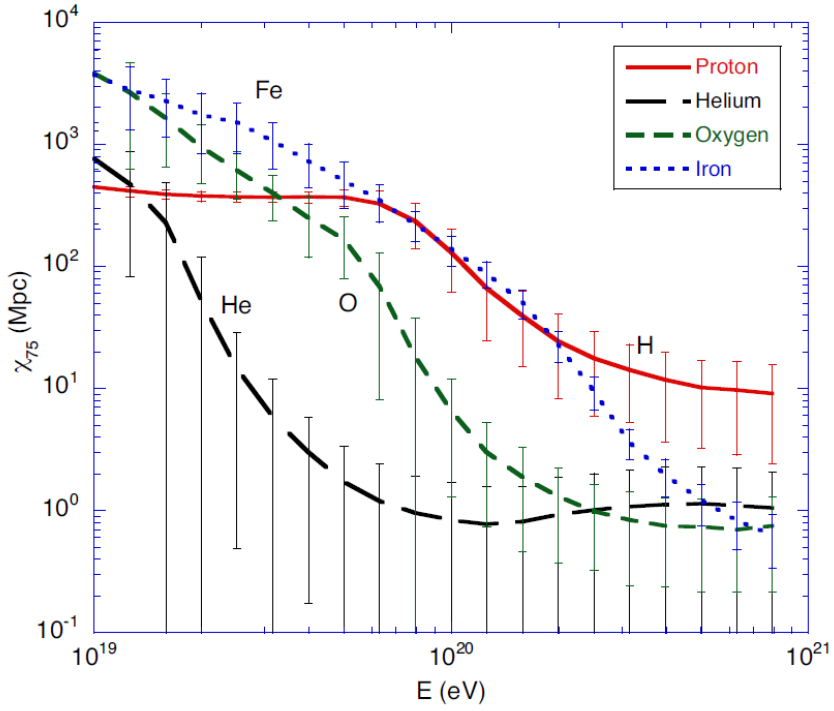


Figure 8.6: $\chi_{75}(E)$ is the propagation length after which a nucleus of initial energy E has lost 25% of its energy to photo-nuclear processes in the CMB and IRB (including photo-pair production). The error bars indicate the dispersion due to the stochastic propagation. From [244].

in a slightly better reproduction of the $\langle X_{\max} \rangle$ of the data. The EPOS-LHC and Sibyll2.3 fits do not reproduce the $\sigma(X_{\max})$ of the data better than the mass fraction only fit, but the reproduction from the QGSJetII-04 $t_{0_{\text{norm}}}$ and σ_{norm} fit is significantly better. The reproduction of the $\sigma(X_{\max})$ of the data may be improved by applying a non-flat acceptance in the central region of the $X_{\max}$ distributions, instead of assuming a flat acceptance in the central region. In terms of reconstructing the correct mass composition, replicating the $\sigma(X_{\max})$ of the data is not as important as replicating the $\langle X_{\max} \rangle$.

Figure 8.5b shows that the $t_{0_{\text{norm}}}$ and σ_{norm} fits reduce the difference between the $X_{\max}$ moment predictions of the EPOS-LHC, QGSJetII-04 and Sibyll2.3 hadronic models. The EPOS-LHC and Sibyll2.3 parameterisation fits result in values of $t_{0_{\text{norm}}}$ which adjust the EPOS-LHC and Sibyll2.3 $\langle X_{\max} \rangle$ prediction towards the initial QGSJetII-04 prediction, while the σ_{norm} values of the QGSJetII-04 and Sibyll2.3 parameterisation fits are adjusted such that the QGSJetII-04 and Sibyll2.3 $\sigma(X_{\max})$ predictions move towards the initial EPOS-LHC $\sigma(X_{\max})$ predictions. The fitted composition is quite sensitive to the $\langle X_{\max} \rangle$ predictions. Given the $t_{0_{\text{norm}}}$ and σ_{norm} fits of the three models predict a $\langle X_{\max} \rangle$ similar to the initial QGSJetII-04 prediction, it is unsurprising that the inferred mass composition is quite similar to the composition that resulted from the QGSJetII-04 fit of only the mass fractions. Our EPOS-LHC, QGSJetII-04 and Sibyll2.3 models have slightly different predictions for how the shape parameters change with mass and energy, but despite this there is strong agreement on the mass composition of the data when we allow their values for $t_{0_{\text{norm}}}$ and σ_{norm} to be fitted to the data.

The $t_{0_{\text{norm}}}$ and σ_{norm} values fitted to the data are shown in Table 8.1. The initial values of the shape parameter coefficients for proton, helium, nitrogen and iron primaries can be found in Table 4.1 (assuming a normalisation energy of $E_0 = 10^{18.24}$ eV). The fitted proton values of $t_{0_{\text{norm}}}$ and σ_{norm} agree between the models. However, for heavier nuclei Sibyll2.3 is in disagreement with EPOS-LHC and QGSJetII-04, as the predicted separation of proton from heavier nuclei in $t_{0_{\text{norm}}}$ and σ_{norm} is much larger for Sibyll2.3, and this separation is conserved in the fit. A shift in

$t_{0_{\text{norm}}}$ and/or σ_{norm} may be due to a model inadequately describing nature, systematics in the measurement of $X_{\max}$, or a combination of both factors.

The first two $X_{\max}$ moments can be parameterised as a function of the first two moments of the $\ln A$ distribution [146]. Using the values of $t_{0_{\text{norm}}}$ and σ_{norm} fitted to the data, the parameters of the equations in [146] have been determined, and are shown in Table F.1 and F.2 of Appendix F.1.

The statistical and systematic errors in the estimated $\langle X_{\max} \rangle$ for proton or iron over the energy range are the same as the statistical and systematic errors in the fitted value of $t_{0_{\text{norm}}}$. For $\sigma(X_{\max})$, the statistical error and systematic error is less than $0.7\,\text{g/cm}^2$ for the three fits.

Fitting $t_{0_{\text{norm}}}$ and σ_{norm}, as opposed to only the mass fractions, drastically improves the goodness of the fit of the QGSJetII-04 and Sibyll2.3 $X_{\max}$ distributions. This is evident by comparing the p-values in Figures 8.4 and 8.1. For the fit of $t_{0_{\text{norm}}}$ and σ_{norm}, there are 8, 8 and 7 p-values below 0.1 for EPOS-LHC, QGSJetII-04 and Sibyll2.3 respectively, compared to 9, 18 and 12 for the mass fraction only fit of these respective models. Overall, the two-coefficient fit results in larger p-values, but these p-values are still poor as only 2.4 p-values on average are expected to be below 0.1 for a model. Comparing the three models, the p-values are reasonably consistent over the energy range, indicating the fitted parameterisations describe the details of the $X_{\max}$ distributions with similar capability in each energy bin. This is not surprising, as we have shown that fitting $t_{0_{\text{norm}}}$ and σ_{norm} reduces the main differences between the $X_{\max}$ parameterisations of the three models.

The fit of $t_{0_{\text{norm}}}$ and σ_{norm} has significantly improved the ability of the QGSJetII-04 parameterisation to describe the data. However, the difference between the estimated composition from the fit of only the mass fractions, compared to the fit of the two coefficients, is smallest for the QGSJetII-04 fit. Compared to $t_{0_{\text{norm}}}$, σ_{norm} has considerably less effect on the estimated composition, but fitting the correct σ_{norm} significantly improves the goodness of the fit. The $t_{0_{\text{norm}}}$ and σ_{norm} fits of QGSJetII-04 and Sibyll2.3 resulted in a significant change in σ_{norm} from the initial predictions, resulting in a better description of the data and therefore the number of p-values below 0.1 reduced considerably compared to the mass fraction only fit of these models.

For the three models, the p-values between $10^{17.8}\,\text{eV}$ to $10^{18.3}\,\text{eV}$ appear to be consistently small, indicating a poor fit. This could be due to the composition being dominated by protons. Chapter 5 demonstrated that proton distributions are more difficult to fit as they have the largest spread. The proton $X_{\max}$ parameterisations are also a less accurate description of their corresponding MC data (see Figure 4.5), and therefore may not describe nature accurately enough. Alternatively, the hadronic model predictions may not describe this energy range as adequately, or the model predictions across the full energy range may be incorrect resulting in a fit that prioritises describing some energy bins better at the expense of other energy bins. Another possibility is that additional corrections need to be applied to the HEAT $X_{\max}$ data to sufficiently align the HEAT $X_{\max}$ distributions with the FD $X_{\max}$ distributions. The p-values are poor between $10^{17.8}\,\text{eV}$ and $10^{18.1}\,\text{eV}$, the energy range where the FD and HEAT data overlap. However, the $t_{0_{\text{norm}}}$ and σ_{norm} fit of only the FD data (see Section 8.2.1) results in similarly poor p-values over this energy range (compare Figures 8.4 and 8.13).

8.1.3 $t_{0_{\text{norm}}}$, σ_{norm}, λ_{norm} and mass fraction fit of the combined Auger FD and HEAT $X_{\max}$ data

Figure 8.7 displays the results from fitting the mass fractions, $t_{0_{\text{norm}}}$, σ_{norm} and λ_{norm} of our EPOS-LHC, QGSJetII-04 and Sibyll2.3 $X_{\max}$ parameterisations to FD and HEAT $X_{\max}$ data.

This three-coefficient fit reconstructs a mass composition which is very similar to that of the two-coefficient fit of $t_{0_{\text{norm}}}$ and σ_{norm}. Table 8.2 shows that the values of $t_{0_{\text{norm}}}$ and σ_{norm} fitted in the three-coefficient fit are consistent with the fitted values from the two-coefficient fit (see Table 8.1). The parameterisations fit a proton λ_{norm} of $\sim 58 \pm 2(stat) \pm 1(sys)\ \text{g/cm}^2$. Taking the errors into account, this value is consistent with the Auger measurement of the proton exponential tail, $\Lambda_\eta = 55.8 \pm 2.3(stat) \pm 1.6(sys)\ \text{g/cm}^2$, which was used in the determination

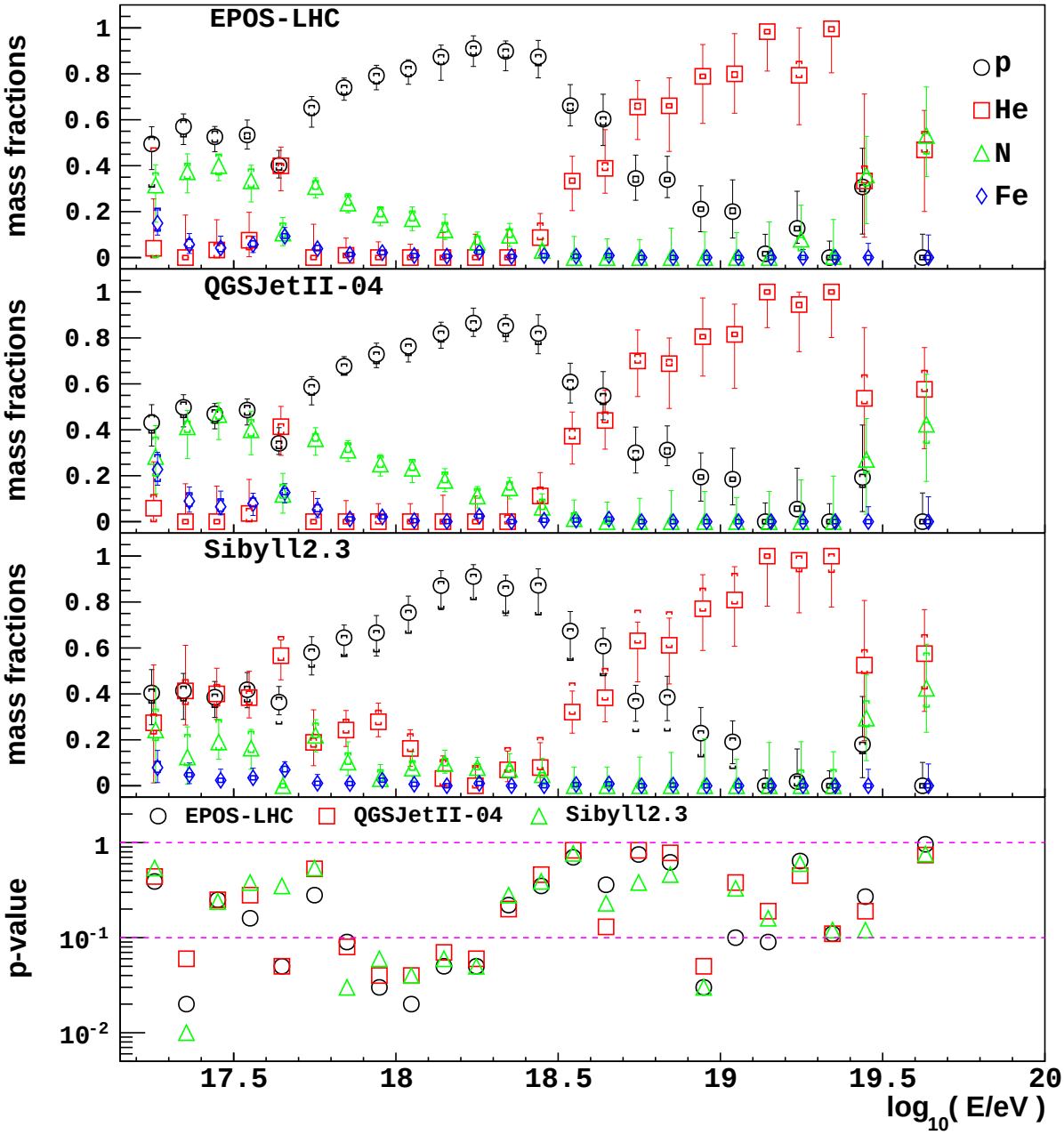
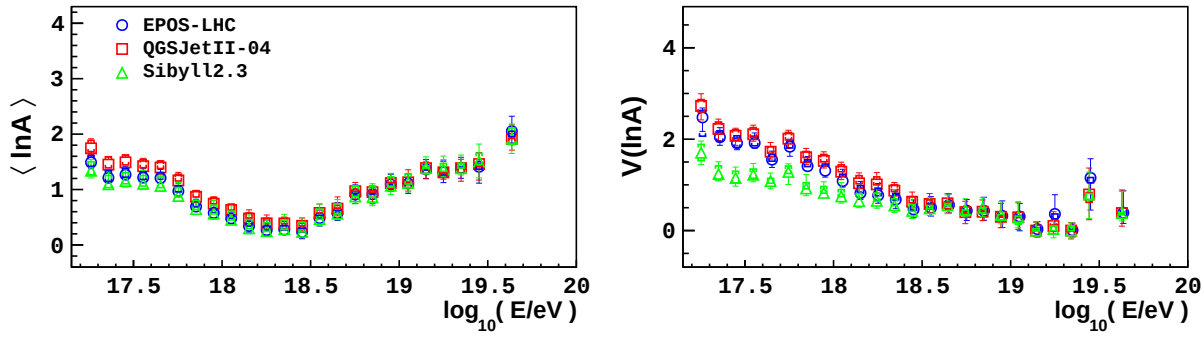


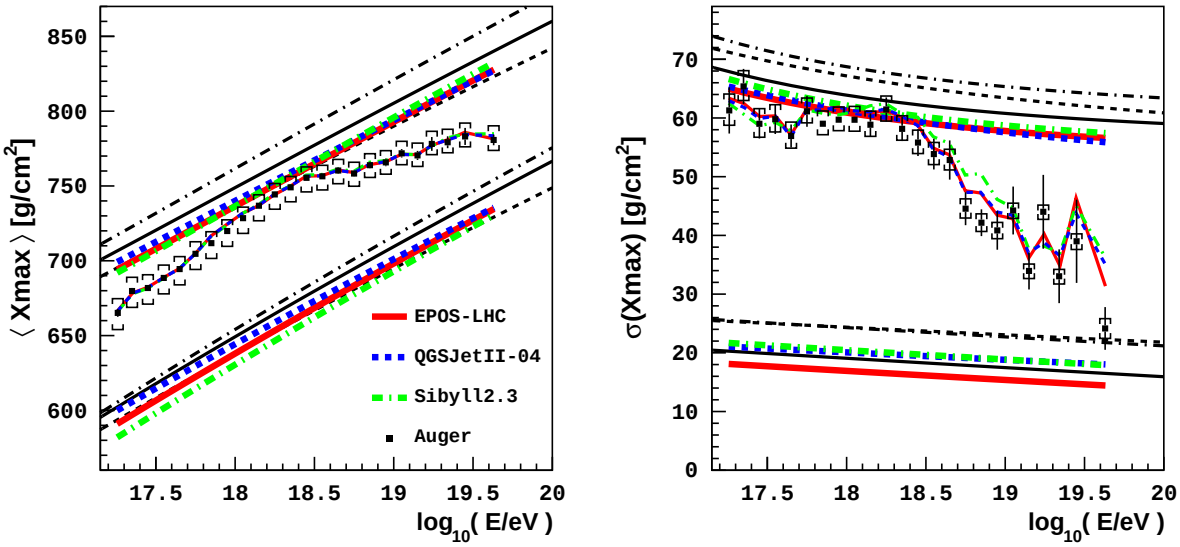
Figure 8.7: Fitting $t_{0_{\text{norm}}}$, σ_{norm}, λ_{norm} and the mass fractions of our parameterisations to the combined FD and HEAT $X_{\max}$ data measured by the Pierre Auger Observatory. The fitted mass fractions and p-values for each fitted model are shown.

of the proton-air interaction cross-section [229]. The fitted proton λ_{norm} is $\sim 2-3\,\text{g/cm}^2$ lower than the initial predictions of the parameterisations. As the fitted $t_{0_{\text{norm}}}$ and σ_{norm} values are consistent between the two and three-coefficient fits, and the mass composition reconstruction of the fits is consistent, we can see that this shift in λ_{norm} does not significantly impact the mass composition reconstruction.

The $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ predicted by the parameterisation fits is sensitive to the fitted value of λ_{norm}. The systematic errors in Table 8.2 shows that systematic shifts in the $X_{\max}$ data are still absorbed primarily by $t_{0_{\text{norm}}}$.



(a) First two moments of the ln A distribution estimated from the fitted fractions.



(b) The red, blue and green lines show the new predictions for the $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ from the fits of $t_{0_{\text{norm}}}$, σ_{norm} and λ_{norm}.

Figure 8.8: Fitting $t_{0_{\text{norm}}}$, σ_{norm}, λ_{norm} and the mass fractions of our parameterisations to FD and HEAT $X_{\max}$ data measured by the Pierre Auger Observatory.

For the fit of $t_{0_{\text{norm}}}$, σ_{norm} and λ_{norm}, there are 10, 8 and 7 p-values below 0.1 for EPOS-LHC, QGSJetII-04 and Sibyll2.3 respectively, compared to 8, 8 and 7 for the fit of $t_{0_{\text{norm}}}$ and σ_{norm}. As λ_{norm} has changed little from the initial model predictions, we do not expect a significant improvement in the quality of the fit. We still observe a set of low p-values for all models in the energy range of $10^{17.8}$ eV to $10^{18.3}$ eV.

The ln A moments estimated from the fitted fractions are shown in Figure 8.8a, and the $X_{\max}$ moment predictions are shown in Figure 8.8b. We see that the predicted proton $\langle X_{\max}\rangle$ of the three fitted parameterisations aligns with the initial proton prediction of the QGSJetII-04 parameterisation. To a lesser degree, the predicted iron $\langle X_{\max}\rangle$ of the three fitted parameterisations are aligned around the initial iron prediction of the QGSJetII-04 parameterisation. The shift in λ_{norm} has resulted in more agreement between the models in the predicted $\langle X_{\max}\rangle$ of proton, compared to the fit of only $t_{0_{\text{norm}}}$ and σ_{norm} (see 8.5b). Similarly, there is more agreement in the predicted proton $\sigma(X_{\max})$. The shift in λ_{norm} has resulted in a proton $\sigma(X_{\max})$ which is now considerably below the initial EPOS-LHC prediction, whereas in Figure 8.5b the fit of only $t_{0_{\text{norm}}}$ and σ_{norm} resulted in a proton $\sigma(X_{\max})$ which aligned with the initial EPOS-LHC prediction. At $10^{18.25}$ eV, the $\sigma(X_{\max})$ of the data is above the proton $\sigma(X_{\max})$ prediction of the fit, as both proton-like and nitrogen-like events are contributing to the distribution.

EPOS-LHC	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\text{norm}}}$	$692.5\ (\text{stat.})^{+4}_{-1}\ (\text{sys.})^{+8}_{-8}$	686.2	669.2	639.7
fitted σ_{norm}	$16.8\ (\text{stat.})^{+1}_{-1}\ (\text{sys.})^{+1}_{-2}$	18.3	14.8	10.2
fitted λ_{norm}	$57.6\ (\text{stat.})^{+1}_{-1}\ (\text{sys.})^{+1}_{-1}$	33.8	19.5	13.1
QGSJetII-04	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\text{norm}}}$	$695.1\ (\text{stat.})^{+3}_{-2}\ (\text{sys.})^{+8}_{-8}$	686.7	667.4	641.7
fitted σ_{norm}	$16.3\ (\text{stat.})^{+2}_{-1}\ (\text{sys.})^{+1}_{-2}$	17.6	15.1	10.8
fitted λ_{norm}	$58.0\ (\text{stat.})^{+1}_{-1}\ (\text{sys.})^{+1}_{-1}$	35.5	24.5	16.5
Sibyll2.3	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\text{norm}}}$	$691.7\ (\text{stat.})^{+6}_{-2}\ (\text{sys.})^{+8}_{-8}$	678.3	655.4	627.4
fitted σ_{norm}	$15.7\ (\text{stat.})^{+2}_{-3}\ (\text{sys.})^{+1}_{-2}$	13.5	10.9	7.9
fitted λ_{norm}	$59.0\ (\text{stat.})^{+2}_{-1}\ (\text{sys.})^{+1}_{-1}$	38.7	28.3	18.4

Table 8.2: The $t_{0_{\text{norm}}}$, σ_{norm} and λ_{norm} coefficients fitted to the Auger FD and HEAT $X_{\max}$ data, assuming a normalisation energy of $E_0 = 10^{18.24}$ eV.

8.1.4 $t_{0_{\text{norm}}}$, B, σ_{norm} and mass fraction fit of the combined Auger FD and HEAT $X_{\max}$ data

EPOS-LHC	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\text{norm}}}$	$688.9\ (\text{stat.})^{+2}_{-2}\ (\text{sys.})^{+8}_{-9}$	682.5	665.6	636.0
fitted B	$3414\ (\text{stat.})^{+121}_{-113}\ (\text{sys.})^{+81}_{-24}$	3396	3429	3484
fitted σ_{norm}	$16.9\ (\text{stat.})^{+1}_{-1}\ (\text{sys.})^{+1}_{-2}$	18.4	14.9	10.2
QGSJetII-04	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\text{norm}}}$	$691.4\ (\text{stat.})^{+1}_{-2}\ (\text{sys.})^{+8}_{-9}$	683.0	663.8	638.0
fitted B	$3470\ (\text{stat.})^{+116}_{-103}\ (\text{sys.})^{+118}_{-13}$	3436	3448	3486
fitted σ_{norm}	$15.7\ (\text{stat.})^{+1}_{-1}\ (\text{sys.})^{+2}_{-2}$	17.0	14.6	10.5
Sibyll2.3	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\text{norm}}}$	$695.3\ (\text{stat.})^{+2}_{-2}\ (\text{sys.})^{+8}_{-9}$	681.9	659.0	631.0
fitted B	$3499\ (\text{stat.})^{+99}_{-123}\ (\text{sys.})^{+73}_{-326}$	3539	3528	3548
fitted σ_{norm}	$16.4\ (\text{stat.})^{+2}_{-2}\ (\text{sys.})^{+2}_{-2}$	14.1	11.4	8.3

Table 8.3: The $t_{0_{\text{norm}}}$, B and σ_{norm} coefficients fitted to the Auger FD and HEAT $X_{\max}$ data, assuming a normalisation energy of $E_0 = 10^{18.24}$ eV.

Figure 8.9 displays the results from fitting the mass fractions, $t_{0_{\text{norm}}}$, B and σ_{norm} of our EPOS-LHC, QGSJetII-04 and Sibyll2.3 $X_{\max}$ parameterisations to FD and HEAT $X_{\max}$ data.

This three-coefficient fit results in a mass composition reconstruction that is consistent between the three models, but quite different to the previous coefficient fits where B was fixed to the data. Below $10^{18.5}$ eV, the nitrogen component from the previous fits has been replaced by a helium component. Proton still peaks around the ankle region, but now there is only a small helium peak around $10^{18.8}$ eV, which is quickly replaced by the growing nitrogen component. An

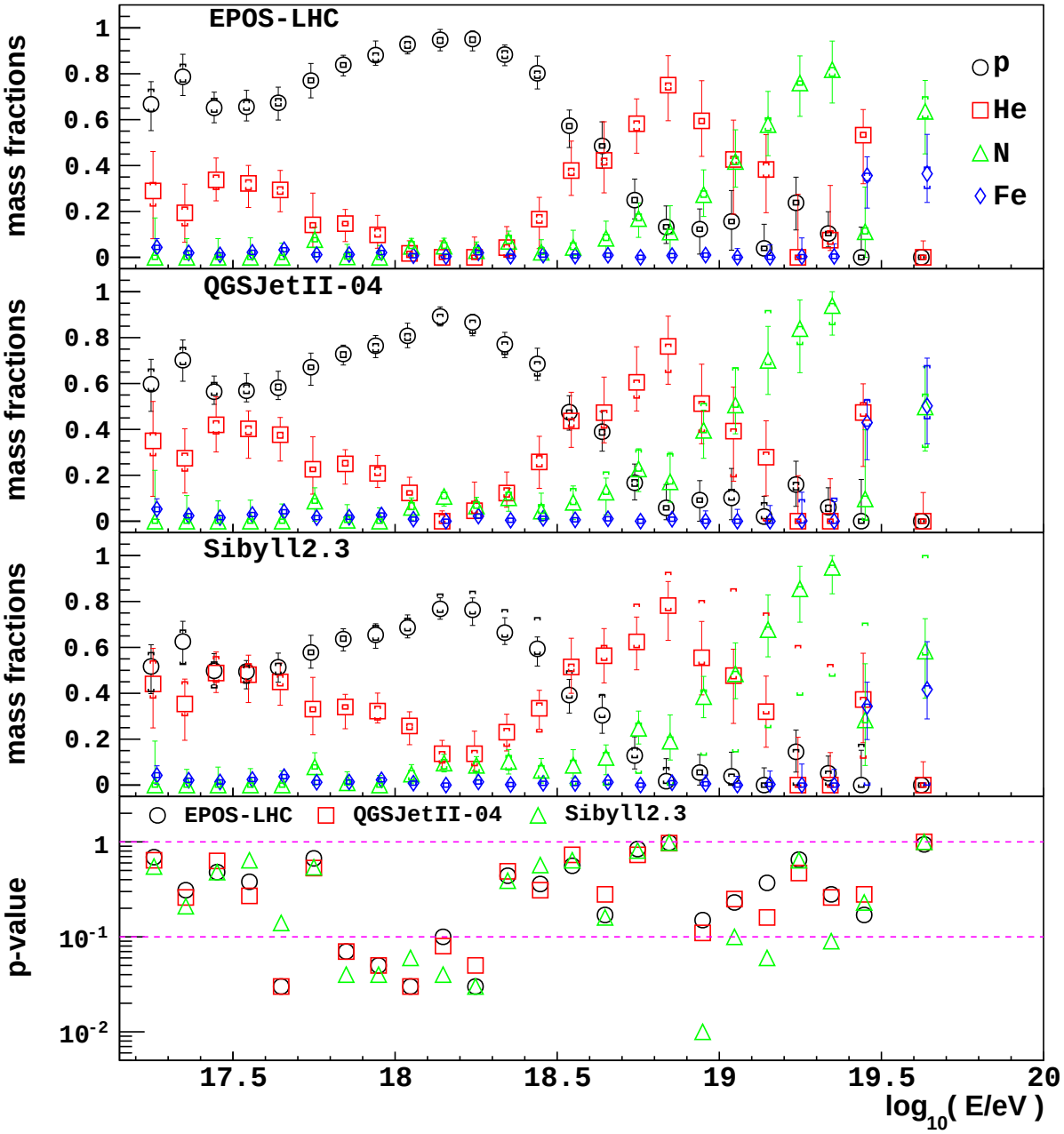
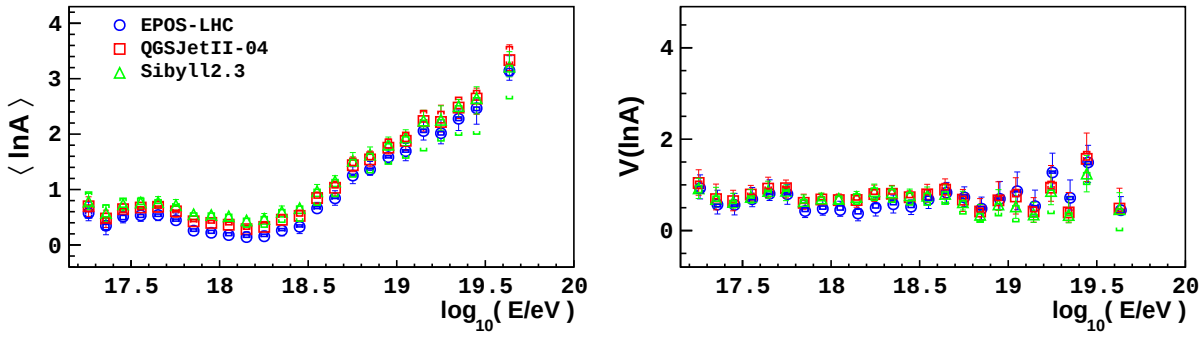


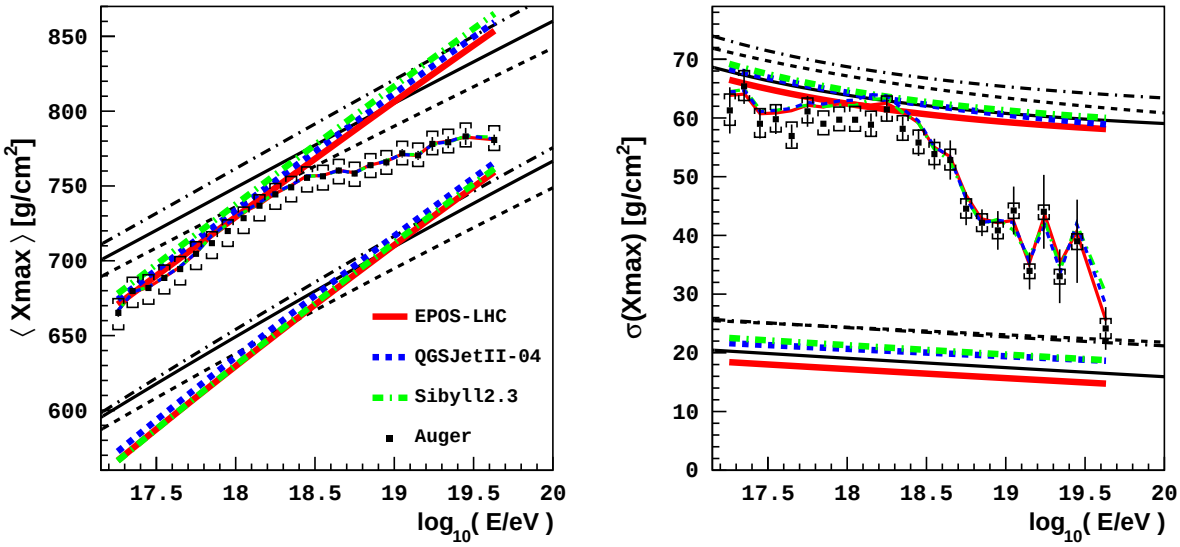
Figure 8.9: Fitting $t_{0_{\text{norm}}}$, B, σ_{norm} and the mass fractions of our parameterisations to the combined FD and HEAT $X_{\max}$ data measured by the Pierre Auger Observatory. The fitted mass fractions and p-values for each fitted model are shown.

iron component appears above $10^{19.4}$ eV. The mass composition transition is reminiscent of a Peters cycle.

For the fit of $t_{0_{\text{norm}}}$, B and σ_{norm}, there are 6, 6 and 9 p-values below 0.1 for EPOS-LHC, QGSJetII-04 and Sibyll2.3 respectively, compared to 8, 8 and 7 for the fit of $t_{0_{\text{norm}}}$ and σ_{norm}. Low p-values for all models in the energy range of $10^{17.8}$ eV to $10^{18.3}$ eV are still observed. By including B in the fit, the $X_{\max}$ parameterisations fitted to the data have changed considerably from the initial predictions, and are considerably different to those predicted by the $t_{0_{\text{norm}}}$ and σ_{norm} fit. However, the p-values of these fits have not improved significantly compared to those



(a) First two moments of the $\ln A$ distribution estimated from the fitted fractions.



(b) The red, blue and green lines show the new predictions for the $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ from the fits of $t_{0_{\text{norm}}}$, B and σ_{norm}.

Figure 8.10: Fitting $t_{0_{\text{norm}}}$, B, σ_{norm} and the mass fractions of our parameterisations to FD and HEAT $X_{\max}$ data measured by the Pierre Auger Observatory.

from the fit of $t_{0_{\text{norm}}}$ and σ_{norm}.

In the case of the Sibyll2.3 fits, the p-values from the fit of B are slightly worse, which could be solely due to statistical randomness in the p-values. However, depending on the contribution of each energy bin to the total log likelihood, a fit which results in a smaller total minimised log likelihood can at the same time have worse p-values overall.

Table 8.3 shows that the values of $t_{0_{\text{norm}}}$, B and σ_{norm} fitted in the three-coefficient fit. Compared to the fit of only $t_{0_{\text{norm}}}$ and σ_{norm} (see Table 8.1), the fitted values of σ_{norm} remain similar. Including B in the fit increases the disagreement between the models in the fitted values of $t_{0_{\text{norm}}}$. In Figure 8.10b, we see that the predicted proton and iron $\langle X_{\max}\rangle$ of the three-coefficient fits are consistent with each other. Compared to the initial predictions of the parameterisations, a significantly larger elongation rate of t_0 with energy has been fitted to the data. Between 10^{18} eV to 10^{19} eV, the rate of change of t_0 with energy is $\sim 80\,\text{g/cm}^2$, compared to the initial prediction of $\sim 60\,\text{g/cm}^2$. This increase directly translates to the predicted $\langle X_{\max}\rangle$ elongation rate of each nuclei. The predicted $\langle X_{\max}\rangle$ elongation rate of each nuclei from these fits would not align with the predictions at lower energies. The drastic elongation rate increase results in an estimated transition from lighter to heavier primaries that is much quicker with energy. The estimated $\ln A$ moments are shown in Figure 8.10a, demonstrating the rapid increase in the mean mass.

Figure 8.10b shows that above $10^{18.4}$ eV, the $\sigma(X_{\max})$ of the data is reproduced much better

than in previous fits, but below $10^{18.4}$ eV the reproduction is worse. The values of σ, λ and the mass fractions determine the total $\sigma(X_{\max})$ of the fitted parameterisation. The fitted value of σ_{norm} is very similar to previous coefficient fits, but the mass composition fitted above $10^{18.4}$ eV is heavier, and below $10^{18.4}$ eV it is lighter, than previous fits.

The analysis of the fit of $t_{0_{\text{norm}}}$, B and σ_{norm} with mock Auger distributions in Section 5.2.1.4 showed that this three-coefficient fit does not fit the coefficients as accurately as the fit of only $t_{0_{\text{norm}}}$ and σ_{norm}, due to the added degeneracy between the fitted parameters. Particularly at the highest energies where there is a lack of statistics, the reconstructed mass fractions are less accurate than the two-coefficient fit. Considering this, and the very large $\langle X_{\max}\rangle$ elongation rate fitted, we do not place much emphasis on the results of this three-coefficient fit. It is possible though that a significantly larger elongation rate was fitted because the hadronic model assumptions are considerably incorrect at least over some portion of the fitted energy range.

8.1.5 B, σ_{norm} and mass fraction fit of the combined Auger FD and HEAT $X_{\max}$ data

EPOS-LHC	Proton	Helium	Nitrogen	Iron
fitted B	$2486\ (\text{stat.})^{+23}_{-11}\ (\text{sys.})^{+59}_{-63}$	2491	2525	2599
fitted σ_{norm}	$16.6\ (\text{stat.})^{+1}_{-1}\ (\text{sys.})^{+1}_{-2}$	18.1	14.7	10.0
QGSJetII-04	Proton	Helium	Nitrogen	Iron
fitted B	$2515\ (\text{stat.})^{+22}_{-15}\ (\text{sys.})^{+53}_{-61}$	2523	2573	2644
fitted σ_{norm}	$15.7\ (\text{stat.})^{+1}_{-1}\ (\text{sys.})^{+2}_{-2}$	17.0	14.6	10.5
Sibyll2.3	Proton	Helium	Nitrogen	Iron
fitted B	$2710\ (\text{stat.})^{+25}_{-20}\ (\text{sys.})^{+67}_{-73}$	2749	2738	2758
fitted σ_{norm}	$26.2\ (\text{stat.})^{+2}_{-3}\ (\text{sys.})^{+2}_{-3}$	22.5	18.2	13.2

Table 8.4: The B and σ_{norm} coefficients fitted to the Auger FD and HEAT $X_{\max}$ data, assuming a t_0 normalisation energy of $E_0 = 10^{14}$ eV, and a σ and λ normalisation energy of $E_0 = 10^{18.24}$ eV.

Figures 8.11 and 8.12 and Table 8.4 displays the results from fitting the mass fractions, B and σ_{norm} of our EPOS-LHC, QGSJetII-04 and Sibyll2.3 parameterisations to FD and HEAT $X_{\max}$ data. This approach fits B in a more restrictive manner than the fit of $t_{0_{\text{norm}}}$, B and σ_{norm}, as the normalisation of t_0 is now fixed at 10^{14} eV. In the previous three-coefficient fit, the resulting $\langle X_{\max}\rangle$ was allowed to take any normalisation value and elongation rate within the limits, and therefore the range of potential $\langle X_{\max}\rangle$ predictions resulting from the fit was greater.

Fixing t_0 at 10^{14} eV and fitting B ensures that the fitted t_0 shape parameter prediction, and consequently the predicted $\langle X_{\max}\rangle$ rails for each primary, align with the predictions at a lower energy where the hadronic models are more certain. This approach assumes the hadronic model predictions at 10^{14} eV are correct, and relies on the functional form of t_0 being correct above 10^{14} eV. The possible t_0 parameterisations fitted to the data between 10^{17} eV and $10^{19.5}$ eV with this fit of B are similar to those possible from the $t_{0_{\text{norm}}}$ fit.

The initial t_0 coefficients of the EPOS-LHC and QGSJetII-04 parameterisations used in this section were determined by fitting CONEX v4r37 data from 10^{14} eV to $10^{19.5}$ eV, whereas the Sibyll2.3 parameterisation was determined by fitting CONEX v4r37 data from 10^{17} eV to $10^{19.5}$ eV, and then the resulting t_0 shape parameter functions were extrapolated to 10^{14} eV.

The EPOS-LHC and QGSJetII-04 fits of B and σ_{norm} result in a mass composition consistent with the results of the $t_{0_{\text{norm}}}$ and σ_{norm} fits. The Sibyll2.3 fit however is not consistent with the

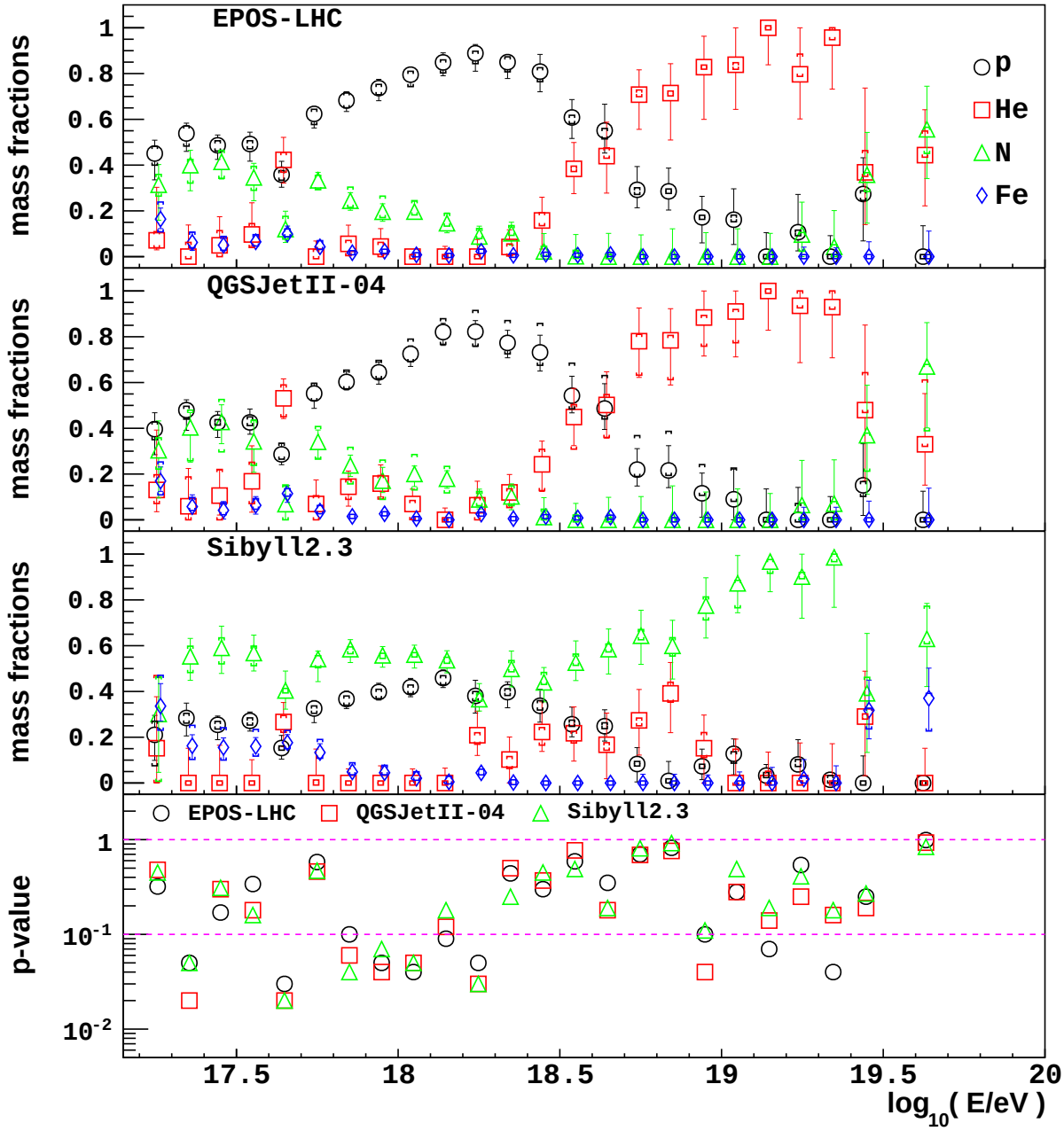
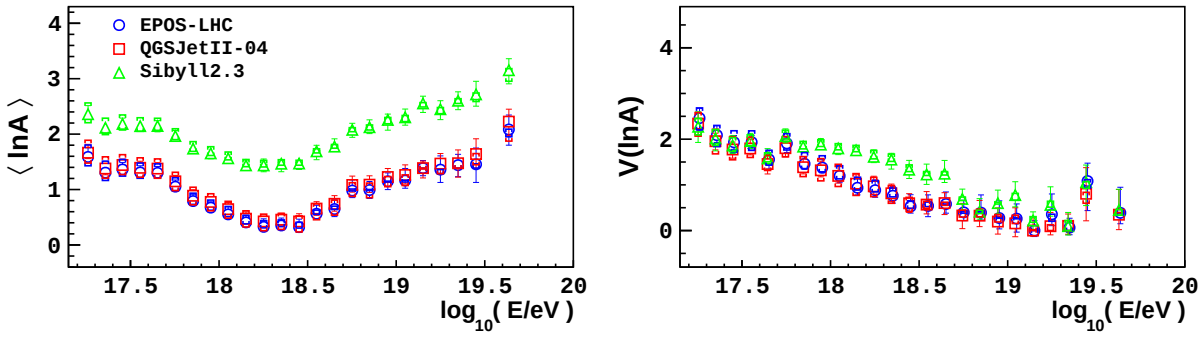
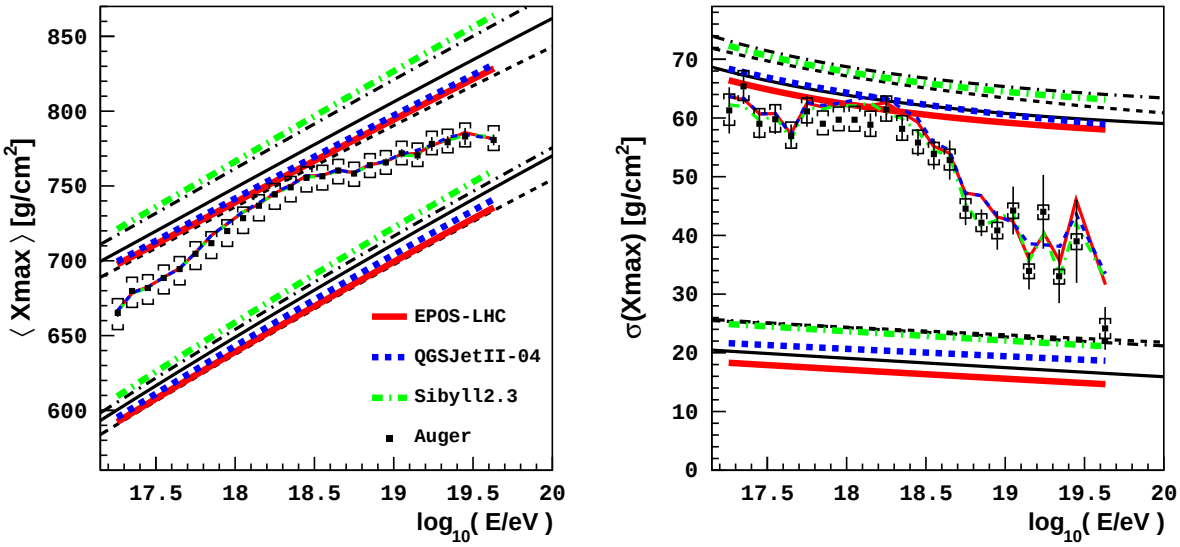


Figure 8.11: Fitting B, σ_{norm} and the mass fractions of our parameterisations to the combined FD and HEAT $X_{\max}$ data measured by the Pierre Auger Observatory. The fitted mass fractions and p-values for each fitted model are shown.

previous fits, instead predicting a mass composition dominated by nitrogen throughout. Table 8.4 shows the fitted value of B for Sibyll2.3 is much larger than the other models, resulting in a heavier composition as the Sibyll2.3 $\langle X_{\max}\rangle$ rails for each primary are shifted up. This is illustrated in Figure 8.12b. The resulting $X_{\max}$ moment predictions of the EPOS-LHC and QGSJetII-04 fits are similar to those from the fits of $t_{0_{\text{norm}}}$ and σ_{norm}. The predicted Sibyll2.3 $\langle X_{\max}\rangle$ for proton and iron illustrates that fitting B with t_0 constrained at 10^{14} eV is quite similar to fitting $t_{0_{\text{norm}}}$. Even though Sibyll2.3 fits a value of B greater than the other models, resulting in a larger $\langle X_{\max}\rangle$ elongation rate, over the energy range of interest, the effect on the predicted $\langle X_{\max}\rangle$ for proton



(a) First two moments of the ln A distribution estimated from the fitted fractions.



(b) The red, blue and green lines show the new predictions for the $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ from the fits of B and σ_{norm}.

Figure 8.12: Fitting B, σ_{norm} and the mass fractions of our parameterisations to FD and HEAT $X_{\max}$ data measured by the Pierre Auger Observatory.

and iron is similar to a shift in $t_{0_{\text{norm}}}$. The p-values of the three fits are similar to each other, despite the Sibyll2.3 solution being considerably different. The p-values are similar to those of the $t_{0_{\text{norm}}}$ and σ_{norm} fits.

The B and σ_{norm} parameter space scan of the FD and HEAT data, shown in Figure 6.8c of Section 6.4, indicates there is a secondary minimum for Sibyll2.3 which coincides with the deepest minimum of the EPOS-LHC and QGSJetII-04 fits. The secondary minimum however is not within 5σ of the deepest Sibyll2.3 minimum.

8.2 FD $X_{\max}$ data composition results

8.2.1 $t_{0_{\rm norm}}$, $\sigma_{\rm norm}$ and mass fraction fit of the Auger FD $X_{\max}$ data

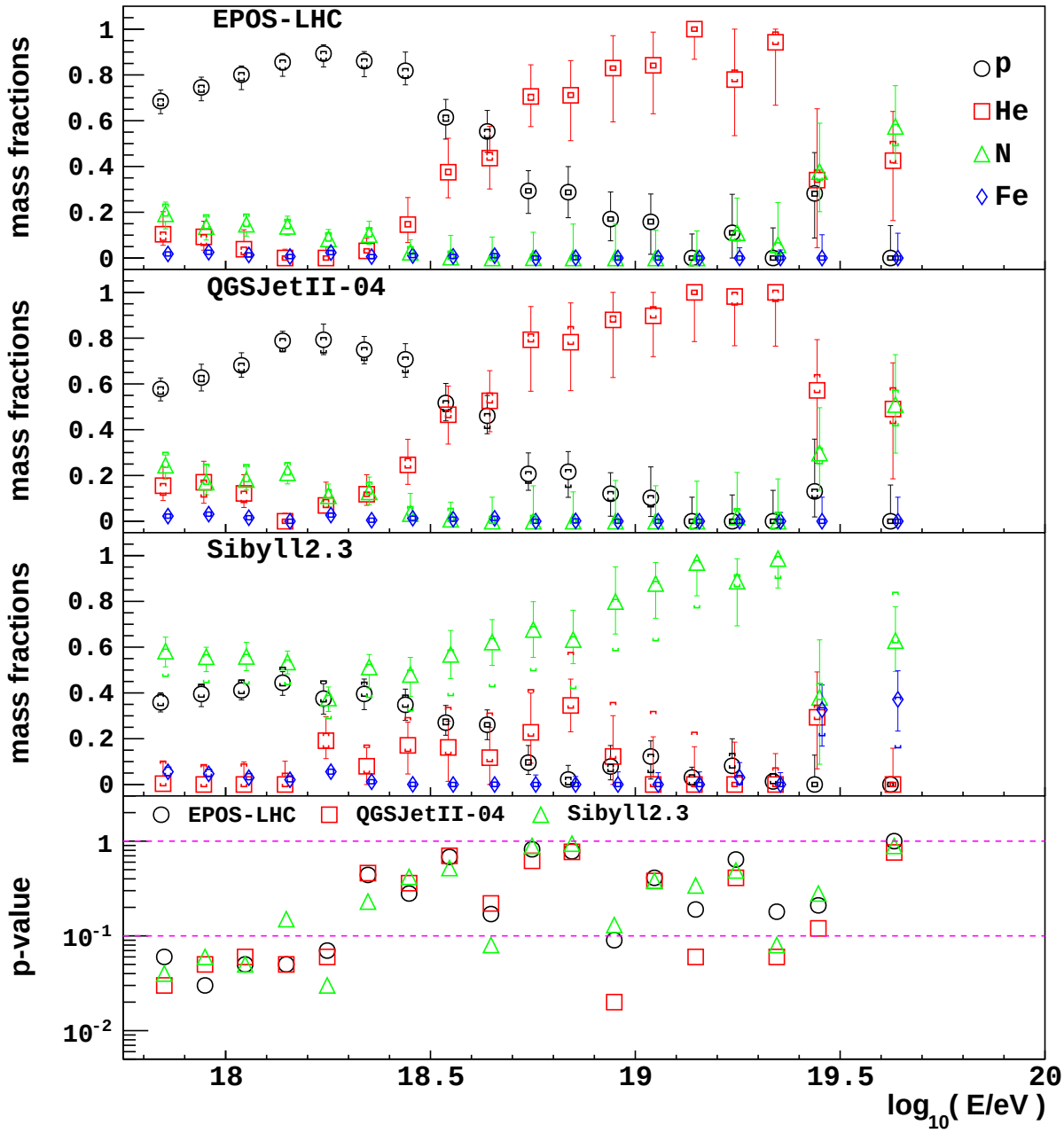
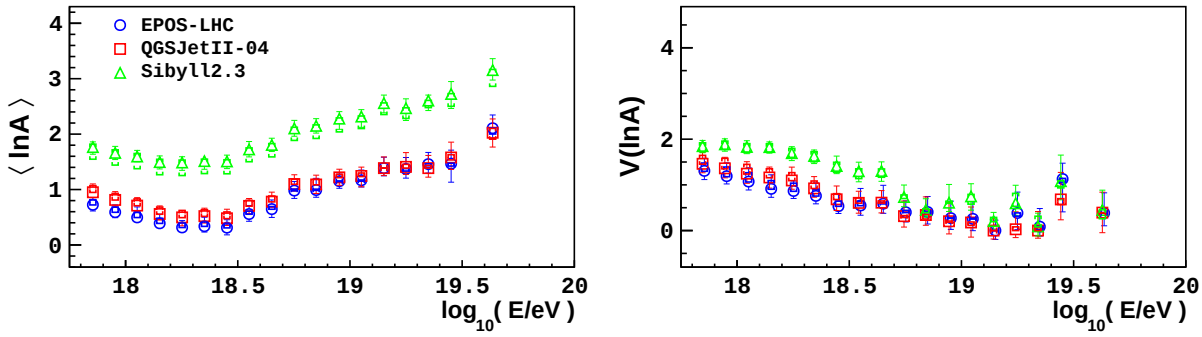


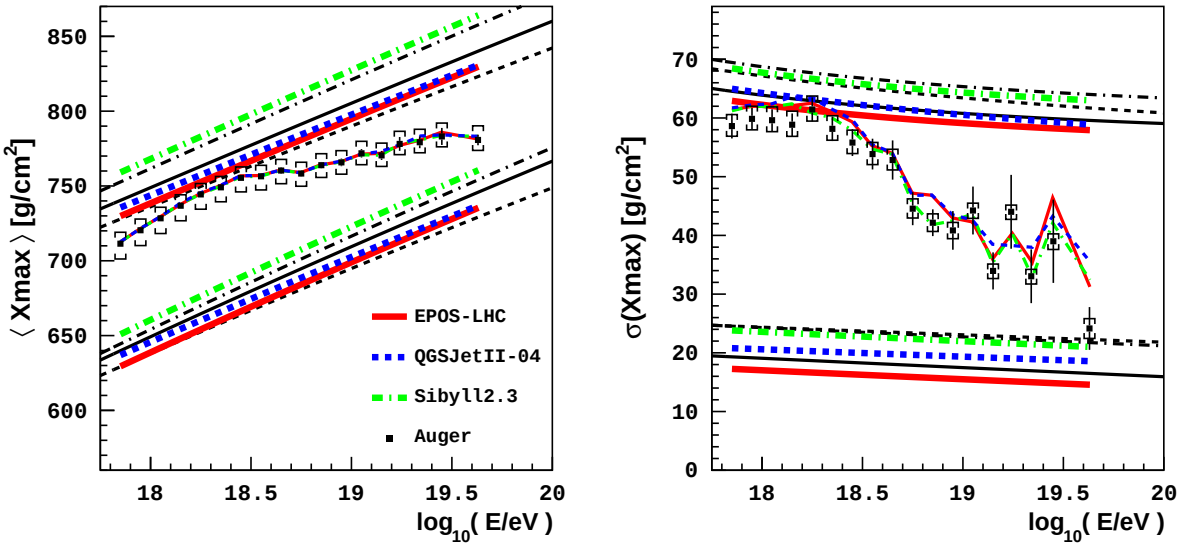
Figure 8.13: Fitting $t_{0_{\rm norm}}$, $\sigma_{\rm norm}$ and the mass fractions of our parameterisations to the FD $X_{\max}$ data measured by the Pierre Auger Observatory. The fitted mass fractions and p-values for each fitted model are shown.

Figure 8.13 displays the results from fitting the mass fractions, $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$ of our EPOS-LHC, QGSJetII-04 and Sibyll2.3 $X_{\max}$ parameterisations to FD $X_{\max}$ data (i.e. without HEAT data) [148]. The estimated $\ln A$ and $X_{\max}$ moments are shown in Figure 8.14.

The fits of the EPOS-LHC and QGSJetII-04 parameterisations result in a mass composition that is consistent between these two models, and one that is consistent with the fits to the



(a) First two moments of the ln A distribution estimated from the fitted fractions.



(b) The red, blue and green lines show the new predictions for the $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ from the fits of $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$.

Figure 8.14: Fitting $t_{0_{\rm norm}}$, $\sigma_{\rm norm}$ and the mass fractions of our parameterisations to FD $X_{\max}$ data measured by the Pierre Auger Observatory.

EPOS-LHC	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\rm norm}}$	692.9 (stat.)$^{+4}_{-2}$ (sys.)$^{+8}_{-8}$	686.6	669.7	640.1
fitted $\sigma_{\rm norm}$	16.2 (stat.)$^{+1}_{-1}$ (sys.)$^{+1}_{-2}$	17.7	14.4	9.9
QGSJetII-04	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\rm norm}}$	695.6 (stat.)$^{+3}_{-3}$ (sys.)$^{+8}_{-8}$	687.1	667.9	642.2
fitted $\sigma_{\rm norm}$	15.5 (stat.)$^{+1}_{-2}$ (sys.)$^{+2}_{-2}$	16.7	14.4	10.3
Sibyll2.3	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\rm norm}}$	720.8 (stat.)$^{+3}_{-3}$ (sys.)$^{+8}_{-9}$	707.4	684.6	656.6
fitted $\sigma_{\rm norm}$	25.8 (stat.)$^{+3}_{-3}$ (sys.)$^{+2}_{-6}$	22.1	17.9	13.0

Table 8.5: The $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$ coefficients fitted to the Auger FD $X_{\max}$ data, assuming a normalisation energy of $E_0 = 10^{18.24}$ eV.

combined FD and HEAT $X_{\max}$ data, as the fitted coefficient values are similar. The Sibyll2.3 fit however predicts a mass composition dominated by nitrogen. The $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$ for proton and iron predicted by the Sibyll2.3 fit are larger.

Figure 6.1b in Section 6.1 displays the $t_{0_{\text{norm}}}$ and σ_{norm} parameter space scan of the FD data, illustrating the Sibyll2.3 fit has a secondary minimum within 4σ of the deepest minimum. This secondary minimum is consistent with the $t_{0_{\text{norm}}}$ and σ_{norm} fits of the EPOS-LHC and QGSJetII-04 parameterisations. As the initial values of $t_{0_{\text{norm}}}$ and σ_{norm} are the main differences between the parameterisations, the mass composition which corresponds to the secondary Sibyll2.3 solution is consistent with the EPOS-LHC and QGSJetII-04 reconstructions.

The p-values from the Sibyll2.3 FD fit which correspond to the deepest minimum are similar to the p-values from the Sibyll2.3 fit of the combined FD and HEAT data, even though the fit of the FD data results in a different solution (compare Figures 8.13 and 8.4). The two Sibyll2.3 solutions describe the measured data with similar accuracy, illustrating the degeneracy between the fitted coefficients and mass fractions. The added information provided by the HEAT data is able to break the degeneracy (see Figure 6.1c). However, even with greater FD statistics, it is possible that nature could inherently result in two plausible solutions given the assumptions of the Sibyll2.3 model, and it is possible the false solution better fits the observed data.

The analysis of mock FD data in Section 5.2.2 showed that $t_{0_{\text{norm}}}$ and σ_{norm} can be fitted reliably, but the fits are not as accurate as those of the FD and HEAT data. The FD $X_{\max}$ distributions do not permit the reliable fit of additional shape parameter coefficients.

Given the reduced energy range of the FD data set compared to the combined FD and HEAT data set, fitting $t_{0_{\text{norm}}}$ and σ_{norm} to only the FD data could result in different coefficient values if the shape parameterisation coefficients that are fixed are incorrect to a significant degree, or the assumed functional form is incorrect for describing nature. For example, if the parameterisation predictions of $\langle X_{\max} \rangle$ above $10^{19}\,\text{eV}$ were significantly different from nature, the fit of only the FD data would be more sensitive to this discrepancy than the fit of the FD and HEAT data, and consequently a different value of $t_{0_{\text{norm}}}$ and σ_{norm} may be fitted to compensate. Taking into consideration that the Sibyll2.3 fit consists of a secondary minimum consistent with the results of the FD and HEAT data, and the fits of mock FD data in Section 5.2.2 showed the two-coefficient fit can be difficult to constrain, we conclude that the fits of the FD data set do not provide strong evidence that the coefficients which are kept fixed incorrectly describe nature to a significant degree.

The results of fitting $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions to the 2014 FD $X_{\max}$ data set [145] are shown in Figures 8.15 and 8.16. The $X_{\max}$ data was organised into $1\,\text{g/cm}^2$ bins. Despite the current FD data set containing 6125 more events than the 2014 FD data set (25884 compared to 19759), the results obtained from the fit of each parameterisation are similar. Using these fitted values of $t_{0_{\text{norm}}}$ and σ_{norm}, the parameters of the equations describing the $X_{\max}$ moments in terms of $\ln A$ in [146] have been determined, and are shown in Table F.3 and F.4 of Appendix F.2.

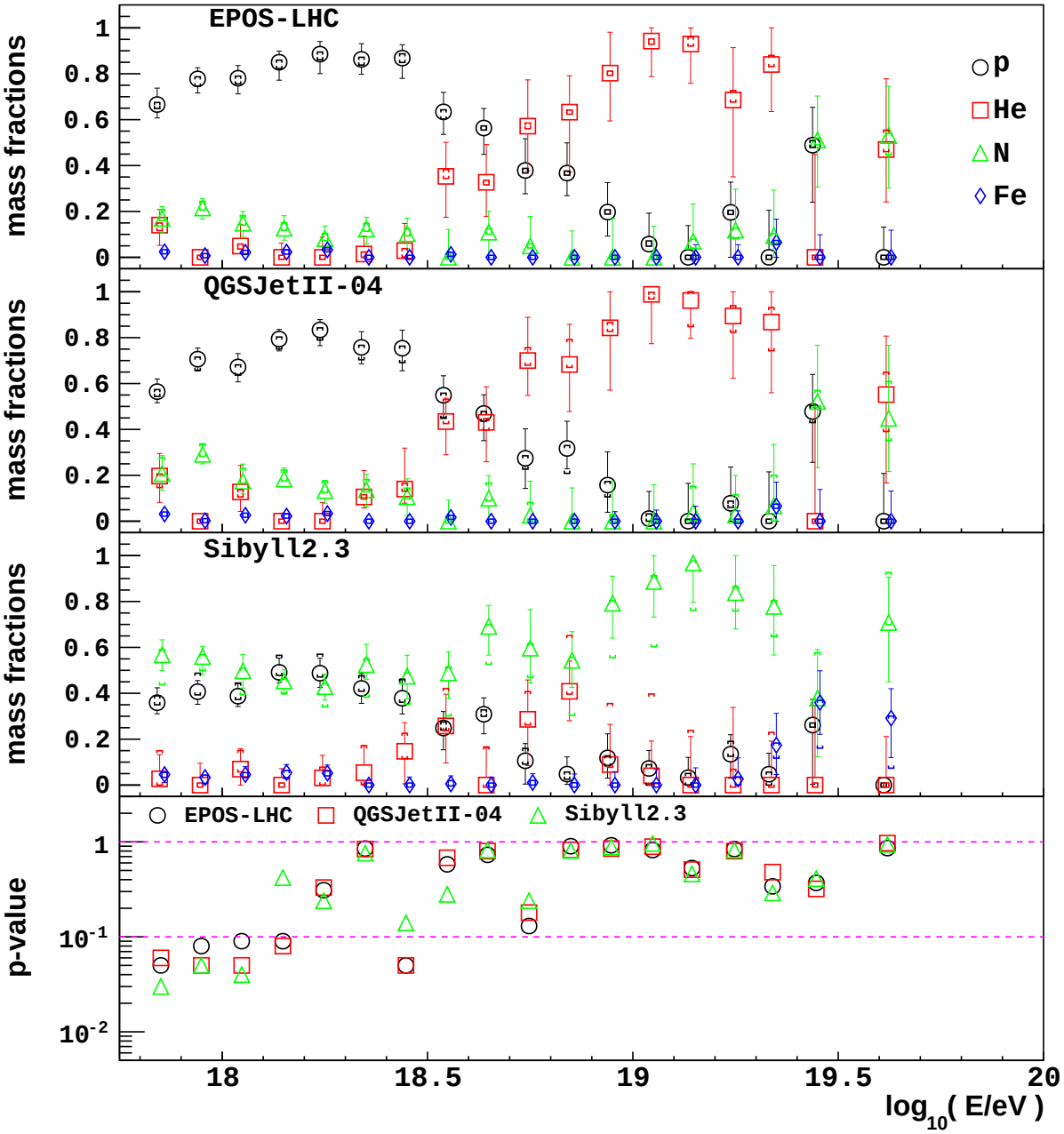
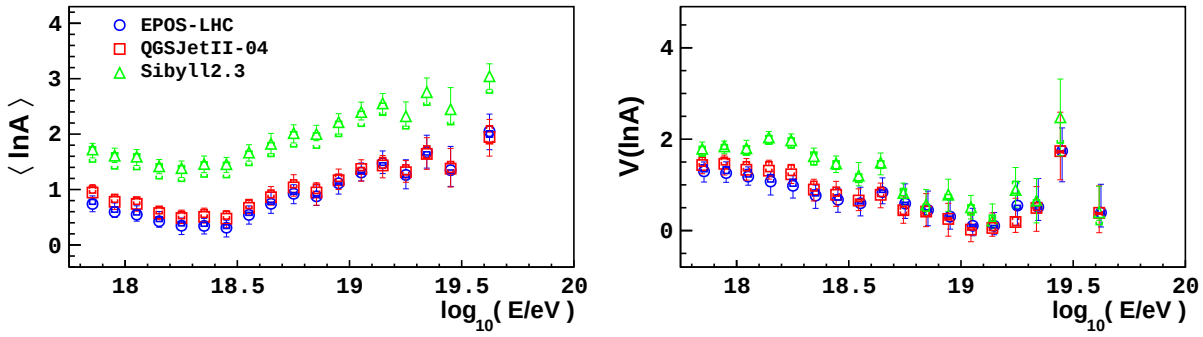
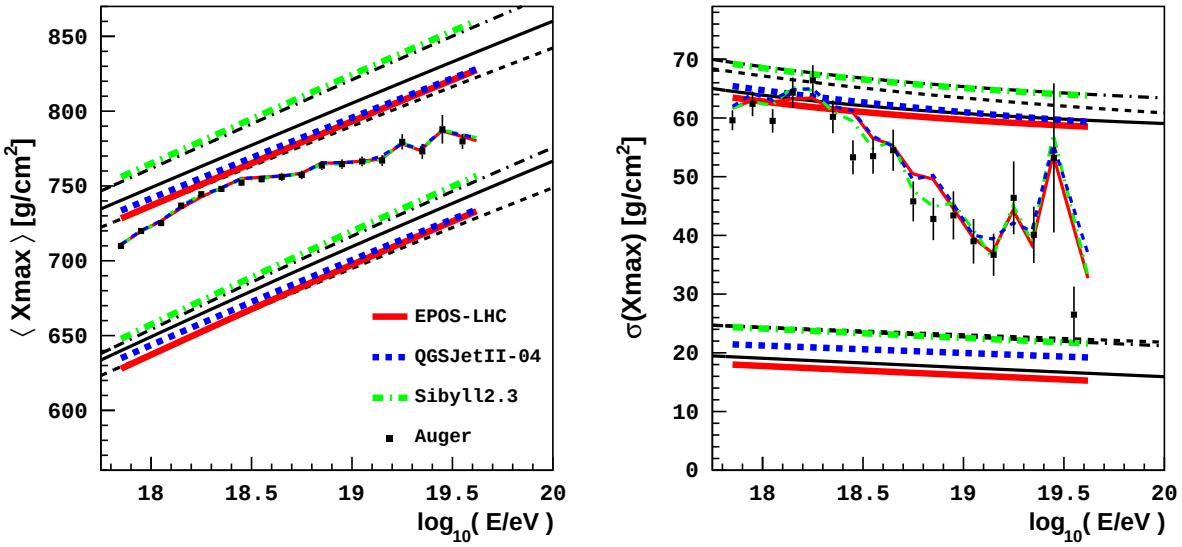


Figure 8.15: Fitting $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions of our parameterisations to the 2014 FD $X_{\max}$ data set [145]. The fitted mass fractions and p-values for each fitted model are shown.



(a) First two moments of the ln A distribution estimated from the fitted fractions.



(b) The red, blue and green lines show the new predictions for the $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ from the fits of $t_{0_{\text{norm}}}$ and σ_{norm}.

Figure 8.16: Fitting $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions of our parameterisations to the 2014 FD $X_{\max}$ data set.

8.3 HEAT $X_{\max}$ data composition results

8.3.1 $t_{0_{\text{norm}}}$, σ_{norm} and mass fraction fit of the Auger HEAT $X_{\max}$ data

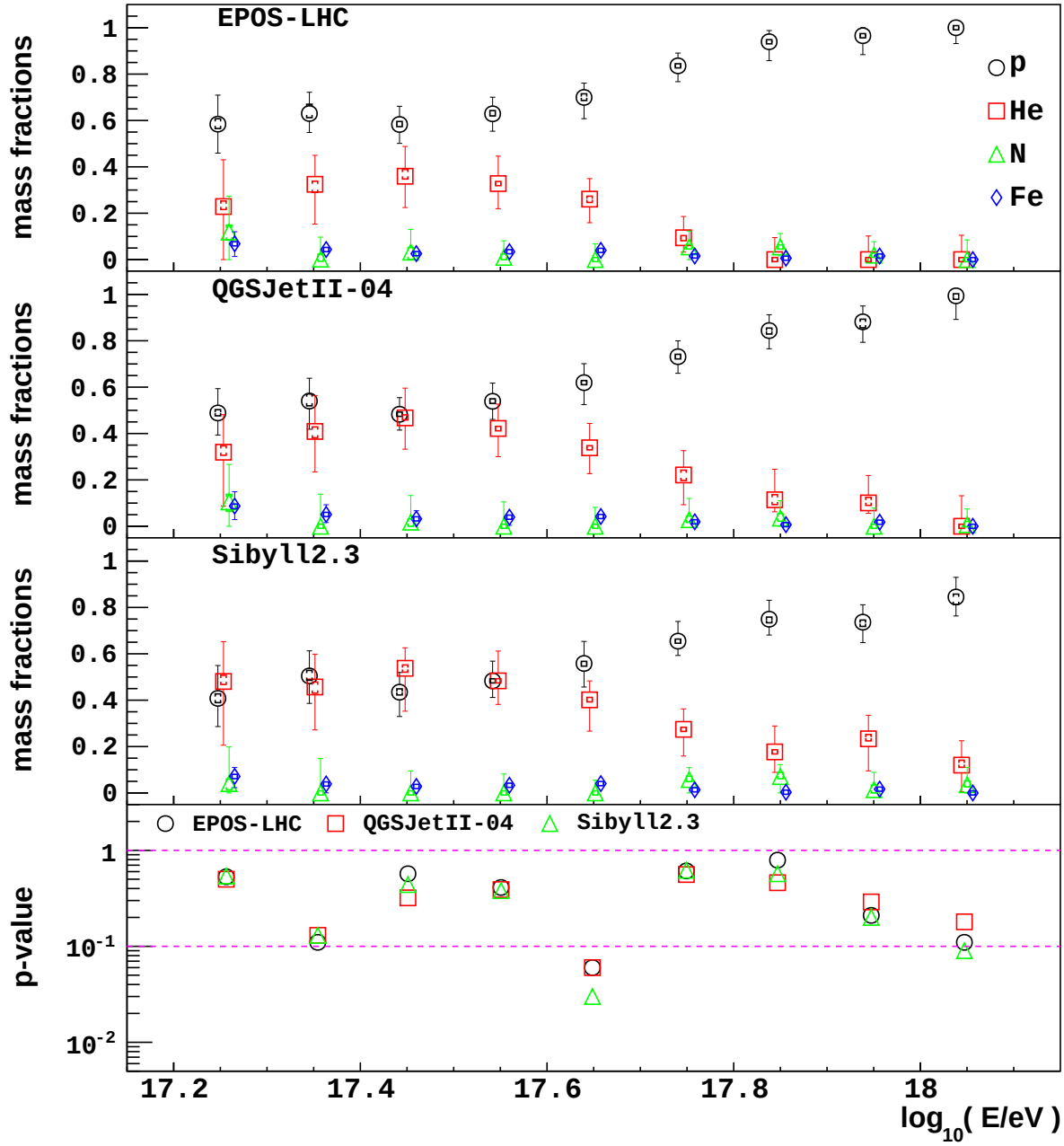
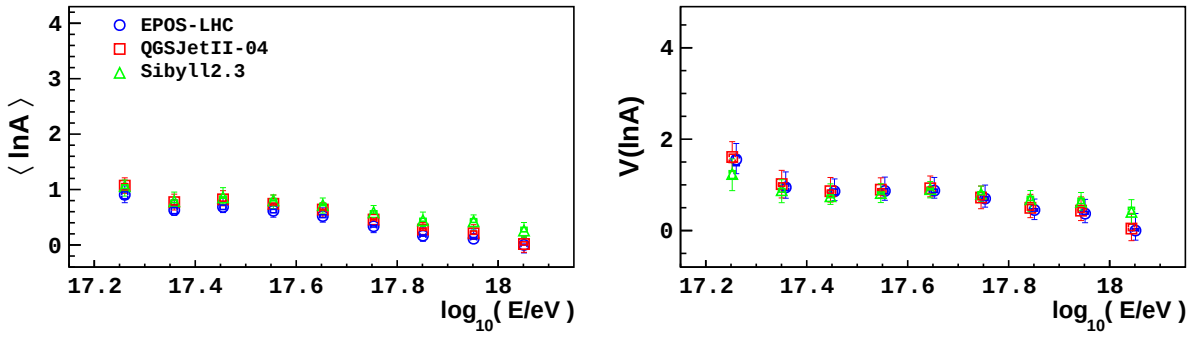
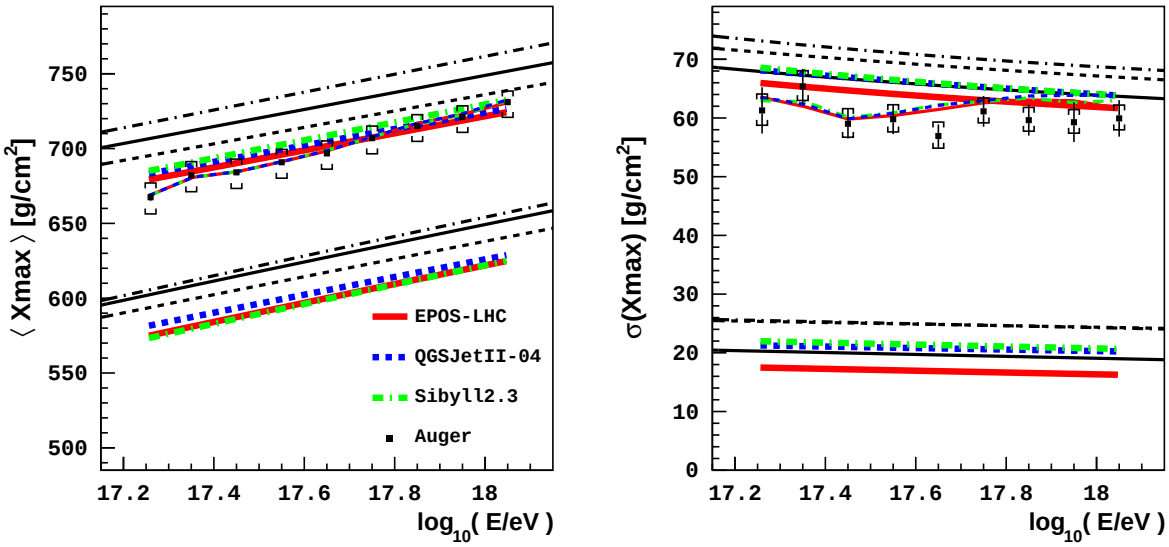


Figure 8.17: Fitting $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions of our parameterisations to the HEAT $X_{\max}$ data measured by the Pierre Auger Observatory. The fitted mass fractions and p-values for each fitted model are shown.

Figures 8.17 and 8.18 displays the results from fitting the mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm} of our EPOS-LHC, QGSJetII-04 and Sibyll2.3 parameterisations to HEAT $X_{\max}$ data [148]. The $X_{\max}$ values of the HEAT data are still shifted by $-2.3\,\text{g/cm}^2$, so that an easy comparison can be made with the earlier results presented. If the HEAT $X_{\max}$ data was left un-shifted, the fitted $t_{0_{\text{norm}}}$ values would be approximately $2.3\,\text{g/cm}^2$ larger and the composition unchanged.



(a) First two moments of the ln A distribution estimated from the fitted fractions.



(b) The red, blue and green lines show the new predictions for the $\langle X_{\max}\rangle$ and $\sigma(X_{\max})$ from the fits of $t_{0_{\text{norm}}}$ and σ_{norm}.

Figure 8.18: Fitting $t_{0_{\text{norm}}}$, σ_{norm} and the mass fractions of our parameterisations to HEAT $X_{\max}$ data measured by the Pierre Auger Observatory.

EPOS-LHC	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\text{norm}}}$	676.1 (stat.)$^{+4}_{-2}$ (sys.)$^{+8}_{-10}$	669.8	652.8	623.3
fitted σ_{norm}	14.7 (stat.)$^{+2}_{-3}$ (sys.)$^{+2}_{-3}$	15.8	12.7	8.7
QGSJetII-04	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\text{norm}}}$	675.8 (stat.)$^{+3}_{-2}$ (sys.)$^{+8}_{-11}$	667.4	648.2	622.4
fitted σ_{norm}	14.6 (stat.)$^{+2}_{-3}$ (sys.)$^{+3}_{-3}$	15.9	13.7	9.7
Sibyll2.3	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\text{norm}}}$	682.3 (stat.)$^{+3}_{-2}$ (sys.)$^{+9}_{-11}$	668.9	646.0	618.0
fitted σ_{norm}	13.6 (stat.)$^{+4}_{-4}$ (sys.)$^{+3}_{-3}$	11.7	9.3	6.8

Table 8.6: The $t_{0_{\text{norm}}}$ and σ_{norm} coefficients fitted to the Auger HEAT $X_{\max}$ data, assuming a normalisation energy of $E_0 = 10^{18.24}$ eV.

The analysis of mock HEAT data in Section 5.2.3 showed that fits of $t_{0_{\text{norm}}}$ and σ_{norm} are considerably less accurate compared to the fits of the combined FD and HEAT data. The fits resulted in absolute offsets in the median mass fractions within 20% of the true mass if the correct parameterisation was fitted to the data. The EPOS-LHC and QGSJetII-04 mock data was fitted with much better accuracy. The HEAT $X_{\max}$ distributions do not permit the reliable fit of additional shape parameter coefficients.

The EPOS-LHC and QGSJetII-04 fitted values of $t_{0_{\text{norm}}}$ and σ_{norm} with the HEAT data are consistent. The fitted $t_{0_{\text{norm}}}$ proton value of Sibyll2.3 is larger than the other models because the Sibyll2.3 t_0 proton separation from heavier nuclei is larger (see Figure 4.8), and the fit of $t_{0_{\text{norm}}}$ conserves the initial separation. Table 8.6 shows that the fitted values of $t_{0_{\text{norm}}}$ for heavier nuclei are consistent between the three parameterisations. Consequently, the fitted mass composition is consistent between the three models, with Sibyll2.3 predicting a less dominant proton composition. Comparing the mass composition reconstruction of the HEAT data with that of the combined FD and HEAT fit of $t_{0_{\text{norm}}}$ and σ_{norm} (compare Figures 8.17 and 8.4), we see that the HEAT fit predicts a significant helium component, which replaces the nitrogen component in the FD and HEAT fit. Consequently, the estimated $\ln A$ moments of the HEAT fit are much smaller than the FD and HEAT fit (compare Figures 8.18a and 8.5a). Despite the reduced statistics of the HEAT data, the mass composition statistical and systematic errors are comparable to the fits of the FD and HEAT data.

The p-values of the HEAT fit are better overall compared to the fit of the FD and HEAT data, which is impressive given the composition is dominated by protons and helium, and Figure 4.5 showed our parameterisations do not describe the proton and helium MC data of these models as adequately as heavier nuclei. The HEAT fit consists of one energy bin where the p-values of at least two of the models are well below the 0.1 line, compared to 4 energy bins for the FD and HEAT fit over this energy range. This illustrates that the coefficients and fractions fitted to the HEAT data are a better description of the measured HEAT distributions than those fitted to the HEAT distributions when the FD and HEAT data are fitted together. However, due to the degeneracy between the fitted parameters when fitting such a data set of minimal statistics and energy range, the fitted parameters do not necessarily describe nature correctly. Alternatively, if the parameterisations do not accurately describe the change in the $X_{\max}$ distributions with energy, fitting a smaller energy range will result in a better fit when only fitting $t_{0_{\text{norm}}}$ and σ_{norm}. Lastly, if the HEAT $X_{\max}$ distributions are not sufficiently aligned with the FD $X_{\max}$ distributions, worse p-values would be obtained from the FD and HEAT $t_{0_{\text{norm}}}$ and σ_{norm} fit compared to the fit of only the HEAT data.

The fitted values of $t_{0_{\text{norm}}}$ are $10 - 20\,\text{g/cm}^2$ lower than the fits of the FD and HEAT data, and also do not align with the fits of only the FD data. Figure 8.18b shows the $X_{\max}$ data at $10^{18.1}$ eV may be preventing an even lower value of $t_{0_{\text{norm}}}$ being fitted. Given the tests on mock data in Section 5.2.3 showed a considerable spread in the fitted values of $t_{0_{\text{norm}}}$ and σ_{norm}, the inconsistent values are likely a consequence of the inability of the HEAT data to constrain the fitted parameters, and/or the susceptibility of the HEAT fits to statistical variation.

Chapter 9

Conclusions

Cosmic ray mass composition information can be extracted by comparing the observed cosmic ray X_{max} data to the expected X_{max} distributions of a hadronic interaction model. At very high cosmic ray energies (10^{17} eV to 10^{20} eV), uncertainties in the hadronic interaction models result in an extracted mass composition that depends strongly on the model assumed. A method of parameterising the X_{max} distributions predicted by a hadronic interaction model has been presented in this thesis, and this method has been applied to the EPOS-LHC, QGSJetII-04 and Sibyll2.3 hadronic interaction models. This thesis has demonstrated that by fitting specific coefficients of the X_{max} parameterisation to the observed data, the cosmic ray mass composition inferred has a reduced dependency on the assumed hadronic interaction model.

The main differences between the X_{max} distribution predictions of the EPOS-LHC, QGSJetII-04 and Sibyll2.3 hadronic models are the normalisation values of the mode and spread of the X_{max} distributions with energy and mass. However, the separation in the mode among primaries, and the separation in the spread among primaries, according to a particular model, is reasonably consistent between these models. By fitting two coefficients (defined as $t_{0_{\mathrm{norm}}}$ and σ_{norm}) to appropriately alter the normalisations of the mode and spread for each primary, while keeping the other shape parameter coefficients of the X_{max} parameterisation fixed, an accurate mass composition can be fitted that is less dependent on the hadronic model assumed, provided that the fitted X_{max} data consists of a sufficient spread of primary masses and statistics over the energy range to constrain the fit. This has been demonstrated by fits to mock X_{max} data that resemble the X_{max} distributions measured by the Pierre Auger Observatory.

The differences between the X_{max} distribution parameterisations of EPOS-LHC, QGSJetII-04 and Sibyll2.3 can be greatly reduced by an appropriate adjustment of $t_{0_{\mathrm{norm}}}$ and σ_{norm}, consequently a fit of these coefficients can provide a consistent mass composition result between these three parameterisations. We have fitted $t_{0_{\mathrm{norm}}}$, σ_{norm} and the mass fractions of our EPOS-LHC, QGSJetII-04 and Sibyll2.3 X_{max} parameterisations to FD and HEAT X_{max} data measured by the Pierre Auger Observatory, revealing a reconstructed mass composition that is consistent between these models. Additionally, systematics in the measured X_{max} values are absorbed by the fits of $t_{0_{\mathrm{norm}}}$ and σ_{norm}. Thus, the mass composition reconstruction is not affected by systematics in X_{max}. Consequently, an observed shift in the fitted values of $t_{0_{\mathrm{norm}}}$ and σ_{norm} from the initial parameterisation prediction could be due to either the initial parameterisation inadequately describing nature, systematics in the measured X_{max} values, or a combination of both factors.

The $t_{0_{\mathrm{norm}}}$ and σ_{norm} fits of the EPOS-LHC, QGSJetII-04 and Sibyll2.3 X_{max} parameterisations to FD and HEAT data suggest that protons are dominant around $10^{18.2}$ eV, with the mass composition then transitioning to a helium dominant composition around $10^{19.2}$ eV. A potentially growing nitrogen composition is hinted around $10^{19.5}$ eV, with more data required at the highest energies to confirm this. The EPOS-LHC and Sibyll2.3 fits result in values for $t_{0_{\mathrm{norm}}}$ which adjust their $\langle X_{\mathrm{max}} \rangle$ predictions towards the initial QGSJetII-04 predictions, while the QGSJetII-04 and Sibyll2.3 parameterisations fit values for σ_{norm} which adjust their $\sigma(X_{\mathrm{max}})$ predictions towards

the initial EPOS-LHC model predictions. The p-values for QGSJetII-04 and Sibyll2.3 illustrate that fitting $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$ results in a significantly better description of the data than a fit of only the mass composition. However, the p-values for the three fits are still relatively poor, suggesting the fits are unable to find a combination of p, He, N and Fe which describes the details of the measured data. This may be due to the particle interaction predictions of the utilised hadronic models inadequately describing the shower physics at ultra high energies and/or an inadequate parameterisation of the detector resolution and acceptance. Additionally, the HEAT $X_{\rm max}$ data may not be aligned sufficiently with the FD $X_{\rm max}$ data. Further work is needed to understand the impact on the fit of combining the HEAT $X_{\rm max}$ data with the FD $X_{\rm max}$ data, given the current adjustments applied to the HEAT data.

The EPOS-LHC and QGSJetII-04 fits of only the FD $X_{\rm max}$ data result in a mass composition reconstruction consistent with the fits of the FD and HEAT data. The Sibyll2.3 fit of only the FD $X_{\rm max}$ data consists of two separate solutions within 4σ of each other. The Sibyll2.3 solution that corresponds to the secondary minimum is consistent with the EPOS-LHC and QGSJetII-04 results. The fits of only the HEAT $X_{\rm max}$ data predicted a helium component declining from $10^{17.2}$ eV to $10^{18.1}$ eV, unlike the EPOS-LHC and QGSJetII-04 fits of the FD and HEAT data which instead predicted a declining nitrogen component. However, the fits of the HEAT data are more uncertain because there is less mass dispersion in this reduced energy range.

Fits to mock FD and HEAT Auger data have shown that obtaining an accurate mass composition is possible when fitting additional shape parameter coefficients. With the current Auger $X_{\rm max}$ data, fitting only $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$ with the mass fractions is the most reliable approach, but adding $\lambda_{\rm norm}$ does not introduce considerable uncertainty. The EPOS-LHC, QGSJetII-04 and Sibyll2.3 fits of $t_{0_{\rm norm}}$, $\sigma_{\rm norm}$ and $\lambda_{\rm norm}$ were consistent with the $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$ fits of these models. The three parameterisations have similar initial predictions for the value of $\lambda_{\rm norm}$, with the proton $\lambda_{\rm norm}$ fitted $2-3\,{\rm g/cm^2}$ lower than the initial predictions. The fitted proton $\lambda_{\rm norm}$ of $\sim 58\pm2(stat)\pm1(sys)$ g/cm^2 is consistent with the Auger measurement of the proton exponential tail, $\Lambda_\eta = 55.8 \pm 2.3(stat) \pm 1.6(sys)$ g/cm^2 [229]. Fitting these three coefficients resulted in a $\langle X_{\rm max}\rangle$ prediction very close to the initial prediction of the QGSJetII-04 parameterisation.

The results of the mass fraction and shape parameter coefficient fits of the $X_{\rm max}$ distributions remain sensitive to the other model parameters that are kept fixed, such as the $\langle X_{\rm max}\rangle$ elongation rate and the $\langle X_{\rm max}\rangle$ separation between nuclei, parameters which significantly impact the reconstructed mass composition. Fitting a shape coefficient which adjusts the predicted $\langle X_{\rm max}\rangle$ elongation rate was attempted. The fits of $t_{0_{\rm norm}}$, B (which is related to the $\langle X_{\rm max}\rangle$ elongation rate) and $\sigma_{\rm norm}$ to the FD and HEAT data resulted in a fitted $\langle X_{\rm max}\rangle$ elongation rate significantly greater than the initial model predictions, and consequently a mass composition quite different to the previously mentioned fits. This fitted $\langle X_{\rm max}\rangle$ elongation rate would not align with the hadronic model predictions at 10^{14} eV. Fitting B in a more restrictive manner, by constraining t_0 at 10^{14} eV and fitting B and $\sigma_{\rm norm}$, resulted in a mass composition that was consistent between the EPOS-LHC and QGSJetII-04 parameterisations, and consistent with the results of the $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$ fit. We do not place any emphasis on the results obtained from the combined fit of $t_{0_{\rm norm}}$, B and $\sigma_{\rm norm}$, as the analysis of mock $X_{\rm max}$ data sets showed these fits of B are considerably less reliable than fits where B is kept fixed. With additional statistics, especially at energies above 10^{19} eV, it may be possible to reliably fit coefficients which adjust the predicted $\langle X_{\rm max}\rangle$ elongation rate.

The mass composition results of the shape coefficient fits are sensitive to not only the fixed parameters, but also the fitting range limits of the fitted shape coefficients. The fitting range limits applied and the shape coefficients which are fitted and fixed will change as our knowledge of the hadronic physics occurring at the highest energies progresses. An increase in the statistics of the Auger $X_{\rm max}$ data, and/or an increased energy range, can also reveal additional information regarding the shape coefficients.

The interaction cross-section of the first interaction is closely associated with the tail of the

X_{max} distribution. Fitting Λ, the exponential slope of the tail of the X_{max} distribution, can provide information on the mass composition of the distribution, in particular the proportion of lighter elements such as proton and helium. An analysis of the tails of X_{max} distributions is more independent of the uncertainties in hadronic interaction models, and experimental systematic uncertainties, compared to an analysis of the full X_{max} distributions [245]. With the collection of more statistics, fits of the X_{max} tail and an examination of the percentage of events in the tail, could provide constraints on the mass composition at the highest energies.

The strategy of fitting some of the coefficients of the predicted X_{max} parameterisations to obtain a less model dependent mass composition interpretation could be applied to other mass sensitive cosmic ray observables measured by the Pierre Auger Observatory. Surface array measurements of X_{max}, the muon number (N_{μ}), and the signal at 1000 m from the shower core (S_{1000}), could be analysed individually with a similar approach. Alternatively, a combined analysis of many different observables could be conducted by parameterising the model expectations of each observable and then fitting the appropriate coefficients in a combined fit. Inconsistency in the mass composition interpretation from current analyses of different observables (for example see Figure 2.18, comparing the analyses of X_{max} fluorescence detector and surface detector measurements by the Pierre Auger Observatory) could be reduced by adopting the approach presented in this thesis.

Appendix A. CONEX v4r37 $X_{\max}$ distribution fits

The fits of Equation (4.1) to energy binned MC $X_{\max}$ data are shown in Figures A.1, A.2 and A.3.

A.1 EPOS-LHC $X_{\max}$ distribution fits

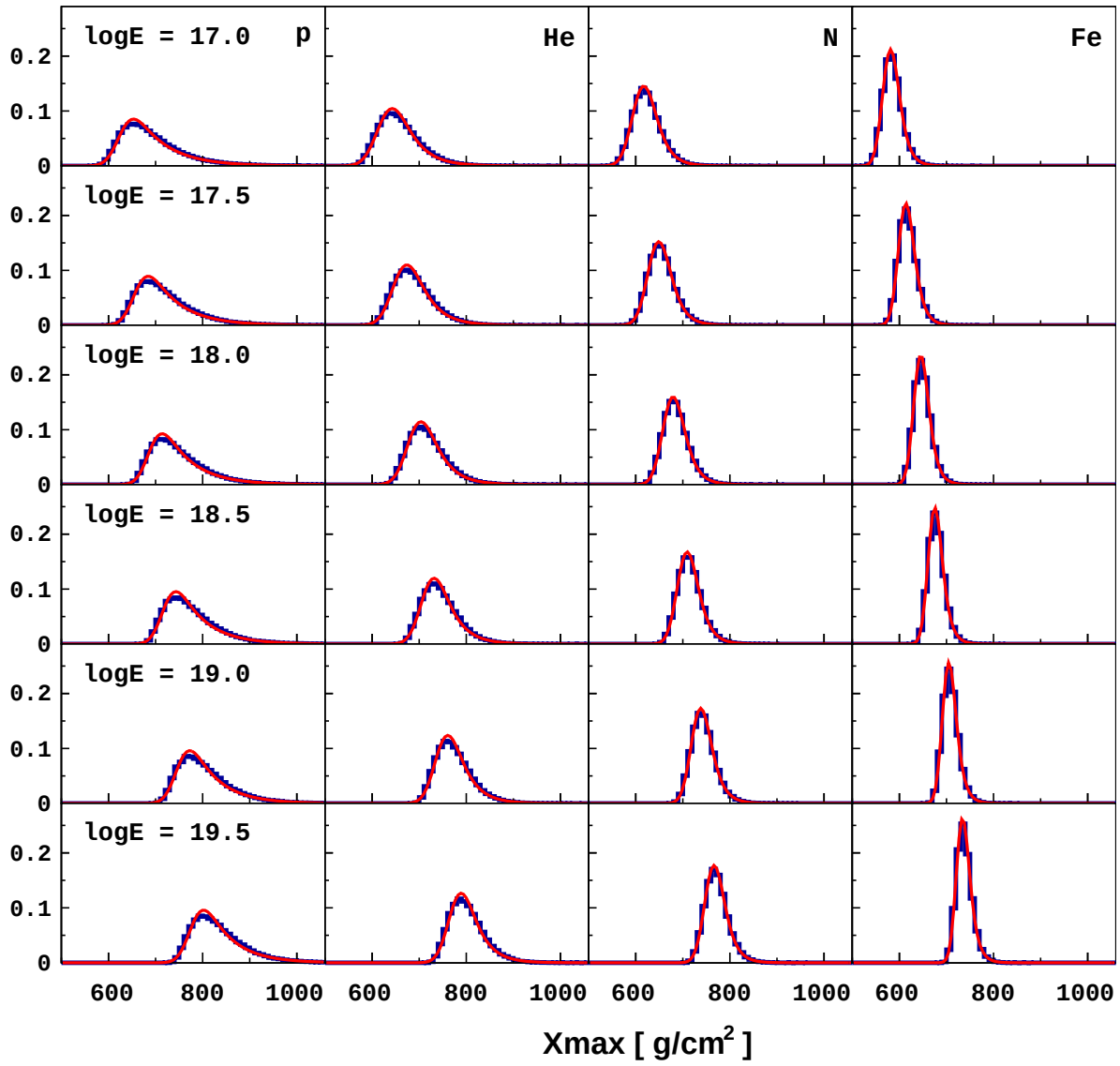


Figure A.1: Energy binned EPOS-LHC $X_{\max}$ distributions (blue line) fitted with Equation (4.1) (red line).

A.2 QGSJetII-04 $X_{\max}$ distribution fits

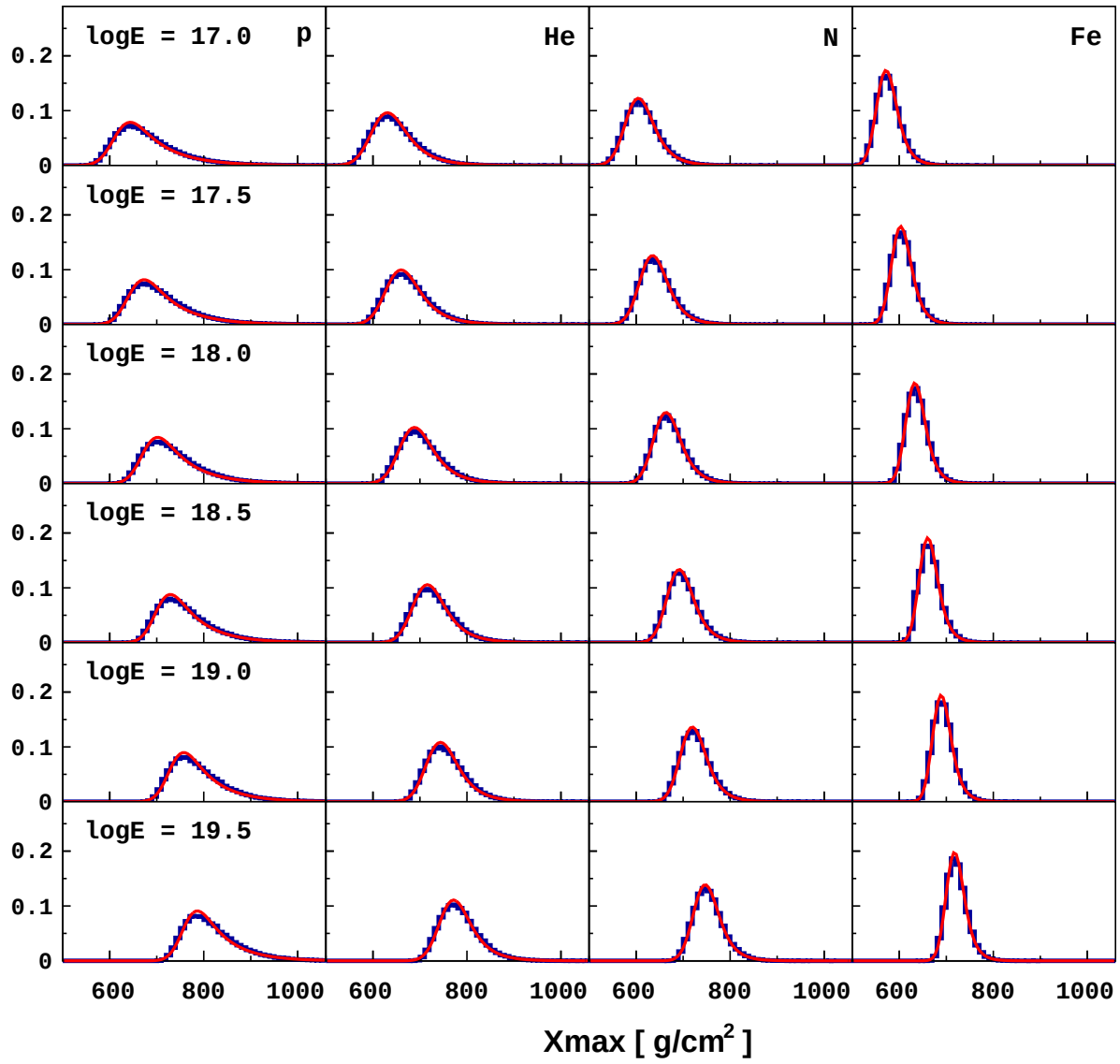


Figure A.2: Energy binned QGSJetII-04 $X_{\max}$ distributions (blue line) fitted with Equation (4.1) (red line).

A.3 Sibyll2.3 $X_{\max}$ distribution fits

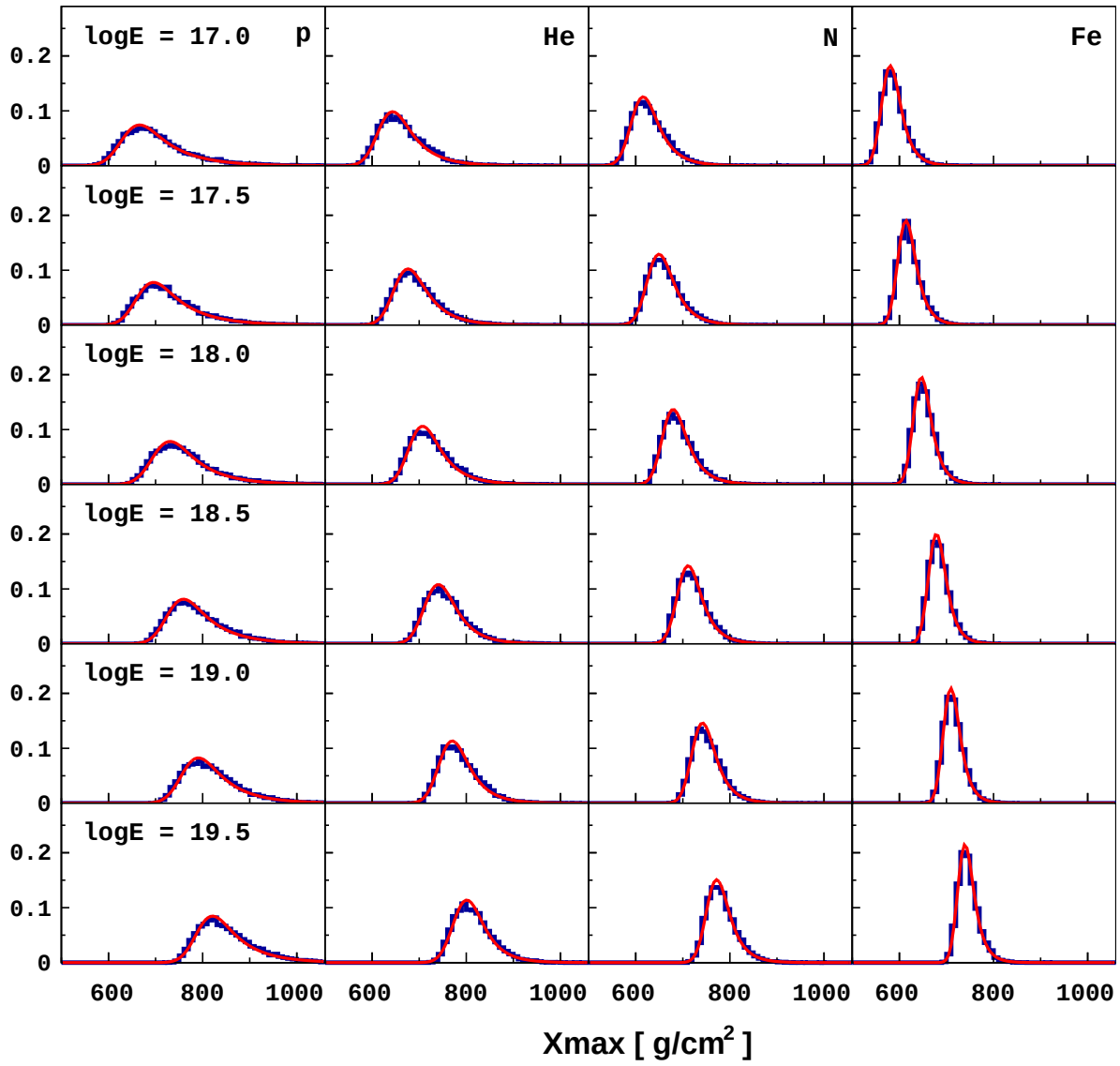


Figure A.3: Energy binned Sibyll2.3 $X_{\max}$ distributions (blue line) fitted with Equation (4.1) (red line).

Appendix B. Statistics of the fitted Auger $X_{\max}$ data

$\log_{10}(E/\mathrm{eV})$	FD Events	HEAT Events
17.2 - 17.3		1052
17.3 - 17.4		1617
17.4 - 17.5		2264
17.5 - 17.6		2565
17.6 - 17.7		2620
17.7 - 17.8		2320
17.8 - 17.9	4608	1827
17.9 - 18.0	3968	1440
18.0 - 18.1	3330	1073
18.1 - 18.2	3437	
18.2 - 18.3	2682	
18.3 - 18.4	2097	
18.4 - 18.5	1585	
18.5 - 18.6	1098	
18.6 - 18.7	842	
18.7 - 18.8	570	
18.8 - 18.9	470	
18.9 - 19.0	355	
19.0 - 19.1	281	
19.1 - 19.2	194	
19.2 - 19.3	128	
19.3 - 19.4	112	
19.4 - 19.5	65	
19.5 - 20.0	62	

Table B.1: Number of events in each energy bin of the fitted FD and HEAT $X_{\max}$ data sets.

Appendix C. $t_{0_{\rm norm}}$, $\sigma_{\rm norm}$ and mass fraction fit of the Auger FD $X_{\rm max}$ data, over an extended $t_{0_{\rm norm}}$ range

The upper limit of $t_{0_{\rm norm}}$ was expanded to 765 g/cm^2 to accommodate the absolute minima identified in the extended $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$ scans of the Auger FD $X_{\rm max}$ data displayed in Section 6.1.1. The results of the mass fraction, $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$ fit of the parameterisations to FD $X_{\rm max}$ data over this extended fit range are shown in Figures C.1, C.2a and C.2b and Table C.1.

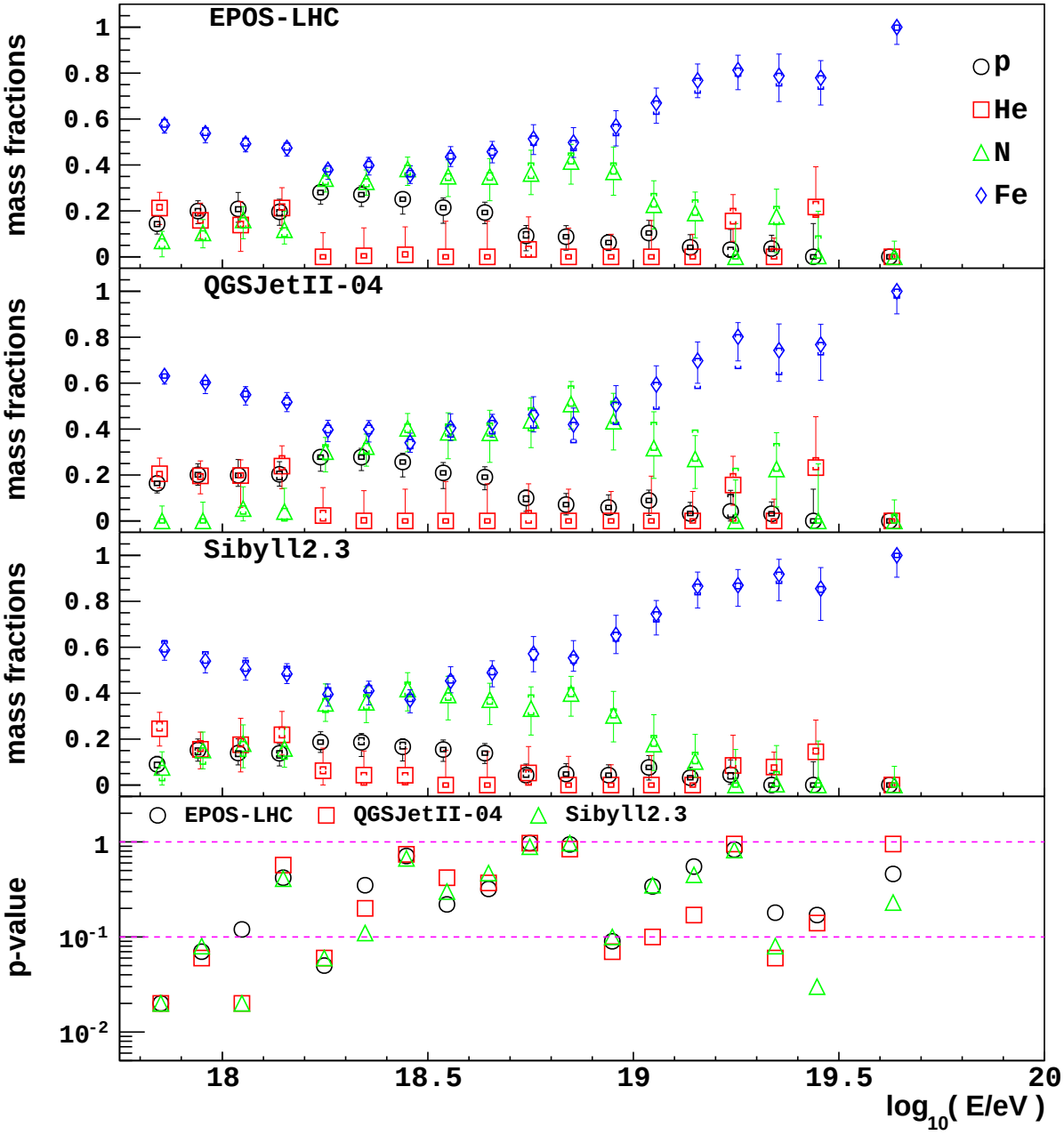
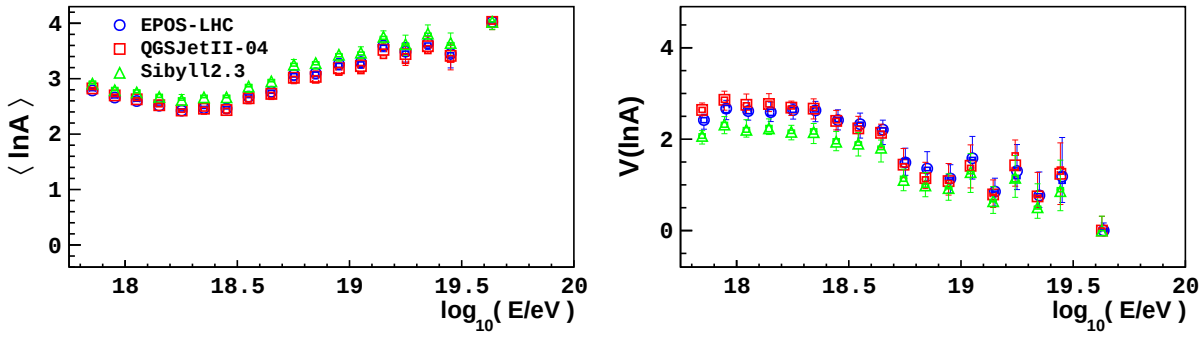
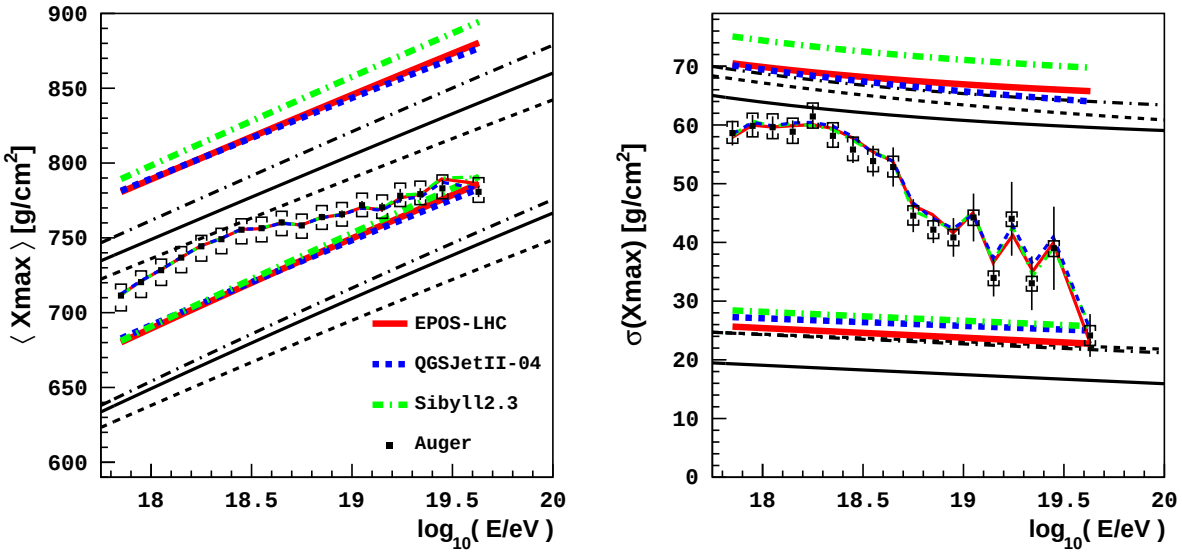


Figure C.1: Fitting $t_{0_{\rm norm}}$, $\sigma_{\rm norm}$ and the mass fractions of our parameterisations to the FD $X_{\rm max}$ data measured by the Pierre Auger Observatory. The fitted mass fractions and p-values for each fitted model are shown.



(a) First two moments of the ln A distribution estimated from the fitted fractions.



(b) The black lines show the $\langle X_{\rm max}\rangle$ and $\sigma(X_{\rm max})$ initially predicted by the $X_{\rm max}$ parameterisations for proton and iron. The red, blue and green lines show the new predictions for the $\langle X_{\rm max}\rangle$ and $\sigma(X_{\rm max})$ from the fits of $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$.

Figure C.2: Fitting $t_{0_{\rm norm}}$, $\sigma_{\rm norm}$ and the mass fractions of our parameterisations to FD $X_{\rm max}$ data measured by the Pierre Auger Observatory.

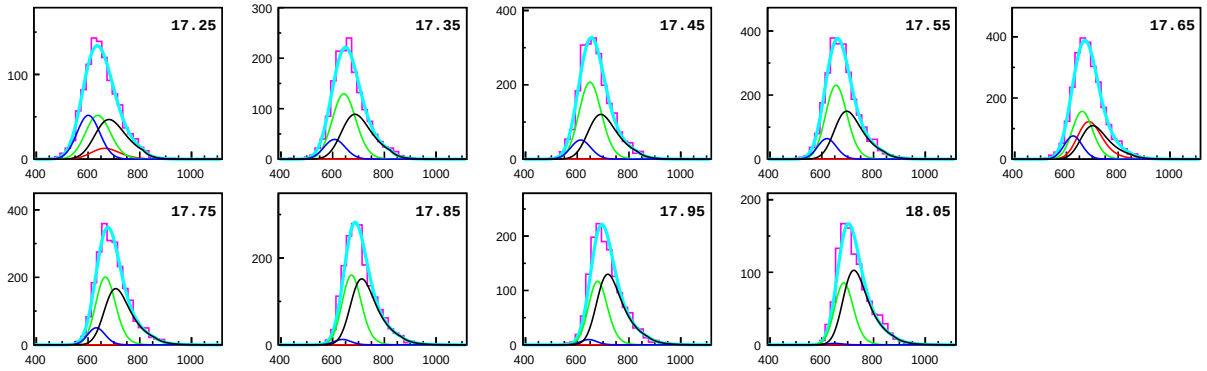
EPOS-LHC	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\rm norm}}$	743.6 (stat.)$^{+1}_{-1}$ (sys.)$^{+8}_{-9}$	737.3	720.4	690.8
fitted $\sigma_{\rm norm}$	35.6 (stat.)$^{+1}_{-1}$ (sys.)$^{+2}_{-2}$	38.3	30.9	21.1
QGSJetII-04	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\rm norm}}$	741.4 (stat.)$^{+1}_{-1}$ (sys.)$^{+8}_{-11}$	733.0	713.8	688.0
fitted $\sigma_{\rm norm}$	30.4 (stat.)$^{+1}_{-1}$ (sys.)$^{+2}_{-2}$	32.9	28.3	20.3
Sibyll2.3	Proton	Helium	Nitrogen	Iron
fitted $t_{0_{\rm norm}}$	751.1 (stat.)$^{+1}_{-2}$ (sys.)$^{+8}_{-9}$	737.7	714.8	686.8
fitted $\sigma_{\rm norm}$	40.0 (stat.)$^{+1}_{-2}$ (sys.)$^{+2}_{-2}$	34.4	27.7	20.1

Table C.1: The $t_{0_{\rm norm}}$ and $\sigma_{\rm norm}$ coefficients fitted to the Auger FD $X_{\rm max}$ data, assuming a normalisation energy of $E_0 = 10^{18.24}$ eV.

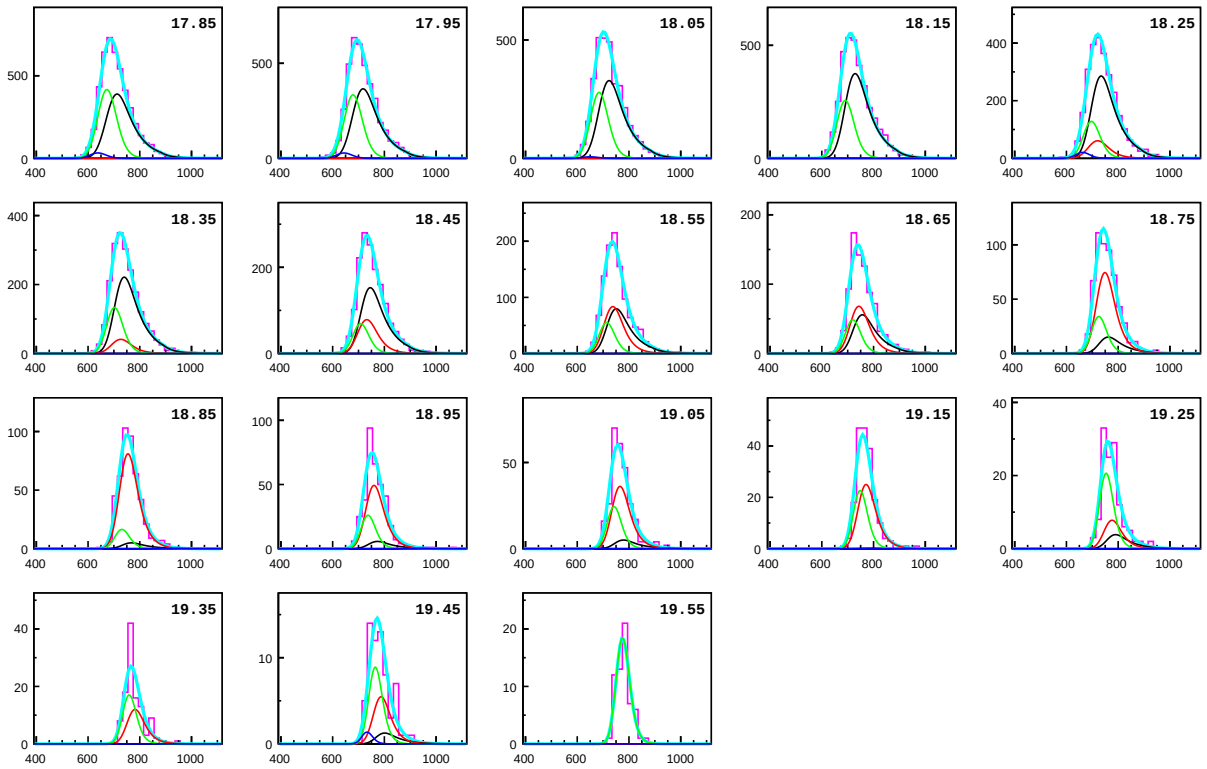
Appendix D. Fits of the Auger $X_{\max}$ distributions

D.1 Mass fraction fits of the Auger FD and HEAT $X_{\max}$ data

The fits of only the mass fractions of each parameterisation to the Auger FD and HEAT $X_{\max}$ distributions [148] are shown in the following plots. The magenta lines illustrate the measured $X_{\max}$ distributions, while the teal lines illustrate the fitted parameterisation. The black, red, green and blue lines are the fitted proton, helium, nitrogen and iron parameterisations respectively. In these figures, the Auger data is organised into $20\,\mathrm{g/cm^2}$ bins for graphical simplicity (but for the fits a bin size of $1\,\mathrm{g/cm^2}$ is used).

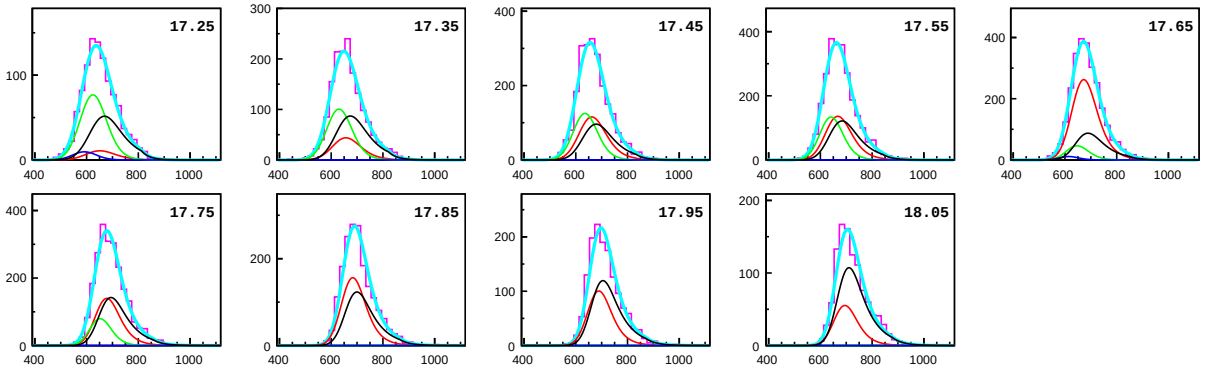


(a) Parameterisations fitted to the HEAT data.

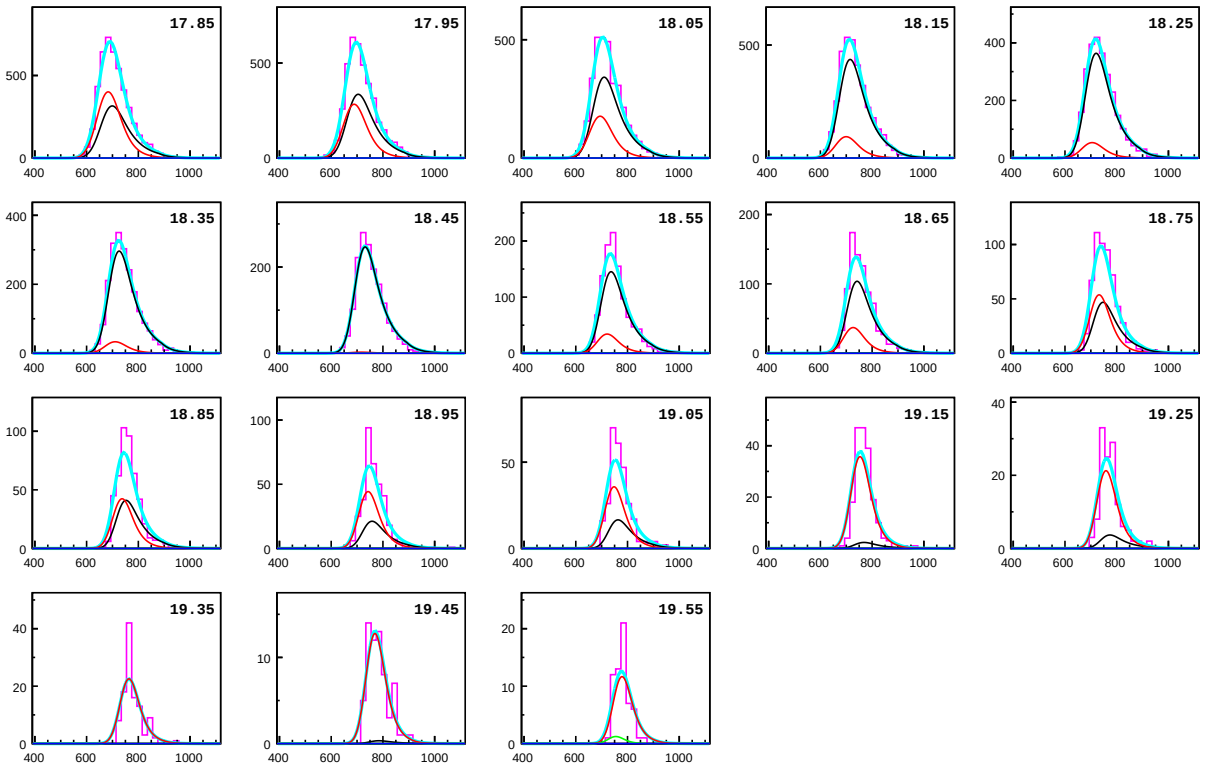


(b) Parameterisations fitted to the FD data.

Figure D.1: Fit of the EPOS-LHC $X_{\max}$ parameterisation. See Section 8.1.1.

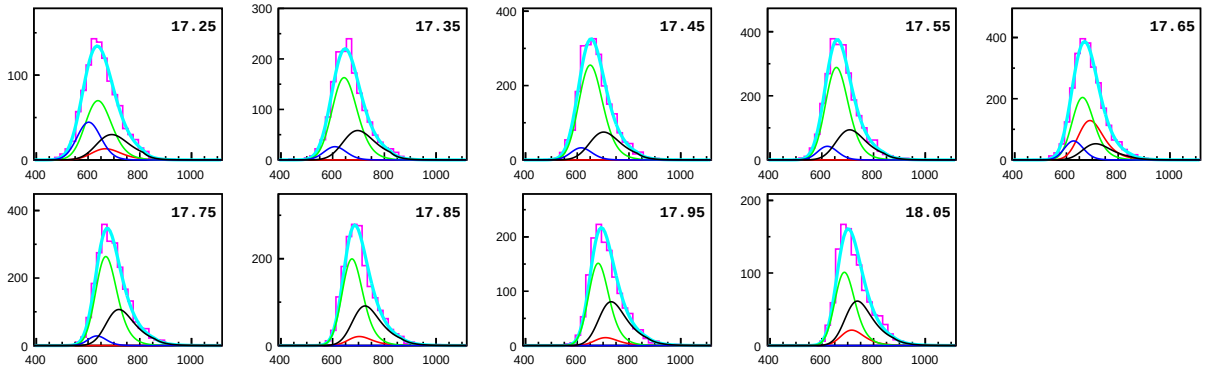


(a) Parameterisations fitted to the HEAT data.

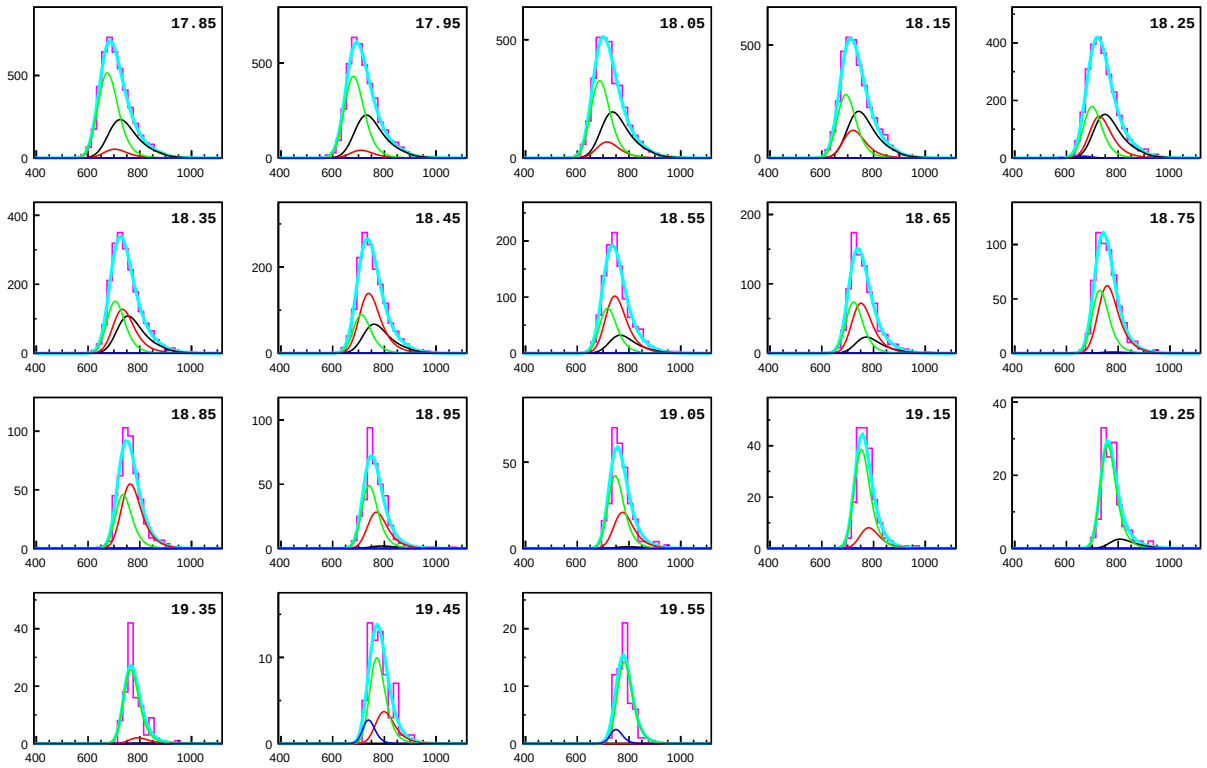


(b) Parameterisations fitted to the FD data.

Figure D.2: Fit of the QGSJetII-04 $X_{\max}$ parameterisation. See Section 8.1.1.



(a) Parameterisations fitted to the HEAT data.

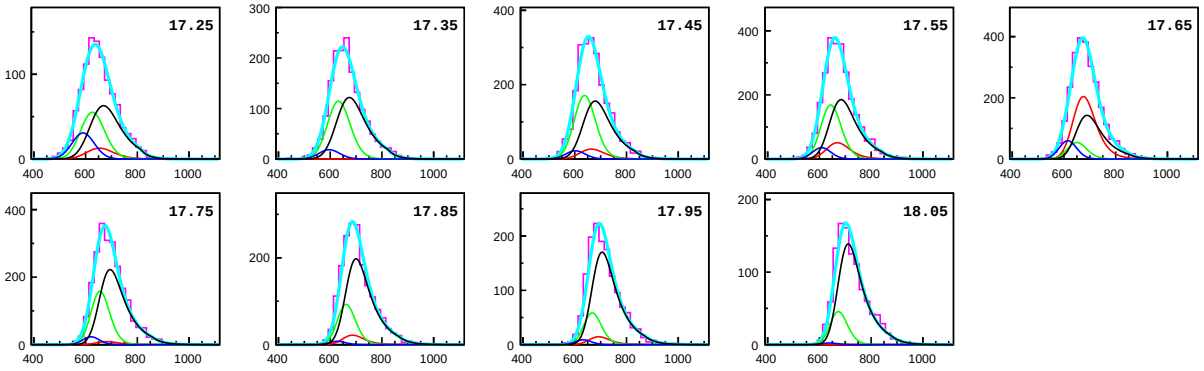


(b) Parameterisations fitted to the FD data.

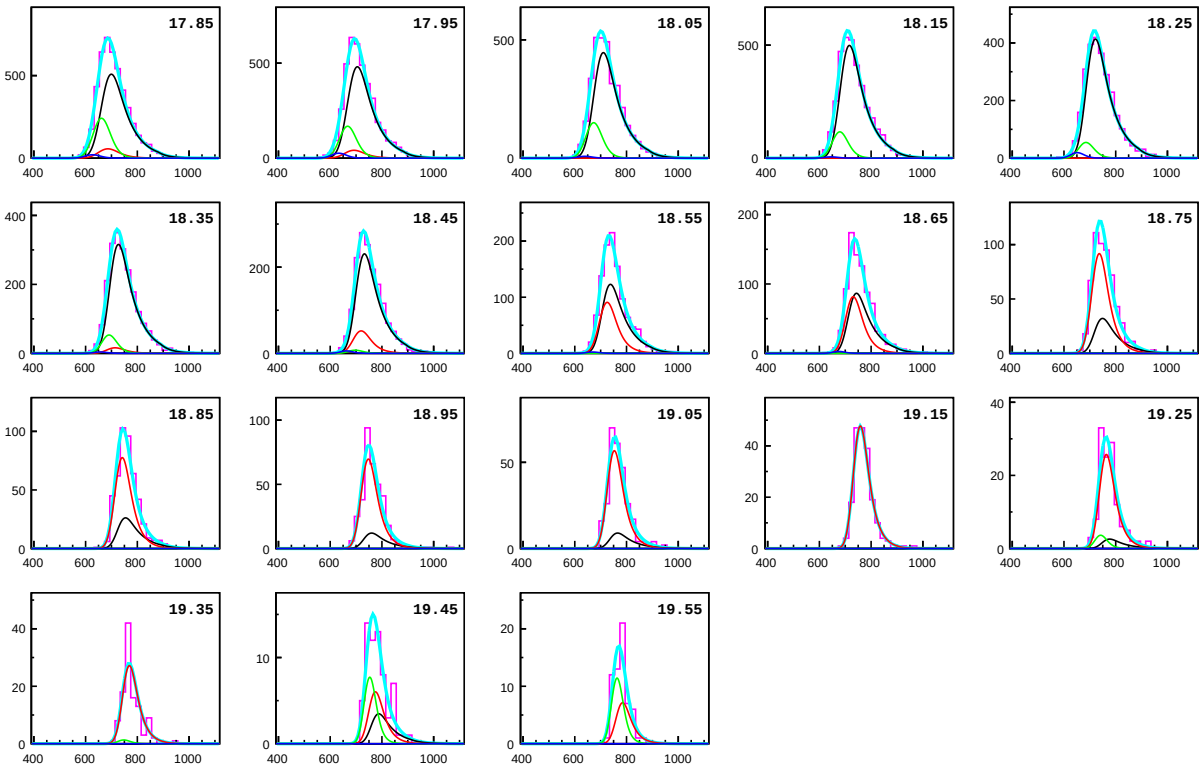
Figure D.3: Fit of the Sibyll2.3 $X_{\max}$ parameterisation. See Section 8.1.1.

D.2 Mass fraction, $t_{0_{\text{norm}}}$ and σ_{norm} fits of the Auger FD and HEAT $X_{\max}$ data

The $t_{0_{\text{norm}}}$, σ_{norm} and mass fraction fits of each parameterisation to the Auger FD and HEAT $X_{\max}$ distributions [148] are shown in the following plots. The magenta lines illustrate the measured $X_{\max}$ distributions, while the teal lines illustrate the fitted parameterisation. The black, red, green and blue lines are the fitted proton, helium, nitrogen and iron parameterisations respectively. In these figures, the Auger data is organised into $20\,\text{g/cm}^2$ bins for graphical simplicity (but for the fits a bin size of $1\,\text{g/cm}^2$ is used).

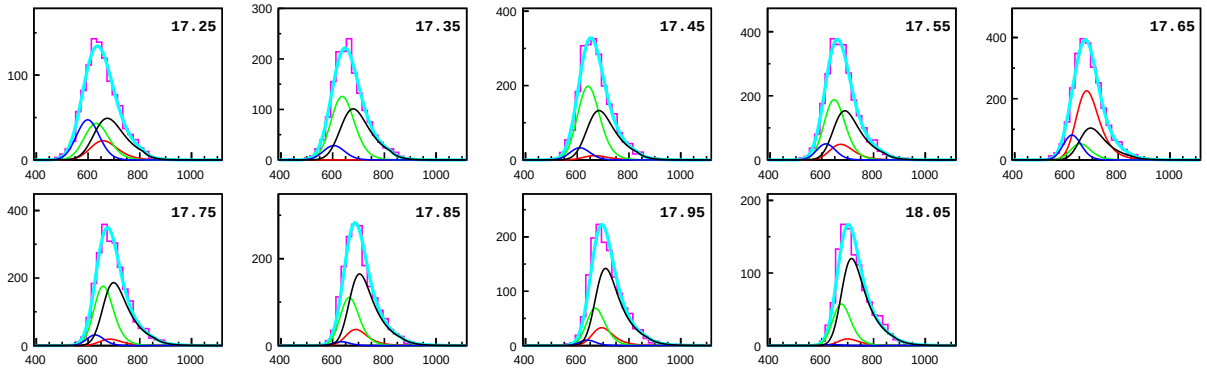


(a) Parameterisations fitted to the HEAT data.

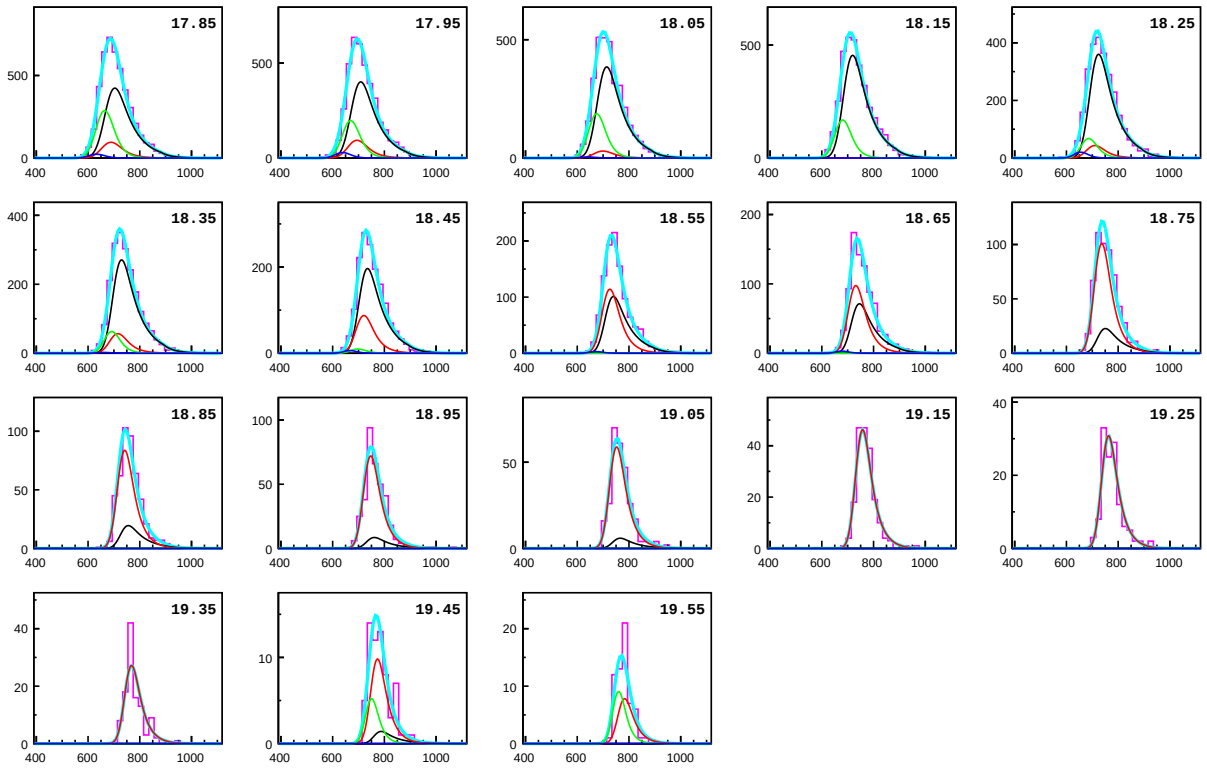


(b) Parameterisations fitted to the FD data.

Figure D.4: Fit of the EPOS-LHC $X_{\max}$ parameterisation. See Section 8.1.2.

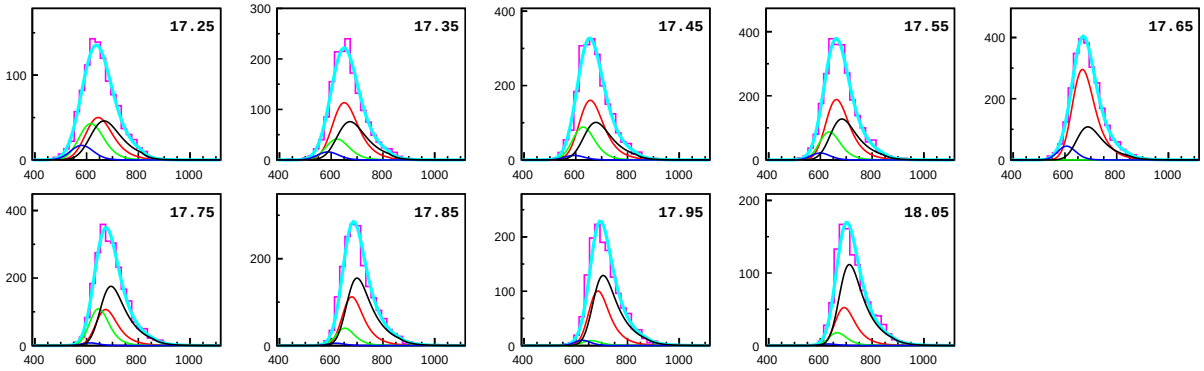


(a) Parameterisations fitted to the HEAT data.

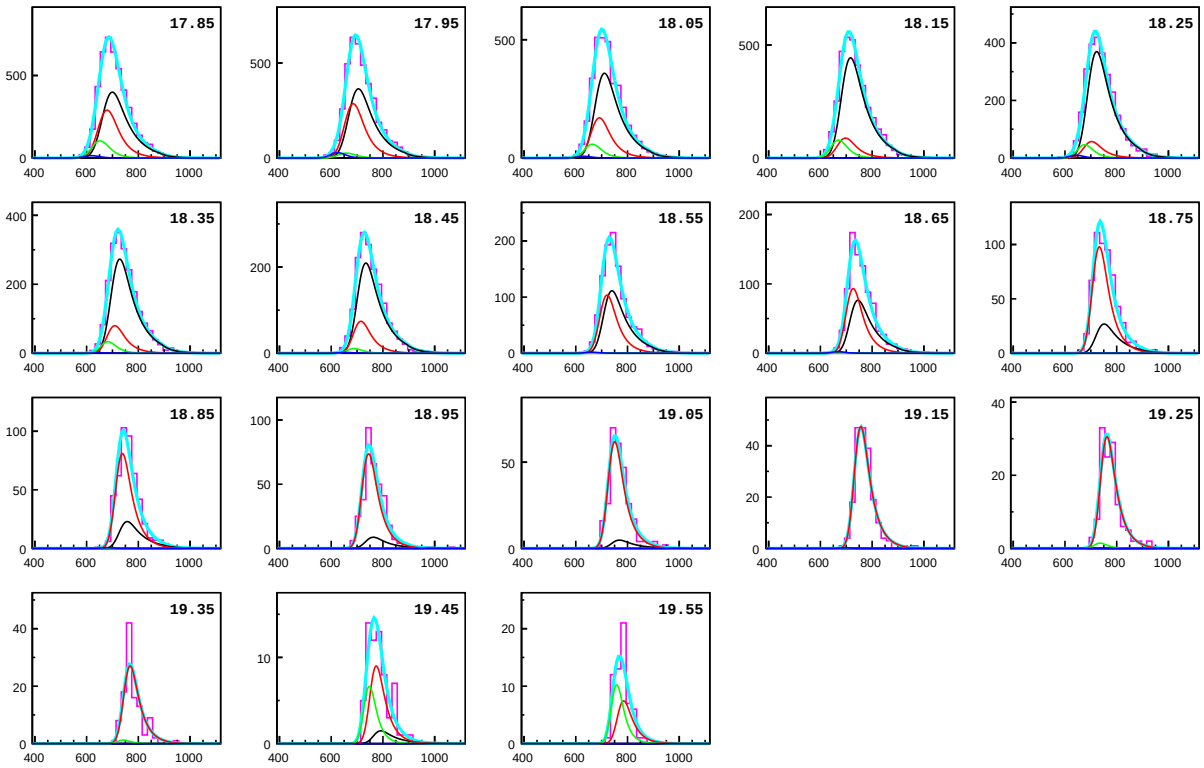


(b) Parameterisations fitted to the FD data.

Figure D.5: Fit of the QGSJetII-04 $X_{\max}$ parameterisation. See Section 8.1.2.



(a) Parameterisations fitted to the HEAT data.

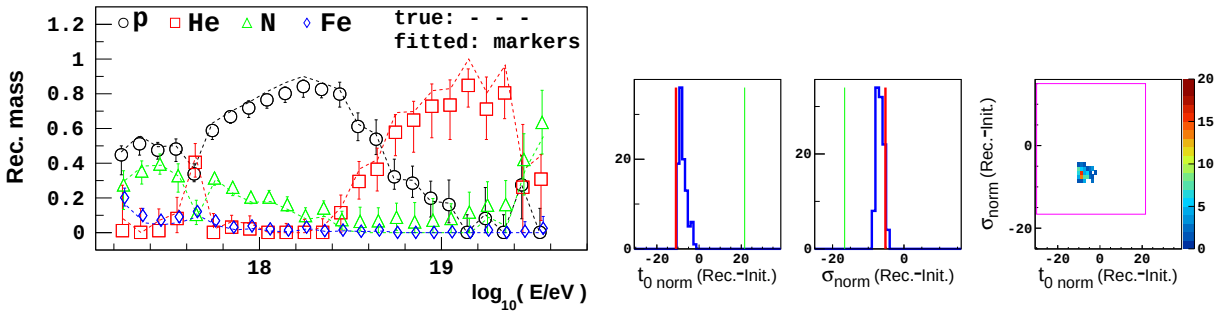


(b) Parameterisations fitted to the FD data.

Figure D.6: Fit of the Sibyll2.3 $X_{\max}$ parameterisation. See Section 8.1.2.

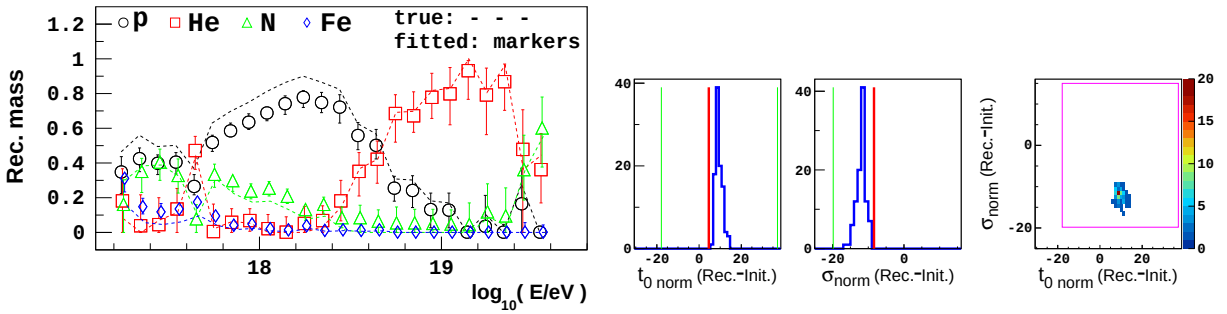
Appendix E. Surplus fits of mock Auger FD and HEAT $X_{\max}$ data

The mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm} of the parameterisations were fitted to the combined FD and HEAT $X_{\max}$ data, the resulting parameterisations used to generate mock data. The mass fractions, $t_{0_{\text{norm}}}$ and σ_{norm} of the parameterisations were then fitted to this mock data. The results are displayed in Figures E.1 to E.9. The mock data fitted is of a distribution that is very similar to the Auger data. Our EPOS-LHC, QGSJetII-04 and Sibyll2.3 fits to this mock data accurately reproduce the true mass composition trend of the mock data. When the mock data is fitted with the correct parameterisation, the absolute offsets in the median reconstructed fractions are less than 20% from the true mass.



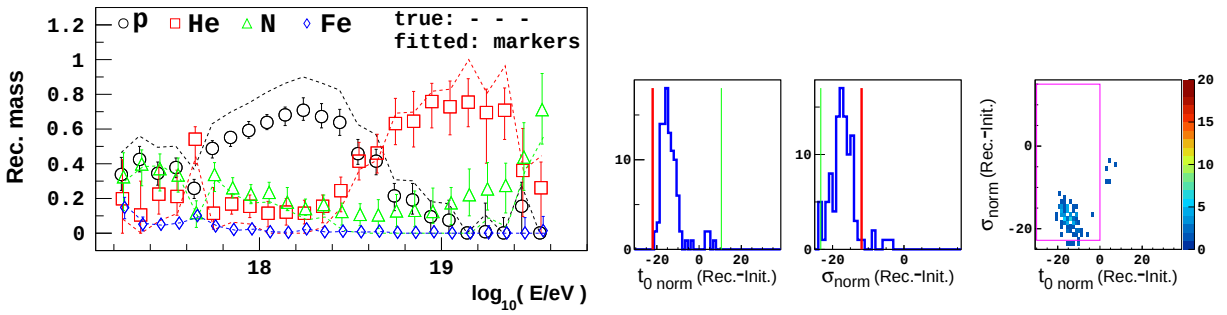
(a) (b)

Figure E.1: Fits of EPOS-LHC based $X_{\max}$ data with the EPOS-LHC parameterisation.



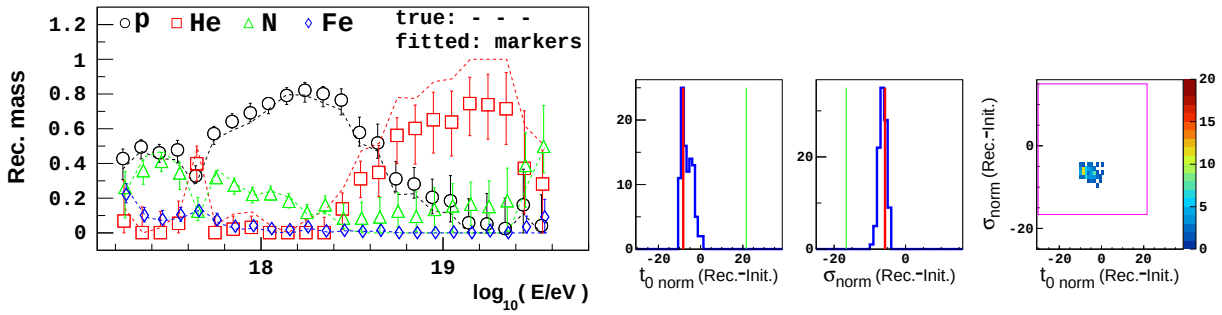
(a) (b)

Figure E.2: Fits of EPOS-LHC based $X_{\max}$ data with the QGSJetII-04 parameterisation.



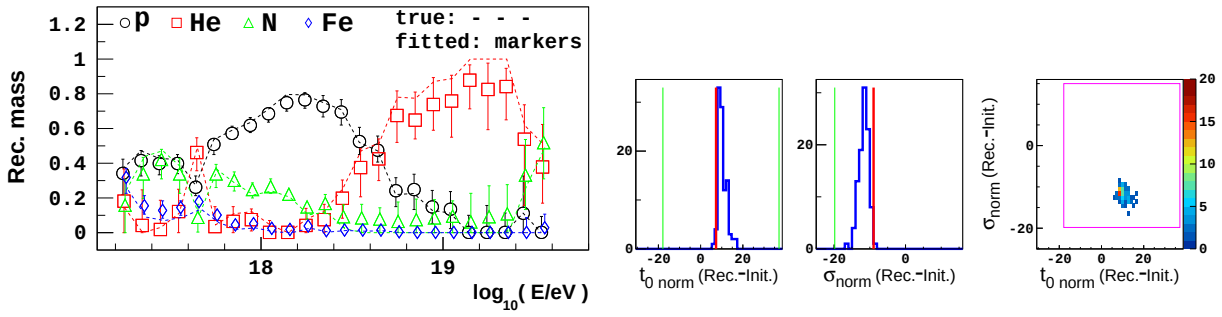
(a) Solution 1. (b)

Figure E.3: Fits of EPOS-LHC based $X_{\max}$ data with the Sibyll2.3 parameterisation.



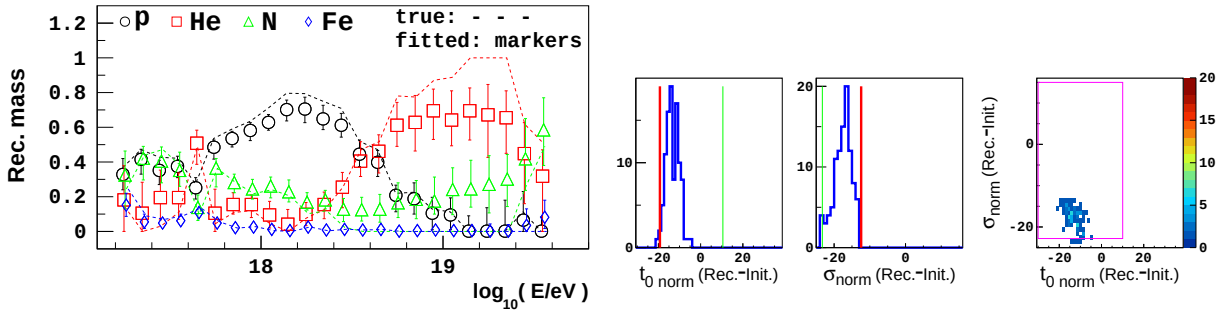
(a) (b)

Figure E.4: Fits of QGSJetII-04 based $X_{\max}$ data with the EPOS-LHC parameterisation.



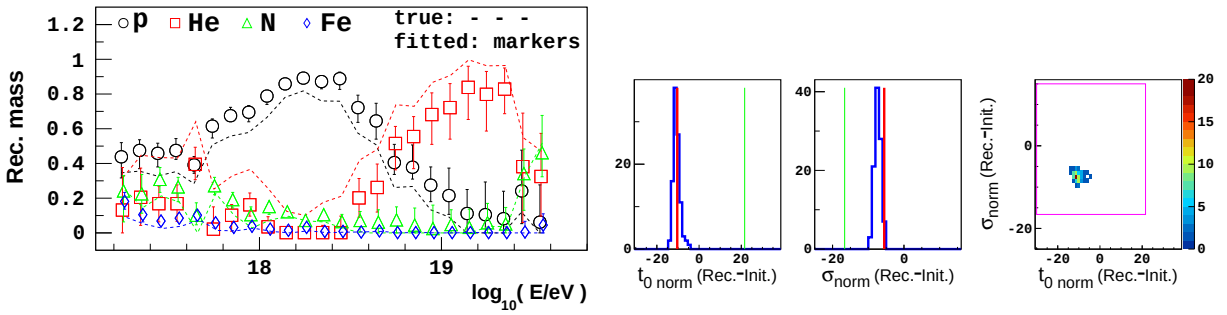
(a) (b)

Figure E.5: Fits of QGSJetII-04 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



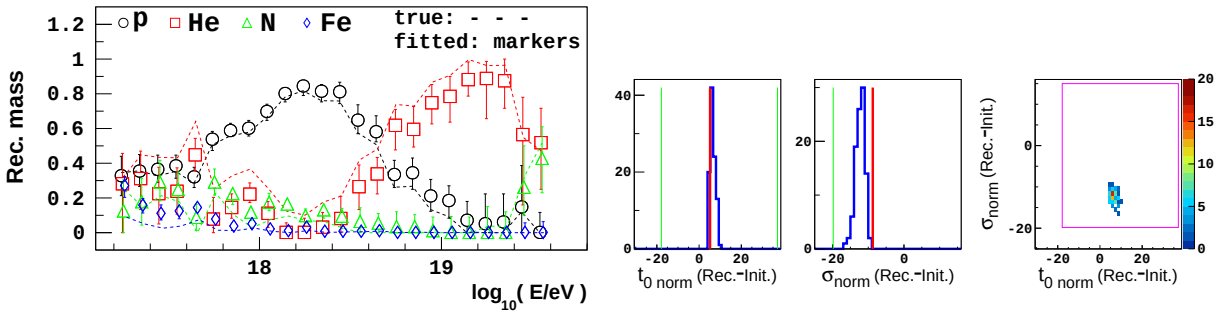
(a) Trimmed. (b)

Figure E.6: Fits of QGSJetII-04 based $X_{\max}$ data with the Sibyll2.3 parameterisation.



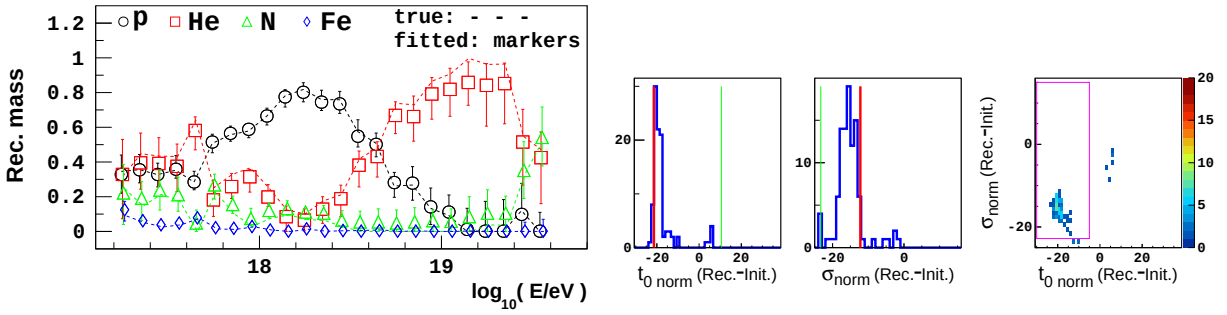
(a) (b)

Figure E.7: Fits of Sibyll2.3 based $X_{\max}$ data with the EPOS-LHC parameterisation.



(a) (b)

Figure E.8: Fits of Sibyll2.3 based $X_{\max}$ data with the QGSJetII-04 parameterisation.



(a) Solution 1. (b)

Figure E.9: Fits of Sibyll2.3 based $X_{\max}$ data with the Sibyll2.3 parameterisation.

Appendix F. $X_{\max}$ moments in terms of $\ln A$ moments.

The first two $X_{\max}$ moments can be parameterised in terms of $\ln A$ as follows [146]:

$$\langle X_{\max}\rangle = X_0 + D\log 10\left(\frac{E}{E_0 A}\right) + \xi \ln A \ + \delta \ln A \ \log 10\left(\frac{E}{E_0}\right), \tag{F.1}$$

and

$$\sigma^2(X_{\max}) = \sigma_p^2[1 + a\langle \ln A\rangle \ + b\langle(\ln A)^2\rangle], \tag{F.2}$$

where

$$\begin{aligned} \sigma_p^2 &= p_0 + p_1 \log 10\left(\frac{E}{E_0}\right) + p_2\left[\log 10\left(\frac{E}{E_0}\right)\right]^2, \\ a &= a_0 + a_1 \log 10\left(\frac{E}{E_0}\right). \end{aligned} \tag{F.3}$$

F.1 FD and HEAT fit parameters

The parameters of Equations (F.1), (F.2) and (F.3) are displayed in Table F.1 and F.2. The parameters were obtained by fitting the respective equations to the predicted $\langle X_{\max}\rangle$ and $\sigma^2(X_{\max})$ of proton, He, N and Fe from the $t_{0_{\text{norm}}}$ and σ_{norm} fit results of the FD and HEAT data in Section 8.1.2. Nine energy bins of width $\Delta \log 10(\text{E/eV}) = 0.25$ ranging from 10^{18} eV to 10^{20} eV were used, and $E_0 = 10^{19}$ eV. The statistical error obtained from the fit is also given in the tables. The mean and maximum $\langle X_{\max}\rangle$ residuals of the fit are $\sim 1\,\text{g/cm}^2$ and $\sim 2.5\,\text{g/cm}^2$ respectively. The mean and maximum $\sigma(X_{\max})$ residuals of the fit are $\sim 1.5\,\text{g/cm}^2$ and $\sim 3\,\text{g/cm}^2$ respectively.

parameter	EPOS-LHC	QGSJetII-04	Sibyll2.3
X_0	794.9 ± 0.3	796.6 ± 0.3	800.7 ± 0.4
D	54.8 ± 0.5	51.9 ± 0.4	59.1±0.6
ξ	-0.10 ± 0.26	-1.52 ± 0.20	0.09 ± 0.33
δ	0.83 ± 0.21	0.13 ± 0.16	1.20 ± 0.26

Table F.1: Parameters of Equation (F.1), obtained by fitting the predicted $\langle X_{\max}\rangle$ of the FD and HEAT $t_{0_{\text{norm}}}$ and σ_{norm} fit. All values are in g/cm^2.

parameter	EPOS-LHC	QGSJetII-04	Sibyll2.3
$p_0 \times \text{g}^{-2}\text{cm}^4$	3473 ± 43	3634 ± 55	3699 ± 64
$p_1 \times \text{g}^{-2}\text{cm}^4$	-319 ± 44	-389 ± 56	-356 ± 65
$p_2 \times \text{g}^{-2}\text{cm}^4$	80 ± 75	77 ± 94	96 ± 110
a_0	-0.527 ± 0.007	-0.481 ± 0.009	-0.484 ± 0.011
a_1	-0.0045 ± 0.0027	-0.0018 ± 0.0034	-0.0051 ± 0.0039
b	0.074 ± 0.002	0.065 ± 0.003	0.065 ± 0.003

Table F.2: Parameters of Equation (F.2) and Equation (F.3), obtained by fitting the predicted $\sigma^2(X_{\max})$ of the FD and HEAT $t_{0_{\text{norm}}}$ and σ_{norm} fit.

F.2 2014 FD fit parameters

Using the $t_{0_{\text{norm}}}$ and σ_{norm} fit results of the 2014 FD dataset (see Figures 8.15 and 8.16 in Section 8.2.1), the parameters of Equations (F.1), (F.2) and (F.3) have been determined, and are displayed in Table F.3 and F.4. The residuals are similar to those from the FD and HEAT results.

parameter	EPOS-LHC	QGSJetII-04	Sibyll2.3
X_0	793.8 ± 0.3	795.0 ± 0.3	825.4 ± 0.4
D	54.8 ± 0.5	51.9 ± 0.4	59.1 ± 0.6
ξ	-0.10 ± 0.26	-1.52 ± 0.20	0.09 ± 0.33
δ	0.83 ± 0.21	0.13 ± 0.16	1.20 ± 0.26

Table F.3: Parameters of Equation (F.1), obtained by fitting the predicted $\langle X_{\max} \rangle$ of the 2014 FD data set $t_{0_{\text{norm}}}$ and σ_{norm} fit. All values are in g/cm^2.

parameter	EPOS-LHC	QGSJetII-04	Sibyll2.3
$p_0 \times \text{g}^{-2}\text{cm}^4$	3536 ± 41	3681 ± 53	4198 ± 56
$p_1 \times \text{g}^{-2}\text{cm}^4$	-323 ± 42	-392 ± 54	-383 ± 57
$p_2 \times \text{g}^{-2}\text{cm}^4$	80 ± 71	77 ± 91	93 ± 95
a_0	-0.516 ± 0.007	-0.472 ± 0.009	-0.441 ± 0.009
a_1	-0.0045 ± 0.0026	-0.0017 ± 0.0033	-0.0045 ± 0.0030
b	0.071 ± 0.002	0.063 ± 0.003	0.056 ± 0.002

Table F.4: Parameters of Equation (F.2) and Equation (F.3), obtained by fitting the predicted $\sigma^2(X_{\max})$ of the 2014 FD data set $t_{0_{\text{norm}}}$ and σ_{norm} fit.

References

[1] V. F. Hess. Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten. *Physikalische Zeitschrift*, 13:1084–1091, November 1912.

[2] R.W. Clay and B.R. Dawson. *Cosmic Bullets*. Allen & Unwin, St Leonards, NSW, 1997.

[3] H. Elliot and Y. Sekido. *Early History of Cosmic Ray Studies: Personal Reminiscences with Old Photographs*. D. Reidel Publishing Company, Dordrecht, Holland, 1985.

[4] P. Auger and M.M. Shapiro. *What Are Cosmic Rays?* University of Chicago Press, Chicago, 1945.

[5] A.M. Hillas. *Cosmic Rays*. Pergamon Press, Oxford; New York, 1972.

[6] J. Cronin. *Proc. of 30th Int. Cosmic Ray Conf.*, 6:3–18, 2009.

[7] T. Stanev. *High Energy Cosmic Rays*. Praxis Publishing Ltd., Chichester, UK, 2004.

[8] P. Auger, P. Ehrenfest, R. Maze, J. Daudin, and R. A. Fréon. Extensive Cosmic-Ray Showers. *Reviews of Modern Physics*, 11:288–291, July 1939.

[9] T.K. Gaisser. *Cosmic rays and Particle Physics*. Cambridge University Press, Cambridge, New York, 1990.

[10] W. Heitler. *The Quantum Theory of Radiation*. Clarendon Press, Oxford, 3 edition, 1954.

[11] B. Whelan. A magnetic spectrometer analysis method for ultra high energy cosmic ray data. *Ph.D. thesis, University of Adelaide*, 2011.

[12] K. Nakamura et al. Review of particle physics. *J. Phys.*, G37:075021, 2010.

[13] A. Huangs, H. Rebel, and M. Roth. Energy spectrum and mass composition of high-energy cosmic rays. *Reports on Progress in Physics*, 66:1145–1206, 2003.

[14] J Blumer, R Engel, and J. R. Horandel. Cosmic Rays from the Knee to the Highest Energies. *Prog. Part. Nucl. Phys.*, 63:293–338, 2009.

[15] T. K. Gaisser and A. M. Hillas. Reliability of the method of constant intensity cuts for reconstructing the average development of vertical showers. *Proc. of 15th Int. Cosmic Ray Conf.*, 8:353–357, 1977.

[16] M. Unger, B. R. Dawson, R. Engel, F. Schussler, and R. Ulrich. Reconstruction of Longitudinal Profiles of Ultra-High Energy Cosmic Ray Showers from Fluorescence and Cherenkov Light Measurements. *Nucl. Instrum. Meth.*, A588:433–441, 2008.

[17] R. U. Abbasi et al. A Study of the composition of ultrahigh energy cosmic rays using the High Resolution Fly's Eye. *Astrophys. J.*, 622:910–926, 2005.

[18] M. Nagano and A. Watson. Observations and implications of the ultrahigh-energy cosmic rays. *Rev. Mod. Phys.*, 72:689–732, 2000.

[19] J. Matthews. A Heitler model of extensive air showers. *Astroparticle Physics*, 22:387–397, January 2005.

[20] P. Abreu et al. The Pierre Auger Observatory II: Studies of Cosmic Ray Composition and Hadronic Interaction models. 3:208, 2011.

[21] J. Abraham et al. Measurement of the Depth of Maximum of Extensive Air Showers above 10^{18} eV. *Phys. Rev. Lett.*, 104:091101, 2010.

[22] L Anchordoqui et al. High energy physics in the atmosphere: Phenomenology of cosmic ray air showers. *Annals Phys.*, 314:145–207, 2004.

[23] F. Arqueros, J. R. Hoerandel, and B. Keilhauer. Air Fluorescence Relevant for Cosmic-Ray Detection - Summary of the 5th Fluorescence Workshop, El Escorial 2007. *Nucl. Instrum. Meth.*, A597:1–22, 2008.

[24] J. Abraham et al. The Fluorescence Detector of the Pierre Auger Observatory. *Nucl. Instrum. Meth.*, A620:227–251, 2010.

[25] P Sokolsky, P. Sommers, and B. Dawson. Extremely high-energy cosmic rays. *Phys. Rept.*, 217:225–277, 1992.

[26] M. Ave et al. Measurement of the pressure dependence of air fluorescence emission induced by electrons. *Astropart. Phys.*, 28:41–57, 2007.

[27] T. K. Gaisser. The cosmic-ray spectrum: From the knee to the ankle. *J. Phys. Conf. Ser.*, 47:15–20, 2006.

[28] P. Sommers and S. Westerhoff. Cosmic Ray Astronomy. *New J. Phys.*, 11:055004, 2009.

[29] D. Newton, J. Knapp, and A. Watson. The Optimum Distance at which to Determine the Size of a Giant Air Shower. *Astropart. Phys.*, 26:414–419, 2007.

[30] A. Aab et al. The Pierre Auger Cosmic Ray Observatory. *Nucl. Instrum. Meth.*, A798:172–213, 2015.

[31] M. Ave. Reconstruction accuracy of the surface detector array of the Pierre Auger Observatory. *International Cosmic Ray Conference*, 4:307–310, 2008.

[32] C. Bonifazi. The angular resolution of the Pierre Auger Observatory. *Nucl. Phys. Proc. Suppl.*, 190:20–25, 2009.

[33] M. Ave et al. The accuracy of signal measurement with the water Cherenkov detectors of the Pierre Auger Observatory. *Nucl. Instrum. Meth.*, A578:180–184, 2007.

[34] A. M. Hillas et al. *Proc. of 12th Int. Cosmic Ray Conf.*, 3:1001–1006, 1971.

[35] M. A. Lawrence, R. J. O. Reid, and A. A. Watson. The Cosmic ray energy spectrum above $4\text{x}10^{17}$ eV as measured by the Haverah Park array. *J. Phys.*, G17:733–757, 1991.

[36] P. Abreu et al. The Pierre Auger Observatory I: The Cosmic Ray Energy Spectrum and Related Measurements. In *Proceedings, 32nd International Cosmic Ray Conference (ICRC 2011): Beijing, China, August 11-18, 2011*, 2011.

[37] B. R. Dawson. Hybrid Performance of the Pierre Auger Observatory. *Proc. of 30th Int. Cosmic Ray Conf.*, 4:425–428, 2008.

[38] K. Shinozaki and M. Teshima. AGASA results. *Nucl. Phys. Proc. Suppl.*, 136:18–27, 2004.

[39] J. Abraham et al. Studies of Cosmic Ray Composition and Air Shower Structure with the Pierre Auger Observatory. In *31st International Cosmic Ray Conference (ICRC 2009) Lodz, Poland, July 7-15, 2009*, 2009.

[40] M. T. Dova, M. E. Mancenido, A. G. Mariazzi, H. Wahlberg, F. Arqueros, and D. Garcia-Pinto. Time asymmetries in extensive air showers: a novel method to identify UHECR species. *Astropart. Phys.*, 31:312–319, 2009.

[41] I. Allekotte et al. The Surface Detector System of the Pierre Auger Observatory. *Nucl. Instrum. Meth.*, A586:409–420, 2008.

[42] J. Abraham et al. Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory. *Physics Letters B*, 685:239–246, March 2010.

[43] J. Abraham et al. Properties and performance of the prototype instrument for the Pierre Auger Observatory. *Nucl. Instrum. Meth.*, A523:50–95, 2004.

[44] J. Abraham et al. A Study of the Effect of Molecular and Aerosol Conditions in the Atmosphere on Air Fluorescence Measurements at the Pierre Auger Observatory. *Astropart. Phys.*, 33:108–129, 2010.

[45] J. Pekala, P. Homola, B. Wilczynska, and H. Wilczynski. Atmospheric multiple scattering of fluorescence and Cherenkov light emitted by extensive air showers. *Nucl. Instrum. Meth.*, A605:388–398, 2009.

[46] J. Belz. Overview of recent HiRes results. *Nucl. Phys. Proc. Suppl.*, 190:5–11, 2009.

[47] J. Blumer et al. The northern site of the Pierre Auger Observatory. *New J. Phys.*, 12:035001, 2010.

[48] P. Sommers. Capabilities of a giant hybrid air shower detector. *Astropart. Phys.*, 3:349–360, 1995.

[49] B. R. Dawson, H. Y. Dai, P. Sommers, and S. Yoshida. Simulations of a giant hybrid air shower detector. *Astropart. Phys.*, 5:239–247, 1996.

[50] P. Abreu et al. The exposure of the hybrid detector of the Pierre Auger Observatory. *Astropart. Phys.*, 34:368–381, 2011.

[51] P. Allison et al. Timing Calibration and Synchronization of Surface and Fluorescence Detectors of the Pierre Auger Observatory. *Proc. of 29th Int. Cosmic Ray Conf*, 8:307, 2005.

[52] P. Sokolsky and G. B. Thomson. Highest energy cosmic-rays and results from the HiRes experiment. *J. Phys. G*, 34(11):R401, 2007.

[53] W. F. Swann. A Mechanism of Acquirement of Cosmic-Ray Energies by Electrons. *Physical Review*, 43:217–220, February 1933.

[54] Enrico Fermi. On the Origin of the Cosmic Radiation. *Phys. Rev.*, 75:1169–1174, 1949.

[55] A. M. Hillas. The Origin of Ultra-High-Energy Cosmic Rays. *Ann. Rev. Astron. Astrophys.*, 22:425–444, 1984.

[56] P. Bhattacharjee and G. Sigl. Origin and propagation of extremely high-energy cosmic rays. *Phys. Rept.*, 327:109–247, 2000.

[57] D. F. Torres and L. A. Anchordoqui. *Rep. Prog. Phys.*, 67:1663–1730, 2004.

[58] R. D. Blandford and J. P. Ostriker. Particle Acceleration by Astrophysical Shocks. *Astrophys. J.*, 221:L29–L32, 1978.

[59] P. Sokolsky. *Introduction to ultrahigh energy cosmic ray physics.* 1989.

[60] Christopher T. Hill. Monopolonium. *Nucl. Phys.*, B224:469–490, 1983.

[61] C. T. Hill, D. N. Schramm, and T. P. Walker. Ultrahigh-Energy Cosmic Rays from Superconducting Cosmic Strings. *Phys. Rev.*, D36:1007, 1987.

[62] P. Bhattacharjee. Cosmic Strings and Ultrahigh-Energy Cosmic Rays. *Phys. Rev.*, D40:3968, 1989.

[63] P. Bhattacharjee and N. C. Rana. Ultrahigh-energy Particle Flux From Cosmic Strings. *Phys. Lett.*, B246:365–370, 1990.

[64] P. Bhattacharjee, C. T. Hill, and David N. Schramm. Grand unified theories, topological defects and ultrahigh-energy cosmic rays. *Phys. Rev. Lett.*, 69:567–570, 1992.

[65] P. Bhattacharjee and G. Sigl. Monopole annihilation and highest energy cosmic rays. *Phys. Rev.*, D51:4079–4091, 1995.

[66] P. Gondolo, G. Gelmini, and S. Sarkar. Cosmic neutrinos from unstable relic particles. *Nucl. Phys.*, B392:111–136, 1993.

[67] V. Berezinsky, M. Kachelriess, and A. Vilenkin. Ultrahigh-energy cosmic rays without GZK cutoff. *Phys. Rev. Lett.*, 79:4302–4305, 1997.

[68] V. A. Kuzmin and V. A. Rubakov. Ultrahigh-energy cosmic rays: A Window to postinflationary reheating epoch of the universe? *Phys. Atom. Nucl.*, 61:1028, 1998.

[69] V. Kuzmin and I. Tkachev. Ultrahigh-energy cosmic rays, superheavy long living particles, and matter creation after inflation. *JETP Lett.*, 68:271–275, 1998.

[70] T. W. Kephart and T. J. Weiler. Magnetic monopoles as the highest energy cosmic ray primaries. *Astropart. Phys.*, 4:271–279, 1996.

[71] S. Bonazzola and P. Peter. Can high-energy cosmic rays be vortons? *Astropart. Phys.*, 7:161–172, 1997.

[72] J. Abraham et al. Upper limit on the cosmic-ray photon flux above 10^{19} eV using the surface detector of the Pierre Auger Observatory. *Astropart. Phys.*, 29:243–256, 2008.

[73] J. Abraham et al. Upper limit on the cosmic-ray photon fraction at EeV energies from the Pierre Auger Observatory. *Astropart. Phys.*, 31:399–406, 2009.

[74] P. Bhattacharjee. Ultrahigh-energy cosmic rays from topological defects: Cosmic strings, monopoles, necklaces, and all that. 1997. [AIP Conf. Proc.433,168(1998)].

[75] V. A. Kuzmin and I. I. Tkachev. Ultrahigh-energy cosmic rays and inflation relics. *Phys. Rept.*, 320:199–221, 1999.

[76] V. Berezinsky. Ultrahigh energy cosmic rays. *Nucl. Phys. Proc. Suppl.*, 81:311–322, 2000.

[77] A. V. Olinto. Ultrahigh-energy cosmic rays: The Theoretical challenge. *Phys. Rept.*, 333:329–348, 2000.

[78] A. Aab et al. Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory. *JCAP*, 1704(04):038, 2017.

[79] D. Hooper and A. M. Taylor. On The Heavy Chemical Composition of the Ultra-High Energy Cosmic Rays. *Astropart. Phys.*, 33:151–159, 2010.

[80] N. Globus, D. Allard, and E. Parizot. A complete model of the cosmic ray spectrum and composition across the Galactic to extragalactic transition. *Phys. Rev.*, D92(2):021302, 2015.

[81] M. Unger, G. R. Farrar, and L. A. Anchordoqui. Origin of the ankle in the ultrahigh energy cosmic ray spectrum, and of the extragalactic protons below it. *Phys. Rev.*, D92(12):123001, 2015.

[82] D. F. Torres and L. A. Anchordoqui. Astrophysical origins of ultrahigh energy cosmic rays. *Rept. Prog. Phys.*, 67:1663–1730, 2004.

[83] F. Halzen and E. Zas. Neutrino fluxes from active galaxies: A Model independent estimate. *Astrophys. J.*, 488:669–674, 1997.

[84] E. Waxman. Cosmological origin for cosmic rays above 10^{19} eV. *Astrophys. J.*, 452:L1–L4, 1995.

[85] M. Vietri. On the acceleration of ultrahigh-energy cosmic rays in gamma-ray bursts. *Astrophys. J.*, 453:883–889, 1995.

[86] H. Kang, D. Ryu, and T. W. Jones. Cluster accretion shocks as possible acceleration sites for ultrahigh-energy protons below the Greisen cutoff. *Astrophys. J.*, 456:422–427, 1996.

[87] T. Stanev. Ultrahigh Energy Cosmic Rays: Origin and propagation. In *Proceedings, 30th International Cosmic Ray Conference (ICRC 2007): Merida, Yucatan, Mexico, July 3-11, 2007*, volume 6, pages 35–49, 2007.

[88] P. Blasi, R. I. Epstein, and A. V. Olinto. Ultrahigh-energy cosmic rays from young neutron star winds. *Astrophys. J.*, 533:L123, 2000.

[89] J. P. Rachen and P. L. Biermann. Extragalactic ultrahigh-energy cosmic rays. 1. Contribution from hot spots in FR-II radio galaxies. *Astron. Astrophys.*, 272:161–175, 1993.

[90] V. A. Acciari et al. A connection between star formation activity and cosmic rays in the starburst galaxy M 82. *Nature*, 462:770–772, 2009.

[91] P. P. Kronberg. Extragalactic magnetic fields. *Rept. Prog. Phys.*, 57:325–382, 1994.

[92] J. W. Cronin. The Highest - energy cosmic rays. *Nucl. Phys. Proc. Suppl.*, 138:465–491, 2005.

[93] T. Stanev. Ultra high energy cosmic rays: Origin and propagation. *Mod. Phys. Lett.*, A25:1467–1481, 2010.

[94] T. Stanev. Ultra-high-energy Cosmic Rays and the Large-scale Structure of the Galactic Magnetic Field. *Astrophys. J.*, 479:290–295, April 1997.

[95] F. W. Stecker and M. H. Salamon. Photodisintegration of ultrahigh-energy cosmic rays: A New determination. *Astrophys. J.*, 512:521–526, 1999.

[96] L. N. Epele and E. Roulet. On the propagation of the highest energy cosmic ray nuclei. *JHEP*, 10:009, 1998.

[97] V. Berezinsky, A. Z. Gazizov, and S. I. Grigorieva. On astrophysical solution to ultrahigh-energy cosmic rays. *Phys. Rev.*, D74:043005, 2006.

[98] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *Astrophys. J.*, 142:419–421, July 1965.

[99] K. Greisen. End to the cosmic ray spectrum? *Phys. Rev. Lett.*, 16:748–750, 1966.

[100] G. T. Zatsepin and V. A. Kuzmin. Upper limit of the spectrum of cosmic rays. *JETP Lett.*, 4:78–80, 1966.

[101] J. W. Cronin. Summary of the workshop. *Nucl. Phys. Proc. Suppl.*, 28B:213–226, 1992. [,213(1992)].

[102] D. Allard, N. G. Busca, G. Decerprit, A. V. Olinto, and E. Parizot. Implications of the cosmic ray spectrum for the mass composition at the highest energies. *JCAP*, 0810:033, 2008.

[103] J. R. Horandel. Cosmic rays from the knee to the second knee: 10^{17} to 10^{18} ev. *Modern Physics Letters A*, 22(21):1533–1551, 2007.

[104] T. K. Gaisser and T. Stanev. High-energy cosmic rays. *Nucl. Phys.*, A777:98–110, 2006.

[105] R. J. Protheroe and R. W. Clay. Ultrahigh energy cosmic rays. *Publ. Astron. Soc. Pac.*, 21(1):1–22, 2004.

[106] R. Aloisio, V. Berezinsky, P. Blasi, A. Gazizov, S. Grigorieva, and B. Hnatyk. A dip in the UHECR spectrum and the transition from galactic to extragalactic cosmic rays. *Astropart. Phys.*, 27:76–91, 2007.

[107] J. R. Horandel. On the knee in the energy spectrum of cosmic rays. *Astropart. Phys.*, 19:193–220, 2003.

[108] F. Francesco for the Pierre Auger Collaboration. The cosmic ray energy spectrum measured using the Pierre Auger Observatory. *PoS(ICRC2017)*, 486, 2017.

[109] R. U. Abbasi et al. Measurement of the Flux of Ultra High Energy Cosmic Rays by the Stereo Technique. *Astropart. Phys.*, 32:53–60, 2009.

[110] D. Ivanov, B. Stokes, and G. Thompson. Measurement of the energy spectrum by the telescope array surface detector. talk presented at 32nd Int. Cosmic Ray Conf.,Beijing, China., 2011.

[111] R. U. Abbasi et al. Observation of the ankle and evidence for a high-energy break in the cosmic ray spectrum. *Phys. Lett.*, B619:271–280, 2005.

[112] R. U. Abbasi et al. First observation of the Greisen-Zatsepin-Kuzmin suppression. *Phys. Rev. Lett.*, 100:101101, 2008.

[113] J. Abraham et al. Observation of the suppression of the flux of cosmic rays above 4×10^{19}eV. *Phys. Rev. Lett.*, 101:061101, 2008.

[114] V. Verzi, D. Ivanov, and Y. Tsunesada. Measurement of Energy Spectrum of Ultra-High Energy Cosmic Rays. 2017.

[115] K. Rawlins and T. Feusels. Latest Results on Cosmic Ray Spectrum and Composition from Three Years of IceTop and IceCube. *PoS(ICRC2015)*, 334, 2016.

[116] S. Knurenko, I. Petrov, Z. Petrov, and I. Sleptsov. The spectrum of cosmic rays in the energy range $10^{16} - 10^{18}$eV according to the Small Cherenkov Array in Yakutsk. *PoS(ICRC2015)*, 252, 2016.

[117] W. D. Apel et al. The spectrum of high-energy cosmic rays measured with KASCADE-Grande. *Astropart. Phys.*, 36:183–194, 2012.

[118] D. Ivanov. TA Spectrum Summary. *PoS(ICRC2015)*, 349, 2016.

[119] I. Valino. The flux of ultra-high energy cosmic rays after ten years of operation of the Pierre Auger Observatory. *PoS(ICRC2015)*, 271, 2016.

[120] M. S. Longair. *High energy astrophysics. Vol.1: Particles, photons and their detection.* March 1992.

[121] M. V. Zombeck. *Handbook of space astronomy and astrophysics.* 1990.

[122] J. J. Engelmann, P. Ferrando, A. Soutoul, P. Goret, and E. Juliusson. Charge composition and energy spectra of cosmic-ray for elements from Be to NI - Results from HEAO-3-C2. *Astron. Astrophys.*, 233:96–111, 1990.

[123] J. Alcaraz et al. Cosmic protons. *Phys. Lett.*, B490:27–35, 2000.

[124] J. Alcaraz et al. Helium in near Earth orbit. *Phys. Lett.*, B494:193–202, 2000.

[125] W. D. Apel et al. Energy Spectra of Elemental Groups of Cosmic Rays: Update on the KASCADE Unfolding Analysis. *Astropart. Phys.*, 31:86–91, 2009.

[126] M. Unger. Study of the Cosmic Ray Composition above 0.4 EeV using the Longitudinal Profiles of Showers observed at the Pierre Auger Observatory. *International Cosmic Ray Conference*, 4:373–376, 2008.

[127] J. W. Fowler, L. F. Fortson, C. C. H. Jui, D. B. Kieda, R. A. Ong, C. L. Pryke, and P. Sommers. A Measurement of the cosmic ray spectrum and composition at the knee. *Astropart. Phys.*, 15:49–64, 2001.

[128] S. Paling et al. Results from the CACTI Experiment: Air Cerenkov and Particle Measurements of PeV Air Showers at Los Alamos. *International Cosmic Ray Conference*, 5:253, 1997.

[129] S. P. Swordy and D. B. Kieda. Elemental composition of cosmic rays near the knee by multiparameter measurements of air showers. *Astropart. Phys.*, 13:137–150, 2000.

[130] D. J. Bird et al. The Cosmic ray energy spectrum observed by the Fly's Eye. *Astrophys. J.*, 424:491–502, 1994.

[131] A. A. Watson. Ultra-high-energy cosmic rays: the experimental situation. , 333:309–327, August 2000.

[132] F. Arqueros et al. Energy spectrum and chemical composition of cosmic rays between 0.3-PeV and 10-PeV determined from the Cherenkov light and charged particle distributions in air showers. *Astron. Astrophys.*, 359:682–694, 2000.

[133] T. Abu-Zayyad et al. Evidence for Changing of Cosmic Ray Composition between 10^{17} eV and 10^{18} eV from Multicomponent Measurements. *Phys. Rev. Lett.*, 84:4276–4279, 2000.

[134] M. Cha et al. *Proc. of 27th Int. Cosmic Ray Conf.*, 1(132), 2001.

[135] J. Dickinson et al. *Proc. of 26th Int. Cosmic Ray Conf.*, 3(136), 1999.

[136] D. Chernov et al. Primary energy spectrum and mass composition determined with the Tunka EAS Cerenkov Array. *Int. J. Mod. Phys.*, A20:6799–6801, 2006.

[137] S Knurenko, V Kolosov, Z Petrov, I Sleptsov, and S Starostin. Cerenkov radiation of cosmic ray extensive air showers. part 1. lateral distribution in the energy region of 1015÷ 1017 ev. In *Proceedings of ICRC*, volume 1, 2001.

[138] N. N. Kalmykov and S. S. Ostapchenko. Comparison of Nucleus- Nucleus Interaction Characteristics in the Model of Quark - Gluon Strings and in the Superposition Model. (In Russian). *Sov. J. Nucl. Phys.*, 50:315–318, 1989. [Yad. Fiz.50,509(1989)].

[139] K. Werner, F. Liu, and T. Pierog. Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron-gold collisions at RHIC. *Phys. Rev.*, C74:044902, 2006.

[140] S. Ostapchenko. Nonlinear screening effects in high energy hadronic interactions. *Phys. Rev.*, D74(1):014026, 2006.

[141] R. Engel. Air Shower Calculations With the New Version of SIBYLL. *International Cosmic Ray Conference*, 1:415, 1999.

[142] A. Herve. Measurement of the Mass Composition of the Highest Energy Cosmic Rays with the Pierre Auger Observatory. *Ph.D. thesis, University of Adelaide*, 2013.

[143] R. U. Abbasi et al. Indications of Proton-Dominated Cosmic Ray Composition above 1.6 EeV. *Phys. Rev. Lett.*, 104:161101, 2010.

[144] D. J. Bird et al. Evidence for correlated changes in the spectrum and composition of cosmic rays at extremely high-energies. *Phys. Rev. Lett.*, 71:3401–3404, 1993.

[145] A. Aab et al. Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV. *Phys. Rev.*, D90(12):122005, 2014.

[146] P. Abreu et al. Interpretation of the Depths of Maximum of Extensive Air Showers Measured by the Pierre Auger Observatory. *JCAP*, 1302:026, 2013.

[147] A. Aab et al. Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications. *Phys. Rev.*, D90(12):122006, 2014.

[148] J. Bellido for the Pierre Auger Collaboration. Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above $10^{17.2}$eV and Composition Implications. *PoS(ICRC2017)*, 506, 2017.

[149] R. U. Abbasi et al. Study of Ultra-High Energy Cosmic Ray composition using Telescope Arrays Middle Drum detector and surface array in hybrid mode. *Astropart. Phys.*, 64:49–62, 2015.

[150] V. De Souza for the Pierre Auger Collaboration and Telescope Array Collaboration. Testing the agreement between the Xmax distributions measured by the Pierre Auger and Telescope Array Observatories. *PoS(ICRC2017)*, 522, 2017.

[151] K. Shinozaki. AGASA results. *Nucl. Phys. Proc. Suppl.*, 151:3–10, 2006.

[152] I. Valino. Measurements of the muon content of air showers at the Pierre Auger Observatory. *J. Phys. Conf. Ser.*, 632(1):012103, 2015.

[153] S. Ostapchenko. QGSJET-II: Towards reliable description of very high energy hadronic interactions. *Nucl. Phys. Proc. Suppl.*, 151:143–146, 2006.

[154] A. A. Watson and J. G. Wilson. Fluctuation studies of large air showers: the composition of primary cosmic ray particles of energy $E_p \sim 10^{18}$ eV. *J. Phys. A*, 7:1199–1212, July 1974.

[155] R. Walker and A. A. Watson. Measurement of the elongation rate of extensive air showers produced by primary cosmic rays of energy above 2×10^{17} eV. *J. Phys. G*, 7:1297–1309, September 1981.

[156] M. D. Ave Pernas, M. Ave, J. Knapp, M. Marchesini, M. Roth, and A. A. Watson. Time Structure of the Shower Front as Measured at Haverah Park above 10^{19}eV. *International Cosmic Ray Conference*, 1:349, July 2003.

[157] X. Bertou and P. Billoir. On the Origin of the Asymmetry of Ground Densities in Inclined Showers. *Pierre Auger GAP-Notes*, 2000-017, 2000.

[158] A. Aab et al. Azimuthal Asymmetry in the Risetime of the Surface Detector Signals of the Pierre Auger Observatory. *Phys. Rev.*, D93(7):072006, 2016.

[159] T. Antoni et al. KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems. *Astropart. Phys.*, 24:1–25, 2005.

[160] M. Bertaina et al. KASCADE-Grande: An overview and first results. *Nucl. Instrum. Meth.*, A588:162–165, 2008.

[161] A. Aab et al. Muons in air showers at the Pierre Auger Observatory: Measurement of atmospheric production depth. *Phys. Rev.*, D90(1):012012, 2014. [Erratum: Phys. Rev.D92,no.1,019903(2015)].

[162] P. Sanchez-Lucas for the Pierre Auger Collaboration. X_{max} measurements and tests of hadronic models using the surface detector of the Pierre Auger Observatory. *PoS(ICRC2017)*, 495, 2017.

[163] M. Mallamaci for the Pierre Auger Collaboration. Measurements of the depth of maximum muon production and of its fluctuations in extensive air showers above 1.510 19 eV at the Pierre Auger Observatory. *PoS(ICRC2017)*, 509, 2017.

[164] M. Unger for the Pierre Auger Collaboration. Highlights from the Pierre Auger Observatory. *PoS(ICRC2017)*, 1102, 2017.

[165] C. Bleve. Updates on the neutrino and photon limits from the Pierre Auger Observatory. *PoS(ICRC2015)*, 1103, 2016.

[166] F. Aharonian et al. Very high-energy gamma rays from the direction of Sagittarius A*. *Astron. Astrophys.*, 425:L13–L17, 2004.

[167] P. Abreu et al. Search for First Harmonic Modulation in the Right Ascension Distribution of Cosmic Rays Detected at the Pierre Auger Observatory. *Astropart. Phys.*, 34:627–639, 2011.

[168] H. Lyberis for the Pierre Auger Collaboration. Analysis of the modulation in the first harmonic analysis of the right ascension distribution of cosmic rays detected at the Pierre Auger Observatory. In *Proceedings, 32nd International Cosmic Ray Conference (ICRC 2011): Beijing, China, August 11-18, 2011*, volume 1, page 255, 2011.

[169] A. Aab et al. Observation of a Large-scale Anisotropy in the Arrival Directions of Cosmic Rays above 8×10^{18} eV. *Science*, 357:1266–1270, 2017.

[170] P. Tinyakov et al. TA Anisotropy Summary. *PoS(ICRC2015)*, 326, 2015.

[171] D. S. Gorbunov, P. G. Tinyakov, I. I. Tkachev, and Sergey V. Troitsky. Evidence for a connection between gamma-ray and highest energy cosmic ray emissions by BL Lacs. *Astrophys. J.*, 577:L93, 2002.

[172] R. U. Abbasi et al. Search for cross-correlations of ultrahigh-energy cosmic rays with BL Lacertae objects. *Astrophys. J.*, 636:680–684, 2006.

[173] D. Harari. Search for correlation of UHECRs and BL Lacs in Pierre Auger Observatory data. *Proc. of 30th Int. Cosmic Ray Conf.*, 4:283–286, 2008.

[174] M. P. Veron-Cetty and P. Veron. A catalogue of quasars and active nuclei: 12th edition. *Astron. Astrophys.*, 455:773–777, 2006.

[175] J. Abraham et al. Correlation of the highest energy cosmic rays with nearby extragalactic objects. *Science*, 318:938–943, 2007.

[176] J. Abraham et al. Correlation of the highest-energy cosmic rays with the positions of nearby active galactic nuclei. *Astropart. Phys.*, 29:188–204, 2008. [Erratum: Astropart. Phys.30,45(2008)].

[177] R. U. Abbasi et al. Search for Correlations between HiRes Stereo Events and Active Galactic Nuclei. *Astropart. Phys.*, 30:175–179, 2008.

[178] P. Abreu et al. Update on the correlation of the highest energy cosmic rays with nearby extragalactic matter. *Astropart. Phys.*, 34:314–326, 2010.

[179] A. Aab et al. Searches for Anisotropies in the Arrival Directions of the Highest Energy Cosmic Rays Detected by the Pierre Auger Observatory. *Astrophys. J.*, 804(1):15, 2015.

[180] U. Giaccari for the Pierre Auger Collaboration. Arrival directions of the highest-energy cosmic rays detected by the Pierre Auger Observatory. *PoS(ICRC2017)*, 483, 2017.

[181] The Pierre Auger Collaboration. Indication of anisotropy in arrival directions of ultra-high-energy cosmic rays through comparison to flux pattern of extragalactic gamma-ray sources. *Submitted to Astrophys. J.*, 2017.

[182] J. Linsley, L. Scarsi, and B. Rossi. Extremely energetic cosmic-ray event. *Phys. Rev. Lett.*, 6:485–487, 1961.

[183] J. Linsley and L. Scarsi. Arrival times of air shower particles at large distances from the axis. *Phys. Rev.*, 128:2384–2392, 1962.

[184] J. Linsley. Evidence for a primary cosmic-ray particle with energy 10^{20}eV. *Phys. Rev. Lett.*, 10:146–148, 1963.

[185] R. M. Tennent. The Haverah Park extensive air shower array. *Proc. Phys. Soc.*, 92:622–631, November 1967.

[186] P. R. Blake, H. Ferguson, and W. F. Nash. Current studies of extensive air showers by the nottingham group at haverah park. *J. Phys. Soc. Jap.*, 33(5):1197–1206, 1972.

[187] R. N. Coy, C. D. England, D. Pearce, R. J. O. Reid, and A. A. Watson. The average properties and fluctuations of the lateral distribution of large air showers. *Proc. of 17th Int. Cosmic Ray Conf*, 6:43–46, 1981.

[188] J. A. Bellido, R. W. Clay, B. R. Dawson, and M. Johnston-Hollitt. Southern hemisphere observations of a 10^{18}eV cosmic ray source near the direction of the galactic center. *Astropart. Phys.*, 15:167–175, 2001.

[189] M. M. Winn, J. Ulrichs, L. S. Peak, C. B. A. Mccusker, and L. Horton. The Arrival Directions of Cosmic Rays Above 10^{17}eV. *J. Phys.*, G12:675–686, 1986.

[190] C. J. Bell et al. The upper end of the observed cosmic ray energy spectrum. *J. Phys.*, A7:990–1009, 1974.

[191] M. M. Winn, J. Ulrichs, L. S. Peak, C. B. A. Mccusker, and L. Horton. The Cosmic Ray Energy Spectrum Above 10^{17}eV. *J. Phys.*, G12:653–674, 1986.

[192] Lisa J. Kewley, R. W. Clay, and B. R. Dawson. Arrival directions of the southern highest energy cosmic rays. *Astropart. Phys.*, 5:69–74, 1996.

[193] A. A. Ivanov, S. P. Knurenko, and I. Ye Sleptsov. Measuring extensive air showers with Cherenkov light detectors of the Yakutsk array: The energy spectrum of cosmic rays. *New J. Phys.*, 11:065008, 2009.

[194] A. A. Ivanov, S. P. Knurenko, Z. E. Petrov, M. I. Pravdin, and I. Ye. Sleptsov. Enhancement of the Yakutsk array by atmospheric Cherenkov telescopes to study cosmic rays above 10^{15} eV. *Astrophys. Space Sci. Trans.*, 6:53–57, 2010.

[195] P. Sokolsky. Results from Fly's Eye and HiRes projects. In J. F. Krizmanic, J. F. Ormes, and R. E. Streitmatter, editors, *Workshop on Observing Giant Cosmic Ray Air Showers From $10^{20}eV$ Particles From Space*, volume 433 of *"AIP Conf. Proc."*, pages 65–75, June 1998.

[196] R. M. Baltrusaitis et al. The Utah Fly's Eye Detector. *Nucl. Instrum. Meth.*, A240:410–428, 1985.

[197] D. J. Bird et al. Results from the fly's eye experiment. In S. J. Seestrom, editor, *Intersections between Particle and Nuclear Physics*, volume 338 of *"AIP Conf. Proc."*, pages 839–854, July 1995.

[198] D. J. Bird et al. Detection of a cosmic ray with measured energy well beyond the expected spectral cutoff due to cosmic microwave radiation. *Astrophys. J.*, 441:144–150, 1995.

[199] H. Ohoka, S. Yoshida, and M. Takeda. Further development of data acquisition system of AGASA. *Nucl. Instrum. Meth.*, A385:268–276, 1997.

[200] M. Takeda et al. Extension of the cosmic ray energy spectrum beyond the predicted Greisen-Zatsepin-Kuz'min cutoff. *Phys. Rev. Lett.*, 81:1163–1166, 1998.

[201] N. Hayashida et al. Observation of a very energetic cosmic ray well beyond the predicted 2.7-K cutoff in the primary energy spectrum. *Phys. Rev. Lett.*, 73:3491–3494, 1994.

[202] P. Sokolsky and G. B. Thomson. Highest Energy Cosmic Rays and results from the HiRes Experiment. *J. Phys.*, G34:R401, 2007.

[203] K. Shinozaki. Observation of UHECRs by AGASA. Talk presented at the Quarks-2006 Conf., St. Petersburg, Russia, 2006.

[204] N. Hayashida et al. The Anisotropy of cosmic ray arrival directions around 10^{18}eV. *Astropart. Phys.*, 10:303–311, 1999.

[205] M. Takeda et al. Small-scale anisotropy of cosmic rays above 10^{19}eV observed with the akeno giant air shower array. *Astrophys. J.*, 522:225–237, 1999.

[206] R. U. Abbasi et al. Measurement of the flux of ultrahigh energy cosmic rays from monocular observations by the High Resolution Fly's Eye experiment. *Phys. Rev. Lett.*, 92:151101, 2004.

[207] H. Kawai, S. Yoshida, H Yoshii, et al. Telescope Array Experiment. *Nuclear Physics B - Proceedings Supplements*, 175-176:221–226, 2008.

[208] T. Abu-Zayyad et al. The surface detector array of the Telescope Array experiment. *Nucl. Instrum. Meth.*, A689:87–97, 2013.

[209] H. Sagawa. Highlights from the Telescope Array Experiment. *Braz. J. Phys.*, 44:589–599, 2014.

[210] A. Aab et al. The Pierre Auger Observatory Upgrade - Preliminary Design Report. arXiv:1604.03637, 2016.

[211] J. Abraham et al. Trigger and aperture of the surface detector array of the Pierre Auger Observatory. *Nucl. Instrum. Meth.*, A613:29–39, 2010.

[212] D. Allard et al. Aperture calculation of the Pierre Auger Observatory surface detector. In *29th International Cosmic Ray Conference (ICRC 2005) Pune, India, August 3-11, 2005*, page 71, 2005.

[213] J. Alberto Abraham et al. Operations of and Future Plans for the Pierre Auger Observatory. In *31st International Cosmic Ray Conference (ICRC 2009) Lodz, Poland, July 7-15, 2009*, 2009.

[214] X. Bertou et al. Calibration of the surface array of the Pierre Auger Observatory. *Nucl. Instrum. Meth.*, A568:839–846, 2006.

[215] D. Allard et al. The trigger system of the Pierre Auger Surface Detector: operation, efficiency and stablility. In *29th International Cosmic Ray Conference (ICRC 2005) Pune, India, August 3-11, 2005*, page 287, 2005.

[216] J. Hersil, I. Escobar, D. Scott, G. Clark, and S. Olbert. Observations of Extensive Air Showers near the Maximum of Their Longitudinal Development. *Phys. Rev. Lett.*, 6:22–23, 1961.

[217] A. Aab et al. The Pierre Auger Observatory: Contributions to the 33rd International Cosmic Ray Conference (ICRC 2013). In *Proceedings, 33rd International Cosmic Ray Conference (ICRC2013): Rio de Janeiro, Brazil, July 2-9, 2013*, 2013.

[218] J. Abraham et al. The Cosmic Ray Energy Spectrum and Related Measurements with the Pierre Auger Observatory. In *31st International Cosmic Ray Conference (ICRC 2009) Lodz, Poland, July 7-15, 2009*, 2009.

[219] P. Abreu et al. The Pierre Auger Observatory V: Enhancements. In *Proceedings, 32nd International Cosmic Ray Conference (ICRC 2011): Beijing, China, August 11-18, 2011*, 2011.

[220] R. Engel. Upgrade of the Pierre Auger Observatory (AugerPrime). *PoS(ICRC2015)*, 686, 2016.

[221] J.M. Figueira for the Pierre Auger Collaboration. An improved reconstruction method for the AMIGA detectors. *PoS(ICRC2017)*, 396, 2017.

[222] L. Wiencke. The Pierre Auger Observatory and interdisciplinary science. *Eur. Phys. J. Plus*, 127:98, 2012.

[223] P. Abreu et al. The Pierre Auger Observatory scaler mode for the study of solar activity modulation of galactic cosmic rays. *JINST*, 6:P01003, 2011.

[224] E. Ruigrok et al. Malargue seismic array: Design and deployment of the temporary array. *Eur. Phys. J. Plus*, 127:126, 2012.

[225] J. Schulz. Status and Prospects of the Auger Engineering Radio Array. *PoS(ICRC2015)*, 615, 2016.

[226] A. Aab et al. The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017). In *Proceedings, 35th International Cosmic Ray Conference*, 2017.

[227] T. K. Gaisser and A. M. Hillas. Reliability of the method of constant intensity cuts for reconstructing the average development of vertical showers. *International Cosmic Ray Conference*, 8:353–357, 1977.

[228] K.-H. Kampert and M. Unger. Measurements of the cosmic ray composition with air shower experiments. *Astroparticle Physics*, 35:660–678, May 2012.

[229] P. Abreu and others. Measurement of the Proton-Air Cross Section at $\sqrt{s} = 57$ TeV with the Pierre Auger Observatory. *Physical Review Letters*, 109(6):062002, August 2012.

[230] T. Bergmann et al. One-dimensional hybrid approach to extensive air shower simulation. *Astropart. Phys.*, 26:420–432, 2007.

[231] T. Pierog et al. First results of fast one-dimensional hybrid simulation of eas using conex. *Nucl. Phys. Proc. Suppl.*, 151:159–162, 2006.

[232] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner. EPOS LHC : test of collective hadronization with LHC data. arXiv:1306.0121, June 2013.

[233] S. Ostapchenko. Monte carlo treatment of hadronic interactions in enhanced pomeron scheme: Qgsjet-ii model. *Phys. Rev. D*, 83:014018, Jan 2011.

[234] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev. A new version of the event generator Sibyll. *PoS(ICRC2015)*, 558, 2016.

[235] R. U. Abbasi and G. B. Thomson. $< X_{max} >$ Uncertainty from Extrapolation of Cosmic Ray Air Shower Parameters. arXiv:1605.05241, May 2016.

[236] T. Pierog. Air shower simulation with a new (first) generation of post-lhc hadronic interaction models in corsika. pages talk presented at 35th Int. Cosmic Ray Conf., Busan, Korea., 2017.

[237] A. Aab et al. Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory. *JCAP*, 1704(04):038, 2017.

[238] C. J. Todero Peixoto, V. de Souza, and J. Alfredo Bellido. Comparison of the moments of the X_{max} distribution predicted by different cosmic ray shower simulation models. *Astroparticle Physics*, 47:18–30, July 2013.

[239] M. Williams. How good are your fits? Unbinned multivariate goodness-of-fit tests in high energy physics. *JINST*, 5:P09004, 2010.

[240] B. Peters. Primary cosmic radiation and extensive air showers. *Il Nuovo Cimento (1955-1965)*, 22(4):800–819, 1961.

[241] V. S. Berezinsky and S. I. Grigor'eva. A Bump in the ultrahigh-energy cosmic ray spectrum. *Astron. Astrophys.*, 199:1–12, 1988.

[242] Pierre Auger Collaboration. Large-scale Distribution of Arrival Directions of Cosmic Rays Detected Above 10^{18} eV at the Pierre Auger Observatory. *Astrophys. J Suppl.*, 203:34, December 2012.

[243] J. Linsley. *Proc. 8th ICRC, Bombay*, 4(77), 1963.

[244] D. Allard, E. Parizot, E. Khan, S. Goriely, and A. V. Olinto. UHE nuclei propagation and the interpretation of the ankle in the cosmic-ray spectrum. *Astron. Astrophys.*, 443:L29–L32, 2005.

[245] A. Yushkov, M. Risse, M. Werner, and J. Krieg. Determination of the proton-to-helium ratio in cosmic rays at ultra-high energies from the tail of the X_{max} distribution. *Astroparticle Physics*, 85:29–34, 2016.