

**UNIVERSIDADE DE LISBOA  
INSTITUTO SUPERIOR TÉCNICO**

**Measurement of the longitudinal profile of  
cosmic ray air-showers at the Pierre Auger  
Observatory**

**Francisco Gonçalves Dias Cardoso Diogo**

**Supervisor:** Doctor Sofia Andringa Dias

**Co-Supervisor:** Doctor Mário João Martins Pimenta

Thesis specifically prepared to obtain the PhD Degree in

**Physics**

**Draft**

**January 2017**



---

## Abstract

The composition of the flux of Ultra-High Energy Cosmic Rays (UHECR) in terms of primary particles is still unknown, as are the properties of their hadronic interactions. As direct detection is currently impossible, both must be inferred through the study of air showers induced by them in the Earth's atmosphere.

The Pierre Auger Observatory, located in the Argentinean Pampa, is the largest cosmic ray detector in the world, with an unprecedented level of precision and statistical significance. It is a hybrid detector, comprising a surface array detecting particles that reach ground level and fluorescence telescopes which track the longitudinal development of showers in the atmosphere. The maximum of this profile,  $X_{\max}$ , is the main composition variable in this field. Comparison of data to the simulation of different primaries, relying on standard hadronic models, indicates a light-intermediate nuclear mass composition. However, other independent measurements, mostly related to the shower muonic component, point to a heavier nuclei composition, which indicates that our models of shower development are not fully consistent.

This work presents the first measurement of the longitudinal profiles of high energy cosmic rays, above  $10^{17.8}$  eV. Its two shape parameters (L and R), in addition to  $X_{\max}$  and  $N_{\max}$ , fully characterize the longitudinal profile, and can be interpreted in terms of primary particle interaction properties and composition. The profile shapes for different energy ranges are all well reproduced by a Gaisser-Hillas function within the range studied. A detailed analysis of the systematic uncertainties is performed using data and a full detector simulation. The parameter L is fully compatible with model predictions at all energies, while R increases with energy more than expected, reaching a value incompatible with any of the considered models at the highest energy. These results are compared to those obtained with other variables and with alternative interaction models. Finally, a study of the effect of changes to the main hadronic parameters (multiplicity, elasticity and charge ratio) is performed using an implementation of the Heitler-Matthews model with added components.

**Keywords:** Ultra High Energy Cosmic Rays; Pierre Auger Observatory; Primary mass composition; Longitudinal shower profile; Hadronic interaction models



---

## Resumo

A composição do fluxo dos Raios Cósmicos de Energia Extrema é ainda desconhecida, tal como o são as propriedades das suas interacções hadrónicas. Sendo a detecção directa actualmente impossível, ambos têm de ser inferidos através do estudo das cascatas induzidas na atmosfera terrestre pelos raios cósmicos .

O Observatório Pierre Auger é o maior detector de raios cósmicos no mundo, com uma precisão e significância estatística inauditas. É composto por um detector de superfície que mede as partículas que chegam ao solo e detectores de fluorescência regista o desenvolvimento longitudinal das cascatas na atmosfera. O máximo deste perfil,  $X_{\max}$ , é a principal variável de composição neste campo. A comparação das dados a simulações de diferentes partículas, baseadas em modelos hadrónicos, indica uma composição nuclear de massa leve-média. No entanto, outras medidas independentes, maioritariamente relacionadas com a componente muónica das cascatas, apontam para uma composição nuclear mais pesada, mostrando que o desenvolvimento das cascatas previstas pelos modelos actuais ainda não é totalmente consistente.

Neste trabalho apresentamos a primeira medida dos perfis longitudinais de raios cósmicos de energias acima de  $10^{17.8}$  eV. Os seus parâmetros de forma, L e R, em adição a  $X_{\max}$  e uma normalização  $N_{\max}$ , caracterizam completamente o perfil longitudinal, e podem ser interpretados em termos da composição e propriedades de interacção dos raios cósmicos de alta energia. Os perfis são todos bem reproduzidas pela função Gaisser-Hillas. Uma análise detalhada das incertezas sistemáticas associadas é feita usando dados e uma simulação completa do detector. O parâmetro L é compatível com as previsões dos modelos a todas as energias, enquanto R cresce com a energia mais do que previsto, atingindo um valor incompatível com qualquer um dos modelos considerados nas energias mais altas. Estes resultados são comparados com aqueles obtidos para outras variáveis e com as previsões de modelos hadrónicos alternativos. Finalmente, um estudo do efeito induzido por mudanças nos principais parâmetros hadrónicos (multiplicidade, elasticidade e rácio de piões carregados) é efectuado usando uma implementação do modelo Heitler-Matthews com componentes adicionais.

**Palavras-chave:** Raios Cósmicos de Energia Extrema; Observatório Pierre Auger; Composição Nuclear Primária; Perfil Longitudinal; Modelos de interacções hadrónicas



---

## Acknowledgments

First and foremost, I would like to thank Sofia Andringa for accepting me as a PhD student. It is the one person without whom this thesis would certainly not exist. Thanks for all the ideas, patience, and essentially for teaching me how to be a physicist.

To Mario Pimenta, for forcing me to think about the big picture when I was too focused in the details of my analysis. Also, I cannot thank him enough for the opportunity to be in the Auger group at LIP and particularly for providing me opportunities to go to summer schools and conferences, even in the early stages of my work. It was a great learning experience.

Thanks to all my work colleagues in the Auger group at LIP. In particular, thanks to Ruben Conceição, who is an integral part of the analysis shown here. Many of the ideas presented here are his, and he was always available for help whenever I was stuck. Also, to Lorenzo Cazon for always supporting my work and believing in its importance, both as a colleague and as the task leader of the Air-Shower Physics group at Auger. Moreover thanks to Pedro Abreu, for always providing a wonderful working environment in our office, Fernando Barão, for accepting to be in my CAT jury and all the suggestions, Pedro Assis, Bernardo, Catarina, Patricia, Eva, João, Micael and Raul.

Also, to all my Auger colleagues that supported me through helpful commentaries in all the meetings.

Finally, many thanks to my family and friends who have provided support and encouragement throughout these last four years.



# Contents

<b>Abstract</b>	i
<b>Resumo</b>	iii
<b>Acknowledgments</b>	v
<b>Introduction</b>	1
<b>1 Air Shower Physics</b>	5
1.1 Ultra High Energy Cosmic Rays . . . . .	6
1.1.1 Energy spectrum . . . . .	6
1.1.2 Acceleration mechanisms . . . . .	10
1.1.3 Arrival directions . . . . .	12
1.1.4 Mass composition . . . . .	13
1.2 Extensive Air Shower Phenomenology . . . . .	15
1.2.1 Electromagnetic shower development . . . . .	16
1.2.2 Hadronic shower development . . . . .	19
1.2.3 Extensive Air shower simulations . . . . .	21
1.2.4 Shower detection techniques . . . . .	26
<b>2 Pierre Auger Observatory</b>	33
<b>3 Longitudinal Profile Shape</b>	35
3.1 Longitudinal profile shape . . . . .	35
3.1.1 Electromagnetic profile parametrization - Gaisser-Hillas . . . . .	37
3.2 Profile shape in shower age vs atmospheric depth . . . . .	40
3.2.1 What is average age measuring? . . . . .	46
3.3 Average profile shape - sensitivity to composition and hadronic models	48
3.3.1 Sensitivity to primary mass composition . . . . .	49
3.3.2 Sensitivity to Hadronic Interaction Models . . . . .	50
3.4 Conclusions . . . . .	53

<b>4 Data Selection &amp; Analysis</b>	<b>55</b>
4.1 Data Set . . . . .	56
4.1.1 Pre-Selection . . . . .	56
4.1.2 Atmosphere . . . . .	58
4.1.3 SD . . . . .	58
4.1.4 Reconstruction & Field-Of-View . . . . .	59
4.1.5 Profile Quality . . . . .	61
4.2 Average Longitudinal Profiles . . . . .	65
4.3 Average Shape Analysis . . . . .	67
4.3.1 Fitting method . . . . .	67
4.4 Analysis of simulated showers . . . . .	72
4.5 Data: first look at results and MC comparison . . . . .	76
4.5.1 Fit to a Gaisser-Hillas: $\chi^2$ and residuals . . . . .	77
4.5.2 Fit results: R and L vs energy . . . . .	77
<b>5 Systematic uncertainties</b>	<b>83</b>
5.1 Reconstruction Bias . . . . .	83
5.2 Energy Scale . . . . .	84
5.3 Detector . . . . .	84
5.3.1 SD-FD Time Offset . . . . .	84
5.3.2 Telescope Alignment . . . . .	85
5.3.3 FD site . . . . .	89
5.3.4 Detector age . . . . .	90
5.4 Constraints on the profile shape . . . . .	91
5.5 Light Component Determination . . . . .	93
5.5.1 Fluorescence Yield . . . . .	93
5.5.2 Rayleigh scattering - $\rho_N$ parameter . . . . .	93
5.5.3 Light propagation model . . . . .	94
5.5.4 Cherenkov model . . . . .	95
5.5.5 Multiple Scattering . . . . .	96
5.5.6 Cross-check: Comparing showers with high and low fluorescence percentage . . . . .	97
5.6 Atmosphere . . . . .	99
5.6.1 Aerosols . . . . .	99
5.6.2 Clouds . . . . .	102

5.6.3	Season of the year . . . . .	103
5.7	Light Integration . . . . .	105
5.7.1	Pixel Integrated vs Circle . . . . .	105
5.7.2	Lateral Width Correction . . . . .	106
5.7.3	Variations in $\zeta_{safety}$ . . . . .	107
5.8	Geometry . . . . .	108
5.8.1	SDP . . . . .	108
5.8.2	Shower Axis . . . . .	109
5.9	Summary . . . . .	110
<b>6</b>	<b>Results and interpretation</b>	<b>115</b>
6.1	Results . . . . .	116
6.2	Energy evolution of shower variables . . . . .	119
6.2.1	The $\langle \ln A \rangle$ derivative, $D_{10}\langle \ln A \rangle$ . . . . .	123
6.2.2	Average shape: $\langle \ln A \rangle$ at $10^{19.4}$ eV and $D_{10}\langle \ln A \rangle$ . . . . .	126
6.2.3	Discussion . . . . .	128
6.2.4	Summary . . . . .	130
6.3	Alternative Interaction Models and New Physics . . . . .	131
6.3.1	String Percolation Model . . . . .	131
6.3.2	Chiral Symmetry Restoration (CSR) Model . . . . .	133
6.3.3	Soft Particle Addition Model . . . . .	137
6.3.4	Comparison with Auger data . . . . .	139
6.4	Impact of hadronic parameters on R and L . . . . .	141
6.4.1	Introduction . . . . .	141
6.4.2	Implementation . . . . .	143
6.4.3	Results . . . . .	145
6.4.4	Impact of hadronic parameters on the longitudinal profile . . . . .	147
6.4.5	Impact on profile shape - R and L . . . . .	149
6.4.6	Comparison with Auger data . . . . .	153
6.5	Discussion . . . . .	155
<b>Summary</b>		<b>157</b>
<b>A</b>	<b>Event-by-event contribution to average shape and outliers</b>	<b>161</b>

<b>B Events with large Time-Fit <math>\chi^2</math> - alternative geometry reconstruction</b>	<b>167</b>
<b>Bibliography</b>	<b>171</b>

# Introduction

Although cosmic rays were discovered more than a hundred years ago, their properties at the highest energies are still mostly unknown. Their origin, chemical composition and production mechanism for cosmic rays with energies above  $10^{18}$  eV, the so-called Ultra High-Energy Cosmic Rays (UHECR), are nowadays hot topics in astrophysics. These are the most energetic particles in the universe, many orders of magnitude above the energies reached at the Large Hadron Collider (LHC) at CERN, and are currently our only open window into probing the hadronic interaction properties at center-of-mass energies in the 100 TeV range.

Measuring UHECR is a challenging task, mainly due to two factors. First, the flux of cosmic rays declines steeply with increasing energy. Secondly, cosmic rays interact with atmospheric nuclei, creating a shower of particles. These secondary particles in turn interact with further nuclei, multiplying as they further penetrate the atmosphere. This is called an Extensive Air Shower (EAS). The properties of the primary incident particles have to be indirectly inferred from the properties of these resulting air showers. There are two main measuring techniques: sampling the particles that reach the ground level or detecting the fluorescence photons emitted by the de-excitation of atmospheric  $N_2$  molecules which interact with the shower particles.

The Pierre Auger Observatory, located in the province of Mendoza, Argentina, is the largest detector ever created for the observation of high-energy cosmic rays. It is a hybrid instrument, comprising a Surface Detector (SD) and Fluorescence Detector (FD). The former is an array of 1600 instrumented water tanks which measure the Cherenkov light emitted by cosmic ray particles as they cross them. The separation between tanks is 1500 m, covering a total area of 3000 km<sup>2</sup>. In the delimitation of this area there are four sites with fluorescence detectors over viewing the array. These track the development of the shower as a function of altitude, allowing the measurement of the longitudinal profile - the energy deposited by the shower in the atmosphere as a function of the amount of traversed matter by it.

One of the missions of the Pierre Auger Observatory is the characterization of the cosmic ray flux in terms of its primary mass composition. Its results put strong limits on the flux of high energy photons and neutrinos, as neither has been yet observed. So, it is now known that high energy cosmic rays are mainly atomic nuclei.

To identify the specific elemental composition, the characteristics of the observed showers have to be compared with those predicted by simulations of showers induced by different primary particles. These simulations are performed using interaction models tested in accelerator experiments, at lower energies and probing different phase spaces than cosmic rays. This introduces an uncertainty associated with the extrapolation to much higher energies and, therefore, the mass composition analysis is intrinsically connected to the study of hadronic interaction properties.

The composition variable most thoroughly studied and commonly used is the depth at which the longitudinal profile reaches its maximum,  $X_{\max}$ . This quantity is directly measured by the fluorescence detector, and is related with the development of the electromagnetic component of the particle cascade - the fraction of energy distributed among electrons, positrons or photons. More recently, variables measured by the surface detector have also been used as composition estimators. The SD has several advantages. One is the larger statistics, as the FD only works at nights with low lunar luminosity and favorable atmospheric conditions. Also, it is sensitive to the signal of muons, which do not contribute to the energy deposit measured by the FD. So, the SD can probe the muonic component of the shower, namely by measuring the total number of muons at ground level and the atmospheric depth at which the largest number of muons is produced,  $X_{\max}^{\mu}$ . A joint analysis reveals an interesting scenario. The comparison of the measured  $X_{\max}$  with hadronic model predictions shows a light composition (mixed and energy dependent, but always light-intermediate) while the analysis of the muonic variables points to a heavier composition, i.e., more muons are produced (and earlier in shower development) than predicted by interaction models. This indicates an inconsistency, shared among all standard models, between the description of the electromagnetic and the muonic components of extensive air showers.

Introducing new, independent variables which may test and constrain hadronic models is fundamental to improve our understanding of air shower development. The longitudinal profile is well described by four parameters:  $X_{\max}$ ,  $N_{\max}$  (normalization, proportional to energy) and two shape parameters,  $\lambda$  and  $X_0$ . A new parametrization is introduced, with two new parameters, R and L. These, besides being less correlated with each other when compared to  $\lambda$  and  $X_0$ , have a clearer interpretation: L is a gaussian width, while R controls the asymmetry - for R=0, the profile follows a normal distribution.

The measurement of the shape of the longitudinal profile of high energy showers,

which has never been done previously by any other experiment, as well as the interpretation of the measured value of its parameters, are the main objective of this work.

This thesis is structured as follows. Chapter 1 presents a general overview of cosmic ray history and the properties of the air showers they induce in the atmosphere. Chapter 2 concerns the Pierre Auger Observatory, with a general description of the reconstruction procedure and a selection of the results which are more relevant for this work. In chapter 3, a study of the behaviour of the two shape parameters describing the longitudinal profiles is done in simulations. Following, in chapter 4, the longitudinal profile shape of air showers measured in Auger is presented, and fitted to retrieve its parameters. These variables are affected by several systematic uncertainties, which are detailed in chapter 5. In chapter 6, the interpretation of these results is done. Finally, this work is summarized and future directions are discussed.



# CHAPTER 1

## Air Shower Physics

---

At the turn of the 19<sup>th</sup> century, a sense of completion had spread among many scientists, who predicted that the grand underlying principles of the universe had been firmly established and there was little else radically new to discover in physics. This could not be further from the truth, and cosmic rays played a leading role in revolutionizing fundamental physics in the subsequent hundred years.

It was known since the late 18<sup>th</sup> century that an electroscope would gradually discharge in an unsheltered location. This indicated there were ions in the air, but the cause for this phenomenon remained elusive for a long time - it was referred to as "spontaneous ionization". The discovery of radioactivity by Becquerel, in 1896, brought forward a new hypothesis: air was ionized by radioactivity from the elements in the ground. In 1912, Victor Hess made several balloon flights where he measured air ionization as a function of the altitude. If the source was radioactivity from the Earth, the ionization rate should decrease as a function of atmospheric height. In fact the opposite was seen, as the discharge was faster at higher altitudes, indicating thus an extraterrestrial origin for this ionization [1]. For the discovery of "Cosmic Radiation", a term coined by Millikan who predicted that the flux was mainly composed of photons, Hess was awarded the Nobel prize in 1936.

Cosmic rays were the main source of discovery of unknown fundamental particles prior to the age of man-made accelerators. Anderson discovered the positron in a cloud chamber in 1932 [2] and four years later, together with Neddermeyer, observed the muon for the first time [3]. In 1948, observations with nuclear emulsions carried balloons to near the top of the atmosphere showed that most cosmic rays are protons, while 10% were  $\alpha$  particles and around 1% heavier elements such as carbon, iron, and lead.

At the same time, the properties of cosmic rays were being studied in a different way. Pierre Auger and its group, in 1939 [4], based on the work of Rossi [5], instrumented an array of Geiger counters at the same altitude but separated widely from each other. They observed an amount of coincidences above random expectations and correctly interpreted it as secondary particles resultant from high

energy cosmic rays incident in the atmosphere. They coined the term Extensive Air Shower to describe this phenomenon. Volcano Ranch was the first large scale array, with an extension of  $8 \text{ km}^2$  and composed by 19 scintillator detectors [6]. Lindsey observed the first  $10^{20}$  eV event in it in 1963 [7], effectively starting the study of ultra high-energy cosmic rays.

## 1.1 Ultra High Energy Cosmic Rays

In addition to the construction of larger arrays, an independent detection method has vastly improved our understanding of high energy air showers. In 1962, K. Suga [8] and A. E. Chudakov [9] suggested measuring the fluorescence light emitted by the  $N_2$  molecules excited by the passage of the air shower particles. The first successful experiment based on the fluorescence technique is the Fly's Eye and its successor HiRes [10]. Presently, the two largest high-energy cosmic ray detectors are the Pierre Auger Observatory [11] and Telescope Array [12], which both us the sampling and fluorescence techniques complementarily.

More than a century after their discovery, some of the main properties of ultra high-energy cosmic rays, like their origin and chemical composition remain largely unknown.

### 1.1.1 Energy spectrum

The main difficulty in studying the properties of cosmic rays at the highest energies lies with the steeply falling particle flux with increasing energy.

Nowadays, the rate of incidence of cosmic rays has been measured by several experiments over more than ten orders of magnitude. Measurements of the flux (see figure 1.1) show a power law like shape

$$\frac{dN}{dE} \propto E^{-\gamma} \quad (1.1)$$

with the  $\gamma$  exponent parameter measured to be around 2.7. However, a closer look at the spectrum shows some yet unexplained features. The regions between the breaking points of the spectrum resemble the anatomy of a human leg: in the first break, called the "knee", the spectrum hardens ( $\gamma \approx 3$ ), getting softer again at higher energies, at the so called "ankle". At the very end of the measurable spectrum there is a suppression, usually defined as the energy at which the spectrum falls to half of the expected value from an extrapolation of the spectral index evaluated from

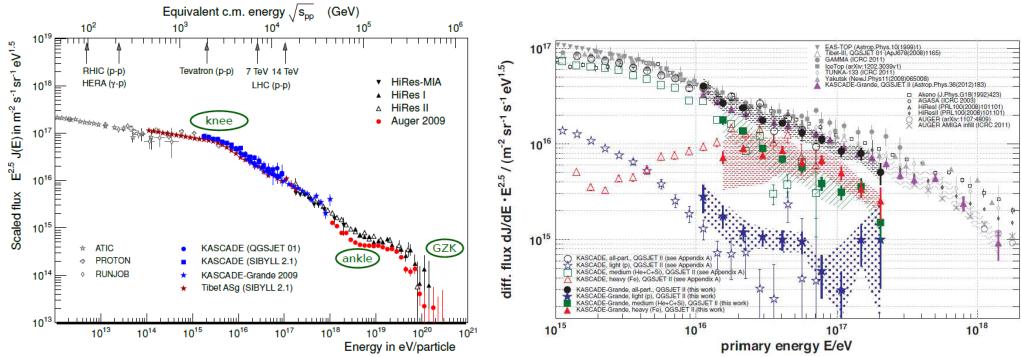


Figure 1.1: Left: Cosmic ray flux over a wide range of particle energies. Measurements from various experiments outlined in the legend are included. Note that the flux is multiplied by a factor  $E^{2.5}$  (from [13]). Right: Energy spectrum measured by KASKADE and KASKADE-Grande experiments compared with other experiments. Different groups of primaries are distinguished by color. Note the spectrum break for protons around 4 EeV, for medium heavy nuclei for 8 EeV and for iron nuclei at 80 EeV (from [14]).

the ankle onwards. The Pierre Auger Collaboration established the existence of this cutoff and measured its energy to be around  $10^{19.6}$  eV [15].

The lower energy cosmic rays have a high flux and can be directly measured by satellites and balloon experiments. This allows us to know precisely their chemical composition. The upper energy limit at which this method is practical is around a few tens of TeV, where the flux drops below 1 particle/ $m^2/day$ . Above it, very large acquisition times or instrumentation areas are required, which is unaffordable in a satellite or a balloon.

Fortunately, high energy cosmic rays interact with the air nuclei starting an avalanche of particles, known as extensive air showers. The showers can be measured to infer properties about the primary cosmic ray, like their energy, mass composition and arrival direction. In figure 1.1, the energy spectrum measured by KASKADE and KASKADE-Grande is shown, together with that of other experiments. The total spectrum is compatible among all of them. Using a simultaneous measurement of the electromagnetic and the muon components, the KASKADE experiments were able to distinguish different mass groups. The lightest one, proton-like particles, has a spectral break at around 4 PeV, while the heavier components only have a break at higher energies, around 8 PeV for medium-heavy particles and 80 PeV for heavy ones. These values increase approximately linearly with  $Z$ :  $\langle Z \rangle \approx 1$  for light,  $\langle Z \rangle \approx 2 - 14$  for medium-heavy and  $\langle Z \rangle \approx 26$  for heavy particles. This indicates a gradual acceleration of particles in a source where the confinement is done by a

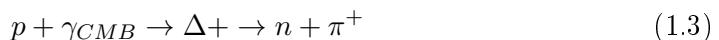
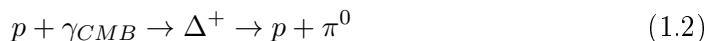
magnetic field, since the gyroradius is proportional to the charge of the accelerated particles.

The proton break around 4 PeV is compatible with the so called knee described above. There is also a less significant break around 80 PeV, where all the mass groups are suppressed, and this is sometimes called the second knee. At the ankle, the spectrum softens again to a power law with a slope of  $\alpha \approx 2.6$  [16]. The cause for this is yet unknown but three theories have been put forward to explain it. In the "dip" model the heavy component dies out at around  $5 \times 10^{17.5}$  eV and the extragalactic protons experience pair production losses ( $p + \gamma_{CMB} \rightarrow p + e^+ + e^-$ ) starting at around  $10^{18}$  eV due to interaction with the Cosmic Microwave Background (CMB) photons. This is, thus, the energy at which the flux from extragalactic cosmic rays is larger. An alternative theory is the ankle model, which predicts a transition to the extragalactic component at  $10^{18.1}$  eV. It requires a spectral index of around  $\alpha = 2$  at the sources, which is consistent with the Fermi acceleration mechanism. Finally, in "mixed-composition" models there are multiple species of nuclei at every energy contributing to the composition. These models have, of course, much more freedom to predict the data features, as each component may die out at different energy thresholds. Current experimental evidence cannot rule out any of these models. However, an ample agreement exists within the community that there must be a transition from galactic to extragalactic cosmic rays, as the galactic magnetic field is not strong enough to contain the highest energy cosmic rays.

#### 1.1.1.1 The GZK cutoff?

The spectral suppression at the end of the spectrum is nowadays an experimental evidence: it has been observed by HiRes [17], Telescope Array [18] and the Pierre Auger Observatory [15]. Multiple interpretations exist for its underlying cause however.

One year after discovery of the Cosmic Microwave Background (CMB) in 1965 [20], Greisen [21], and independently Zatsepin and Kuzmin [22] proposed that cosmic rays with sufficiently high energies could lose energy by interacting with those CMB photons. The lower threshold energy for this effect is around  $10^{19.8}$  eV, and the mechanism is the pion production via  $\Delta^+$  resonance



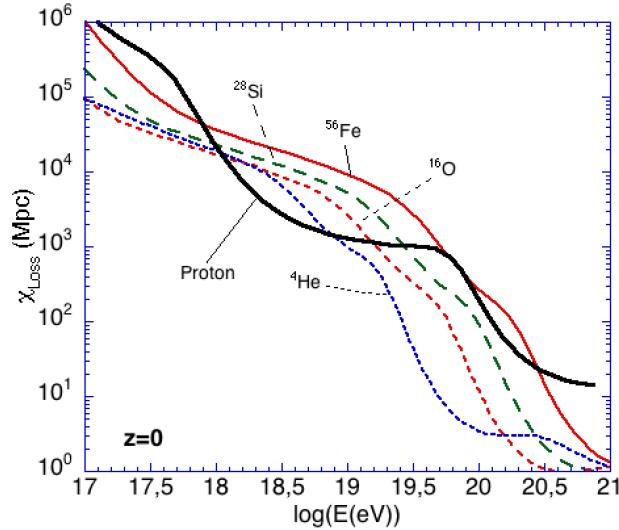


Figure 1.2: Attenuation length depending on energy for different nuclei considering the redshift  $z=0$  (from [19]).

These processes have a cross-section that steeply increase with energy, rendering infeasible the propagation of protons of energies above  $10^{20}$  eV over very large distances (of the order of 100 Mpc). This is known as the GZK effect. Another process by which protons crossing interstellar space can lose energy through interaction with the CMB is pair production

$$p + \gamma_{CMB} \rightarrow p + e^+ + e^- \quad (1.4)$$

This process has a lower energy threshold than GZK, around  $10^{18}$  eV, and is an important energy loss mechanism around this energy, but above  $10^{20}$  eV the GZK effect is by far the dominant one. Interactions with the optical, infra-red and ultra-violet photons are sub-dominant throughout the whole energy range.

If cosmic rays are instead dominated by heavier nuclei, the interactions with background photons are different. In this case, nuclei suffer photo-disintegration through the Giant Dipole Resonance (GDR), resulting in the emission of either one or two nucleons, or  $\alpha$  particles

$$A + \gamma_{CMB} \rightarrow GDR \rightarrow (A - 1) + n \quad (1.5)$$

$$A + \gamma_{CMB} \rightarrow GDR \rightarrow (A - 2) + 2n \quad (1.6)$$

$$A + \gamma_{CMB} \rightarrow GDR \rightarrow (A - 4) + \alpha \quad (1.7)$$

Figure 1.2 shows this process can be very efficient for most nuclei at high energies (above  $10^{20}$  eV): the attenuation length of medium mass nuclei (He to Si) is smaller than those of proton and iron nuclei in that energy range. For iron, the attenuation is very similar to that of proton around the GZK energy range ( $10^{19.6}$  to  $10^{20}$  eV). Therefore, it is expected that in the suppression region the main contributions will come from either proton or iron primaries, unless the extragalactic sources are within a radius of the order of a hundred Mpc of the Earth.

### 1.1.2 Acceleration mechanisms

For particles with energies exceeding  $10^{20}$  eV to exist, there need to be very powerful acceleration sites in the universe. It is, however, hard to devise mechanisms by which a source can efficiently accelerate particles to such extreme energies.

The bottom-up<sup>1</sup> mechanisms for cosmic ray production can be separated in two type: **stochastic** acceleration in magnetic clouds or **direct** acceleration by an electric field, usually by means of strong shock waves propagating through interstellar space.

The first one goes back to Enrico Fermi, when in 1949 he proposed a mechanism in which particles gain energy gradually by numerous encounters with moving magnetized plasma [24]. Charged particles reflected by randomly moving, interstellar magnetic clouds may either gain or lose energy in the process. The average gain, however, is non-zero as the probability of head-on collisions is larger than that of head-tail collisions. On average, the energy gain is proportional to the relative velocity  $\beta^2 = \frac{v^2}{c^2}$ . Due to this dependence on the square of  $\beta$ , this is known as second order Fermi acceleration.

The big disadvantage of this mechanism is the very slow acceleration process. A more efficient mechanism was proposed in 1978, in which the energy gain is proportional to  $\beta$  [25]. In the presence of shock waves such as those in supernova remnants, a particle moving through the shock either up- or downstream gains energy if it is reflected back by interacting with the magnetic fields of the shock front. Since the increase is proportional to  $\frac{v}{c}$ , this is referred to as first order Fermi acceleration. This model naturally gives a power law spectrum to the high produced cosmic rays, although its slope ( $\gamma \approx 2$ ) is softer than the observed one.

---

<sup>1</sup>In bottom-up models the cosmic ray starts with low energy and is accelerated. In top-down models very heavy exotic particles decay to the observed cosmic ray particles. More information on top-down models as possible UHECR can be found, for example, in [23].

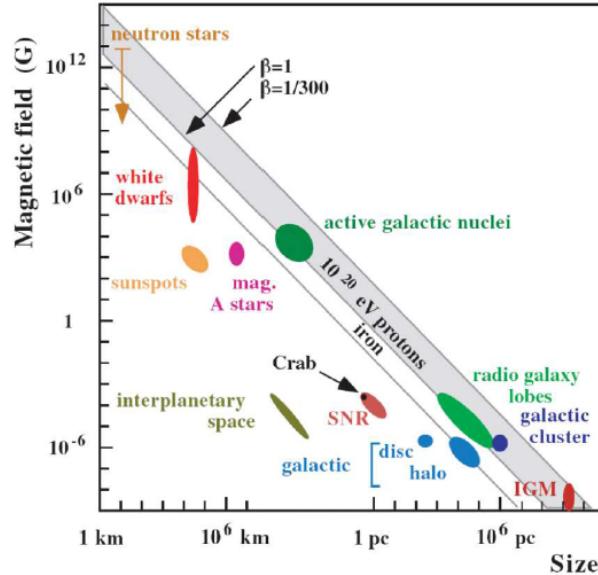


Figure 1.3: Hillas diagram. On the  $x$  axis, the size of the accelerating region in the Fermi process is displayed, and on the  $y$  axis. Diagonal lines are the limits for th indicated particles at  $10^{20}$  eV. From [26].

For these models, the maximum achievable energy for a particle of charge  $Z$  is

$$E_{\max} \propto \beta ZBR \quad (1.8)$$

where  $\beta$  is the velocity of the shock front relative to the speed of light. Thus, the maximum energy of the cosmic rays produced at a source is limited by its radius,  $R$ , and its magnetic field. Hillas argued there were very few astrophysical objects capable of accelerating particle up to  $10^{20}$  eV [27]. Of these, three standout:

- **AGNs** (Active Galactic Nuclei) can exist in galaxies which contain a supermassive black hole in the center. These can produce long-ranged jets of particles, coming from the accretion of mass into the black hole. UHECR can be accelerated along the jets [28] as they produce magnetic fields six orders of magnitude larger than Supernova remnants. The closest AGN is Cen A, which is located 4 Mpc from the Earth in the Centaurus constellation.
- **Magnetars** are neutron stars which have an extremely intense magnetic field (of the order of  $10^{11}$  T). Despite their small radius, the magnetic field may be sufficient to contain charged particles during the acceleration process up to UHECR [29]. Their spatial distribution should follow the distribution of luminous matter at the cosmological scale.

- **Galactic Clusters** are regions where many galaxies are concentrated. If they are large enough to accelerate particles during large-scale shocks, e.g. [30], they may produce UHECR. The closest known cluster is located at around 50 Mpc from Earth in the direction of Cen A.

So, sources and acceleration mechanisms remain a field of active research. To get the full picture, we have to understand the interplay of many observables measurable at the Earth, like the distance and overall space distribution of sources, metallicity of the source environment, spectral shape for each component at emission and detection, which are different due to propagation effects, and anisotropies which may help identify the astrophysical structures where accelerators are located.

### 1.1.3 Arrival directions

A proton of  $10^{20}$  eV in a microgauss-scale magnetic field (such as the one in the Milky Way) has a gyroradius in the order of a hundred kiloparsec. It is then evident that a galaxy such as ours, with a radius of around 20 kpc, is not capable of containing protons at this energy, and as such they must come from larger structures outside our galaxy. This hypothesis can be tested by checking whether high energy cosmic rays arrive in the most part from the direction of the galactic plane. They do not, as the Pierre Auger Collaboration reports no significant flux excesses at energies above 40 EeV from either the Galactic Center or Galactic Plane, when compared with isotropic source distribution [31]. This indicates cosmic rays at these energies are unlikely to be galactic in origin.

To look for extragalactic sources, the arrival directions of neutral particles (like photons and neutrinos) are perhaps the most logical place to start. They are not deflected by magnetic fields, and so their arrival direction points to their sources. At these energies, neutrinos are mainly expected to be the result of the decay of pions produced in UHECR during acceleration. They may also arrive from the decay of very heavy exotic particles in many top-down models, as may high-energy photons. The Pierre Auger Observatory and Telescope Array collaborations have looked extensively for both, but neither found a clear signal of ultra high energy photons or neutrinos [32].

Cosmic rays of sufficiently high energies are not substantially affected by the intergalactic magnetic fields. This allows us to search for anisotropies in their arrival direction even if they are protons or charged nuclei. In 2007, the Pierre Auger

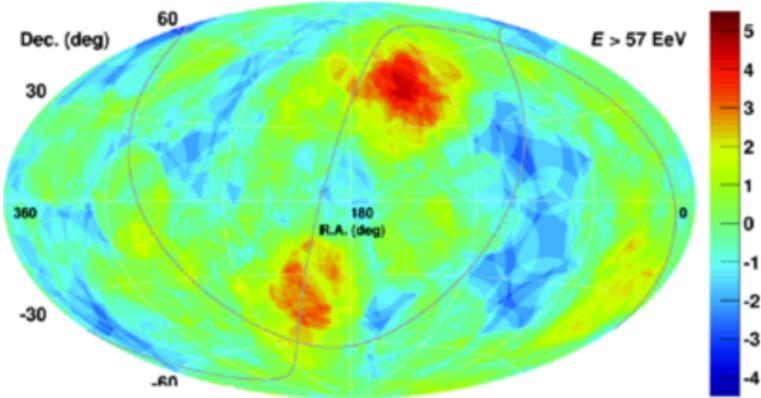


Figure 1.4: Map of the statistical significance of excesses of events in a  $20^\circ$  circle as measured by TA and Auger. The TA hot spot in the northern hemisphere and the Auger hot spot near Cen A are visible. From [33].

Collaboration reported a strong correlation between the observed arrival directions of cosmic rays with  $E > 57$  EeV and known nearby Active Galactic Nuclei (AGN) in the vicinity of our galaxy [34] (within 200 Mpc). Eight out of 13 observed events were within a  $3.1^\circ$  radius around an AGN, which corresponds to a chance probability in an isotropic source distribution scenario of  $1.7 \times 10^{-3}$ . The most recent published analysis on the degree of correlation [35], however, puts its value only  $2\sigma$  above the isotropic expectation. So, currently, this test does not indicate a significant presence of small-scale anisotropy.

At larger scales, Telescope Array finds an excess of events above 57 EeV (hot spot) in the northern hemisphere, in a  $20^\circ$  circle around  $(\alpha, \delta) = (148.4^\circ, 44.6^\circ)$  - 24 observed, 6.88 expected from isotropy;  $3.4\sigma$  deviation [36]. Auger observes an excess above 58 EeV in a  $15^\circ$  circle around Centaurus A (14 observed, 4.5 expected;  $p=1.4\%$ ) [37]. This indicates that although no individual sources can be identified, there are hotter regions of space for ultra high-energy cosmic ray production. If the TA hotspot is confirmed with increasing statistics, the interpretation is hard as no astrophysical structure in the list of known possible sources is along this trajectory.

#### 1.1.4 Mass composition

It is impossible to precisely measure the composition of UHECRs on an event-by-event basis with current detectors. But fortunately, there are observables whose statistical distribution depends on the logarithm of the mass number  $\ln A$  of the

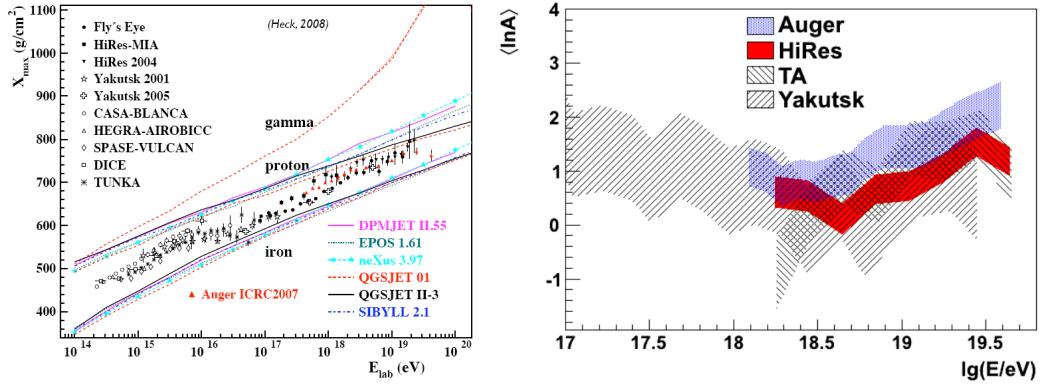


Figure 1.5: Left:  $X_{\max}$  over a large energy range and for several experiments, indicated in the legend along with the used hadronic interactionn models. Right:Comparison of average composition  $\langle \ln A \rangle$  measured by four experiments at the highest energies. From [38]

primary particles. One such composition estimator is  $X_{\max}$ , the maximum of the longitudinal profile measured by fluorescence detectors. Its average is, on first approximation, a linear function of  $\log(E/A)$  because a shower initiated by a nucleus with mass number  $A$  and energy  $E$  can be approximated as  $A$  showers, each initiated by a proton of energy  $E/A$ . Also, the variance between showers gets smaller as  $A$  increases, as the fluctuations in individual sub-showers are less important. The fact that  $X_{\max}$  is very different between primaries ( $100 \text{ g cm}^{-2}$  between proton and iron initiated showers), and that this difference is larger than both the typical experimental resolution and the spread in model predictions for a given primary make it the foremost estimator for the energy evolution of the cosmic ray flux composition.

The measurement of the shower maximum by different experiments over a large energy range is shown in figure 1.5. The average mass of cosmic ray primaries changes dramatically with energy. At  $10^{16}$  eV the composition is very light, but gets heavier during the next energy decade, almost reaching the model predictions for iron initiated showers at  $10^{17}$  eV. This is qualitatively consistent with the previous interpretation of the knee at 4 PeV being the energy at which the galactic source of protons start to die out, leaving only increasingly heavier elements as the energy rises. At around  $10^{18}$  eV the extragalactic, and lighter, component starts to dominate. At the highest energies, uncertainty in model extrapolations and lower statistics make the analysis more difficult. Recent results from the Pierre Auger [90] and Telescope Array [91] Collaborations show a different trend: the latter sees a dominant light component at energies above  $10^{19}$  eV, while the former clearly sees an increase with energy in the average primary mass. The comparison is not

straightforward, as both experiments perform the analysis differently. In Telescope Array, in order to maximize the number of events, the detectors effects are not unfolded, while Auger reports the unbiased  $X_{\max}$  value.

One way to circumvent this is to translate the values of the variable, in this case  $X_{\max}$ , into a model dependent estimator of the mass,  $\langle \ln A \rangle$ . This variable can be easily compared between different experiments, as is shown in figure 1.5. The TA experiment and HiRes find consistent values with each other. However, both of them favour a lighter composition than Auger. This could be caused either by a different flux composition in the southern and northern hemisphere or by detector effects. The collaborations of the two running experiments worked together and concluded that the Telescope Array analysis is not capable of distinguishing between a pure proton composition and a mixture of four components that describe the mass composition seen by the Pierre Auger Observatory [92]. More statistics are needed before a definite conclusion can be drawn about the compatibility of  $X_{\max}$  as measured in both experiments. On the other hand, the spread of the predictions of hadronic models also hamper the interpretation of composition. These differences emerge from the underlying theory of how the hadronic interaction models describe air shower collisions. So, the simulation of showers is a fundamental part of the analysis chain, and an understanding of the processes that determine shower development is necessary in order to improve our knowledge of the cosmic ray properties at the highest energies. The theoretical framework of shower cascade development will be explained in the following section.

## 1.2 Extensive Air Shower Phenomenology

An extensive air shower initiated by the interaction of a hadronic particle can be separated in three components: hadronic, muonic and electromagnetic component. A schematic overview of this process can be seen in figure 1.6. These three components develop differently as they are associated with distinct processes.

The hadronic component can be the backbone of the cascade: it stays close to the shower core and is composed mainly by charged pions, and fast baryons in smaller number. Its percentage of total shower energy decreases with the cascade development, as the products of hadron interaction can feed the other components. One of this processes is the pion decay, which is the main factor controlling the muonic component of air shower. Most of muons can travel to ground level, as they

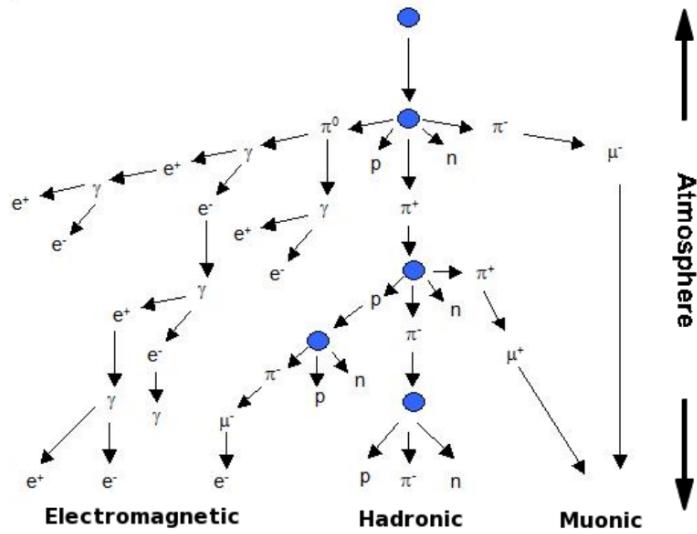


Figure 1.6: Schematic view of a shower initiated by a hadron.

are relatively stable particles with low interaction cross-sections. Each pion decay results not only in a muon, but also its corresponding neutrino. These particles have a virtually null cross-section with the atmosphere, meaning the energy carried by them is undetectable. They account for around 5% of total shower energy for a proton primary at  $E = 10^{19}$  eV, as shown in figure 1.7.

The majority of the shower energy is in the electromagnetic component, comprising electrons, positrons and photons. It is mostly the result of the decay of neutral pions into two photons, and of muon decay into electrons in a much smaller degree. As the shower enters the atmosphere, more and more energy goes into this component, and by the time particles arrive at ground around 90% of the shower energy is in electromagnetic particles.

### 1.2.1 Electromagnetic shower development

Cascades initiated by electrons or photons were discovered in cloud chambers by Blackett and Occhialini [40]. Their behaviour is governed at high energies by well known processes. High energy electrons (or positrons) interact with matter mostly through bremsstrahlung, emitting a photon as they are slowed down by their passage in the vicinity of atomic nuclei. As a result, they also undergo scattering. The Molière radius is defined as the radius of a cylinder in which 90% of the total shower energy is contained, and is around 100 meters at ground level for UHECR.

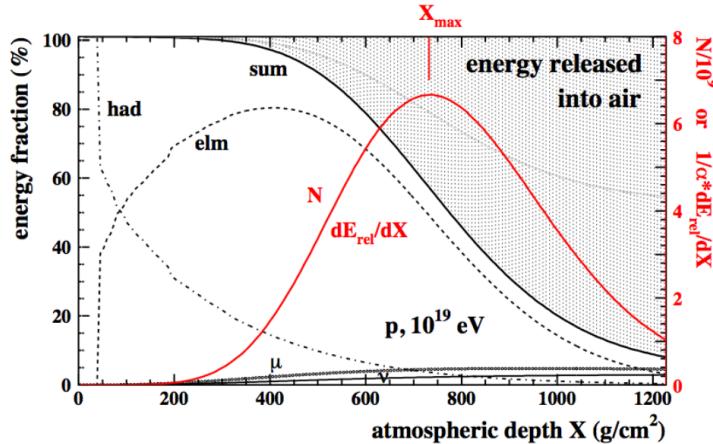


Figure 1.7: Energy flow in an individual proton simulated event at  $10^{19}$  eV. Left scale shows the energy fractions stored in hadron, electromagnetic particles, muons and neutrinos. The right scale shows the longitudinal profile  $N(X)$  (from [39]).

High energy photons are predominantly converted into an electron-positron pair in the presence of nuclei. Both processes have a similar and approximately energy independent radiation length,  $X_0$ , as [41]

$$X_0 \|_{brem} \approx 0.78 X_0 \|_{pair} \quad (1.9)$$

For lower energies, ionization becomes the dominant process and electromagnetic particles are absorbed by atmospheric molecules. This is called the critical energy, and for electrons is around  $E_c = 82$  MeV.

One important property of these processes is that they are independent of the distance traveled by the particles, depending only on the quantity of nuclei in the vicinity of the shower particles, i.e., the quantity of matter traversed. This is known as the slant depth and can be written in good approximation (for  $\theta < 60^\circ$ ) as

$$X = \frac{X_v}{\cos \theta} \quad (1.10)$$

where  $X_v$  is the vertical atmospheric depth, measured in  $\text{g cm}^{-2}$  and given as a function of atmospheric density  $\rho_{atm}$  by

$$X_v = \int_h^\infty \rho_{atm}(z) dz \quad (1.11)$$

### 1.2.1.1 Heitler model

Most general features of an electromagnetic cascade can be explained by a simple model, first proposed by Heitler [42]. He described a cascade consisting of only

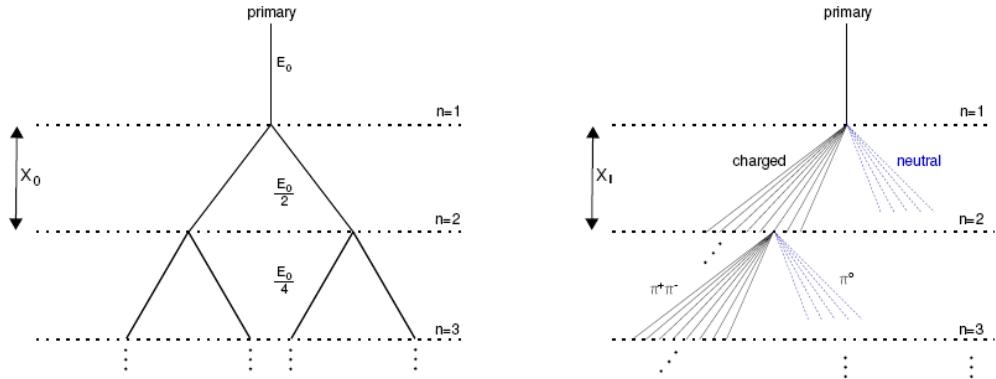


Figure 1.8: Electromagnetic (left) and hadronic (right) Heitler model

electrons and photons, each interacting after the same length  $\lambda = X_0 \ln 2^2$ , making use of equation 1.9. Every interaction results in two particles, each sharing equally the energy of its ancestor. So, after  $n$  splitting lengths, the shower front is at depth  $X = N\lambda$  and contains  $2^n$  particles, each with energy  $E_n = E_0/2^n$ . This is shown schematically in figure 1.8.

Multiplication ceases when pair reach the critical energy,  $E_c$ , below which electromagnetic start being absorbed by the molecules in the atmosphere. After this, the cascade stops growing at a maximum number of particles given by

$$N_{\max} = \frac{E_0}{E_c} \quad (1.12)$$

occurring at depth

$$X_{\max} = X_0 \ln \left( \frac{E_0}{E_c} \right) \quad (1.13)$$

Although this model is based on quite crude assumptions, it reproduces well two observed features of air-shower development -  $N_{\max} \propto E_0$  and  $X_{\max} \propto \ln(\frac{E_0}{E_c})$ . Another important quantity is the elongation rate of  $X_{\max}$ , defined as its evolution with the logarithm of energy

$$D_{10} = \frac{dX_{\max}}{d \log_{10}(E_0)} = X_0 \ln(10) \approx 85 \text{ g cm}^{-2} \quad (1.14)$$

This is also in agreement with full shower simulations up to energies around  $10^{17}$  eV.

Above this value small hadronic subshowers start to appear, making showers more

---

<sup>2</sup>Note that  $\lambda$  is not an interaction length in the usual sense (i.e., the mean path length required to reduce the number of non-interacting particles by a factor of  $1/e$ ) but the distance over which an electron loses, on average, half of its energy.

penetrating, and at the highest energies  $X_{\max}$  increases dramatically due to the LPM effect [43, 44], which predicts a reduction of the cross sections for bremsstrahlung and pair production at high matter densities due to interference from successive scattering centers.

### 1.2.1.2 Longitudinal Parametrization

Even though it agrees qualitatively with observations and is good to illustrate shower behaviour, the Heitler model is rather simplistic and does not give a profile of shower development. In particular, it predicts particles stop being created at shower maximum. In 1956, Greisen proposed an analytical solution for the number of electrons as a function of traversed depth, given by [45]

$$N_e = \frac{0.31}{\sqrt{\beta_0}} e^{X(1-3/2 \ln s)} \quad (1.15)$$

where  $\beta_0 = E_0/E_c$  and  $s$  is the shower age

$$s = \frac{3X}{X + 2X_{\max}} \quad (1.16)$$

### 1.2.2 Hadronic shower development

When a nucleus enters the atmosphere, a first hadronic interaction occurs. This is very different from the electromagnetic process described above, as the interaction is mediated by the strong force and pions, kaons and several other hadrons can be created. Although these hadrons are mostly unstable, the majority of them interact before decaying due to their very high energies. A chain of several strong interactions starts, giving rise to the hadronic cascade. Although these cascades are more complex than their electromagnetic counterpart, they can also be understood in terms of a simple toy model.

#### 1.2.2.1 Heitler-Matthews model

The same principles by the Heitler model explained above can be used to build a model that explains the development of showers started by an hadron. This was first done in 2004 by Matthews [46], and is thus usually called the Heitler-Matthews model. The scheme is depicted in figure 1.8. A hadronic particle interacts with a nucleus from an air molecule, producing  $N_{\text{mult}}$  particles. Within this model, the resulting particles are all pions, evenly distributed between their types: 2/3 of charged pions ( $\pi^\pm$ ) and 1/3 of  $\pi^0$ . This is an oversimplification, but a reasonable

one: all current hadronic interaction models predict that more than 80% of created particles are pions.

The neutral pions do not interact, immediately decaying through the  $\pi \rightarrow \gamma\gamma$  channel, as  $c\tau \approx 25$  nm. The resulting photons yield electromagnetic subshowers, as described previously.

The complexity of this model comes from the behaviour of charged pions, as they undergo two competing processes. If their energy is high, they interact every hadronic interaction length,  $X_h \approx 120\text{ g cm}^{-2}$  in air. Moreover, the primary is not split into two particles anymore, but into  $N_{mult}$  to account for the multiplicity of the hadronic processes. Each resulting particle is assumed to have an energy  $E_0/N_{mult}$ . Once the charged pions are below the critical energy,  $E_c^h$ , they are assumed to decay into muons and corresponding neutrinos. This critical depends on the atmospheric density and on the energy changing pion decay length, but can be approximated to be 20 GeV [47].

With this simple set of assumptions, it is possible to find expressions for the number of muons and the shower maximum. In each step,  $N_{mult}$  charged pions are created. After  $n$  interactions, there are  $N_{\pi^\pm} = (2/3N_{mult})^n$  charged pions, each with energy  $E_0/(N_{mult})^n$ . The critical energy is reached at the step

$$n_c = \frac{\ln(E_0/E_c^h)}{\ln(N_{mult})} \quad (1.17)$$

and the number of muons at shower maximum is given by

$$N_\mu = N_{\pi^\pm} = (N_{mult})_c^n = \left(\frac{E_0}{E_c^h}\right)^\beta \quad (1.18)$$

where  $\beta$  is given by

$$\beta = \frac{\ln(\frac{2}{3}N_{mult})}{\ln(N_{mult})} \quad (1.19)$$

Assuming multiplicities between 10 and 100, one obtains value between  $\beta = 0.84$  and  $\beta = 0.92$  [47]. In full shower simulations the value obtained,  $\beta \approx 0.9$ , is within this range.

The derivation of the shower maximum is similar to the one from the electromagnetic model. Noting that the first generation photons, with energy  $E = (1/2)E_0/N_{mult}$ , are the most deeply penetrating ones, the depth of the maximum of shower development is given by

$$X_{\max}^h = \lambda^{h-air} \ln 2 + X_0 \ln \left( \frac{E_0/E_c}{2N_{mult}} \right) \quad (1.20)$$

where  $\lambda^{h-air}$  is the hadron interaction length with air and  $E_c$  is the electromagnetic critical energy. Comparing this equation with the one for the electromagnetic Heitler model, it is possible to see that a dependence on the average multiplicity appears,  $\propto 1/\ln(N_{mult})$ . Thus, hadronic showers are expected to have a shallower maximum when compared to photon initiated ones.

For interacting atomic nuclei, we consider the superposition model, which considers that a nucleus with  $A$  nucleons can be seen as a bunch of  $A$  protons with energy  $E_0/A$ . Substituting this into equation 1.20, the general expression for the hadronic  $X_{\max}$  is derived

$$X_{\max}^h = \lambda^{h-air} \ln 2 + X_0 \ln \left( \frac{(E_0/A)/E_c}{2N_{mult}} \right) \propto \ln \left( \frac{E_0}{N_{mult}A} \right) \quad (1.21)$$

So, the maximum decreases with the number of nucleons, and protons are more penetrating than other nuclei, as seen in more complex shower simulations.

This model qualitatively reproduces some observables, like  $N_\mu$  and  $X_{\max}$ , and works remarkably well as an instrument for gaining sensitivity to the shower behaviour. However, it neglects several important interaction parameters: the fraction of energy carried by the leading particle, inelasticity, the production of kaons and other hadrons, the energy spectrum of produced particles and pion decay below the critical energy, among many others. Also, the stochastic nature of the secondary particle production leads to shower-to-shower fluctuations that are not predicted by this model. So, full shower simulations where the hadronic interaction details are parametrized in detail are needed for a full understanding of cascade development

### 1.2.3 Extensive Air shower simulations

Simulations of the observed phenomena are a crucial ingredient in the analysis of experimental results in high energy physics. Extensive air showers are no exception. Many experiments rely on the predictions of muons and electromagnetic particles at ground for energy reconstruction. Even those who can measure the energy more directly still require accurate simulations for testing the reconstruction algorithm. Finally, the interpretation of the observables is often made in terms of the expectations of simulations for different composition and interaction properties, so an inaccurate description of shower development can result in misleading conclusions.

The full simulation of extensive air showers requires a three-dimensional tracking of particles as they travel through the atmosphere. This requires a description of particle interactions over up to 14 orders of magnitude in energy, as well as a

thorough description of the Earth's atmospheric density and magnetic field. In the cosmic rays community, the most commonly used programs are MOCCA [48], AIRES [49] and especially CORSIKA [50].

Since experiments have become large enough to detect cosmic rays with energies above  $10^{20}$  eV, the full Monte Carlo treatment of air showers has become prohibitively costly in terms of computing resources. One way to deal with this is to apply thinning algorithms, in which instead of tracking each particle a weight is added to a bunch of particles once they go below a certain energy threshold, as the programs mentioned above do. Another solution was the development of hybrid simulation packages. The most famous codes using this technique are CONEX [51] and SENECA [52]. These programs do a full simulation of the first interactions and then numerically solves cascade equations for the high-statistics part of the shower. These cascade equations are parametrizations of showers, which means that no fluctuations and timing and geometry information are saved, and in CONEX only the longitudinal profile is obtained.

To simulate extensive air showers, these simulators need to know the interaction properties of the particles with atmospheric matter as they traverse it. While the interaction properties of electromagnetic particles are well known, this is not true for hadronic interactions, particularly at the highest energies, where no accelerator data is available. So, there are hadronic interaction models which are plugged into the shower simulators and whose values are used for the interactions in which they apply. Here, we will briefly describe them.

### 1.2.3.1 Low-energy hadronic interactions

Low-energy models typically cover the range from the particle production threshold up to 200 GeV. The most well known in the cosmic rays field are Gheisha, UrQMD and FLUKA.

These models will not be discussed in detail as the longitudinal profile does not change significantly between low energy hadronic models. It is mainly determined by the speed of energy transference to the electromagnetic component which occurs on the first high energy interactions. Therefore, the main observables related to it, like  $X_{\max}$  and its variance, are virtually unchanged by the uncertainty in low energy interactions.

It should be noted however that the particles distribution at ground, at large distances from the core ( $> 1000$  m), can vary significantly. In particular, the number

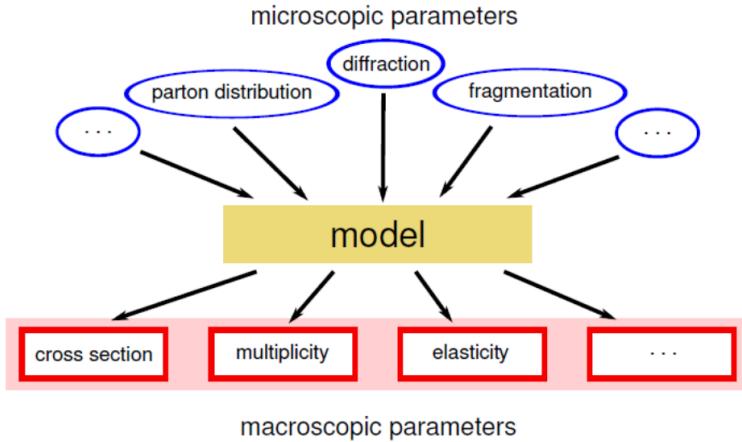


Figure 1.9: Schematic view of high energy hadronic models. The physical and phenomenological inputs of a model (internal parameters) produces a set of correlated characteristic features (macroscopic parameters) of the secondary particle distributions relevant to air shower development (from [66]).

of muons at distances larger than 3 km (relevant distance to UHECR in the Auger) is around 15-20% larger in Gheisha than in FLUKA [?].

### 1.2.3.2 High-energy hadronic interaction models

Although QCD is a well established theory governing strong interactions, it cannot be solved exactly and a perturbative approach has to be used. Processes with a large momentum transfer ("hard" processes) can be calculated within perturbative QCD (pQCD). This works for most of high energy accelerator's data where the collision between two accelerated beams results in most particles being in the transverse plane. In cosmic rays, the region of interest is the forward region where soft processes dominate, so pQCD cannot be used for most of them.

The simulation of the full scope of interactions in high energy cosmic rays is very difficult as models are needed that describe both soft and hard processes. The currently most successful high energy interaction models for cosmic ray analysis - QGSJET ([54, 55]), EPOS ([56, 57]) and SYBILL ([58, 59, 60]) - use perturbative QCD in its region of validity to predict the cross-section of minijet production [61], and Gribov's Reggeon field Theory (RFT) [62] for the treatment of soft processes. Since confinement doesn't allow a gluon or parton exchange for this processes, but an interaction requires some particle to be exchanged, in this framework it is considered that quasi-particles - pomerons - are exchanged. The parametrization of this exchange is subject to a large uncertainty, since it has to be parametrized from accel-

erator data and extrapolated to much higher energies. As a consequence, the models used for hadronic interactions give rise to the main source of uncertainties in EAS simulations. Since some parameters of the primary cosmic rays are inferred from the comparison of shower data and Monte Carlo simulations, in particular those related to the primary mass, the lack of detailed knowledge about hadronic interactions is the main source of systematic uncertainties in air shower analysis. Below follows a brief description of the three mostly used high energy hadronic interaction models.

**QGSJetII-04** is based on the Quark Gluon String (QGS) model, with an added minijet component. A microscopic model of multi-pomeron interaction is implemented, leaving the smallest number of free parameters out of all models [63]. The version used in this work has been tuned to describe LHC data.

**EPOS-LHC** is optimized to give a very good description of accelerator data, both for p-p and Au-Au collisions. It has been tuned to new measurements by LHC experiments [64]. Not only is the pomeron microscopic model calculated, but nuclear effects are also included, making EPOS suited for heavy ion physics. It predicts a much harder pion spectrum from proton-induced interactions, and a larger number of muons at ground level compared to the other models.

**SYBILL 2.1** Sybill was initially a pure minijet model, and does not implement pomeron interactions. It is the simplest model described here [65], containing a minimum number of assumptions in order to describe the features of interactions needed for air-shower simulation. It is not suited for simulations of heavy ion collisions.

The inner workings of the hadronic models are very complex, as they have several internal parameters (parton distribution, string fragmentation, diffraction, ...) that define the outcome of the hadronic collision. The main features of the collision itself however, can be characterized in terms of macroscopic parameters, as shown in figure 1.9. These have the advantage of being less in number and having a much clearer effect on shower observables. Below we discuss some of the main ones.

**Multiplicity** is defined as the number of secondary particles produced in a collision. This parameter has a large impact on the shower development speed, as having a larger number of less energetic particles leads to the shower reaching its maximum faster. Also, an increased number of secondaries leads to smaller

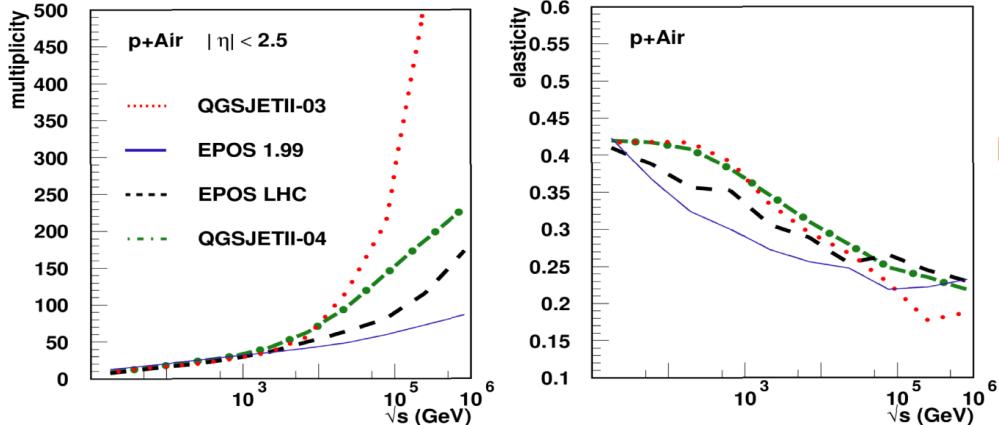


Figure 1.10: Prediction of pre- and post-LHC hadronic interaction models for the multiplicity (left) and elasticity (right) parameters for proton-air collisions, as a function of center-of-mass energy (from [67]).

overall fluctuations of the subsequent EAS development. As a nuclei is a set of  $N$  nucleons, the interaction of heavier nuclei results in a much larger multiplicity than for lighter ones.

**Elasticity** is defined as  $E_{max}/E_{tot}$  and quantifies how much energy is carried by the leading particle and does not contribute to the secondary particle production. Larger elasticity leads to a deeper shower maximum and larger fluctuations between showers. Protons, unlike irons, can have quasi-elastic interactions, and so have much larger elasticity.

**Cross-section** determines atmospheric depth at which the shower starts. As can be seen in figure 1.11, the differences between models are not very large (< 20%), with only SYBILL having a higher cross section, already in tension with the measured one by Auger. Although the cross-section of nuclei doesn't scale exactly with the number of nucleons contained in it, it is approximately 4 times larger for iron nuclei than for proton.

It is interesting to see in figure 1.10 that although there are still sizable differences in predicted multiplicity and elasticity between models, this has severely decreased after tuning for the new LHC data. This, in turn, has lead to a subsequent decrease in the difference between model predictions of shower observables like  $X_{max}$ , and shows the potential of the LHC data to help the analysis of high energy cosmic rays showers by constraining interaction models. The effect of these macro-parameters on the shower observables will be further studied in section 6.4.4.

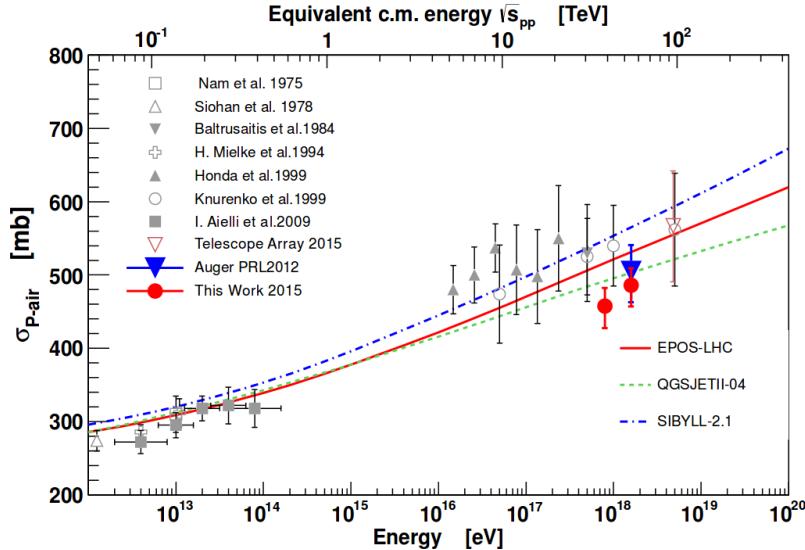


Figure 1.11: The  $\sigma_{p\text{-air}}$ -measurement compared to previous data and model predictions (from [68]).

### 1.2.4 Shower detection techniques

As described previously, there are mainly two well developed and complimentary ways of measuring air showers from high energy cosmic rays. The most common method is to detect the shower particles that reach the ground, which are mainly photons, electrons, positrons and muons. Another well-established method is to measure the fluorescence light emitted by the nitrogen nuclei excited by the shower particles. In modern high energy cosmic rays experiments, both detection methods are usually implemented, allowing a hybrid reconstruction of events.

#### 1.2.4.1 Surface Detectors

A surface detector (SD) array consists of several particle detectors, such as Cherenkov radiators or plastic scintillators, deployed in a grid with regular spacing.

A SD array is mainly defined by three parameters, characteristic of the energy range measured: detector site altitude, detector spacing and array area. The atmospheric depth for which the number of particles in an air shower reaches a maximum increases with energy, resulting in a decreasing optimal detector site altitude. The area of an air shower front increases with energy, resulting in an increasing separation between detectors for a constant sampling of the shower's density profile. Finally, the array area is proportional to the number of events observed per unit of time. For energies around  $10^{18.5}$  eV the flux is around  $1 \text{ particle} \cdot \text{km}^{-2} \text{year}^{-1}$ , and

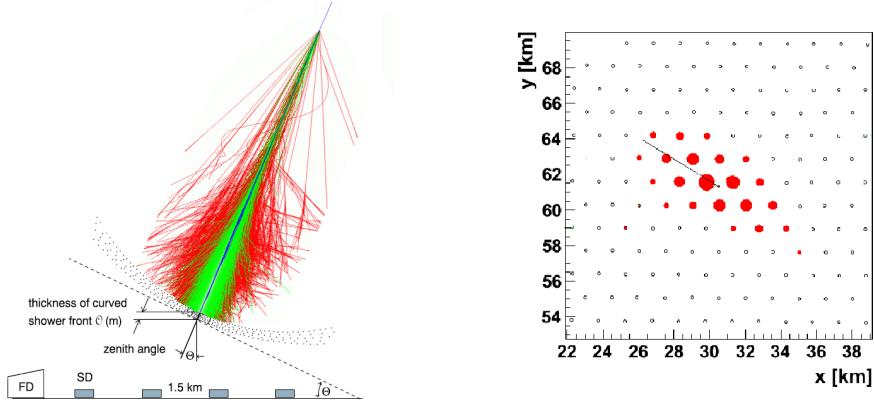


Figure 1.12: Left: Schematic of an air shower and hybrid cosmic ray experiment, with a surface and fluorescence detector shown. Right: Event measured in the Pierre Auger Observatory, for an estimated energy of  $3.3 \times 10^{19}$  eV. The red dots represent SD stations, and the dot size is proportional to the signal measured. The black line corresponds to the reconstructed direction of the air shower axis, and its tip to the reconstructed shower core.

an increase in the maximum particle energy by 10 requires a 1000 times increase in area. So, SD arrays built for the detection of ultra high energy cosmic rays cover thousands of square kilometers.

The aim is to measure the spatial distribution of energy deposit of the EAS as a function of time. With the energy density measured at ground and the relative timing of hits in the different detectors one can estimate the energy and direction of the primary cosmic ray. Since an air shower front can be seen as a cylinder traveling at the speed of light, the angle with respect to the vertical is approximately given by  $\cos\theta \approx \frac{c\delta t}{d}$ , where  $d$  is the distance between detectors and  $\delta t$  the time difference between the triggering of two detectors. The shower energy and core can be determined by the measured particle density distribution at the ground. This density profile is well described by the NKG function

$$\rho(r) = k \left( \frac{r}{r_M} \right)^{-\alpha} \left( 1 + \frac{r}{r_M} \right)^{-(\eta-\alpha)} \quad (1.22)$$

where  $r_M \approx 78$  m is the Moliere radius and  $k, \eta$  and  $\alpha$  are empirical parameters. The energy is estimated from the parameter  $S_{1000}/CIC_\theta$ , which is the signal at 1000 m from the core, corrected by a function of the zenith angle  $\theta$ . Although this value is found to be the most robust in simulations, for a constant energy there are still relevant oscillations for different primaries and hadronic models, around 20%.

The muon content depends on the composition of the primary cosmic ray. Sur-

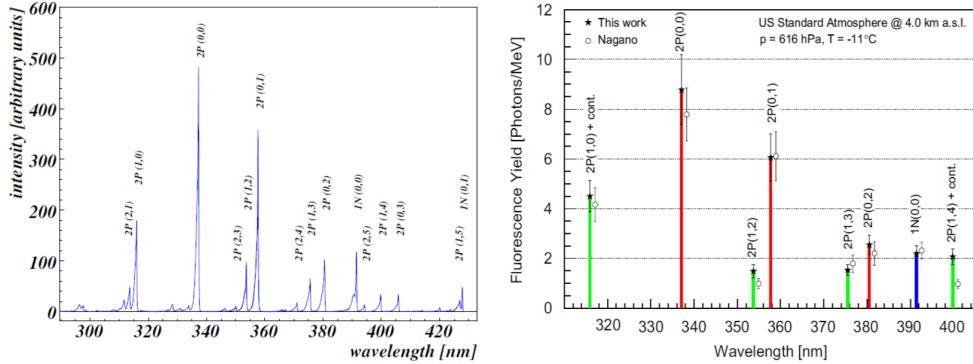


Figure 1.13: Left: Air fluorescence spectrum excited by 3 MeV electron at 800 hPa as measured by AirFly [69]. The most prominent bands are labeled. Right: Absolute fluorescence yield spectrum measured by AirLight [70] of some of the investigated nitrogen transitions in dry air at 4 km height a.s.l. and compared to the values of Nagano et. al. [71].

face arrays with the ability to distinguish muons from electrons and photons have thus the ability to provide some hints about the composition of the incident cosmic ray flux.

#### 1.2.4.2 Fluorescence detectors

When a cosmic ray hits the atmosphere and the air shower ensues, its constituent electromagnetic particles excite the atmospheric nitrogen molecules. These subsequently return to their ground state by isotropic emission of fluorescence photons.

A fluorescence telescope consists of an array of photomultipliers (PMTs) where the light is collected, and can be focused either by a mirror placed behind the array, or a lens placed in front of it. As the shower traverses the atmosphere, the fluorescence light produces a pattern in the PMT array. From the pixels time and direction it is possible to reconstruct the geometry, and from the intensity of the signal the collected number of photons. After correcting for the attenuation in the atmosphere, the shower profile, defined as the ionizing energy per atmospheric depth bin, is determined. The integral of this profile,  $\int_0^\infty N_e(X)dX$  is proportional to the electromagnetic energy of the shower. The constant of proportionality is the average rate of energy loss for ionization in the shower and is approximately  $2.2 \text{ MeV} \cdot g \cdot cm^{-2}$ .

Here we will briefly review the determinant processes for the fluorescence technique: the emission and transmission effects for fluorescence and Cherenkov light.

**Nitrogen fluorescence in air** The vast majority of secondary particles in EAS are electrons and positrons below 1 GeV. These particles at low energies interact with atmospheric matter mainly through ionization. There are two main electronic transitions in the nitrogen molecule,  $N_2$ , and the molecular ion  $N_2^+$ , that have been experimentally observed: those coming from the  $N_2$  second positive system and the  $N_2^+$  first negative system (see figure 1.13). These emit in a wavelength band between 300 and 400 nm.

The most important quantity for shower reconstruction is the relation between the number of emitted photons and the energy deposit by shower particles in the atmosphere. This can be written as [72]

$$n[\text{photons}] = \varepsilon_\lambda(P, T, \lambda) \frac{\lambda}{hc} E_{\text{dep}} \quad (1.23)$$

where  $\varepsilon_\lambda$  is the fluorescence efficiency, i.e., the fraction of the total energy loss that is emitted as fluorescence light. The fluorescence yield,  $FLY/E_{\text{dep}} = \varepsilon_\lambda(P, T, \lambda)\lambda/(hc)$ , has been measured between 300 and 400 nm in dry air using electron beams at several energies. The most precise value currently has been measured by the AIRFLY Collaboration, and gives  $FLY/E_{\text{dep},337nm} = 5.61 \pm 0.06_{\text{stat}} \pm 0.21_{\text{syst}}$  [73] (around 4% systematic uncertainty).

**Cherenkov radiation** Unlike fluorescence light, Cherenkov radiation produced by relativistic particles is not emitted isotropically but strongly beamed in the forward direction. In the incoming shower direction, the Cherenkov contribution can easily overcome the fluorescence light component. But, even outside the beam direction scattered Cherenkov light can contribute for the measured photons in a telescope. Therefore, air shower reconstruction needs to take this component into account.

When a particle crosses a dielectric medium, the medium becomes polarized. It can be considered as an excitation, as electrons will assume a given configuration and subsequently return to their initial, lower energy, position. The electromagnetic waves caused by this process interacting destructively if the velocity of the crossing particle is less than the speed of light in that medium. If the particle is faster than the local speed of light, the waves interfere in a constructive way, giving rise to the Cherenkov effect. The emitted light is emitted in a very narrow forward cone, with an opening angle given by

$$\cos \theta = \frac{ct/n}{vt} = \frac{1}{\beta n} \quad (1.24)$$

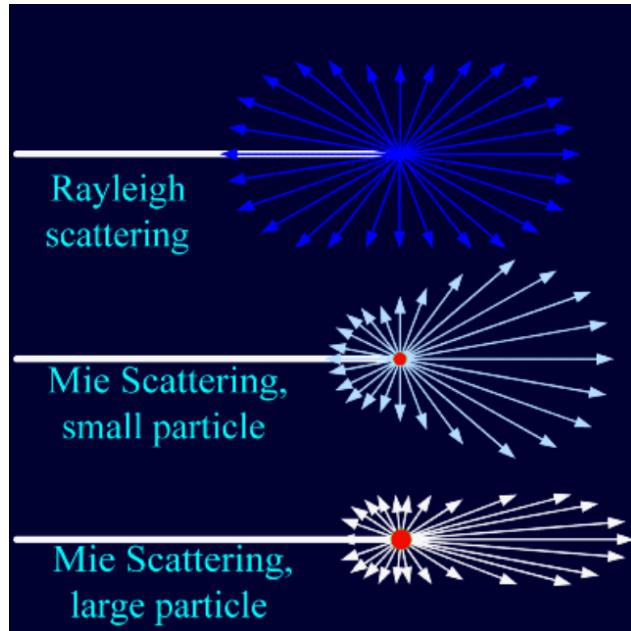


Figure 1.14: Schematic view of the Rayleigh and Mie scattering angular distribution

with  $\beta = v/c$  and  $\theta$  the angle relative to the incident particle direction. At sea level, for ultra-relativistic particles, the opening angle is around 1.5 deg. The energy threshold for electrons in air is around 21 MeV, so the towering majority of shower electrons will produce Cherenkov light as they traverse the atmosphere.

The number of emitted photons per unit length  $l$  in the wavelength region from  $\lambda_1$  to  $\lambda_2$  is

$$\frac{dN}{dl} = 2\pi\alpha \int_{\lambda_1}^{\lambda_2} \left(1 - \frac{1}{\beta^2 n^2(\lambda)}\right) \frac{1}{\lambda^2} d\lambda \quad (1.25)$$

where  $\alpha$  is the fine structure constant. Due to the  $\lambda^{-2}$  proportionality, most of the radiated photons are on the short-wavelength region.

**Scattering of photons in the atmosphere** The reconstruction of the shower from the measured light involves not only understanding the production mechanisms of those photons, but also their propagation through the atmosphere. Both Cherenkov and fluorescence light traveling through the atmosphere are scattered essentially by two processes: Rayleigh and Mie scattering.

**Rayleigh Scattering** Named after physicist Lord Rayleigh, this process is the elastic scattering of the light on particles that are much smaller than the light

wavelength. The total cross-section per molecule of air is given by

$$\sigma(\lambda) = \frac{24\pi^3(n_s^2 - 1)^2}{\lambda^4 N_s^2(n_s^2 + 2)^2} \left( \frac{6 + 3\rho_n}{6 - 7\rho_n} \right) \quad (1.26)$$

where  $n_s$  is the air refractive index,  $N_s$  the molecular number density and  $\rho_n$  the depolarization factor, that accounts for the anisotropy of the air molecules. The differential cross section, usually called the phase function, is given by

$$\frac{1}{\sigma} \frac{d\sigma}{d\Omega} = \frac{3}{16\pi(1+2\gamma)} [(1+3\gamma) + (1-\gamma)\cos^2\theta] \quad (1.27)$$

where  $\gamma = \rho_n/(2 - \rho_n)$ . In the classical case, when  $\rho_n = 0$ , the well known  $1 + \cos^2\theta$  angular distribution is recovered. This is depicted in figure 1.14, where the maxima at the forward and backward direction can be seen.

**Mie scattering** At high altitudes, the atmosphere is only composed of molecules, mainly nitrogen and oxygen. However, closer to the ground level, it is possible to start encountering bigger particles, like dust or liquid droplets, suspended in the air. These particles are called aerosols and their typical radius can go from 0.1 to 10  $\mu\text{m}$ .

These particles are of the order of the typical fluorescence light wavelength, and thus the scattering of light on them cannot be described by the Rayleigh scattering. The Mie theory is used, a far more complex one, as it has to deal with the overall size and shape of the scattering centers.

Its phase function is mostly forward peaked, also with a weaker backward peak. The relative strength of each is dependent on the particle size and shape, as can be seen in figure 1.14. A good effective parametrization for the phase function is given by

$$\frac{1}{\sigma} \frac{d\sigma}{d\Omega} = \frac{1-g^2}{4\pi} \left( \frac{1}{(1+g^2-2g\cos\theta)^{3/2}} + f \frac{3\cos^2\theta - 1}{2(1+g^2)^{3/2}} \right) \quad (1.28)$$

where  $g$  controls the relative strength of the forward peak, and  $f$  controlling the weaker backscattering component.

The aerosol density falls exponentially with altitude, so the amount of photons diffused due to Mie scattering can be approximated by [72]

$$\frac{dN_\gamma}{dl} = -\frac{N_\gamma}{L_M} e^{-h/H_M} \quad (1.29)$$

where  $L_M$  is the aerosol horizontal attenuation length (at ground level), and  $H_M$  the scale height factor which controls the speed of aerosol extinction with height.

These parameters can quickly change, within one hour, and therefore have to be constantly monitored in cosmic ray experiments.

Having the light production and attenuation mechanisms, all the theoretical ingredients for the reconstruction of cosmic ray showers with the fluorescence technique are laid out.

## CHAPTER 2

# Pierre Auger Observatory

---

In this chapter an overview of the Pierre Auger Observatory, with its experimental setup and reconstruction methods will be done. Also, the most recent results will be described.



## CHAPTER 3

# Longitudinal Profile Shape

---

The development of high energy cosmic ray showers in the atmosphere gives indirect information on the primary particle and its interaction with atmospheric nuclei. In general, the maximum number of particles and the depth at which that maximum occurs are used to characterize the showers. The first is a measurement of energy, while the second statistically distinguishes primary particles. More information is, however, hidden in the details of the shower longitudinal profile shape [80, 130, 131]. Accessing that information, especially in a way in which it can be physically interpreted, is not straightforward.

In this chapter, we will introduce a parametrization of the shower shape in terms of the variables  $L$  (shower width) and  $R$  (asymmetry) [80]. This is a parametrization of the shower as a function of its depth,  $X$ . Following our work [133], the reasoning for the choice of working in atmospheric traversed depth instead of in shower age (as the HiRes/MIA [78] and Hires-II [79] collaborations did) will be presented. In data, determining the longitudinal profile shape is very hard: the region far from the maximum is either poorly measured due to the smaller number of measured photons compared to the ones at maximum, or not measured at all due to quality selection requirements and field-of-view constraints. We will present the idea of measuring the longitudinal profile shape averaged over several hundred or thousands of events, and present the parameters  $R$  and  $L$  resultant from its fit. The interpretation of the parameters in terms of primary composition and hadronic interaction model used will also be done. This is mostly based on the work at [99].

### 3.1 Longitudinal profile shape

The electromagnetic profile of air showers is mostly determined by the maximum number of particles,  $N_{\max}$ , and its corresponding depth,  $X_{\max}$ . After translation to  $X' \equiv X - X_{\max}$  and normalization  $N' \equiv N/N_{\max}$ , the profile for different energies, primaries and hadronic models has a quasi-universal shape [134, 135, 136], hence its common denomination Universal Shower Profile (USP). However, a closer

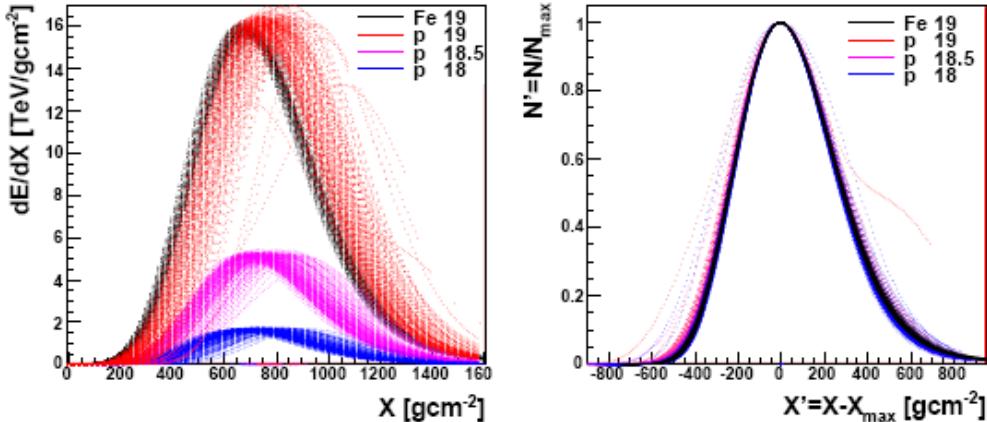


Figure 3.1: Longitudinal profile for 150 showers of each of the energy and composition indicated. Left: Longitudinal shower profile, shown as  $dE/dX$  vs  $X$ . Right: USP, shown as  $N'$  vs  $X'$

inspection of figure 3.1 shows that the shower development is faster for iron initiated showers than proton ones. This is expected from the fact that heavier nuclei have larger multiplicities on the first interaction, and shows that extra information can be retrieved from the electromagnetic profile even when the influence from its two main parameters has been removed.

The shower profile is typically parametrized by a four parameter Gaisser-Hillas (GH) [76] function to describe the energy deposit versus traversed depth,  $X[\text{g cm}^{-2}]$ , with the shape being described by the parameters  $\lambda$  and  $X_0$ . However, several different parametrizations have been proposed in the literature. Historically, one of the first was a three parameter Gaussian in  $X$  [137], which was appealing from an experimental viewpoint as it only depended on one parameter to characterize the shower shape. This hypothesis has been discarded as it did not describe the first averaged profiles presented by the HiRes/MIA collaboration [78]. Interestingly, this work, as well as a subsequent one by the HiRes-II collaboration [79], found that three functional forms described data comparably well: the previously noted Gaisser-Hillas [76], the Greisen function [45], motivated by analytical shower theory, and a Gaussian in Age (GIA) [78], motivated by the observation that showers were symmetric when plotted in shower age,  $s = \frac{3X}{X+2X_{\max}}$ . A strong correlation between the GH shape parameters was also found in these papers, which together with the fact that a simpler (1 parameter) function described the average profile shape equally well, led the authors to adopt the Gaussian in age as the functional form for

their future analysis, with its width being the measured variable to be physically interpreted. These different parametrizations and the differences between averaging showers in depth or in shower age will be detailed in this chapter.

### 3.1.1 Electromagnetic profile parametrization - Gaisser-Hillas

The most common and widely accepted longitudinal profile parametrization (and the one used at the Pierre Auger Observatory) is the Gaisser-Hillas function, which is conventionally written:

$$N(X) = N_{max} \left( \frac{X - X_0}{X_{max} - X_0} \right)^{\frac{X_{max} - X_0}{\lambda}} \exp \left( \frac{X_{max} - X}{\lambda} \right) \quad (3.1)$$

where  $\lambda$  is usually identified with an effective interaction length and  $X_0$  is related to the point of first interaction. This function has been shown to describe showers very well, at the percent level [138]. However, it remains an open question whether physical information can be extracted from the shape parameters.

Since the idea is to isolate the shape information by translating and normalizing the profile, from now on we will mostly write the Gaisser-Hillas function in terms of the USP variables,  $N'$  and  $X'$

$$N' = \left( 1 - \frac{X'}{X'_0} \right)^{-\frac{X'_0}{\lambda}} \exp \left( -\frac{X'}{\lambda} \right) \quad (3.2)$$

where  $X'_0$  is the translated  $X_0$ , i.e.,  $X'_0 = X_0 - X_{max}$ . The main drawbacks of this parametrization are the difficulty in interpreting the meaning of the variables and, more importantly, the correlation between both of them [79]. In [53], a simple test was performed to quantify this correlation: different realizations of a single profile with statistical fluctuations in the number of photons and finite binning in  $X$  were fitted. The pair of values retrieved from all these realizations are shown in figure 3.2, where it can be seen that both parameters have variations of up to 100% and are strongly correlated. This was the main motivation behind a different parametrization of equation 3.2.

The Gaisser-Hillas equation can be written as:

$$\begin{aligned} N' &= \exp \left( -\frac{X'_0}{\lambda} \log \left[ 1 - \frac{X'}{X'_0} \right] - \frac{X'}{\lambda} \right) \\ &= \exp \left( -\frac{X'^2}{2|X'_0\lambda|} \right) \prod_{n=3}^{\infty} \exp \left( \frac{1}{n} \frac{X'_0}{\lambda} \left( \frac{X'}{X'_0} \right)^n \right) \\ &= \exp \left( -\frac{1}{2} \left( \frac{X'}{L} \right)^2 \right) \prod_{n=3}^{\infty} \exp \left( \frac{R^{n-2}}{n} \left( -\frac{X'}{L} \right)^n \right) \end{aligned} \quad (3.3)$$

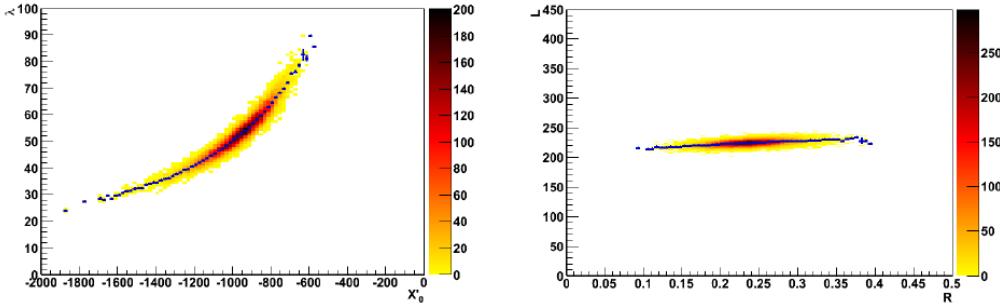


Figure 3.2: Correlation between USP variables for several realizations of the same shower with 200 photons at maximum. Left: Correlation between  $\lambda$  and  $X'_0$ . Right: Correlation between  $R$  and  $L$ . From [53].

where the Taylor expansion of the logarithm around the maximum  $X' = 0$  evinces the profile shape properties. The first term in  $X'$  cancels leaving a Gaussian of characteristic width  $L = \sqrt{|X'_0 \lambda|}$ , and a distortion term governed by the parameter  $R = \sqrt{\lambda / |X'_0|}$ . For positive  $X'$ , even and odd terms in  $R$  have partial cancellation, while in the negative region all terms contribute to the distortion, so there is more sensitivity to the  $R$  parameter for  $X' < 0$ .

The exercise made previously to test the correlation of the standard Gaisser-Hillas function parameters was repeated for this parametrization. The results are shown in figure 3.2, where the scale in  $L$  was chosen to be the same as in the  $X'_0$  and  $\lambda$  plot<sup>1</sup>.  $R$  and  $L$  are more stable than the standard variables, with  $L$  in particular varying by less than 10%, and their correlation much weaker. Still, there is still a positive correlation between them as can be seen in the superimposed profile in figure 3.2.

We thus prefer to write equation 3.2 as

$$N' = \left(1 + R \frac{X'}{L}\right)^{R^{-2}} \exp\left(-\frac{X'}{RL}\right) \quad (3.4)$$

where  $L$  is associated with the profile width and  $R$  with the asymmetry between its rising and falling regions. This is shown pictorially in figure 3.3. Increases in  $L$  lead to a larger profile on both sides, while increases in  $R$  lead to a faster shower rise (larger asymmetry): less energy is deposited before the shower maximum, which is compensated by a slower decrease on the positive relative depth region.

<sup>1</sup>Since  $L = \sqrt{|X'_0 \lambda|}$ , the maximum in the  $L$  scale was defined with the maximum absolute value of both standard variables in their plot, i.e., -2000 and 100 for  $X'_0$  and  $\lambda$  respectively

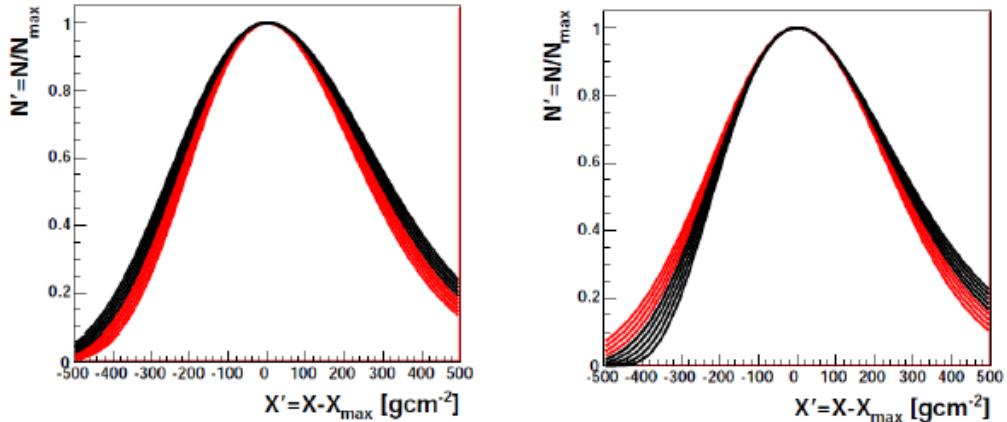


Figure 3.3: USP for varying  $R$  and  $L$ . Red lines corresponds to smaller values of the parameter, and black to bigger ones. Left: USP for varying  $L$  around  $225 \text{ g/cm}^2$  in  $5 \text{ g/cm}^2$  steps. Right: USP for varying  $R$  around  $0.25$  in  $0.05$  steps [?]

The shower energy is given by the integral of this profile, corrected for the invisible energy, which is carried by muons and neutrinos. This varies accordingly to primary particle and hadronic model used, and is expected to decrease as energy increases, but is small compared to the total energy, around 5-10%. The integral can be written in terms of the Gamma function as:

$$\frac{E}{dE/dX|_{X_{max}}} = (LR)A^A \exp(A) \cdot \Gamma(A + 1) \approx \sqrt{2\pi}L \quad (3.5)$$

with  $A = |X'_0|/\lambda = R^{-2}$ . This approximation, for typical values of  $R$  ( $< 0.35$ ) results in an underestimation of the integral smaller than 1%. So, it can be considered that the electromagnetic shower energy is determined by  $L$  and the energy deposit at maximum, while  $R$  determines the balance between the left and right side of the USP <sup>2</sup>.

Finally, the composition information contained in the profile shape can be extracted. Showers initiated by iron and proton primaries at  $10^{19}$  eV using QGSJet-II model were simulated (model and energy dependence discussed ahead, in 3.3), and the longitudinal profile of each fitted with equation 3.4. The distributions of both parameters for proton and iron primaries at  $10^{18.5}$  eV are shown in figure 3.4.  $L$  is a quasi-measurement of the electromagnetic energy. It varies very little from

<sup>2</sup>It should be noted that a similar parametrization was independently developed [82], where two parameters  $fhwm$  (shower width at half maximum) and  $f$  (asymmetry parameter, equal to the fraction of that width that is to the left of the maximum -  $1/2$  for a symmetric profile).

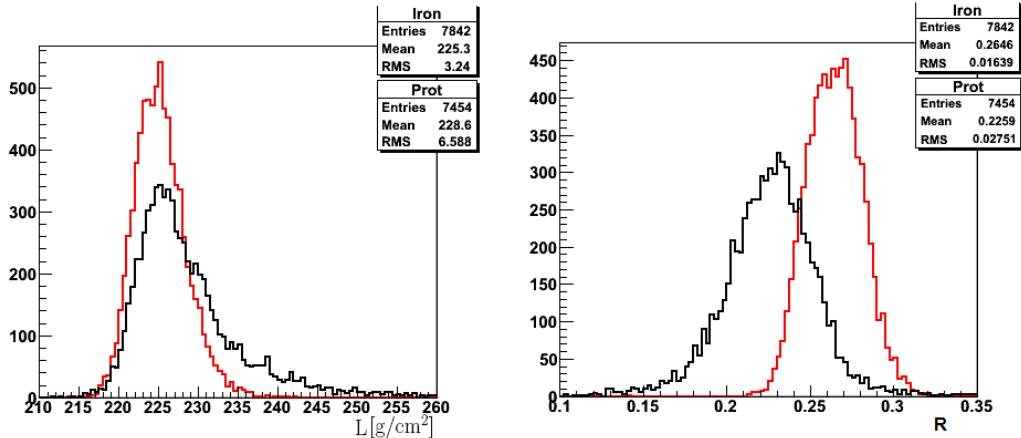


Figure 3.4:  $L$  (left) and  $R$  (right) distribution for  $\log(E/\text{eV})=18.5$  events simulated with QGSJET-II. Mean and RMS indicated. Red lines represent iron nuclei, black lines protons, simulated with QGSJet-II as the hadronic interaction model.

shower to shower (around 2%, and larger for proton due to larger variations in the first interactions) and is very similar between both primaries. The value for iron is slightly smaller than the one for proton due to the larger fraction of energy going to muons and neutrinos.  $R$  can be considered a composition variable. While there is some overlap in the proton and iron distributions, the lower and upper tails can yield fairly a pure proton and iron sample respectively.

### 3.2 Profile shape in shower age vs atmospheric depth

An alternative method of profile shape analysis is to do it as a function of the shower development stage with respect to the profile maximum - the shower age  $s = \frac{3X}{X+2X_{\max}}$  - instead of traversed atmospheric relative depth,  $X'$ . The concept of shower age first emerged in the study of the longitudinal development of photon or electron initiated showers [139], but it has also been shown that distributions of electromagnetic particles in hadronic showers (initiated by proton and iron primaries) show universal behaviour when expressed in age [131].

In figure 3.5, the average normalized energy deposit ( $dEdX' \equiv dEdX/dEdX_{\max}$ ) profiles in shower age and relative depth,  $X'$ , are shown, for proton and iron primaries using QGSJetII-03 as the high energy hadronic model. The difference in energy deposit between primaries is also plotted in figure 3.5. It is clear that the profiles are much more similar between primaries when they are plotted in relative

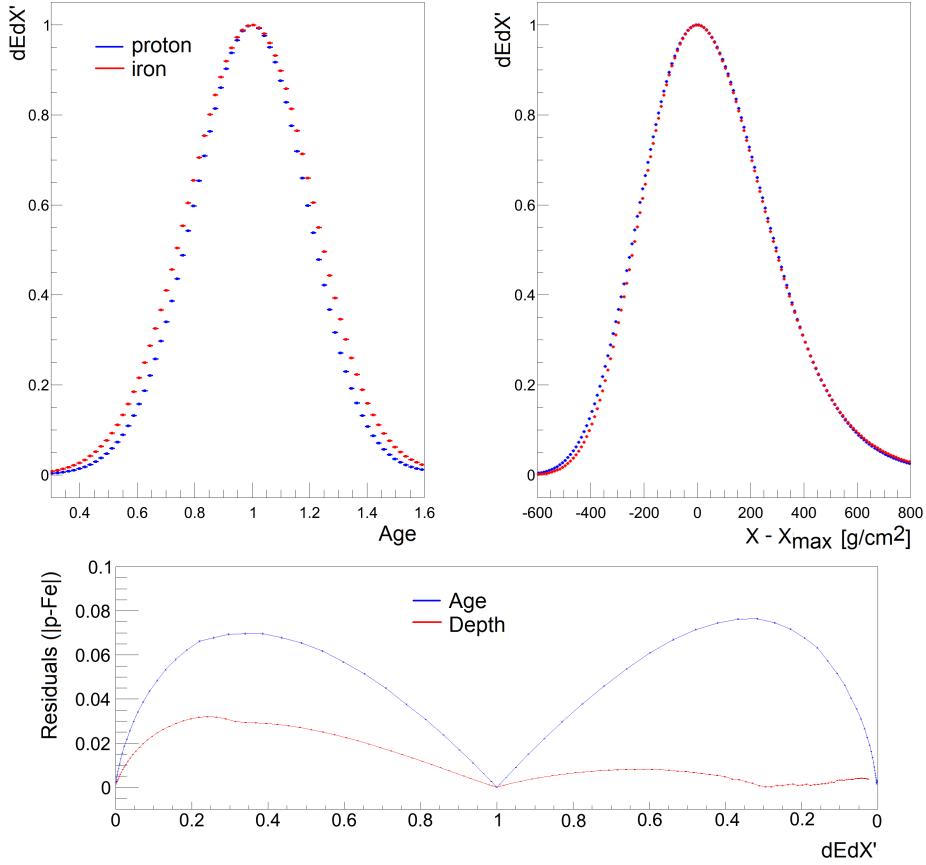


Figure 3.5: Top: Average profile in age (left) and depth (right) for proton (blue) and iron (red) at  $10^{19}$  eV. Bottom: Absolute value of the difference between proton and iron in the profiles in age (blue) and in depth (red) as a function of  $dE/dX'$  of the proton for the left and right side of the maximum

depth, particularly after the maximum. This has already been noted in [136], where the universality of the shower shape in  $X'$  (not only between primaries but also energy and chosen hadronic model) was used to study whether the Cherenkov contribution was correctly estimated in the Pierre Auger Observatory data. However, we also note that while profiles in  $X'$  are indeed very similar after the maximum, there is a visible difference in the negative region of  $X'$ , particularly between -400 and -200  $g\text{ cm}^{-2}$  (0.1 to 0.7 in  $dE/dX'$ ). This is consistent with the discussion in the previous section, as the rising region of the profile is more sensitive to differences in the asymmetry parameter  $R$ . The profiles in shower age in turn, have a large difference between primaries (and also hadronic model used) and are fairly symmetric.

In the first stages of fluorescence detector profiles analysis, the shower was de-

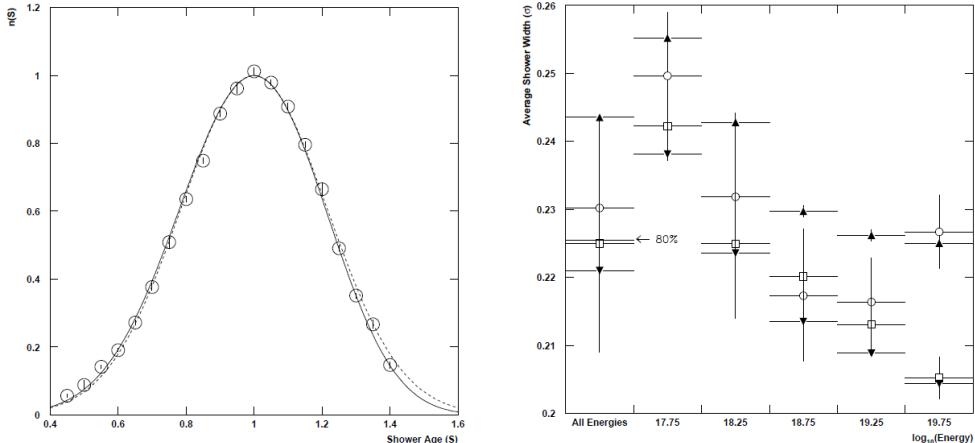


Figure 3.6: Left: average Shower  $10^{19} - 10^{19.5}$  for the HiRes-II data. Fit to GH (Solid line) and gaussian in age (Dotted line). Right: average Shower widths for the HiRes-II data. Empty Circles: Data. Empty Squares: MonteCarlo. Lower Triangles: Proton. Upper Triangles: Iron (from [79])

scribed by a three parameter (just one for the shape) Gaussian in X. The first published high quality longitudinal profile by the HiRes/MIA collaboration [78] discarded this hypothesis, but it was noted that the average profile for the same event sample, when written in shower age, could be described by a Gaussian in Age:

$$dEdX'(s) = \exp \left[ \frac{1}{2\sigma_{age}^2} (s - 1)^2 \right] \quad (3.6)$$

where  $\sigma_{age}$  is the sole shape parameter. This is a much simpler function than the Gaisser-Hillas written in shower age:

$$dEdX'(s) = \left( 1 - \frac{1-s}{3-s} \frac{3T_{max}}{T_{max} - T_0} \right)^{T_{max} - T_0} \exp \left[ 3T_{max} \frac{1-s}{3-s} \right] \quad (3.7)$$

with two shape parameters,  $T_{max} = X_{max}/\lambda_{crit}$  and  $T_0 = X_0/\lambda_{crit}$ , with  $\lambda_{crit} = 36.7 \text{ g cm}^{-2}$ .

In [79], both functional forms were tested against measured profiles, yielding similar results, and both compatible with data (figure 3.6). Hence, the simpler function was chosen for the subsequent analysis. The results of the Gaussian width,  $\sigma_{age}$ , of profiles in five energy bins between  $10^{17.5}$  and  $10^{20}$  eV for the HiRes-II data are shown in figure 3.6, together with the expectations for a full detector simulation of proton and iron showers. They are consistent with an 80% proton composition (from the  $X_{max}$  analysis) for all energy bins except the last one, from  $10^{19.5}$  to  $10^{20}$  eV, where data points to a composition closer to iron.

Unfortunately, since in this work data is compared to the full detector simulation of that experiment, a direct comparison is not possible. So, we build the average

profile for Auger data trying to follow as much as possible the same method used in [79], and compare them to our simulations using the same criteria. The same set of selection cuts were applied, except for the atmospheric quality ones where, since the methods for requiring proper atmospheric conditions are different, we decided to keep the usual ones in the Auger analysis, described later in this work 4.1.2. The resultant profile for energies between  $10^{19}$  and  $10^{19.5}$  eV is shown in figure 3.7. Superimposed is the profile for the same energy range measured by the HiRes-II experiment, and the one using the standard Auger data selection (as detailed in 4.1). The HiRes-II average profile is visibly wider than the Auger one, even if the same selection cuts are used. However, in figure 3.6 we can see that this width agrees with the full detector simulations, so the cause of this widening of the profile has already been folded into the simulation.

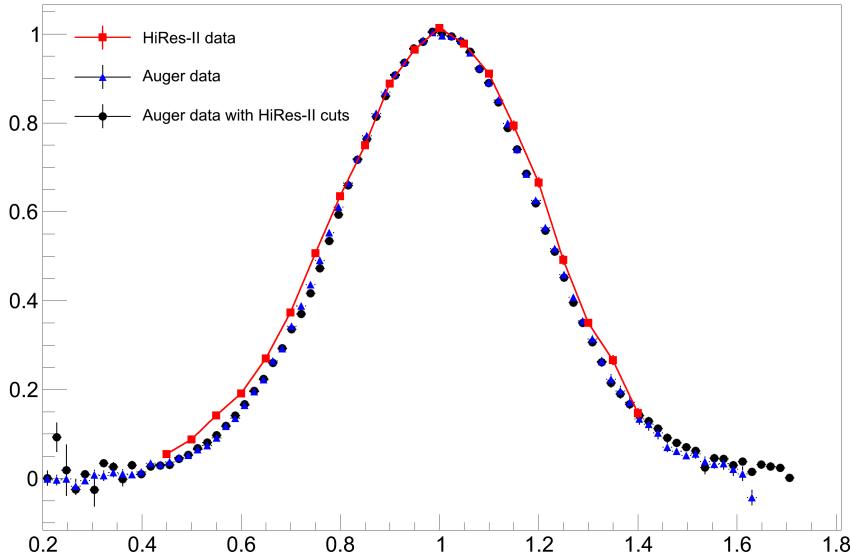


Figure 3.7: Average profile in age for the energy bin  $10^{19}$  to  $10^{19.5}$  eV. The HiRes-II profile (from [?]) is shown in red, Auger data with Xmax cuts in blue and with HiRes-II cuts in black.

The shower profiles in age have been parametrized by a Gaisser-Hillas 3.7 and a Gaussian 3.6, as previously described. Another popular parametrization is the Greisen function [45], which is motivated by analytic shower theory, and can be written as:

$$dEdX'(s) = \frac{0.31}{\sqrt{s}} \exp \left[ s \left( 1 - \frac{3}{2} \log[s] \right) \right] \quad (3.8)$$

with only one shape parameter,  $y$ . An overview of the functional shapes used to describe the shower development, and the relationship between can be found in [148]. Finally, just for the purpose of fully reproducing the results of the HiRes/MIA collaboration we test the hypothesis that the profile is a Gaussian in relative depth.

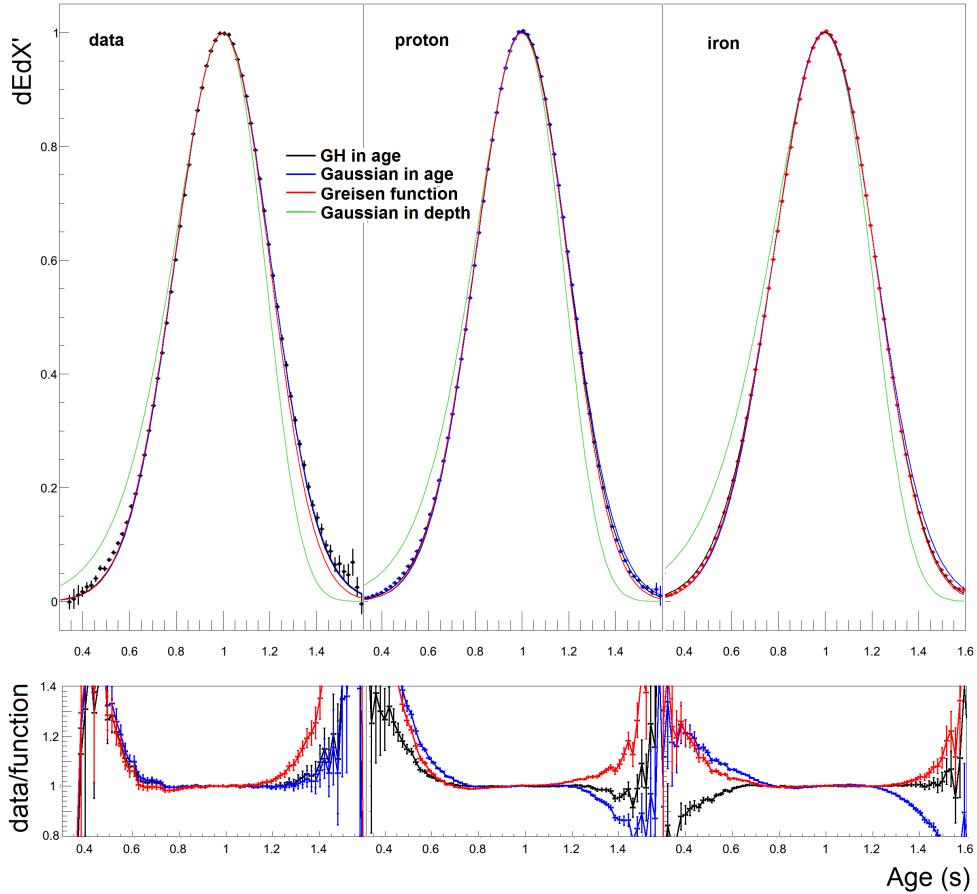


Figure 3.8: Average profile in age for the energy bin  $10^{19}$  to  $10^{19.5}$  eV. Auger data is shown on the left and real MC for proton and iron on the middle and right, respectively. The functions shown in the legend are also drawn with the values resulting from their fit to data. On the bottom plots we show data divided by the value of the fitted function in each bin. Gaussian in depth is not shown because the fit is very poor.

The fits of these four functions to data and simulated profiles are shown in figure 3.8. It can be seen that there is a very good agreement in data for the GH and Gaussian in age functions. The Greisen function has a fixed asymmetry which fits well the simulations but not the higher values of age in data. Note that this is the part where the uncertainty due to the increasing Cherenkov contribution is larger. The Gaussian in depth does not describe data nor simulations, which is in agreement with the findings on [78]. Since the fit to a Gaussian in Age is compatible with data

and simulations, we can retrieve the shape parameter from the fits and compute its energy evolution, comparing it with the one for proton and iron primaries. The  $\sigma_{age}$  as a function of energy is shown in figure 3.9, for both the case in which the HiRes-II cuts and the case where the Auger ones were implemented. It should be noted that in the lowest energy bin the event selection criteria was only studied down to  $10^{17.8}$  eV, so the bin starting at  $10^{17.5}$  eV may be subject to larger uncertainties.

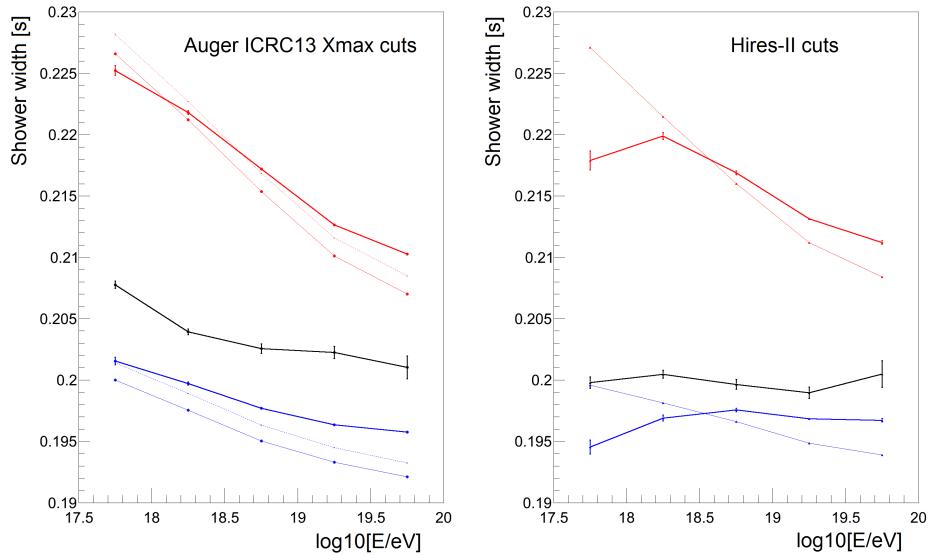


Figure 3.9: Gaussian in age  $\sigma_{age}$  parameter as a function of energy. Auger data in black, proton in blue and iron in red. The solid thick line is *realMC*, solid thin line is generated profiles and thin dashed line is when a smearing of  $19 \text{ g/cm}^2$  in  $X_{max}$  is added to the generated average.

The reconstructed shower width for Auger *realMC* is consistent with the generated one, but in general slightly larger. One of the effects that contributes to this widening is the resolution in  $X_{max}$ . So, we smeared it  $19 \text{ g/cm}^2$  (from stereo showers) and calculated the age for each individual profile with this new  $X_{max}$ . The results are shown as dashed lines in figure 3.9. This effect partly explains the effect in  $\sigma_{age}$ , and the remaining bias w.r.t. reconstruction is small (much smaller than the proton iron separation). Auger data lies between proton and iron, always closer to proton and moving slightly closer to iron with increasing energy. There is a linear relation between the width  $\sigma_{age}$  and the average mass of the primary particles,  $\langle \ln A \rangle$ . So, noting that for proton  $\ln[A] = 0$  and for iron  $\ln[A] \approx 4$ , we can use the distance of data to both lines (for a given model) to estimate  $\langle \ln A \rangle$ . The  $\langle \ln A \rangle$  plot for the model used in this work (QGSJet-II 03) is shown in figure 3.10.

The average width,  $\sigma_{age}$ , measured by both experiments is consistent (when

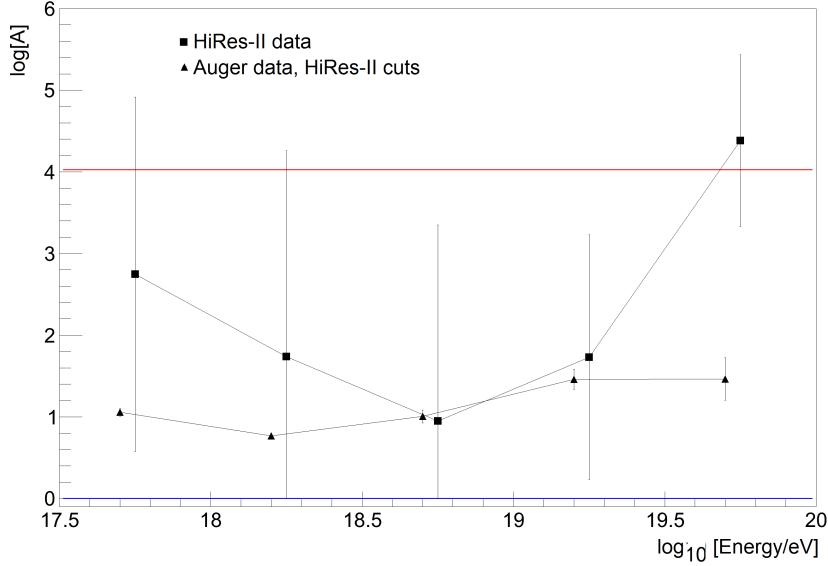


Figure 3.10:  $\ln(\mathcal{A})$  from the  $\sigma_{age}$  measurement from HiRes-II (points) and Auger with HiRes-II cuts (triangles). Both measurements are in the same energy bin, the shift in energy between them is just for easier visualization. Note that the Auger points only have the statistical error, while the HiRes-II points are transformed by us from figure 3.6, and the errors included in making the plot are not enumerated on [79]. Therefore, a comparison between the errors cannot be made.

compared to each *realMC*) for all energy bins except the last one, where HiRes-II has wider showers than Auger.

### 3.2.1 What is average age measuring?

We can measure the "width"  $L$  in relative depth (although the profile is not really a gaussian),  $X_{max}$  and  $\sigma_{age}$  in age. However, it can be useful to get the relationship between these variables. We can express the age in  $X'$  as:

$$s = \frac{3X}{X + 2 \cdot X_{max}} = 1 + \frac{2X'}{3 \cdot X_{max} - X'} \quad (3.9)$$

Assuming that at around the maximum the difference in asymmetry between profiles is negligible it is possible to define variables  $\sigma_{X' \pm}$

$$\sigma_{X'+} = \left| \frac{2L}{3 \cdot X_{max} - L} \right| \quad (3.10)$$

$$\sigma_{X'-} = \left| \frac{-2L}{3 \cdot X_{max} + L} \right| \quad (3.11)$$

by substituting  $X'$  by its  $\pm 1$  "σ" value,  $\pm L$ . Then the value of  $\sigma_{age}$  from the fit in age is comparable to the measured "average width" in depth when corrected by

$X_{max}$ :

$$\sigma_{X' \pm} = (\sigma_{X'+} + \sigma_{X'-})/2 \quad (3.12)$$

An approximated  $L_\sigma$  can equivalently be calculated by inverting equation 3.12:

$$L_\sigma = \frac{3X_{max}}{\sigma_{age}} (\sqrt{(1 + \sigma_{age}^2)} - 1) \quad (3.13)$$

In figure 3.11, the value of  $L$  measured from the average profile in depth and  $L_\sigma$  are plotted. The agreement between both is very good, both in absolute value and slope with energy. This indicates that equation 3.12 is a valid approximation, and the width in age,  $\sigma_{age}$ , is mostly a function of  $L$  and the average  $X_{max}$ .

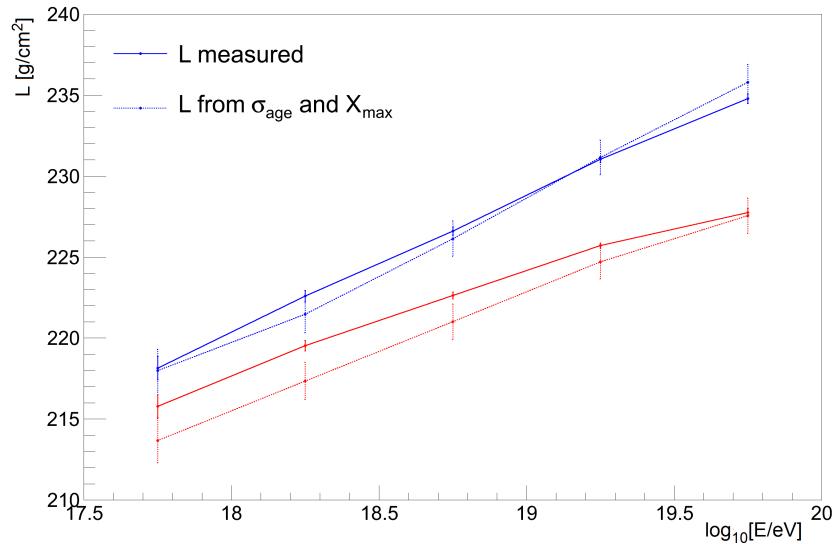


Figure 3.11:  $L$  measured (full line) and  $L$  calculated from  $X_{max}$  and  $\sigma_{age}$  (equation 3.13) as a function of energy for proton (blue) and iron (red) primaries

Now we can see what governs the behaviour of  $\sigma_{age}$  by differentiating it with respect to these two variables:

$$\frac{\partial \sigma_{age}}{\partial X_{max}} = -6L \left( \frac{1}{(3X_{max} - L)^2} + \frac{1}{(3X_{max} + L)^2} \right) \quad (3.14)$$

$$\frac{\partial \sigma_{age}}{\partial L} = -6X_{max} \left( \frac{1}{(3X_{max} - L)^2} + \frac{1}{(3X_{max} + L)^2} \right) \quad (3.15)$$

Noting that the separation in  $X_{max}$  between proton and iron is  $\approx 80 \text{ g/cm}^2$ , in  $L$  is  $\approx 4 \text{ g/cm}^2$ , and typical values for  $L$  and  $X_{max}$  are 220 and  $700 \text{ g/cm}^2$  respectively:

$$\frac{\Delta\sigma_L}{\Delta\sigma_{X_{max}}} = \frac{\frac{\partial\sigma_{age}}{\partial L} \cdot \Delta_L}{\frac{\partial\sigma_{age}}{\partial X_{max}} \cdot \Delta_{X_{max}}} \approx \frac{X_{max} \cdot 4}{L \cdot 80} \approx \frac{1}{6} \quad (3.16)$$

This means that  $\sigma_{age}$  is governed by the value of  $X_{max}$ , hiding the physical width of the shower. This is consistent with what we see by looking at the proton-iron difference in  $L$  and  $\sigma_{age}$ : the separation in  $L$  is around 1-2% (221 for iron and 224 g cm<sup>-2</sup>) and the separation in width in age is closer to 10% (0.20-0.22), with this difference coming from the different average  $X_{max}$  between primaries.

Since we are trying to derive two independent measurable variables, we choose to use atmospheric depth as a transformation for the universal shower profile analysis, as the profiles in age are strongly influenced by the value of the shower maximum, and thus any variable retrieved from it will be correlated with  $X_{max}$  and add little information to the previously measured elongation rate.

### 3.3 Average profile shape - sensitivity to composition and hadronic models

The full shape of the longitudinal profile of each shower has been shown to be sensitive to primary mass composition, in an event-by-event basis, using simulated showers. However, from the experimental point of view this is not an easy measurement. On currently measured profiles, particularly those at the Pierre Auger Observatory, on which this work is focused, the Gaisser-Hillas fit to extract the shape parameters is affected by the statistical uncertainties on the light collection by the fluorescence detectors and the uncertainty on the determination of the shower maximum. Also, the telescopes have a finite field-of-view, which means that while all showers that pass the quality selection have a region around the maximum well measured, the rising part of the shower (which is the most sensitive to the proton-iron separation) is very often not seen. Moreover, this kind of optical experiments have several systematic uncertainties such as Cherenkov direct and scattered light estimation, telescope alignment, atmospheric conditions, among others [126, 140]. This makes it extremely difficult to extract the information on the shape, in particular to access the two shape parameters at the same time.

One way to minimize this limitations is to take a different approach: measuring the shape of several showers averaged within an energy range [98]. By doing so, the statistical uncertainties are reduced, so that the two shape parameters,  $L$  and  $R$ , can

be measured simultaneously. This increased resolution can also help to control the experimental systematic uncertainties on the profile measurement, enabling future event-by-event analysis (for instance, the results between different telescopes and years should be statistically consistent).

Such strategies rely on the assumption that the small shape differences introduced by the changing hadronic interaction properties is not washed away on the averaging, i.e., that the average universal electromagnetic longitudinal shower profile still has information about the primary mass composition. To investigate this, 10000 showers of proton, helium, nitrogen and iron, for different hadronic interaction models (Sibyll 2.1 [143], QGSJet-II.04 [144] and EPOS-LHC [145, 146]) were generated, using CONEX [141, 142], and different energies, being the reference energy throughout this section  $E = 10^{19}$  eV. The zenith angle was fixed at  $40^\circ$ , but it has been shown previously [80, 147] that the shape does not depend significantly on this quantity.

The first test to the method is averaging several longitudinal profiles, fitting them with a Gaisser-Hillas function and looking at the stability of  $R$  and  $L$  as the number of averaged shower grows. The fitting range was chosen as  $X' \in [-300, 200]$  g cm $^{-2}$ . Although the choice of this range is somewhat arbitrary, it is motivated by experimental constraints, as we try to include a range as big as possible to increase the significance of the procedure while avoiding tails which are subject to selection bias and have large systematic uncertainties, namely due to the Cherenkov scattering light component estimation. More details about the choice of the fitting range in 4.3.1.1. Figure 3.12 shows  $R$  and  $L$  for different primaries as a function of the number of events that was used to perform the average. It can be seen that in fact, the average electromagnetic profile is sensitive to the difference between primaries. Moreover, the values of  $L$  and  $R$  converge after averaging a few hundred events allowing to distinguish between the different primaries.

### 3.3.1 Sensitivity to primary mass composition

It has been seen in the previous sections that the average longitudinal electromagnetic shape variables are both individually sensitive to the nature of the cosmic ray that induces the shower. More information can be retrieved if both variable are simultaneously plotted in function of each other for a fixed energy. This is shown in figure 3.13. All possible primary mass combinations for the chosen energy and hadronic model are shown using four primaries: proton, helium, nitrogen and iron.

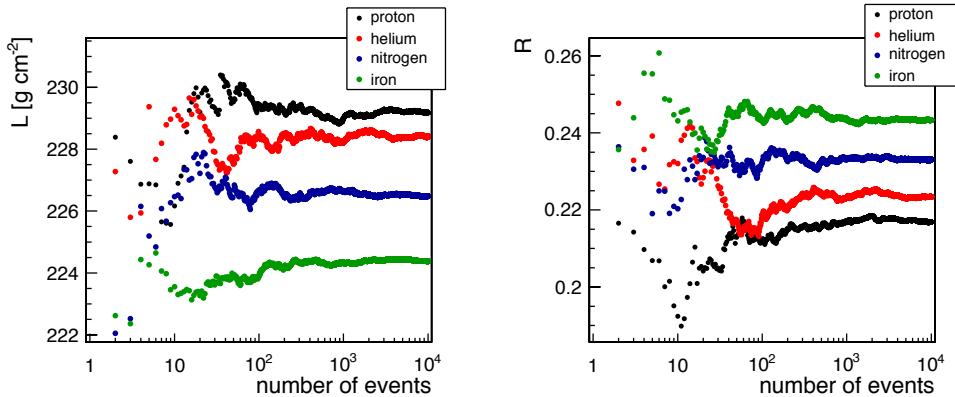


Figure 3.12: Shape parameters  $L$  (left) and  $R$  (right) as a function of the number of events used to get the average shower. The showers were generated using QGSJet-II.04 as the high energy hadronic interaction model. Color codes show different primaries at  $E = 10^{19}$  eV.

Superimposed are crosses representing the values for pure mass composition samples.

Each coloured dot represents a particular mass composition scenario and the colour is the average mass logarithm given by

$$\langle \ln A \rangle = f_p \ln(1) + f_{He} \ln(4) + f_N \ln(14) + f_{Fe} \ln(56) \quad (3.17)$$

where  $f_x$  is the fraction of the element present in the sample.

The values of both parameters give two measurement of  $\langle \ln A \rangle$ , with small variations according to the exact composition of the considered mixture: a pure sample of helium primaries will have an only slightly lower value of  $R$  and larger value of  $L$  than a proton and iron mixture with the same average mass. This difference is in fact very small, and we can conclude that both variables are for most purposes linear in  $\langle \ln A \rangle$ . This is very useful, as a way to test the validity of hadronic models is to compute the  $\langle \ln A \rangle$  inferred from the maximum number of variables and see if they are all consistent, as explained in section 6.2.

### 3.3.2 Sensitivity to Hadronic Interaction Models

The dependence of the shape variables with the hadronic interaction models was also investigated. It can be seen, in figure 3.14 that there is some discrimination power for the standard hadronic interaction models. Although EPOS-LHC and QGSJet-II.04 have some overlap there are mass composition choices for which the models occupy a different phase space, for instance for pure proton.

This becomes even more clear when we look at predictions for the shape variables

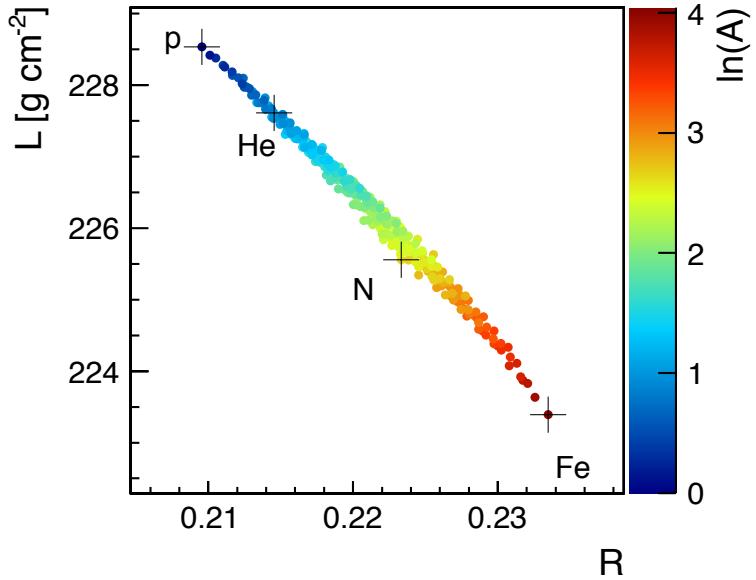


Figure 3.13: Relation of the shape parameters of the average shower profile with the  $\langle \ln A \rangle$ . The showers were generated using QGSJet-II.04 at  $E = 10^{19}$  eV. Color codes show the  $\langle \ln A \rangle$  for different combinations of primaries and crosses mark the results for pure composition samples. Each point was generated averaging 10000 shower longitudinal profiles.

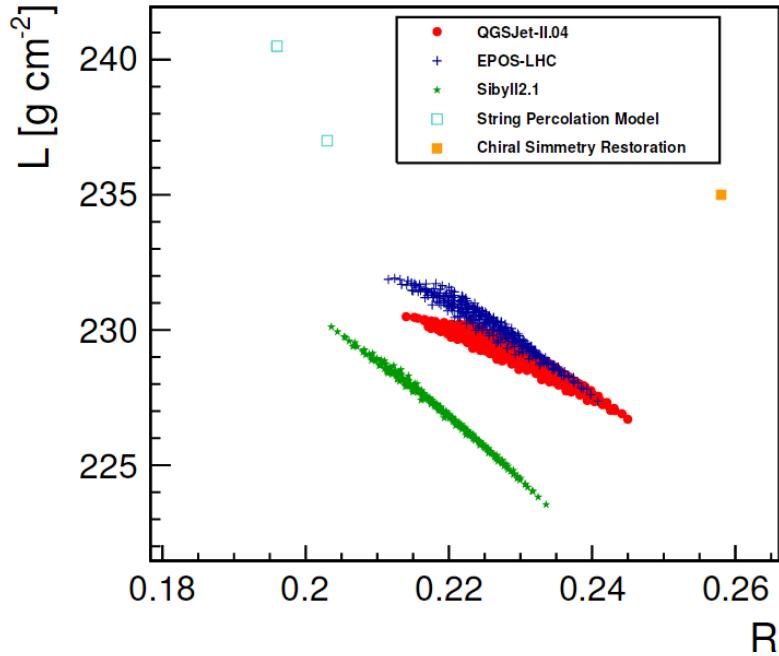


Figure 3.14: Shape parameters  $L$  and  $R$  for  $E = 10^{19}$  eV, for all possible different mass composition combinations with standard hadronic interaction models and two models with alternative interactions at the highest energies (details in legend).

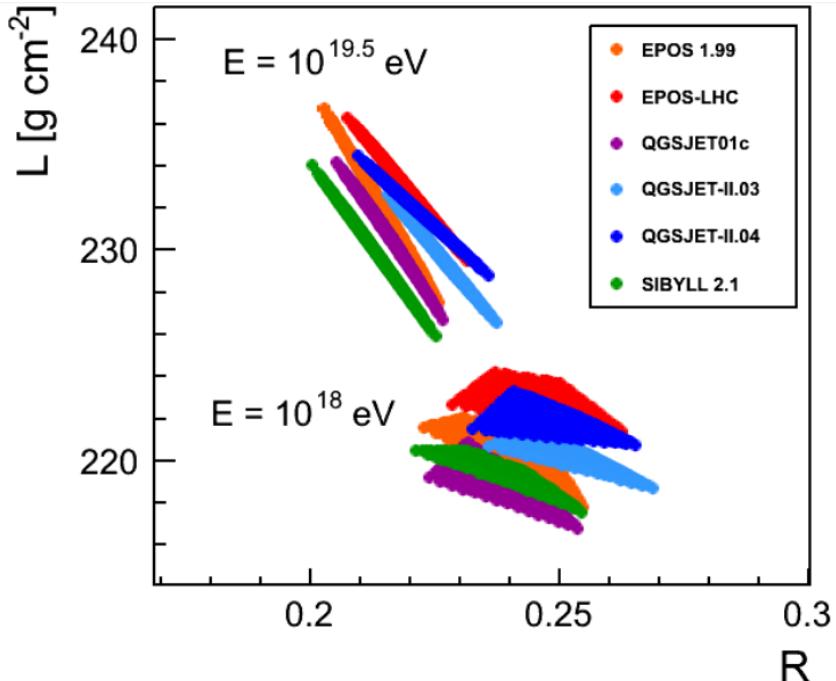


Figure 3.15: Shape parameters  $L$  and  $R$  for  $E = 10^{19.5}$  eV (top) and  $E = 10^{18}$  eV (bottom) for all possible different mass composition combinations and all standard hadronic interaction models (see legend for details).

of the electromagnetic average profile from "alternative" models (models that try to explain the UHECR data through changes on the hadronic interaction physics, while the mass composition remains constant) which are displayed in figure 3.14. It can be seen from this plot that the shape variables have a strong discrimination power for this kind of models. Moreover, note that these models were tuned to reproduce the  $X_{\max}$  and the number of muons at ground. Therefore, the measurement of the shape of the average electromagnetic profile is an important test to these exotic scenarios: changing the number of muons, for example, inevitably implies changes in the pion distributions, which affect the electromagnetic profile through the decay of the neutral ones - keeping the  $X_{\max}$  more or less constant means that the shape has to be affected.

The energy evolution of the phase space of the shape variables for different models was also investigated and is presented in figure 3.15 (for all the available high energy hadronic interaction models, including older versions). All the models display a similar behaviour: as the energy increases so does  $L$ , and  $R$  decreases. The first is expected from the previous section, since  $L$  is very correlated with

energy, and within a few percent  $E_{em} = \sqrt{2\pi}L\frac{dE}{dX}|_{max}$ . On the other hand, the decrease of  $R$  is an indication that the shower is getting less asymmetrical (more gaussian). Moreover, the distance between the models phase space is being reduced as the energy increases. This suggests that the shower is becoming more universal and, apart from sudden changes to the hadronic interaction properties at a certain threshold (as in the alternative models plotted), there is less freedom for the hadronic parameters to impact the longitudinal shower development. Finally, it is worth mentioning that the post-LHC version of the models seems to occupy a different phase space region, favouring higher  $L$  and  $R$  parameters. These models have been re-tuned not only to describe better the new Large Hadron Collider (LHC) data but also to reproduce the number of muons at ground measured by UHECR experiments, which is larger than the one predicted by standard models. The increase of the muon content on simulated showers changes in the pion sector of the shower, and this should be reflected in the shower average longitudinal shape, which is governed by the decay of neutral pions. This means that the assessment of the  $L$  and  $R$  of the average profile may be used to evaluate some of the tunings of the hadronic interaction models.

### 3.4 Conclusions

The interpretation of UHECRs in terms of mass composition is bounded to our understanding of hadronic interactions at the highest energies. It has been shown that the shape of the shower profile, when measured in relative depth and not on shower age, provides a new insight on shower mass composition and hadronic interaction physics. Although very promising, these quantities are not easy to access experimentally on an event-by-event basis. Hence, in this work we propose to measure instead the shape of the average electromagnetic shower profile. In this way one could measure simultaneously the two shape parameters. It has been demonstrated here that the shape of the average electromagnetic longitudinal profile is both sensitive to mass composition and to hadronic interaction models.

In the next chapter, the details of the implementation of this analysis on the profiles measured at the Pierre Auger Observatory will be described.



# CHAPTER 4

## Data Selection & Analysis

---

The Pierre Auger Observatory FD detectors have measured more than 5 million events since they started operating in 2004. However, not all of them can be used in this analysis. To measure the average profile shape, a stringent event selection is required, namely to ensure:

- good atmospheric conditions and up-to-date aerosol measurements
- correct calibration constants and no timing issues
- a well-measured (within FOV and non-biased)  $X_{max}$  and energy, since the shape is a second order measurement in relation to these
- a well determined longitudinal profile - for example requiring long profiles with no large holes and with a good fit to a Gaisser-Hillas function.

The event selection in this work is based on the one in the  $X_{max}$  analysis [83], as the objectives are mostly the same: ensuring a non-biased  $X_{max}$  sample with good energy resolution, which in turns requires a good control on the overall energy deposit shape. Of course the two analysis are not entirely overlapping, so two more cuts were introduced to address some problems found in data profiles during the analysis. They will be described in section 4.1. Having defined the dataset, we will describe the method of averaging and show the resulting average longitudinal profiles in data. This will be presented in section 4.2. Finally, we describe the analysis that is the main focus of this work: retrieving the two shape parameters from these average profiles in different energy bins. The fitting function and range used, as well as the validation of the process with full detector simulation will be shown in sections 4.3 and 4.4. The first results, without systematic uncertainties yet, of the energy evolution of the shape parameters will also be shown at the end of this chapter.

## 4.1 Data Set

This work is based on data collected from December 2004 to the end of 2014. The vast majority of the selection criteria applied was the same as used on the  $X_{max}$  analysis paper. We will describe it briefly here, as further details can be found in [83] and [84]. Also, two extra cuts were applied, which will be explained in more detail. The number of shower candidates that are selected by the data acquisition system is about  $3.35 \times 10^6$ . The criteria used for this previous selection is very loose (signal spacial coincidence in contiguous time bins), so firstly, we need to ensure a sample with minimum quality requirements for physics analysis.

### 4.1.1 Pre-Selection

The CLF (and also XLF more recently) routinely shoot lasers (at a much lower rate than showers) which trigger the FD. Since we know their timestamp we can discard them, which is done by the **isCLF** and **isXLF** cuts. The **badFDPeriodRejection** rejects events acquired in periods where there were problems in data aquisition of the fluorescence detector. This encompasses several problems, as not having a calibration available, GPS-clock glitches, unstable baseline and bad alignment of telescopes (telescope 3 in Loma Amarilla was realigned in 2012, so the previous period is not used).

The FD reference time is based on a 1 PPS (pulse per second) signal. However, as we are measuring 100 ns time bins with a 10 MHz oscillator, it is possible that after several cycles there is a slight time delay, which can be stored and corrected [86] to ensure good synchronization between clocks (SD and FD). Candidate showers in periods in which this does not happen are discarded by the **good10MHzCorrection** cut.

Good conditions for all the photomultipliers ("pixels" in the camera) in the shower track are also required. Pixels can be saturated, in which case they are removed from the reconstruction chain (**skipSaturated** cut) or "bad", which can be for several reasons, but mostly malfunctioning or bad calibration. In the latter case, the event is discarded by the **badPixels** cut.

Finally, two fundamental basic requirements to define a reconstructed shower are demanded: reconstruction direction pointing downwards **maxZenithFD(90°)** and having a reconstructed energy (**hasFDEnergy**).

Cut (criteria)	No. of events	Efficiency (%)
Total	3.35e+06	-
Pre-Selection		
isCLF	3.28e+06	97.8
isXLF	3.26e+06	99.5
badFDPeriodRejection	2.92e+06	89.5
good10MHzCorrection	2.90e+06	99.3
badPixels	2.71e+06	93.7
skipSaturated	2.71e+06	99.7
maxZenithFD (90°)	2.69e+06	99.3
hasFDEnergy	2.42e+06	90.0
Atmosphere		
hasMieDatabase	1.80e+06	74.3
maxVAOD (0.1)	1.68e+06	93.9
cloudCut	1.27e+06	75.4
SD		
maxCoreTankDist (1500m)	506424	39.8
ambiguousHybridRejection	506211	100.0
minPBrass (0.95)	467233	92.3
Reconstruction & Field-Of-View Cuts		
minLgEnergyFD (17.8)	102791	22.0
xMaxInFOV	77607	75.5
xMaxObsInExpectedFOV	58826	75.8
FidFOVICRC13prel	21177	36.0
maxDepthHole (20%)	21008	99.5
profileChi2Sigma	20398	97.1
depthTrackLength (300 g/cm <sup>2</sup> )	20216	99.1
Profile Quality Cuts		
timeFitChi2(5)	20214	100.0
noTel6Coihueco	19547	96.7

Table 4.1: Summary of the cuts used in this work. Note that the low efficiency of the cut maxCoreTankDist is dominated by the fact that most FD events ( $\approx 60\%$ ) do not have a reconstructed hybrid geometry and not for having a core outside of a 1500 m radius.

### 4.1.2 Atmosphere

To correctly determine the transmission of light from the shower to the camera, a measurement of the aerosol content of the atmosphere is fundamental, as it varies substantially not only day-to-day, but also within the same night. The **hasMieDatabase** cut only accepts events which are within one hour of the last measurement. For this work the official period for the Auger collaboration contribution to the ICRC2015 was used, which only includes data up to December 2013. This is due to some inconsistencies found in the definition of a reference night for 2014, so to be safe that no bias was being introduced that year was not used. Events measured in unfavorable atmospheric conditions are also discarded. An aerosol content integrated from the ground to 3 km (VAOD) smaller than 0.1 is required (**maxVAOD(0.1)**).

Finally, clouds can absorb and scatter non-uniformly the fluorescence and Cherenkov light from the shower. Information from the Lidar stations and cloud cameras in each site and the Geostationary Operational Environmental Satellites (GOES) can be combined to estimate a cloud coverage. This allows us to calculate, when these measurements are available, the cloud fraction in the light integration  $\zeta$  angle. If this fraction is below 10%, it is considered cloud-free. In the case where the cloud fraction in the shower path cannot be precisely calculated we rely on the cloud height measurement. If the cloud is less than  $400 \text{ g/cm}^2$  above the minimum fiducial field of view or below the detector field of view, the event is discarded. Finally, when neither the cloud in the shower path or the cloud height can be calculated, the cloud coverage value measured by the Lidar is used. First, the Lidar of the site in which the shower was measured is tried, and if no data is available for that one, the value from one of the other sites is used. Events measured with cloud coverages above 25% are discarded (**cloudCut**).

### 4.1.3 SD

For a good shower geometry reconstruction an hybrid measurement is necessary. First, we require a measured shower core at a distance of less than 1500 m from the closest station (**maxCoreTankDist (1500m)**). The array is fully efficient above  $< 10^{18} \text{ eV}$ , but its efficiency decreases rapidly as we go down in energy. So, the inefficiency of this cut is dominated by low energy ( $< 10^{17.5} \text{ eV}$ ) nearby showers, for which the probability of triggering at least one tank in a 1.5 km triangular grid is

very low.

Not only is the array not fully efficient at lower energies and higher zenith angles, it is also not uniformly efficient for showers created by different primaries. To avoid a mass compositions bias, the **minPBrass(0.95)** cut requires the expected SD trigger probability for proton and iron induced showers to be at least 95%.

There are also "ambiguous" events, in which the station with the hottest signal is not near the core. This can be caused by coincidental muons or showers. These events are identified by the **ambiguousHybridRejection** cut.

#### 4.1.4 Reconstruction & Field-Of-View

This group of selection criteria rejects events with poor quality profiles and biased  $X_{\max}$  due to the limited field-of-view of the telescope.

The first step of this selection is applying the lower threshold of this analysis,  $E > 10^{17.8}$  eV, alike [90]. At this level it is a very stringent cut (22% efficiency), however most of the events discarded here do not pass the following cuts.

Secondly, it is required that the fitted  $X_{\max}$  is within the observed range (**xMaxInFOV** cut). This is fundamental for a good profile measurement: if only the rising or falling edge of the profile is detected, the energy and shower maximum cannot be measured reliably.

Not only is it required that  $X_{\max}$  is observed, but also that its resolution is good. Showers with a poorly measured maximum are rejected based on the expected precision of  $X_{\max}$ ,  $\hat{\sigma}$ . For each measured profile, several ones are simulated with different  $X_{\max}$ , but the same geometry and energy. The expected signal in the camera is determined by propagating the energy deposit through the atmosphere, using our knowledge of the light yield and transmission through the atmosphere. From that, the uncertainty of  $X_{\max}$  in each of these virtual profiles is calculated. For a given shower, this only depends on the proximity of the maximum to the field-of-view limits - showers with  $X_{\max}$  close to either the lower or upper limit have an expected poorer resolution, since a smaller region around it is observed. Only showers with  $\hat{\sigma} < 40 \text{ g cm}^{-2}$  are accepted.

Furthermore, showers with small viewing angles,  $\alpha_{\min}$ <sup>1</sup>, are disregarded in this analysis. The reason for this is twofold. First, the geometry uncertainty becomes larger as  $\alpha_{\min}$  decreases below 20° [113], since showers are pointing towards the eye and thus are more compressed in time. Second, these events also have a larger

---

<sup>1</sup>the angle between the shower axis and each pixel FOV direction

Cherenkov fraction, and are therefore more susceptible to systematic uncertainties in the reconstructed profile. So, showers with an expected minimum viewing angle below  $20^\circ$  are also rejected by the **xMaxObsInExpectedFOV** cut.

Requiring the  $X_{\max}$  to be in the field-of-view of the detector leads to a biased event selection, since showers with small  $X_{\max}$  may have their maximum below the field-of-view and showers with large  $X_{\max}$  will often reach their maximum below it. The resultant distribution of observed shower maxima will thus be lacking events in both the lower and upper tails when compared to the true  $X_{\max}$  distribution. The standard Auger FD telescopes (excluding HEAT) operate in elevation angles between  $\omega_{\text{low}} = 1.5^\circ$  and  $\omega_{\text{high}} = 30^\circ$ . This defines the measurable height interval for vertical showers:  $R \tan[\omega_{\text{low}}] < h < R \tan[\omega_{\text{high}}]$ . So, as  $R$  increases both the upper and lower geometric F.O.V. decrease, and a different region of the  $X_{\max}$  distribution is being sampled. It gets even more difficult when the incident angle  $\theta$  is considered, as  $X \approx h \cdot \cos \theta$ , and therefore deepens the upper field-of-view and increases its difference to the lower one. A schematic picture of this bias is shown in figure 4.1.

The magnitude of the field of view bias depends on the geometry of the event sample, which is detector dependent, and on the distribution of the depth of shower maximum of incoming showers at each energy interval, which is not known a priori. One way to deal with this problem could be to study the effect in simulations and correct a posteriori in data, but since the shift is very different between primaries (up to  $12 \text{ g cm}^{-2}$  for proton and  $1 \text{ g cm}^{-2}$  for iron, from [113]) and the composition is not known, this could yield a large uncertainty. The solution is to derive from data a fiducial volume where deep and shallow showers can be detected with equal probability.

For this, the drift of the mean value of the  $X_{\max}$  distribution as a function of the upper and lower field of view limits is studied. The results for a lower energy bin are shown in figure 4.2. When the field of view is wide enough, there is no bias - for showers with large lower FOV and high lower FOV values the measured  $X_{\max}$  drifts away from the value of the true one.

The  $X_{\max}$  distribution can be approximated by the convolution of a Gaussian and an exponential functions,  $G * E$ . The first describes shower development while the second corresponds to the first interaction point. The value of  $\langle X_{\max} \rangle$  is approximately

given by the truncated mean function:

$$\langle X_{\max}^{\text{trunc}} \rangle = \mu_{\text{trunk}}(x_1, x_2) = \frac{\int_{x_1}^{x_2} x \cdot G * E(x) dx}{\int_{x_1}^{x_2} G * E(x) dx} \quad (4.1)$$

where  $x_1$  is the lower FOV limit and  $x_2$  the upper one. When  $x_1 = 0$  and  $x_2 = \infty$ , we retrieve the function for the unbiased mean. As they deviate from this ideal detector value, the bias in the measured maximum increases. The **FidFOVI-CRC13prel** cut selects events with field-of-view limits for which  $X_{\max}$  deviates less than  $5 \text{ g cm}^{-2}$  from its asymptotic value, i.e.,

$$|\mu_{\text{trunk}}(x_1, x_2) - \mu_{\text{trunk}}(0, \infty)| \leq 5 \text{ g cm}^{-2} \quad (4.2)$$

So, events with  $X_{low} < X_{low}^{fid}$  and  $X_{up} > X_{up}^{fid}$  are accepted, which ensures an unbiased distribution of shower maximum.

Finally, three further requirements on the measured profiles are applied. The reconstruction algorithm can produce "holes" in the profile for sections which cross telescopes, particularly close showers, since the light in each camera is only integrated for angles more than  $\zeta_{opt}$  away from the camera. So, events with profile gaps larger than 20% of the track length are rejected as well. Also, although there are cuts specific to ensure a low percentage of cloud coverage and small aerosol content, not all events measured in poor atmospheric conditions are rejected. A usual signal of cloud presence is profiles which present some distortions, which can be identified by the goodness of the Gaisser-Hillas fit. The standard-normal transformation of the  $\chi^2$ ,  $z = (\chi^2 - ndf)/\sqrt{2ndf}$ , is applied and showers in the non-Gaussian tail, above  $2.2\sigma$ , are rejected. Finally, a long profile is required - showers with track lengths below  $300 \text{ g cm}^{-2}$  are rejected.

#### 4.1.5 Profile Quality

While analyzing stereo showers for cases in which there were large differences between profiles measured at different sites we found a particularly strange event, depicted in figure ???. It is a stereo event observed in Coihueco and Loma Amarilla, but there is a large disagreement in measured energy between both stations. Also, the shape of both profiles is clearly incompatible.

Looking at the Coihueco reconstruction, nothing strange can be seen, however the one in Loma Amarilla seems to have some problems. A detailed look at the

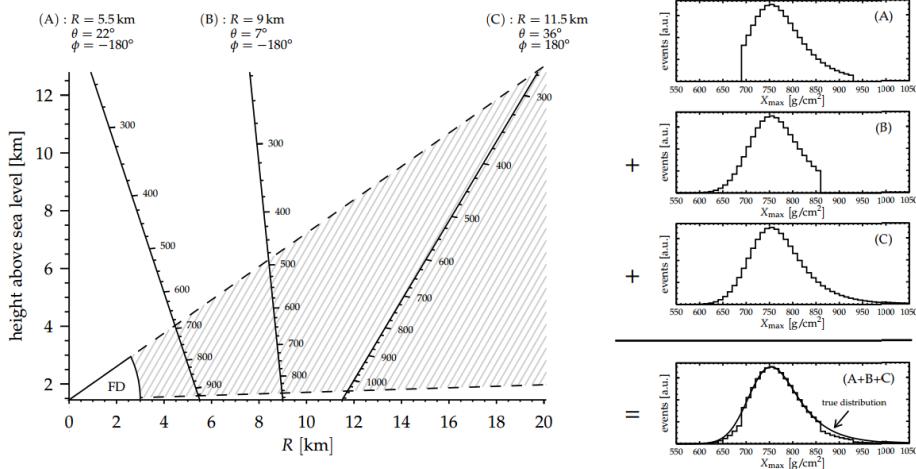


Figure 4.1: From [90]. Illustration of the influence of the FD field of view on the sampling of the  $X_{\max}$  distribution. On the left three different examples of geometries are shown. The field of view is represented by the filled dashed area. The correspondent  $X_{\max}$  distributions for each of the geometries are shown in the right, together with their sum.

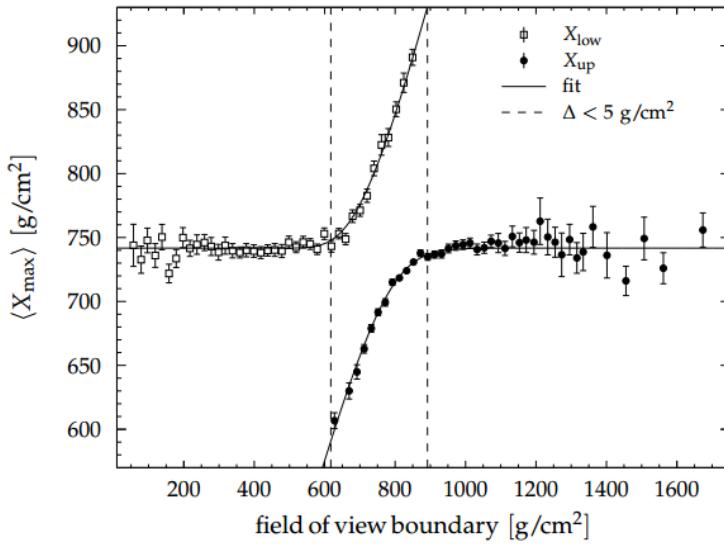


Figure 4.2: From [90].  $\langle X_{\max} \rangle$  as a function of the lower and upper limits of the field-of-view, in the energy interval  $10^{18.1}$  to  $10^{18.2}$  eV. The solid line shows a fit with the truncated mean of an exponential function folded with a Gaussian, and the dashed line indicates the field-of-view value at which this function deviates  $5 \text{ g cm}^{-2}$  from its asymptotic value.

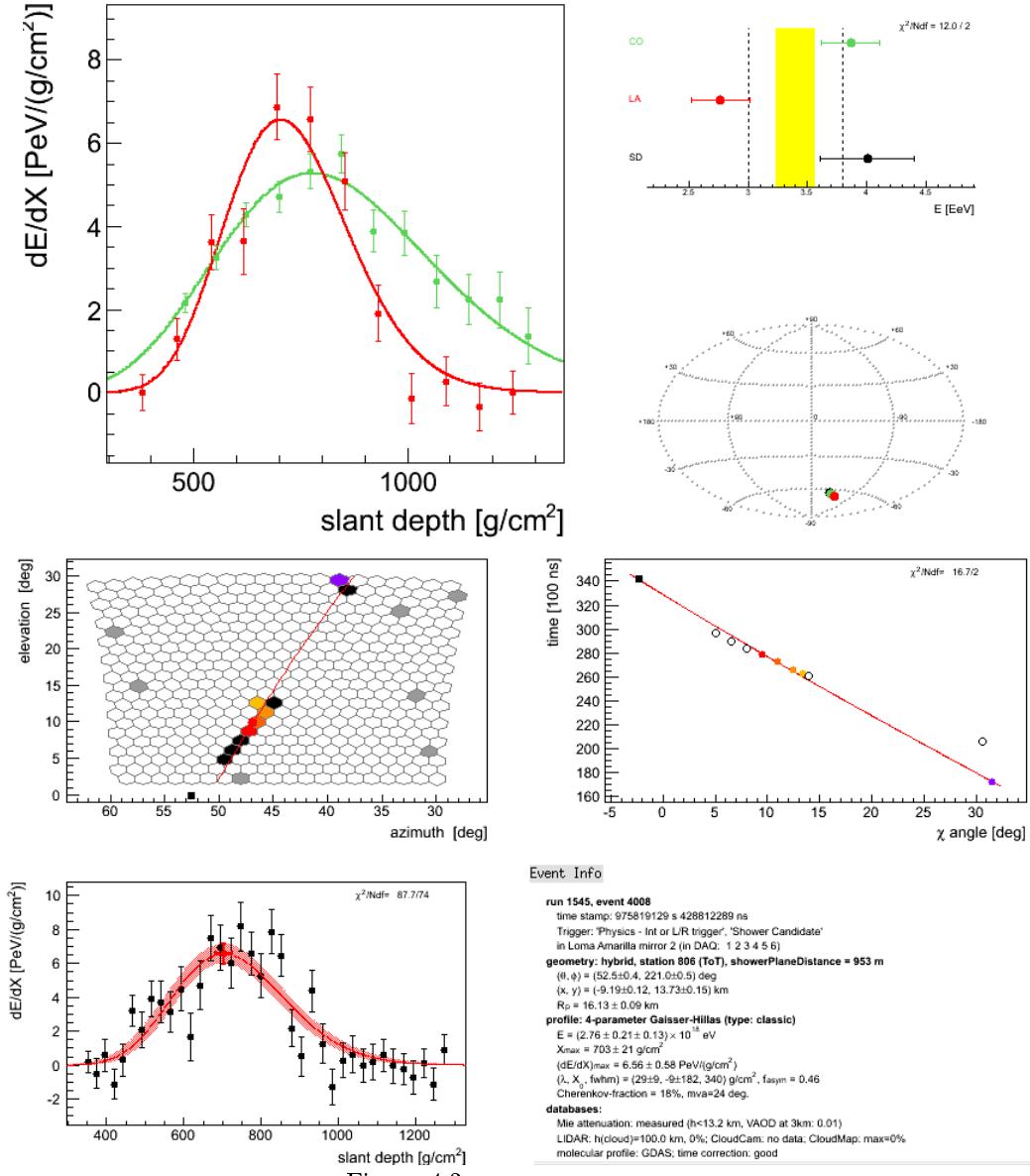


Figure 4.3: SD event 10732320

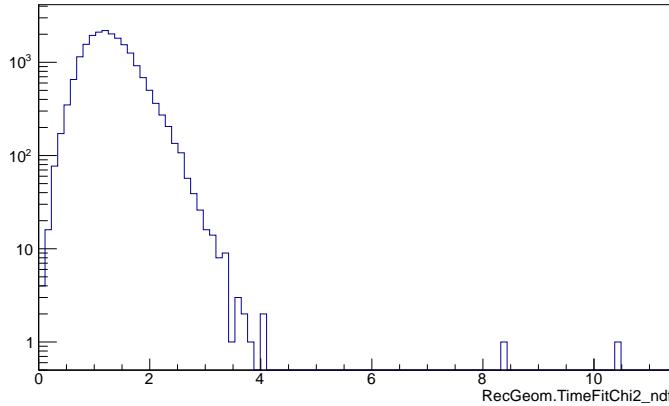


Figure 4.4: Time fit reduced  $\chi^2$  distribution for all data events analyzed

reconstruction steps for this eye is seen in figure 4.3. On the top left side the triggered pixels are shown colored, with the ones in black being the ones discarded by incompatibility with the time fit. The red line is the reconstructed axis, which clearly does not follow the line of triggered pixels. The reason for this can be seen on the  $\chi_i(t)$  fit on top-right, as the fit discards five pixels to be able to accommodate the pixel on the top of the camera, which by eye looks like the outlier. One possible alteration to the standard geometry reconstruction algorithm that correctly selects pixels for these two events is described in the appendix B.

On the bottom of figure 4.3 the profile and its fit results can be seen. Although it looks fairly continuous, it is much "thinner" than normal profiles, and in particular has half the width of the one measured for the same event in Coihueco. On a previous work presented at the last ICRC [100] we cut this event based on the percentage of pixels that had to be removed from the time fit. This is one of only two events for which this percentage is above 50%. In this work we decided to use the time fit reduced  $\chi^2$ , as it is a more physical variable and it has a more peaked distribution with clear outliers. Also, it rejects the same type of events.

The  $\chi^2$  distribution is presented in figure 4.4. It follows a perfect log-normal distribution except for two clear outliers: the first at around 8 is the one just shown, the other is shown in the appendix (figure ??) and not only has a reduced  $\chi^2$  larger than 10, but also had to have 20 out of its 24 pixels discarded, and has a spiky longitudinal profile. Therefore, these two events with  $\chi^2 > 5$  were discarded in this analysis (**timeFitChi2(5)** cut).

Finally, while comparing the profiles in different telescopes it was found that the bay 6 of Coihueco yielded more asymmetric profiles than all other telescopes. Previous studies [112] have found that, for showers that cross bay 5 and 6 of this site, there is a residual in the time fit for pixels in telescope 6 larger than present in any other crossing. This can be mitigated by changing the alignment, particularly in elevation, but even then there is always an unexplained shift in the  $\chi_i(t)$  curve in the telescope border. For this reason, and since each border telescope (telescopes 1 or 6) only accounts for about 3% of the total number of events, we decided to not analyze showers measured by the telescope 6 of the Coihueco site (**noTel6Coihueco** cut).

## 4.2 Average Longitudinal Profiles

With a dataset of high quality hybrid events we can finally reconstruct the average longitudinal shape profiles for different energy bins. The process of averaging the profiles is very straight forward but it will be detailed here for clarity.

First, the events were divided in six energy bins in  $\log_{10}[E/eV]$ : 17.8-18, 18-18.2, 18.2-18.5, 18.5-18.8, 18.8-19.2 and above 19.2. The rationale in choosing the binning was that for all energies the statistical error in L and R has to be below the proton-iron separation. For the first energy bin the proton and iron values of  $L$  almost overlap so it was not possible regardless of the binning, but otherwise this is accomplished.

For each event in a given energy bin we calculate the vector of translated depths  $X' \equiv X - X_{max}$ , normalized energy deposits  $(dE/dX)' \equiv (dE/dX)/(dE/dX)_{max} \equiv N'$  and normalized energy deposit error.

The energy deposit error is a fundamental ingredient to this method, and its calculation is not straightforward, since there are at least two sensible ways of doing it. There are two Offline modules used in Auger for profile reconstruction: Fd-ProfileReconstructorKG (FdPR) and FdEnergyDepositFinderKG (FdEDF). While they follow mostly the same procedure, they differ in some details, one of them the calculation of the energy deposit error.

In FdPR, the error is proportional to the square root of the number of detected (bin-by-bin) photons. This means that for the same average value, bins with upward fluctuations have larger errors than bins with downward fluctuations. In FdEDF, after the profile fit is done and  $dE/dX$  (as well as each individual light contribution) is

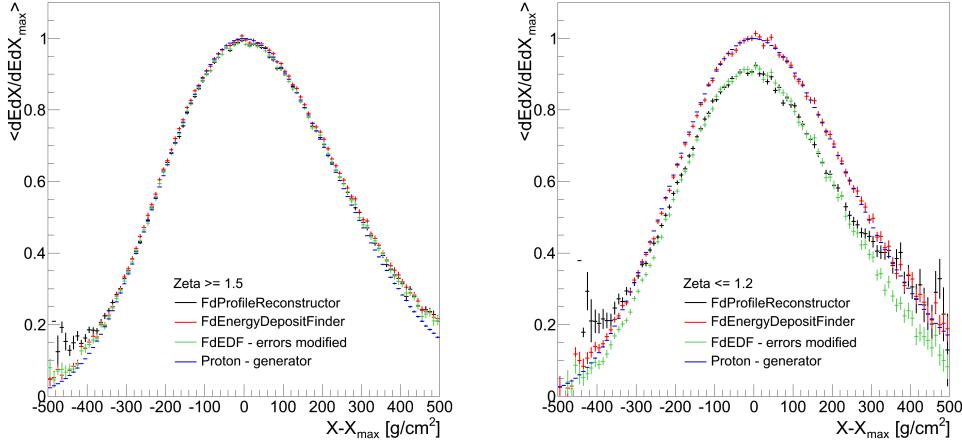


Figure 4.5: Average profile for the 4 different samples described (FdEDF, FdEDF modified, FdPR and proton (MC)), for high zeta (left) and low zeta (right)

calculated, the energy deposit error is recalculated, using this time not the measured number of photons, but the sum of all the **fitted** light contributions. Therefore, the error is a continuous function, and depends (for a given shower) only on the depth, and not on  $dEdX$ .

Since we are averaging each energy deposit weighed by the inverse of the error to build our profiles, the first method leads to a downward bias in our average  $dEdX$ , while the second will not. Also, clearly the importance of this effect decreases as the average number of photons in the shower increases, as the importance of statistical fluctuations dwindles. This can be seen in figure 4.5 which shows average profiles for high and low  $\zeta$  (light integration angle size) showers, with  $\zeta$  being very correlated with the number of measured photons. For high  $\zeta$ , all methods give a correct average profile, centered at  $X' = 0$  with a maximum at  $N' = 1$ . For low  $\zeta$ , however, different methods give vastly different average profiles. Not only the normalization changes between methods, but also the shape itself, since the number of photons is not the same for all depths.

To be sure that the cause of the difference in average profiles between the modules used comes from this, we reconstructed the data with the FdEDF module but without the recalculation of the error post-fit, and the results can be seen in figure 4.5, perfectly compatible with the FdPR module values. As a sidenote, it should be said that the error calculation discussed here is not used in any step of the reconstruction, since the fitting method is a maximumLikelihood on the number of photons, and not a chi-square. Thus, the post-fit recalculation of the energy de-

posit error introduces no effect on any reconstructed shower variable. Therefore, we choose to use the error based on the sum of the fitted light contributions, as it gives a stable shape with respect to the number of measured photons.

Finally, having the depth, energy deposit and its error, a profile is filled in  $10 \text{ g/cm}^2$  bins in  $X'$ . For each depth the corresponding  $(dE/dX)'$  is accumulated with a weight corresponding to the inverse of its squared error. The value of  $10 \text{ g/cm}^2$  was chosen mostly for visual clarity: it is large enough for the statistical errors to nearly invisible anywhere but the tails, but small enough so the relevant features can be seen. The effect of this choice on the fit results was studied: no impact in any of the variables was measured for binning below  $50 \text{ g/cm}^2$ .

The average longitudinal profiles used in this work are shown in figure 4.6. They are continuous in a large region around the center ( $N' > 0.2$  approximately) at all energies. At the tails the statistical uncertainty becomes large as there are only a few events with long shower tracks. Since higher energy showers leave a stronger signal on the PMTs, the high energy bins have in average longer showers, which is why they have better defined tails even as the number of events contributing is a order of magnitude below when compared to the lower energy bins.

In the next section the profiles will be discussed in more detail as they are compared to a fitting function and Monte Carlo simulations.

## 4.3 Average Shape Analysis

The final objective of this work is to extract the two shape parameters from the average profiles presented in the last section. In this section the fitting method will be detailed.

### 4.3.1 Fitting method

The parameters are going to be estimated using a standard integral  $\chi^2$  minimization. The function used to fit the profiles is a Gaisser-Hillas written as a function of R and L, i.e.:

$$\left( \frac{dE}{dX} \right)' = \left( \frac{dE}{dX} \right)'_{max} \left( 1 + R \frac{X' - X'_{max}}{L} \right)^{R^{-2}} \exp \left( -\frac{X' - X'_{max}}{RL} \right) \quad (4.3)$$

This is the most general case of that function, which allows for a normalization different from one ( $(\frac{dE}{dX})'_{max}$ ) and for it to be centered in a value other than  $0 \text{ g cm}^{-2}$  ( $X'_{max}$ ).

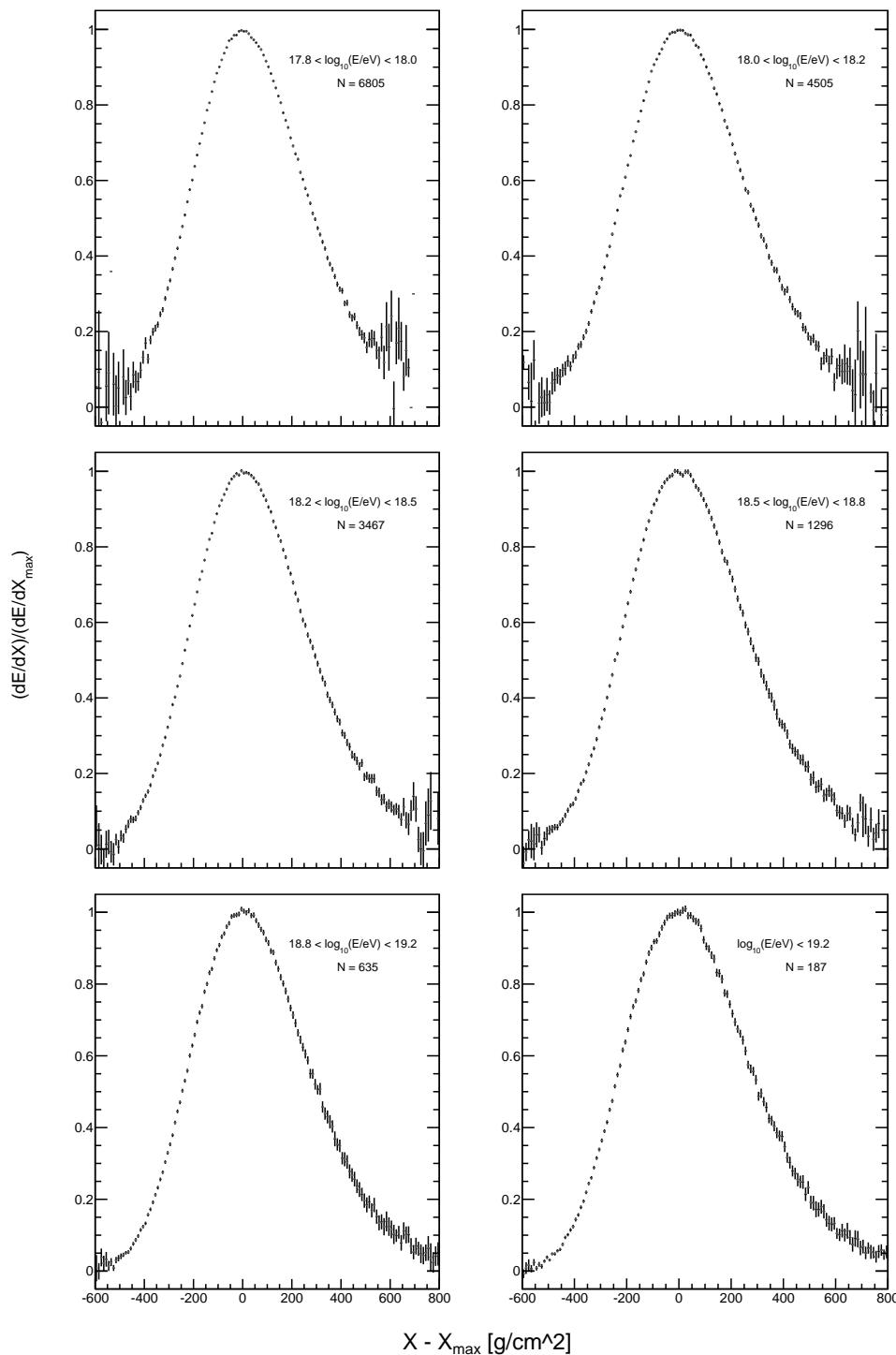


Figure 4.6: Average profiles for all the energy bins used in this work

### 4.3.1.1 Fitting Range

In this work, it was decided not to fit the whole average profile, but to define a lower and upper limit in the depth,  $X'$ , which is used for the fit. This definition is not straightforward, as it is the result of a balance between the need to enlarge the fitting region to lower the statistical uncertainty on the variables, and that of excluding depths closer to the profiles tails which are more sensitive to experimental uncertainties.

The region around  $X' = -300 \text{ g cm}^{-2}$  was found to be the most sensitive to mass composition, as it is where the proton-iron difference is larger. In [80], where the values of the shape parameters in simulations of proton and iron primaries was studied, the range chosen was from  $-350$  to  $+100 \text{ g cm}^{-2}$ , as the shower before the maximum contains more composition dependent information, and not much is gained by going above  $+100 \text{ g cm}^{-2}$ . However, when we consider reconstructed events, the  $dEdX_{rec}/dEdX_{MC}$  strays from 1 for depth below  $\approx -300 \text{ g/cm}^2$  (figure 4.7). More specifically, the average reconstructed energy deposit is increasingly larger than the simulated one. This is expected as there are two effects causing it. First, for the same simulated event, in a time bin where the light reaching the pixel is near the measurement threshold, if there is an upward fluctuation it will be considered as an event pixel, while if there is a downward fluctuation it will not. So, upward fluctuations are preferred in time bins with low signal. Second, profiles with larger widths are more likely to be measured further from the maximum, as they have a larger energy deposit for any given depth<sup>2</sup>. So, while around the maximum there is a fair average of all type of events (the event selection used in this work virtually guarantees all profiles are measured in a  $\pm 100 \text{ g cm}^{-2}$  window), this bias becomes more important as we get further away from it.

In this work, the lower fitting range was defined to be  $-300 \text{ g cm}^{-2}$ , which in full detector simulations yields a good compromise between measuring the sensitive part to proton-iron separation while avoiding reconstruction and/or acceptance problems. In particular, we guarantee the bias in measured energy deposit is negligible in both proton and iron simulations, while below  $-300 \text{ g cm}^{-2}$  it would not be.

The upper range used in this work is harder to define. For a given energy, all

---

<sup>2</sup>In practice this is even more complex, as the F.O.V. of the telescope also plays a part. The ground in particular is usually just a few hundred  $\text{g cm}^{-2}$  after  $X_{\max}$ , which leads to a earlier cutoff in the profile of wide profiles with deep  $X_{\max}$ . This does not seem to be important, even in protons, before 400-500  $\text{g cm}^{-2}$  (figure 4.12), so it will not be further analyzed in this work.

showers are very similar above  $X_{\max}$  (figure 4.12), so not much composition information is added by extending the range upward. This is the main reason why the upper range was defined as  $+100 \text{ g cm}^{-2}$  in [80]. The region above  $+300 \text{ g/cm}^2$ , in particular, should be avoided, as the profile is increasingly distorted by the contribution to  $dEdX$  of the muon decay. The upper region is also the most subject to experimental uncertainties in the energy deposit measurement. The percentage of collected light that is due to fluorescence emission falls very rapidly above  $X_{\max}$  (figure 4.8). So, the determination of the fluorescence profile is more dependent on our description of other light components, particularly Mie and Rayleigh scattering. This requires a very precise description of the atmospheric profiles, mainly its aerosol content (average and fluctuations), which as we will see in more detail when discussing systematics is not trivial. So, our objective is to let the profile rise dominate our measurement, as the region after  $X_{\max}$  is less interesting in simulations and more prone to experimental systematic uncertainties. However, a significant range around  $X_{\max}$  needs to be used: the statistical uncertainty in fitted parameters, particularly the width  $L$ , falls when we increase the range upward - fitting up to  $+200 \text{ g cm}^{-2}$  leads to a reasonable statistical error in the parameters determination (half of the one we get using only  $+100 \text{ g cm}^{-2}$ ), and allows  $L$  to be below proton-iron separation in models for all energy bins.

So, the fitting range  $[-300, 200] \text{ g cm}^{-2}$  was chosen. It allows a good determination of the shape parameters while minimizing the expected systematics in the tails. Also, it requires no corrections for comparisons between energies, like using limits varying with energy would.

#### 4.3.1.2 Fit parameters: free or constrained

The method in this work consists of determining the average profile by translating each profile by its fitted  $X_{\max}$  and normalizing it by its fitted maximum. Then, it will be fitted with equation 4.3. It might seem evident that the resultant average profile from this method will be centered at  $0 \text{ g cm}^{-2}$  with a maximum of 1. If this was the case, the  $X'_{\max}$  and normalization  $((\frac{dE}{dX})'_{\max})$  variables could be constrained (or even fixed) at 0 and 1 correspondingly. However, it is not.

The reconstructed maximum is the result of a fit in which the shape is constrained. To accommodate this preferred shape the other two variables have to be altered from their true value. Also, in the real reconstruction the maximum is determined as a fit to a profile whose lower and upper limits change event-to-event. This

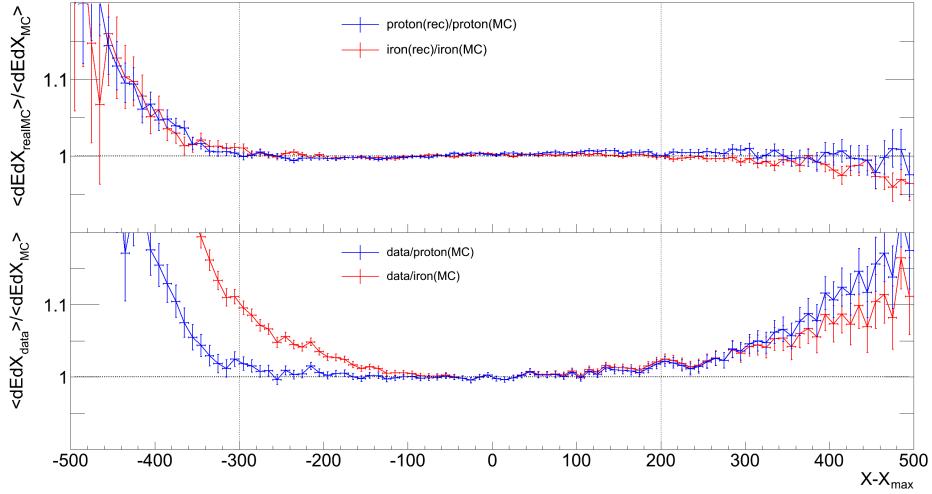


Figure 4.7: Average  $dEdX_{rec}$  over  $dEdX_{MC}$  as a function  $X'$  for the primaries described and energies  $18.25 < \log(E/eV) < 18.75$ . Top: realMC reconstructed showers over MC showers for proton and iron simulations. Bottom: data showers over MC showers for proton and iron simulations.

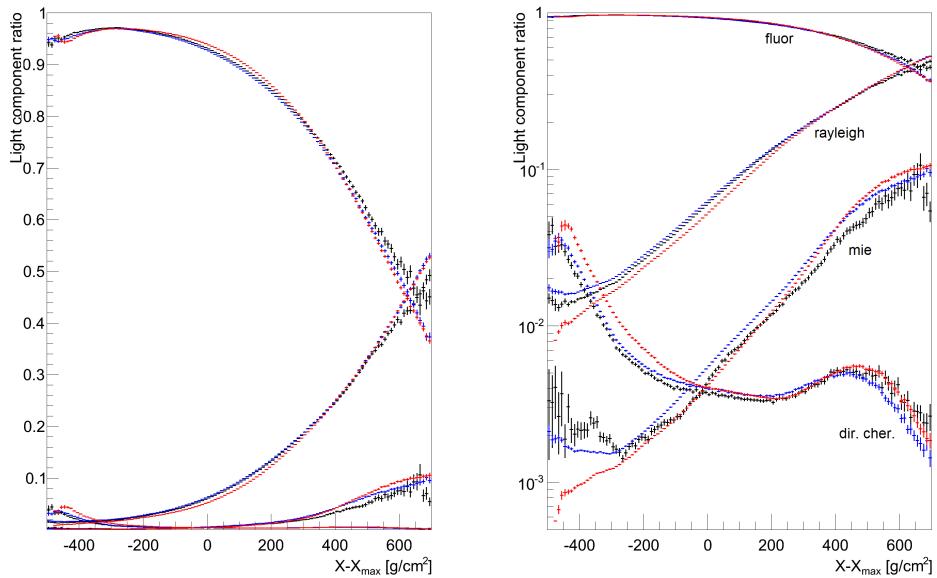


Figure 4.8: Fraction of photons, for showers between  $10^{18.25}$  and  $10^{18.75}$  eV, reconstructed to each of the indicated components: fluorescence or direct, Mie or Rayleigh Cherenkov. Data is in black, proton in blue and iron in red. Linear scale on the left and logarithmic on the right to better visualize some features.

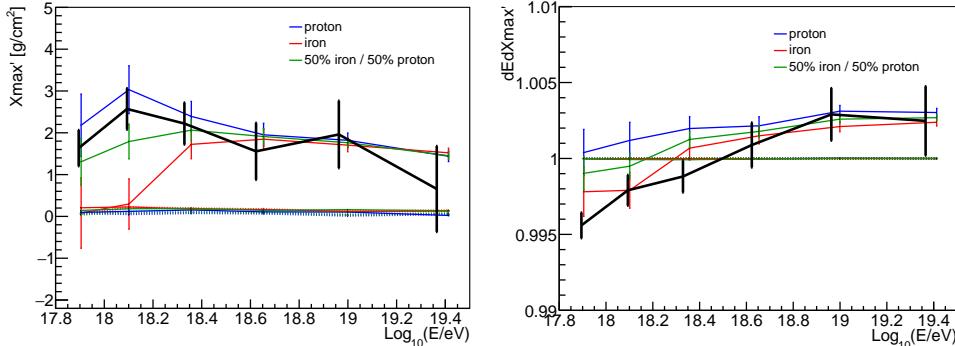


Figure 4.9: Value of the fit to the translated and normalized average profile:  $X_{\max}'$  (left) and  $N'_{\max}$  (right). Shown are the values for simulated (dashed colored lines), reconstructed (full colored lines) and data profiles (black).

yields, even in simulations, a different maximum than the one determined by the standard method in simulations (a parabolic fit to the three bins with the largest energy deposit). The bias introduced in  $X_{\max}$  value can be around  $3 \text{ g cm}^{-2}$  ([113]) in relation to the real maximum of the profile.

In the value of the maximum, not only are there these two effects, which tend to make the maximum value slightly above 1 (around 0.5%), but also a third one that brings  $N'_{\max}$  while not affecting  $X'_{\max}$ : the resolution in  $X_{\max}$ , which is around  $15 \text{ g cm}^{-2}$  (energy dependent). When two equal profiles (at simulation) are added, where in one the  $X_{\max}$  was overestimated and the other underestimated, the resultant profile is one that is wider and has a maximum below one. Since the resolution is a symmetric effect around  $0 \text{ g cm}^{-2}$ , this value is not altered. In simulations, it was found that  $N'_{\max}$  decreases around 0.5% for a resolution of  $15 \text{ g cm}^{-2}$ .

So, the four parameters ( $X'_{\max}$ ,  $N'_{\max}$ , R and L) were left free in the fit, which is the most general case. Since the allowed fluctuations in the first two are known, we will use them as a further control plot for our method. The results in simulations and data are shown in figure 4.9. Both variables are within the expected difference of  $X'_{\max} = 0 \text{ g cm}^{-2}$  and  $N'_{\max} = 1$  ( $3 \text{ g cm}^{-2}$  and 0.5% respectively), and in both cases the energy variation is compatible between data and simulations. Also, as expected,  $N'_{\max}$  increases with energy as the  $X_{\max}$  resolution decreases.

## 4.4 Analysis of simulated showers

The next step in the validation of this analysis is applying the steps just described to simulated showers processed through the full detector simulation for Auger, and

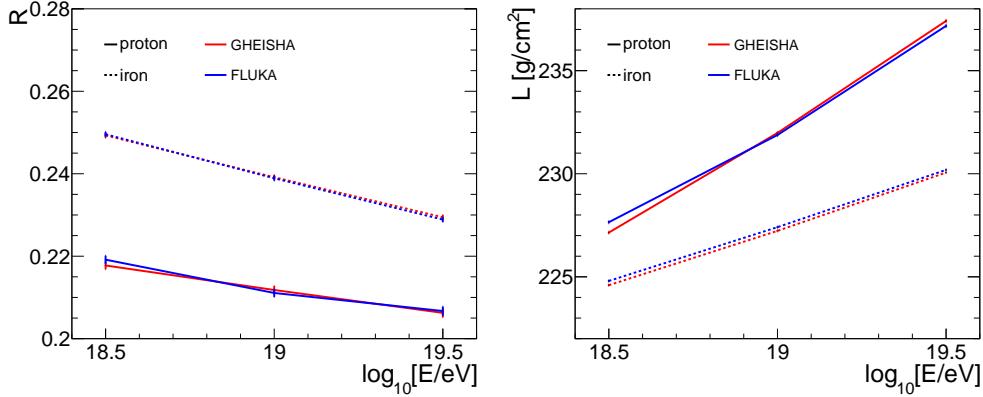


Figure 4.10: L (left) and R (right) as a function of energy for proton (full lines) and iron (dashed lines) primaries, simulated in CONEX using the FLUKA and GHEISHA low energy models.

comparing the results to those at generator level.

The longitudinal profile can be simulated by different codes for generating air showers, such as CORSIKA, SENECA, AIRES or Conex. They take an incoming particle (typically a photon, proton or nucleus) with a given energy, entry point and angle, and simulate its resulting shower as it traverses the earth's atmosphere. The first interactions happen at very high energies and are modeled by high-energy hadronic models (section ??). As the energy of particles decreases, the modeling of their interaction shifts to low-energy hadronic models (GHEISHA and FLUKA are the most commonly used ones). The uncertainty in the simulation of shower development in general, and in longitudinal profiles in particular, is dominated by the primary particle and high energy hadronic model used - the difference in shape between the low energy models is negligible (figure 4.10).

In this section showers simulated with Conex are used, employing QGSJetII-03 as the high energy hadronic model for proton and iron primaries. Showers are simulated in flat energy bins from  $10^{17.625}$  to  $10^{19.625}$  eV, with geometric properties (zenith angle and core position) taken from the ones observed in data. These simulated showers are then used as input for the Auger reconstruction framework, Offline. Fluorescence and Cherenkov photons at each point in the atmosphere are propagated and attenuated to the telescope camera. In this step, information about atmospheric conditions and calibrations, for example, is used from the data databases. So, the simulated shower can be reconstructed with a given date, where it will be subject to exactly the same reconstruction chain as data from that period. This full detector simulation is the so called *realMC*.

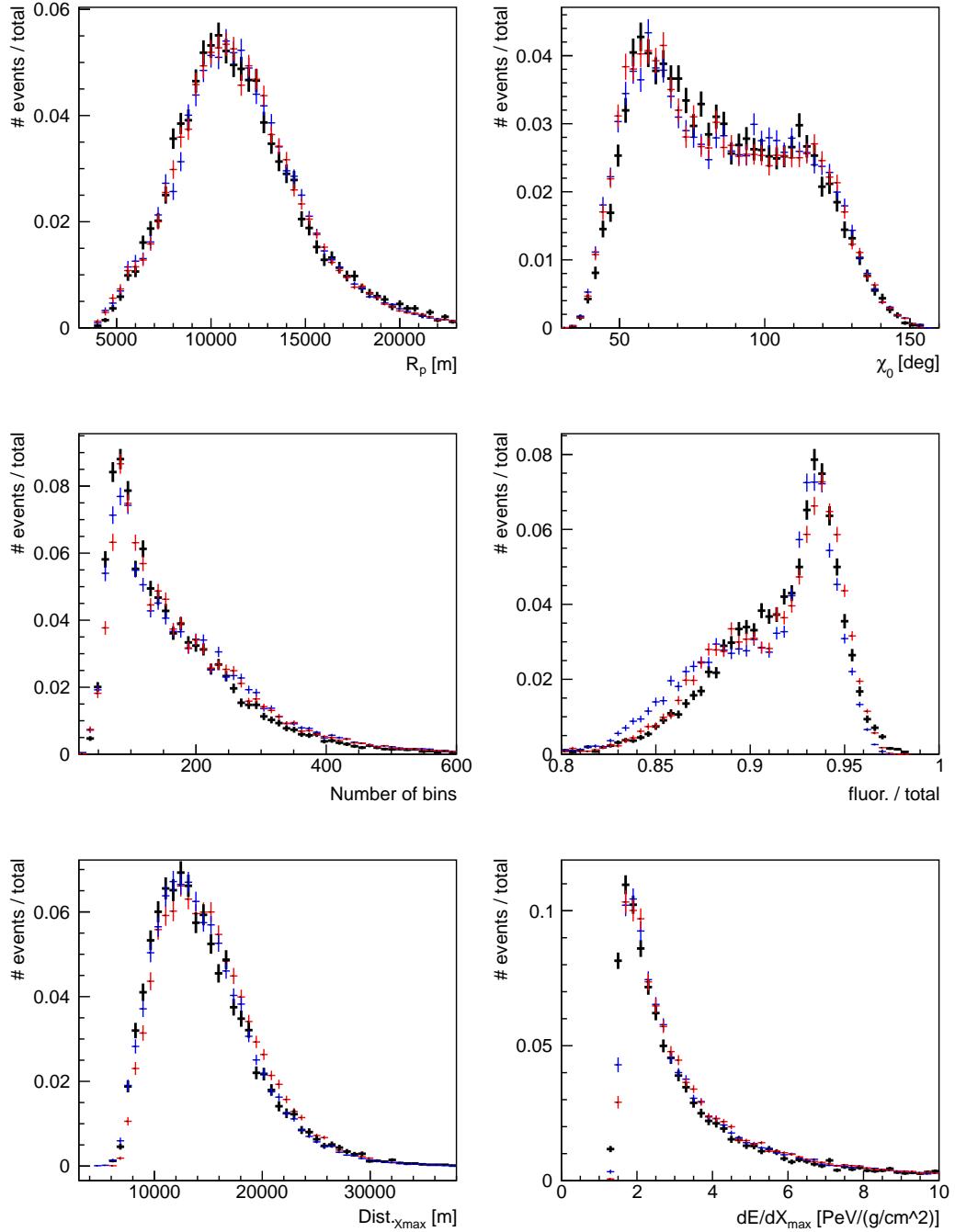


Figure 4.11: Control plots for the 6 variables indicated, from top left to bottom right:  $R_p$ ,  $\chi_0$ ,  $N_{bins}$ , %fluor.,  $d_{X_{\max}}$  and  $dE/dX_{\max}$ . Data in black, proton in blue and iron in red. The distributions are for all energies used in this work ( $E > 10^{17.8}$  eV).

The reconstruction used in this section was performed with `Offline` version v2r7p6 and the standard module sequence, apart from replacing the `FdProfileReconstructorKG` module with `FdEnergyDepositFinderKG`, as explained previously. These simulations go through the same cuts as data, detailed in the beginning of this chapter.

Before constructing at the longitudinal profiles we validate the simulations by looking at the distributions of some key physical parameters in the `realMC` to check whether they are compatible with data. They are shown in figure 4.11. We chose two geometric variables,  $R_p$  and  $\chi_0$ , as well as more convoluted ones like the number of time bins, fraction of fluorescence photons, distance of the FD detector to  $X_{max}$  and the fitted  $dE/dX_{max}$  per event. With these six distributions we can probe well the compatibility for the full reconstruction chain, as they go from basic geometry to final profile fit results. All six are fully compatible with data.

The average longitudinal profiles for generated and reconstructed proton and iron showers are shown in figure 4.12, for a low and high energy bins. The similarity between both is remarkable, as no difference can be seen by eye. However, if their ratio is computed we can see the reconstructed energy deposit is larger than the generated one at lower depths. This is consistent with the expected statistical bias: at tails the number of photons is small so only upwards fluctuations can rise above the background threshold, while downward fluctuations are neglected. For very high depths, the reconstructed energy deposit appears to be below simulation. This is mostly a selection bias, as for a typical  $10^{19.5}$  eV shower the ground is around  $200 \text{ g/cm}^2$  after  $X_{max}$  for a vertical shower and  $450 \text{ g/cm}^2$  for a  $40^\circ$  one. This means that above  $500 \text{ g/cm}^2$ , where we see a difference, we are increasingly choosing special shower shapes and very inclined geometries in reconstruction, while the generated showers are simulated without considering ground, so they contribute evenly. If the simulated profiles are filled only in the field of view depth this effect is largely mitigated.

The values of R and L resultant from the fit of a Gaisser-Hillas to these profiles is shown in figures 4.13a ( $R$ ) and 4.13b ( $L$ ). The reconstruction is quite accurate. At the highest energies the values of both variables agree perfectly with simulation, and only for the lowest energy a composition dependent bias on L ( $1$  to  $3 \text{ g cm}^{-2}$  in proton and iron respectively) can be seen. This appears to be related with the fact that the shape constraints used on the energy deposit fit are optimized for higher energies, so for lower energies the integral is miscalculated. Since the fitting

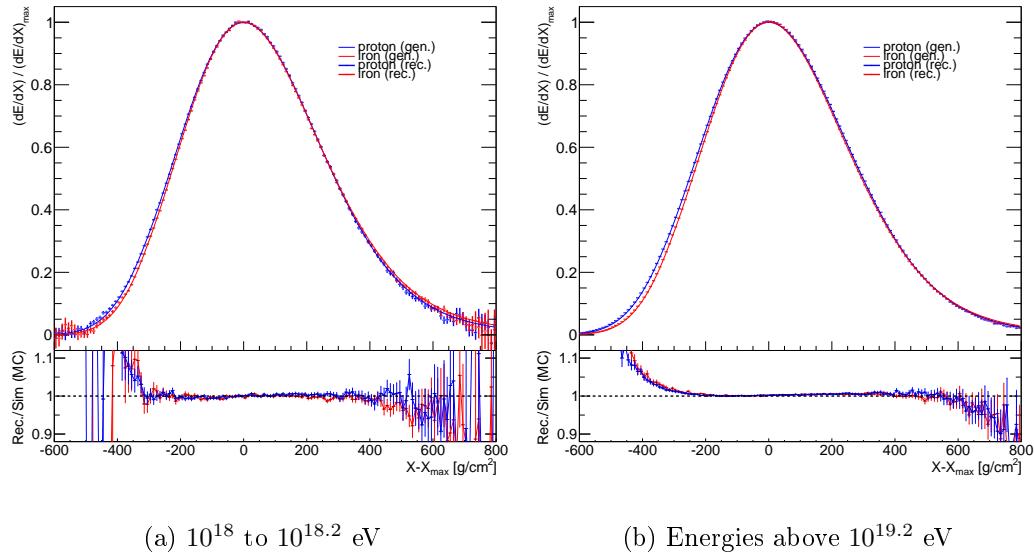


Figure 4.12: Average profiles for proton and iron generated and reconstructed profiles at a low energy (left) and high energy bin (right). Top:  $dE/dX$  as a function of  $X'$ . Proton is shown in blue and iron in red. Reconstructed showers are the thicker points, while generated ones are overlapped as a curve. Bottom: Ratio of reconstruction to simulated showers.

process is iterative, this affects the whole shape. For the asymmetry, R, proton is reconstructed almost without bias for all energies, but at lower energy iron and mixed compositions are below the simulated value. This, however, is a small effect and smaller than the statistical uncertainty on data. This bias will be corrected and the maximum difference in composition added as a systematic uncertainty, as explained in the next chapter.

## 4.5 Data: first look at results and MC comparison

To end this chapter, the average profiles of data are shown, as well as the results of the fit, together with some brief comments on their compatibility with simulations.

Figure 4.14 shows data profiles overlapped with the reconstructed proton and iron ones (also with individual components). The precision of data is high at all energies, as all profiles have a continuous shape with small uncertainties for all depths except the end of the tails. Also, at least visually it is compatible with simulations for hundreds of  $\text{g}/\text{cm}^2$  around the maximum. However, it can be seen that for high depths it starts increasing above MC predictions. This is an area in which the scattered Cherenkov light contribution, particularly Rayleigh, rises exponentially and can even dominate over fluorescence. This could indicate there is

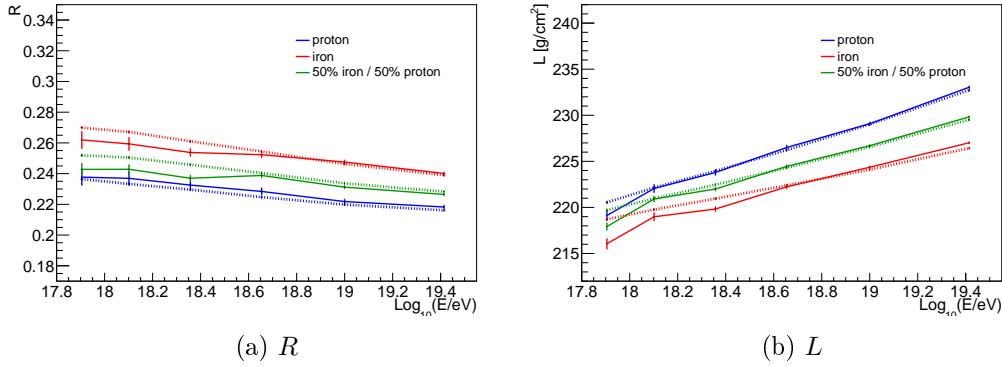


Figure 4.13: Fitted variables as a function of energy for reconstructed (full) and simulated (dashed) profiles

some miscalculation of light components during the fit, as more light is attributed to fluorescence in data then it should when the Rayleigh contribution is very high. This effect is more pronounced at the highest energy, where the data profile starts to visibly be above simulation at around 300 g/cm<sup>2</sup>.

#### 4.5.1 Fit to a Gaisser-Hillas: $\chi^2$ and residuals

Having validated the measured longitudinal profiles and the fitting method chosen, the data profiles are fitted. The results are shown in figure 4.15.

The figure shows that shower longitudinal profiles are described by a Gaisser-Hillas function. It had already been known that individual showers can be described by it, however in here the uncertainty in each point is very small, and still the compatibility is very good, with  $\chi^2$  never being above 2, and in most cases around or below 1. Looking in more detail at the residuals as a function of the depth, no region of the profile where the fit strays from data is seen - the agreement is good throughout the entire fitting range.

#### 4.5.2 Fit results: R and L vs energy

Finally, the main objective of this work can be presented - the shape parameters of the electromagnetic shower profiles.

The asymmetry parameter for data profiles, R, is shown in figure 4.16. The energy evolution observed is very surprising. While all models predict a decrease of R, data indicates an increase at the highest energies. Below  $10^{18.5}$  eV, data is compatible with proton predictions for the three models, but above this value it

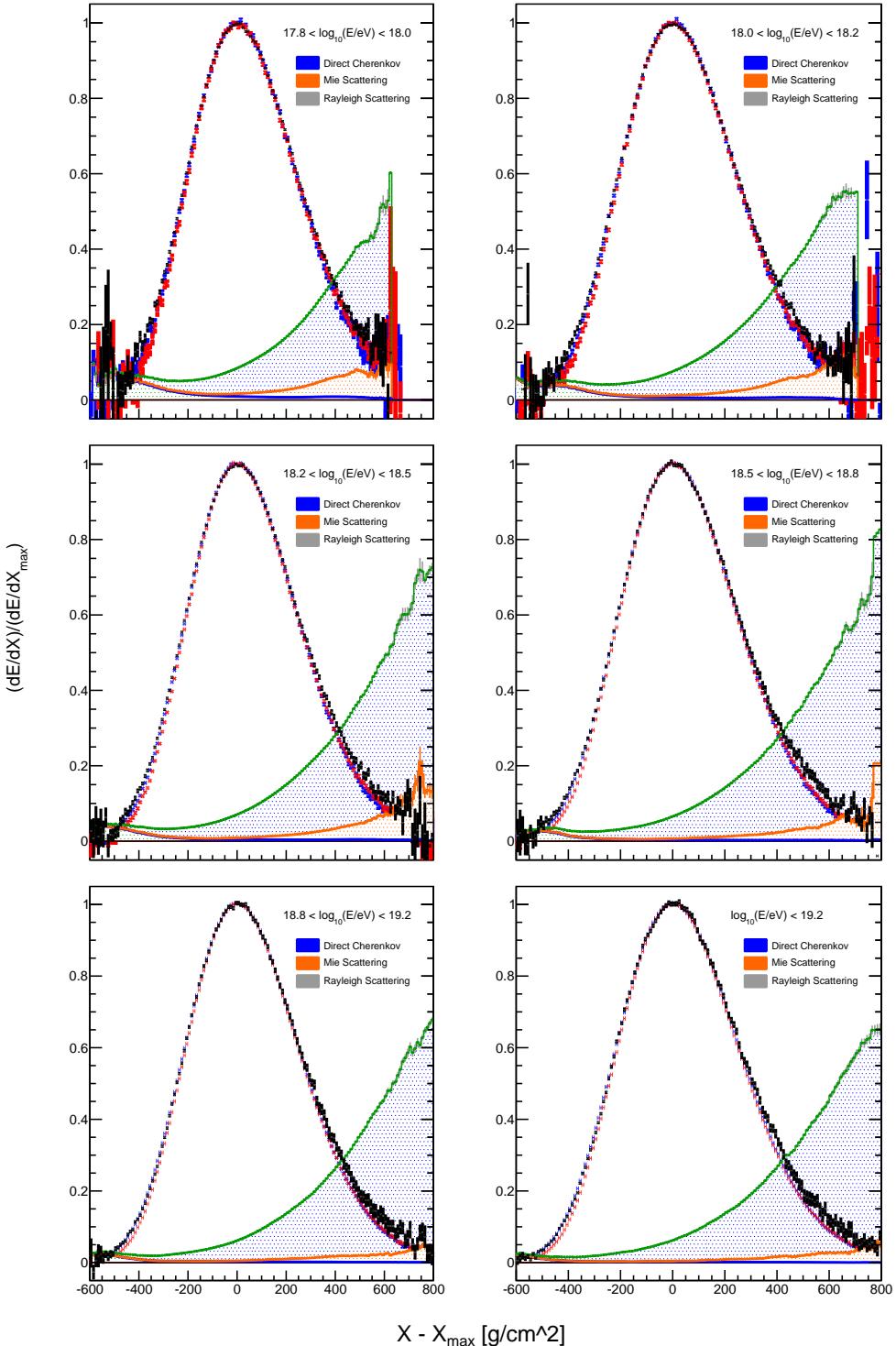


Figure 4.14: Average longitudinal profiles for data (black), reconstructed proton (blue) and iron (red). The average light component contribution for data on each depth bin is shown as a filled area.

increases to values compatible with iron, and above  $10^{19.2}$  eV the fitted asymmetry in data is larger than what is predicted by any of the standard high energy hadronic models. This is a very interesting feature, which we will pay special attention in the next section where we study the several effects that can influence the profile asymmetry. The width,  $L$ , is shown in figure 4.17. Here, the energy evolution is compatible with the one predicted by all models.  $L$  increases monotonically, being compatible with proton simulations at all energies except the lowest one.

The next chapter will be dedicated to studying which steps of the reconstruction most affect the profile and estimating the systematic uncertainty of our measurement.

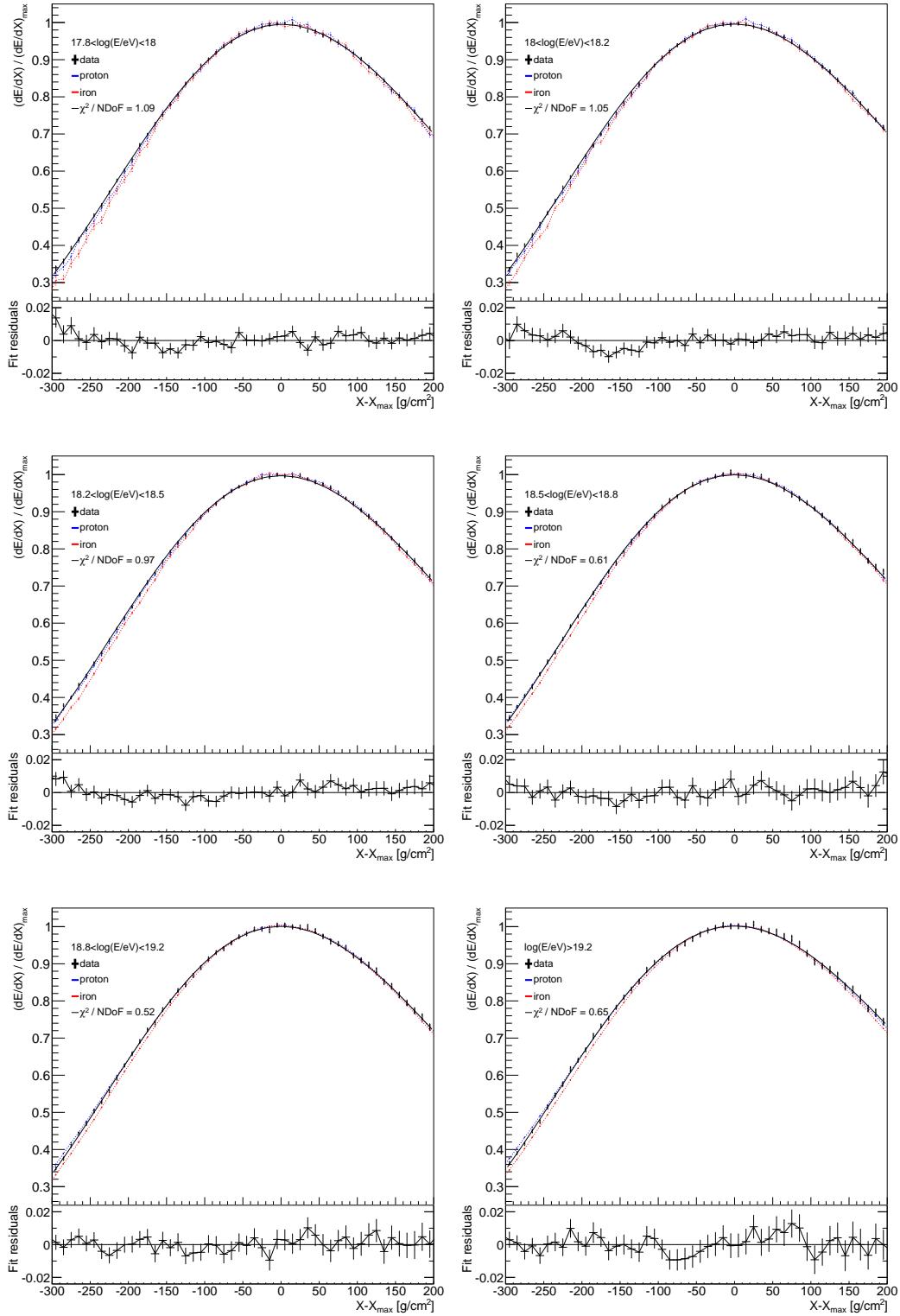
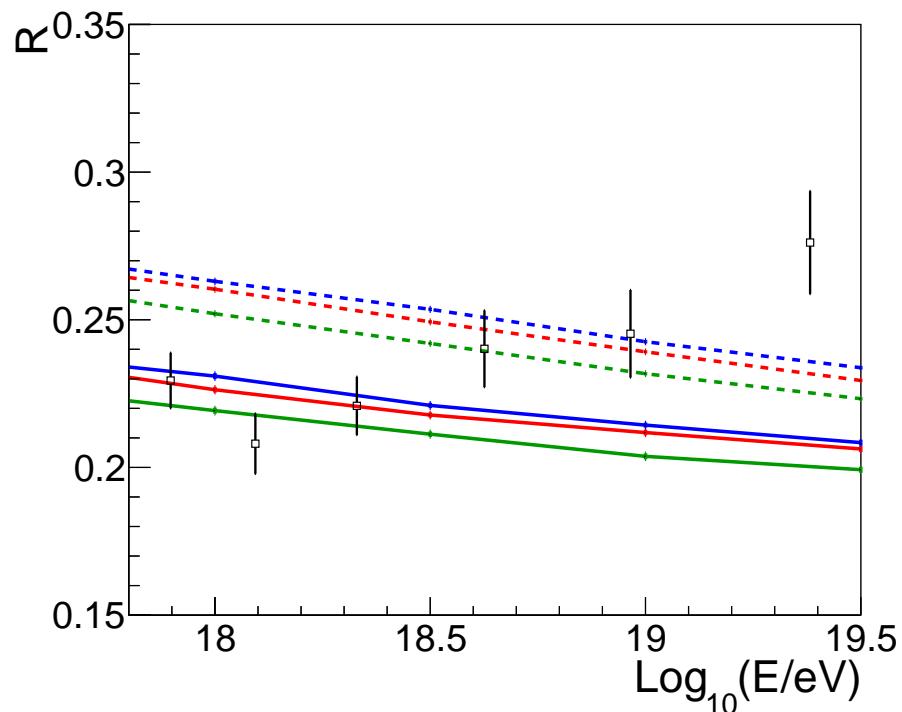
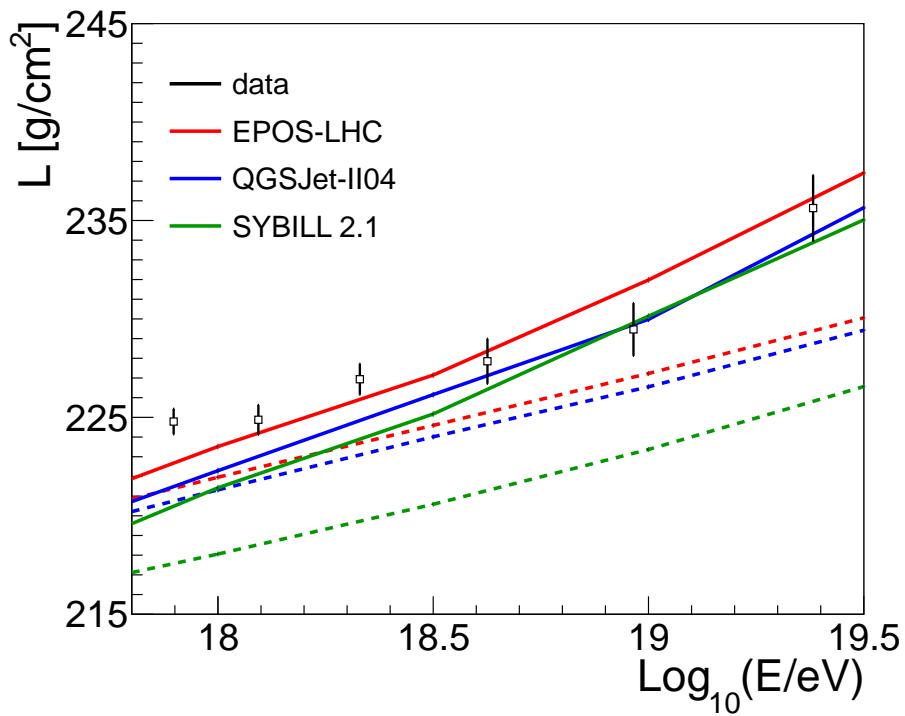


Figure 4.15: Average profiles for data (black points), proton (blue line) and iron (red line). Superimposed is the result of the fit of a Gaisser-Hillas function to data (black line). In the bottom of each plot the residuals of the fit to data are shown as a function of  $X'$ .

Figure 4.16:  $R$  as a function of energyFigure 4.17:  $L$  as a function of energy



## CHAPTER 5

# Systematic uncertainties

---

In this section the systematic uncertainties in the measurement of the  $R$  and  $L$  variables are estimated.

## 5.1 Reconstruction Bias

To estimate the effect of the full reconstruction algorithm on our variables, we can use *realMC*. As can be seen in figure 5.1, for all energy bins except the lowest one, the difference between simulated and reconstructed showers is very small. In fact, above  $10^{18.5}$  eV, it is compatible with 0 g cm $^{-2}$  for all primaries in L and below 0.003 in R. At the highest energy bins, in R, the pure profiles have no bias while mixed ones have a small negative one. The 50% composition mix had the largest bias out of all the combinations tested, so it was taken as an upper value for the uncertainty.

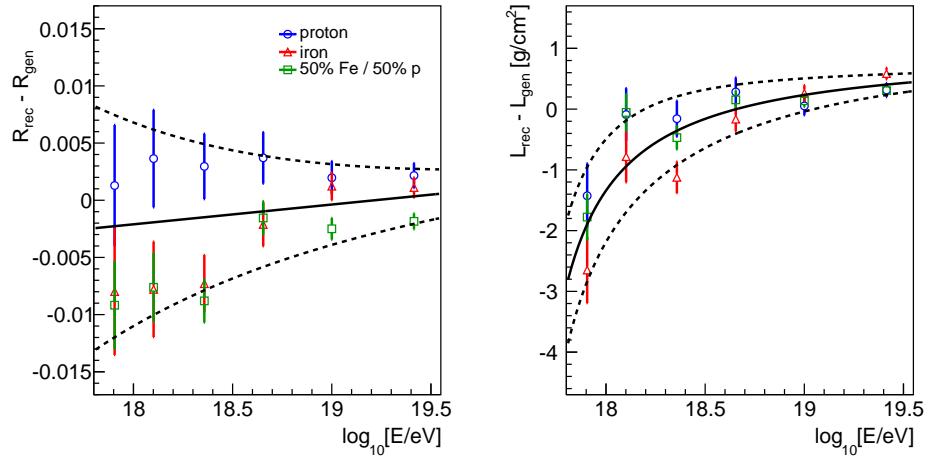


Figure 5.1: Difference in R (left) and L (right) between reconstructed and simulated profiles of proton and iron primaries, as well as a 50% mix of both. The full line is the bias correction , fitted as the average between the most extreme points at each energy, while the dashed lines represent the systematic uncertainty bounds

From figure 5.1 we fit the bias correction as half of the maximum and minimum

(solid line):

$$\begin{aligned}\text{Bias}_{\text{corr}}(R) &= (-2.10 + 1.77 \log_{10}[\text{E}/\text{EeV}]) \times 10^{-3} \\ \text{Bias}_{\text{corr}}(L) &= 1.02 + 1.18 / (\log_{10}[\text{E}/\text{eV}] - 17.5) \text{ g/cm}^2\end{aligned}\quad (5.1)$$

and the systematic uncertainty associated with it enfolds the largest differences (dashed line):

$$\begin{aligned}\Delta R_{\text{bias}} &= \exp[-4.72 - 0.93 \log_{10}(\text{E}/\text{EeV})] \\ \Delta L_{\text{bias}} &= \exp[-0.16 - 1.13 \log_{10}(\text{E}/\text{EeV})] \text{ g/cm}^2\end{aligned}\quad (5.2)$$

## 5.2 Energy Scale

The uncertainty of the energy scale in Auger is an energy independent  $\pm 14\%$ . This uncertainty is a constant shift for all events. Events at the energy bins borders can migrate as energy is shifted by this value. As the energy evolution is very slow for both variables we do not expect this to result in a significant change. In figure 5.2 it can be seen that the influence is indeed not large, and no consistent energy drift is observed. However, the statistical error bars on the difference are over-estimated, as the number of events overlapping in both samples is around 90%, so we decided to consider a systematic uncertainty that encompasses all points. Since no clear energy trend can be seen, a constant uncertainty band was fitted:

$$\begin{aligned}\Delta R_{\text{bias}} &= 0.01 \\ \Delta L_{\text{bias}} &= 1 \text{ g/cm}^2\end{aligned}\quad (5.3)$$

## 5.3 Detector

In this section we discuss how variables related to the physical properties of the detector influence the measured shape. This includes timing offsets, alignment problems, eye-to-eye differences and detector aging.

### 5.3.1 SD-FD Time Offset

The SDP reconstruction uses only information from the FD. The shower axis within the SDP however, is reconstructed using one SD station timing information to constrain the  $\chi(t)$  fit. This means that an offset in the FD and SD clocks can result in a systematic mis-reconstruction of the shower axis. From inclined laser studies and golden hybrid events analysis it is known that a timing offset, if it exists, is less

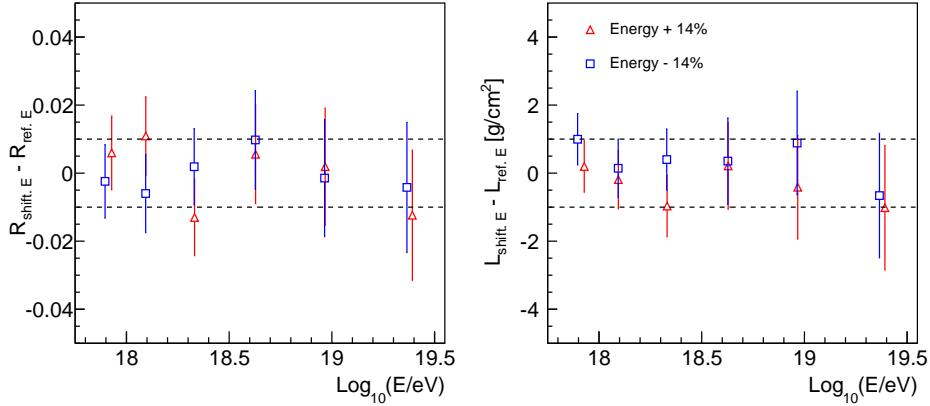


Figure 5.2: Difference in R (left) and L (right) between profiles measured with the reference energy and changing it by  $\pm 14\%$

than 50 ns. Therefore, data was reconstructed with the two extreme hypothesis, a SD-FD offset of -50 ns and +50 ns, and it was assessed whether the profile shape was significantly affected.

Looking at figure 5.3, it can be seen the difference between both reconstructions in R is negligible at all energies, and always smaller than 0.001, so that is taken as an upper limit for the systematic uncertainty

$$\Delta R_{\text{offset}} = 0.001 \quad (5.4)$$

The uncertainty in L is also small comparing to other systematics but non-negligible. It decreases with energy and fits well with an exponential:

$$\Delta L_{\text{offset}} = \exp[0.08 - 0.87 \log_{10}(E/\text{EeV})] \text{ g cm}^{-2} \quad (5.5)$$

### 5.3.2 Telescope Alignment

To accurately determine the geometry of the shower, the alignment of the FD telescopes must be well known. In particular, changes in elevation affect the shape (as well as  $X_{\max}$  and energy) more significantly than equal sized variations in azimuth. To determine the telescope alignment, four different methods have been used at Auger.

#### Timing

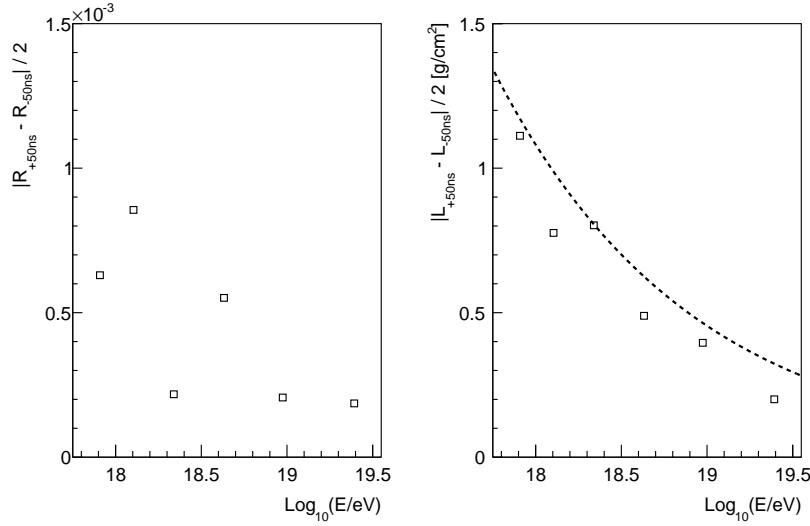


Figure 5.3: Half of the difference in R (left) and L (right) between events reconstructed with a SD-FD offset of +50 ns and -50 ns. Since the same events are used in both cases no statistical error was considered. The dashed line on the right is the fitted uncertainty on L.

This is the method that currently gives the standard alignment for the FD detector. Given a shower axis, there is a trivial relation between the time of the signal and the  $\chi$  angle corresponding to the pixel direction with respect to the shower. Conversely, we can attribute a corresponding  $\chi$  value to the SD core given its location and time,  $\chi_{core}$ . This method consists in finding the alignment that minimizes

$$\Delta\chi = \chi_{FD}(t_{core}) - \chi_{core} \quad (5.6)$$

where  $\chi_{FD}(t_{core})$  is the value of the fit to the FD  $\chi(t)$  curve at  $t_{core}$ .

### SDP method

For a perfectly aligned telescope, the average SDP should contain the true shower core. This method consists in minimizing the angular distance between the SDP and the core-FD vector,

$$\Delta\phi = \overrightarrow{n} \cdot \overrightarrow{core - fd} \quad (5.7)$$

This method has been applied in two different analyses with similar results. The first, *SDP*<sub>1</sub> [118], minimizes this distance for pairs of events with SDP orientations differing at least 24°, for which the orientation is very constrained. The other, *SDP*<sub>2</sub> [119], minimizes equation 5.7 for all events simultaneously.

The other difference lies in the correction for the bias in the core determination.

Since the reconstructed core is not the true one, and has a known dependence on the zenith angle, it must be corrected. This correction in the  $SDP_1$  method is MC based and the one in  $SDP_2$  method is data based.

### Star tracking

The FD PMTs pickup the light from the brightest stars in the sky. Since the stars position is well known, we can determine each pixel pointing direction by knowing the time at which the star enters and leaves the PMT field of view. More details can be found in [114] and [115].

### CLF/XLF laser shots

The CLF and XLF perform routinely during the night for atmospheric monitoring purposes, giving both sweeping and vertical pulses with an orientation accuracy of the order of  $0.01^\circ$ . For these shots, the reconstructed SDP can be compared with the true SDP given by the position and direction of the C/XLF pulse. The main drawback of this method is that only the telescopes facing the C/XLF (center of the array) can be aligned accurately. So for the border telescopes one of the other methods has to be used, which in turn can be cross-checked at the center telescopes by the C/XLF. More details on the methods can be found in [116, 117, 118].

## Results

The differences between the standard alignment (timing method) and the others is summarized in table 5.1. We can see that the methods agree fairly well, within approximately  $0.1^\circ$ .

	RMS in elevation [ $^\circ$ ]	RMS in azimuth [ $^\circ$ ]
Star Tracking	0.13	0.21
$SDP_{Adel.}$	0.09	0.11
$SDP_{Karls.}$	0.12	0.10

Table 5.1: Mean and  $\sigma$  of the distributions of the difference between the elevation and azimuth values of the reference method (timing) and the ones described.

To estimate our systematic due to alignment we reconstructed data with the three alternative alignments. The results are shown in figure 5.4. The functions that bracket the data points are:

$$\begin{aligned} \Delta R_{\text{const}}^- &= -0003 \\ \Delta R_{\text{const}}^+ &= 0.004 + \exp[-10.4 + 3.8 \log_{10}(E/\text{EeV})] \end{aligned} \tag{5.8}$$

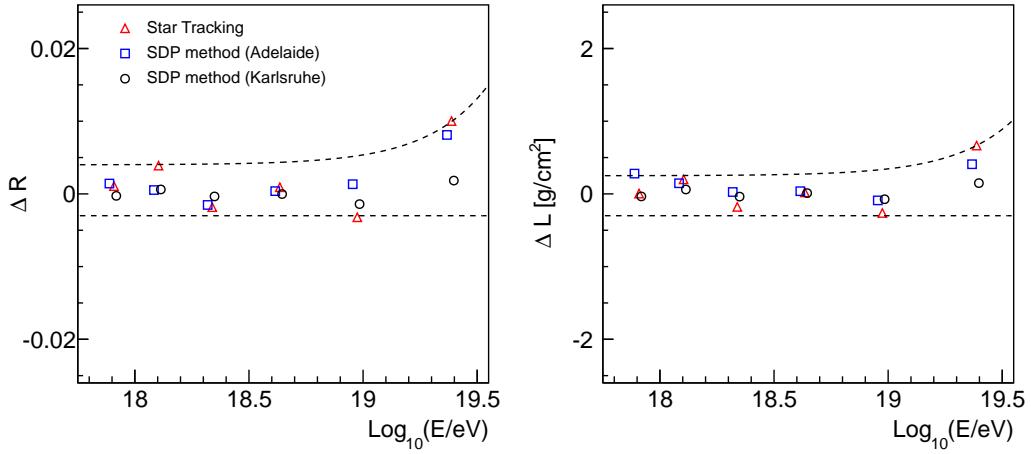


Figure 5.4: Difference in R (left) and L (right) between the standard alignment method used in Auger (timing) and the ones shown in the legend.

and

$$\begin{aligned}\Delta L_{\text{const}}^- &= 0.3 \text{ g/cm}^2 \\ \Delta L_{\text{const}}^+ &= 0.25 + \exp[-6.2 + 3.82 \log_{10}(E/\text{EeV})] \text{ g/cm}^2\end{aligned}\quad (5.9)$$

The effect of the different alignment methods is, as expected, more important at high energy since farther away showers are more sensitive to differences in geometry. However, even at the highest energy, this effect is fairly small when compared to other uncertainties considered.

### 5.3.2.1 Cross-check: Multiple telescopes vs single telescope shower tracks

For the shower profile reconstruction the pixels in all telescopes from the same FD site in which there is signal are considered as part of the same track for every step in the reconstruction. This means that in a significant fraction of the events (energy dependent but around 35% at high energy) there are tracks that cross two telescopes contributing to the same profile. This yields an opportunity to simultaneously cross-check the main effects that can change from telescope to telescope:

- alignment: if two contiguous telescopes are misaligned the reconstructed shape of one is going to be distorted w.r.t. the other, since the distance for a given pixel pointing angle is going to be systematically over- or under-corrected.
- calibration: telescopes are independently calibrated. A mis-calibration between both telescopes leads to a discontinuity in the estimated number of

photons at aperture at the border between the two telescopes. Since we are averaging for several  $\theta$  angles and  $X_{max}$  values this discontinuity can vary significantly from shower to shower, but if the difference is large it can appear as a different shape between events that are contained in one telescope compared to events that cross two telescopes.

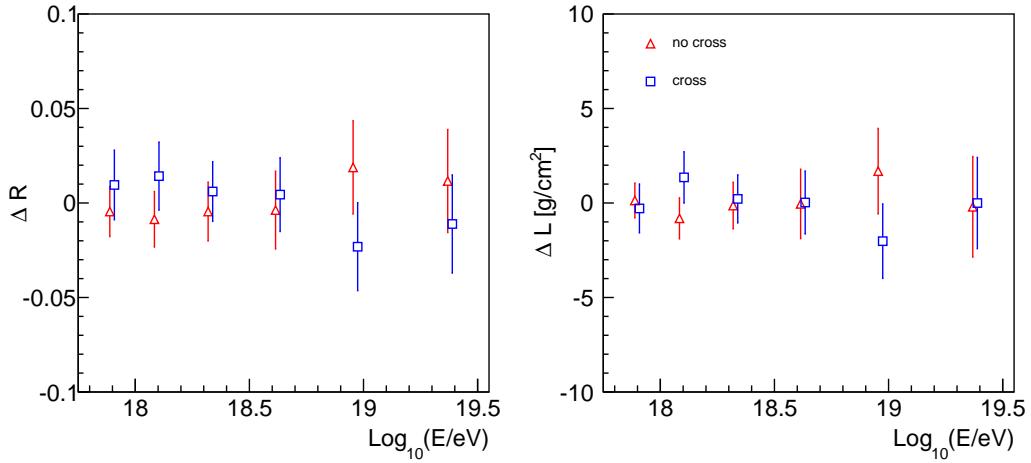


Figure 5.5: Difference in  $R$  (left) and  $L$  (right) between the average profile for all events and for events that cross (blue) or do not cross (red) at least two telescopes.

No clear offset is seen between both classes of shower profiles (figure 5.5). A fit to a straight line at the origin yields a  $\chi^2$  around 1 for both variables. While it is true that we are not very sensitive due to lack of statistics, it is comforting sign that no large systematic is seen either at the eye-to-eye or telescope-to-telescope level.

### 5.3.3 FD site

A basic test that can be made on our data is to check whether the four FD sites are measuring the same profiles in average. This is a very good cross-check for the consistency of our reconstruction since several things can change for different sites: mirror shape, independent Lidar stations, height, alignment, mirror cleaning, etc. In figure 5.6 the average value of both variables for the four FD eyes are plotted. We can see they are perfectly compatible, and both fit to a line at the origin with a  $\chi^2$  below 1, so no uncertainty was assigned.

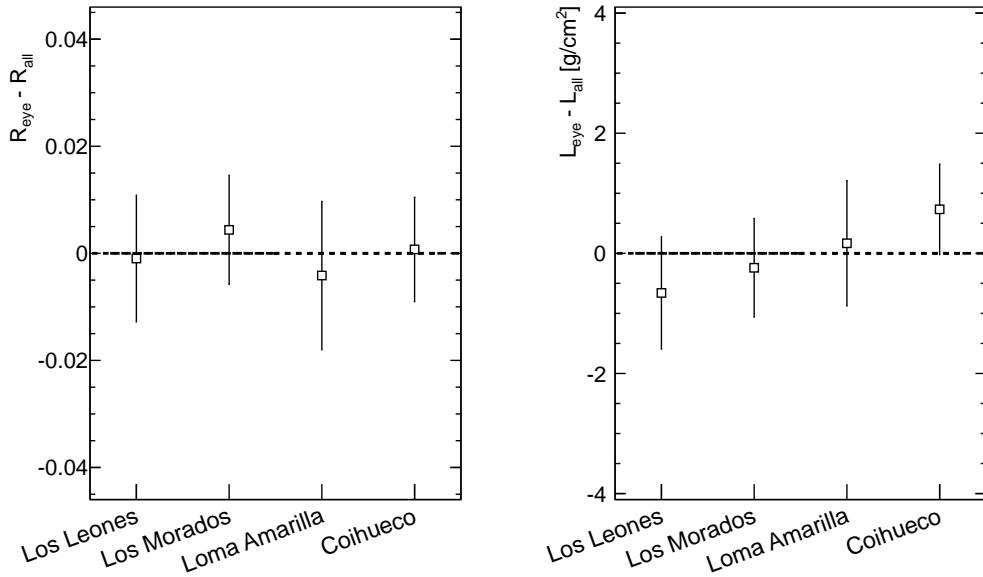


Figure 5.6: Difference in R (left) and L (right) as a function of the season

### 5.3.4 Detector age

The detector performance as a function of time is affected by several factors, as for example dust in the aperture filter and mirror and worsening PMT performance. The latter drift should be corrected by the calibration. The former should be mostly corrected (in normalization) by the drum calibration, however the widening of the signal in the camera due to the dust particles is very difficult to simulate. In addition, there can be unknown time dependent effects that worsen the detector performance. We divided data in two year bins for the latter years when the observatory was fully built, and three year bins otherwise. The results are shown in figure 5.7.

If we only look at the first three bins there seems to be an upward drift with age which, however, is not compatible with the measured value for the last time bin. The values are compatible with a constant R and L value as a function of age, with  $\chi^2$  below 2 for both variables. However, to be safe, we added an age systematic that would bring the  $\chi^2$  to 1. The uncertainty values are:

$$\Delta R_{age} \approx 0.005 \quad (5.10)$$

$$\Delta L_{age} \approx 1 \text{ g}/\text{cm}^2 \quad (5.11)$$

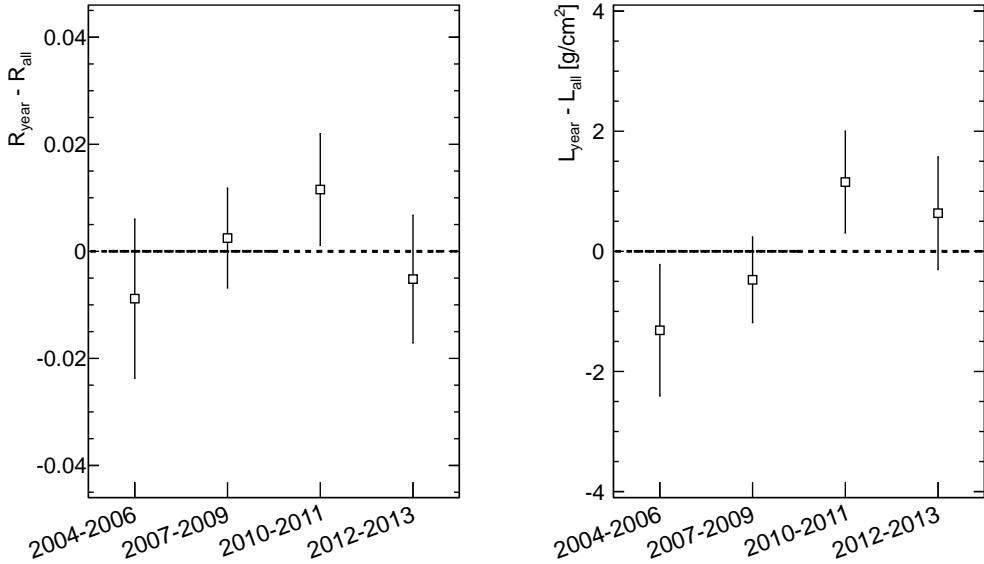


Figure 5.7: Difference in  $R$  (left) and  $L$  (right) as a function of the data taking period

## 5.4 Constraints on the profile shape

As described in section ??, the fit of the  $dE/dX$  profile uses a log-likelihood minimization with constraints on the profile shape. These are taken from high quality data events, where the shape parameters were found to be:

$$X_0 = -121 \pm 172 \text{ g/cm}^2 \text{ and } \lambda = 61 \pm 13 \text{ g/cm}^2 \quad (5.12)$$

Since the variation with energy of the parameters in air shower simulations (around  $100 \text{ g/cm}^2$  in  $X_0$  and  $10 \text{ g/cm}^2$  in  $\lambda$  between  $10^{17.8}$  and  $10^{20}$  eV) is within the resolution achieved in data, we decided to re-reconstruct the showers by changing them  $\pm 1 \sigma$  [83]. This is a conservative approach. As  $X_0$  and  $\lambda$  are correlated, considering the 2 anti-correlation options will lead to an over-estimation of the systematic uncertainty.

There are two main effects when changing the constraints.

First, we get a shift in  $X_{\text{max}}$  and  $dEdX_{\text{max}}$  to accommodate the different shape. This does not change the profile itself, but introduces a bias, of less than  $3 \text{ g/cm}^2$  in  $X_{\text{max}}$  and 2% in  $dEdX_{\text{max}}$ . Since we leave the maximum free in our fit, we can get back these values, and  $R$  and  $L$  do not depend on them. However, it is worth noting that when we use the anti-correlated pairs ([51, 48] and [-293, 73]) we retrieve, for

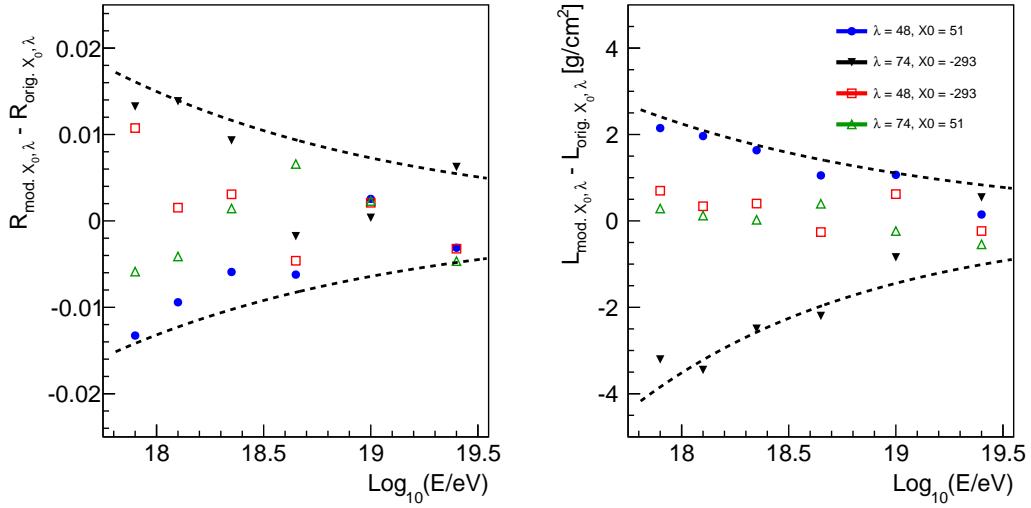


Figure 5.8: Difference in R (left) and L (right) between events reconstructed with normal shape parameters or with ones changed by  $1\sigma$  (see legend) as a function of energy. Dashed lines indicate upper and lower systematic uncertainty.

the lower energy bin, a  $dEdX'_{\max}$  of  $1.02 \pm 0.002$  and  $0.98 \pm 0.002$  respectively. This shows that the shape constraints used do not describe the real shape, since the fitted maximum is not compatible with the true one.

A second order effect is the change of the light components during the (iterative) fit process, where the Cherenkov contribution to each bin depends on the integral of the fitted function up to that point. This can alter not just the fitted parameters but the profile itself, and therefore R and L.

After normalizing and translating the profiles as usual, we get very similar profiles. The results for R and L are shown in figure 5.8. They are consistent with expectations: constraints are more important at lower energies, where showers leave less photons and a shorter track in the camera. An exponential function is fitted to the lower and upper points in each energy bin and shown dashed.

The upper and downwards fitted uncertainties are:

$$\begin{aligned}\Delta R_{\text{const}}^- &= -\exp[-4.34 - 0.99 \log_{10}(E/\text{EeV})] \\ \Delta R_{\text{const}}^+ &= \exp[-4.20 - 0.72 \log_{10}(E/\text{EeV})]\end{aligned}\quad (5.13)$$

and

$$\begin{aligned}\Delta L_{\text{const}}^- &= \exp[1.26 - 0.89 \log_{10}(E/\text{EeV})] \text{ g/cm}^2 \\ \Delta L_{\text{const}}^+ &= \exp[0.81 - 0.71 \log_{10}(E/\text{EeV})] \text{ g/cm}^2\end{aligned}\quad (5.14)$$

## 5.5 Light Component Determination

In this section we deal with uncertainties related to the classification of light as either fluorescence or Cherenkov (direct, Mie or Rayleigh). In fact, most variables change the fluorescence/Cherenkov ratio, from geometry to shape constraints or aerosol determination for example. Here we discuss just the ones that change the ratio in a direct and clear way: fluorescence yield, Rayleigh distribution parameter, multiple scattering calculation and Cherenkov model used.

### 5.5.1 Fluorescence Yield

In the FD reconstruction, the number of photons emitted per MeV of energy deposited is proportional to the atmospheric fluorescence yield. This value has been measured very accurately by the AirFly experiment, which is the reference used in Auger. The effect of the fluorescence yield in the shape determination by changing it in reconstruction by its uncertainty ( $\pm 4\%$ ) was studied. The effect in width (figure 5.9) is the expected one: profiles reconstructed with a higher yield are wider, as they have a larger fluorescence content. However, the size of the effect is small, around  $0.5 \text{ g/cm}^2$ . The shower asymmetry was not expected to change, as the profiles are affected evenly by this change. However, R and L are positively correlated, so this introduces a residual effect in the asymmetry, even smaller than the one in L. A conservative uncertainty band is given by:

$$\begin{aligned}\Delta R_{Yield} &= 0.003 + 0.001 \log_{10}(E/\text{EeV}) \\ \Delta L_{Yield} &= 0.5 + 0.2 \log_{10}(E/\text{EeV}) \text{ g/cm}^2\end{aligned}\tag{5.15}$$

### 5.5.2 Rayleigh scattering - $\rho_N$ parameter

Rayleigh scattering is the elastic scattering of electromagnetic radiation by particles much smaller than the radiation wavelength. Its cross-section per molecule of air is given by [74]

$$\sigma(\lambda) = \frac{24\pi^3(n_s^2 - 1)^2}{\lambda^4 N_s^2(N_s^2 + 2)^2} \left( \frac{6 + 3\rho_n}{6 - 7\rho_n} \right) m^2\tag{5.16}$$

where  $n_s$  is the refractive index of air at wavelength  $\lambda$ ,  $N_s$  is the molecular number density, and  $\rho_n$  is the depolarization factor, which takes into account the anisotropy of the air molecules (for point-like scatterers,  $\rho_n = 0$  and the classical equation is retrieved [120]). This is the only uncertainty factor in the theory - in [74],  $\rho_n$  varies slowly with wavelength, between  $\rho_n = 0.03178$  and  $\rho_n = 0.02955$  in the [300-400]

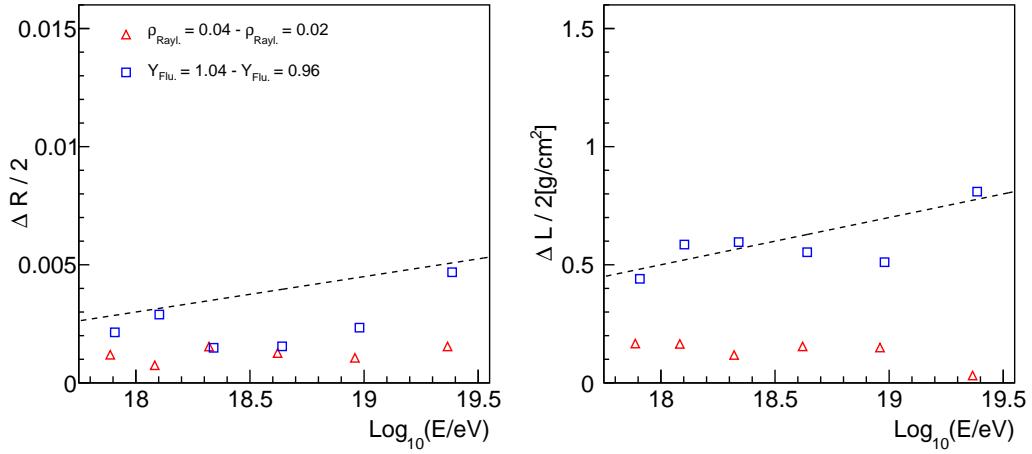


Figure 5.9: Difference in R (left) and L (right) between profiles reconstructed with a fluorescence yield of  $\pm 4\%$  and Rayleigh parameter of 0.02 and 0.04.

nm range. However, the uncertainty in  $\rho_n$  in the literature is at the level of  $\pm 0.01$  [121].

Since the wavelength dependence is negligible, the depolarization factor is implemented in the reconstruction as a constant  $\rho_n = 0.03$ . To study the impact of its uncertainty on the average shape, this value was shifted by  $\pm 0.01$  in `Offline`. The results are shown in figure 5.9. This is the smallest systematic uncertainty considered in this work, and can be parametrized by

$$\begin{aligned}\Delta R_{\rho_N} &\approx 0.001 \\ \Delta L_{\rho_N} &\approx 0.15 \text{ g/cm}^2\end{aligned}\tag{5.17}$$

### 5.5.3 Light propagation model

In the standard reconstruction, excited atmosphere molecules emit fluorescence photons instantaneously, and these photons travel at the speed of light in vacuum,  $c$ . However, both are approximations. The de-excitation time of nitrogen in the air is around 40 ns (height dependent). Also, the atmosphere is not a vacuum, but can be modeled as a stack of layers with decreasing density, and therefore  $n$ . So, the speed of light on its way to the detector varies, and is lower near the ground. Both these effects have already been implemented in the `Offline` framework [122]. These corrections can be switched on and off and check whether it affects the profile shape. It does not in a significant way. Only a small widening of the profile (smaller than

$0.4 \text{ g cm}^{-2}$ ) can be seen, along with a correspondent increase in  $R$ . We take as uncertainty the largest differences observed for all energies, i.e.,  $0.4 \text{ g cm}^{-2}$  in  $L$  and 0.003 in  $R$ .

#### 5.5.4 Cherenkov model

In the standard Auger reconstruction the parametrization used for the Cherenkov photon production (both normalization and angular distribution) is the one by F. Nerling et al [123]. A parametrization based on the work of A. Hillas [124] has also been implemented, so we can switch to the latter in the reconstruction algorithm and verify whether there is any implication on the final reconstructed profile shape. The difference for  $R$  and  $L$  between showers reconstructed using both parameterizations is shown in figure 5.10.

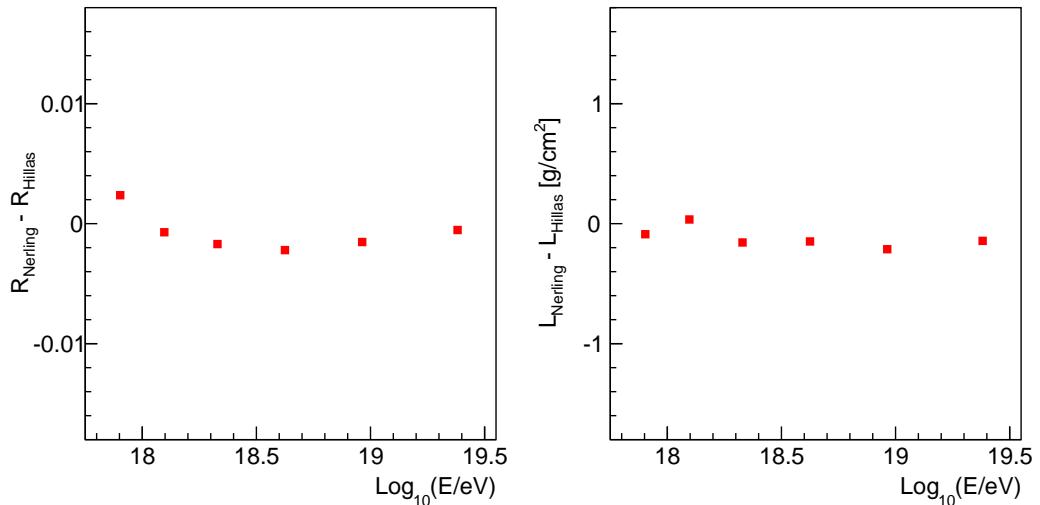


Figure 5.10: Difference in  $R$  (left) and  $L$  (right) between profiles reconstructed with the usual Nerling parametrization and the Hillas one.

The resulting shift is negligible for both variables, but for completeness we will consider it in our uncertainties with a value that encompasses all the points in figure 5.10, i.e.:

$$\Delta R_{\text{Cher.Par.}} < 0.002 \quad (5.18)$$

$$\Delta L_{\text{Cher.Par.}} < 0.2 \text{ g/cm}^2 \quad (5.19)$$

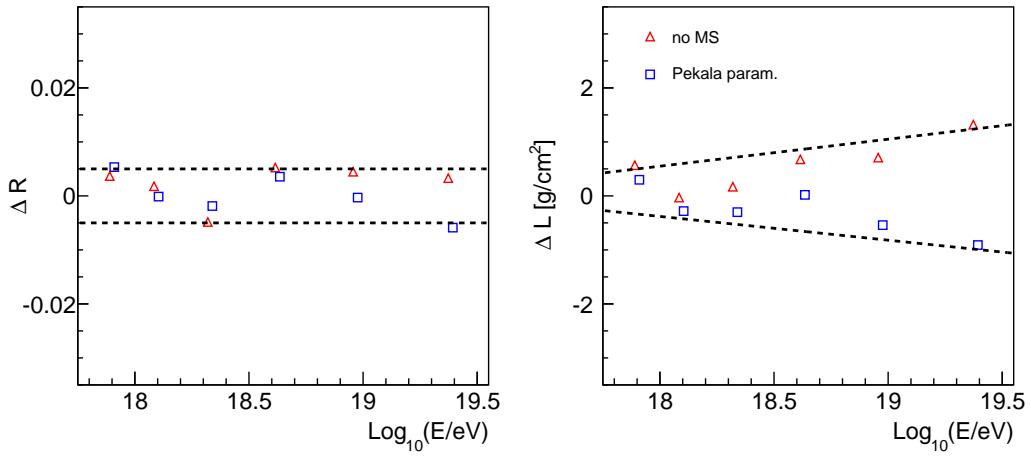


Figure 5.11: Difference in R (left) and L (right) between profiles reconstructed with the usual Roberts parametrization and the Pekala one (blue) or no multi-scattering calculation (red).

### 5.5.5 Multiple Scattering

Multiple scattering can also occur on the atmosphere, although it is a second order effect in relation to Rayleigh. This diffusion is taken into consideration in the Auger standard reconstruction, where the percentage of fluorescence and Cherenkov light in each bin due to multiple scattering is calculated by the Roberts parametrization [125]. An alternative parametrization is derived by Pekala et al [126], which is also implemented in `Offline`. Finally, another alternative is to turn the multiple scattering correction off, so the amount of light due to fluorescence and "regular" scattering rises. The results for the dataset reconstructed with these two alternative hypothesis are on figure 5.11.

In the asymmetry, R, no clear systematic effect is seen with energy, so the uncertainty was defined as a constant (symmetric around 0) value that encompasses all the difference points:

$$\Delta R_{\text{const}} = 0.005 \quad (5.20)$$

For L, clear increase of the importance of multiple scattering correction with energy is seen. This is also qualitatively expected as the average distance to the FD increases with energy, so the scattering corrections get increasingly important. Also according to our expectations, when we do not consider the multiple scattering correction at all, showers are reconstructed larger, due to the fact that multiple

scattered photons end up being mislabeled as fluorescence. The fitted uncertainty in L from figure 5.11 is given by:

$$\begin{aligned}\Delta L_{\text{const}}^- &= -0.38 - 0.44 \log_{10}(E/\text{EeV}) \text{ g/cm}^2 \\ \Delta L_{\text{const}}^+ &= 0.55 + 0.5 \log_{10}(E/\text{EeV}) \text{ g/cm}^2\end{aligned}\quad (5.21)$$

### 5.5.6 Cross-check: Comparing showers with high and low fluorescence percentage

The effect of shifting models of the description of different light components calculations was studied. However, this does not necessarily encompass the uncertainty in the fit of the direct and scattered light components.

To check the consistency of the reconstruction, events in each energy bin are separated in half, according to the reconstructed fluorescence percentage (number of fluorescence photons divided by the total number of measured photons).

The results are shown in figure 5.12. This is the single largest difference seen in width out of all the reconstruction effects and cross-checks studied. The difference between the half of events with the largest fluorescence percentage versus the other half is around  $7 \text{ g cm}^{-2}$ , and slightly increases with energy. This difference is difficult to pin down to a single cause. The percentage of photons allocated to fluorescence depends on the reconstructed geometry, atmosphere, the energy deposit profile itself (through the constrained fit procedure), and in general on the whole reconstruction chain, not only on the factors described in this section.

The reconstruction procedure is the same for MC as for data, so we can check if some effect is seen there, thus making it easier to study and understand. In figure 5.13 is plotted the difference for proton and iron showers of the average shower to those with reconstructed high and low fluorescence contributions. Qualitatively, the same effect is observed: when the fluorescence light is overestimated showers are wider, with a much smaller effect on the asymmetry. However, the size of the shift is much smaller, only around  $\pm 1 \text{ g cm}^{-2}$  compared to around  $\pm 3.5 \text{ g cm}^{-2}$  in data.

Since this difference is larger than the ones found in this sub-section, the upper and lower differences to the average profile were added as a systematic uncertainty. The fit results are shown in figure 5.12:

$$\begin{aligned}\Delta L_{\text{const}}^- &= -2.30 \text{ g/cm}^2 \\ \Delta L_{\text{const}}^+ &= 3.30 + 1.25 \log_{10}(E/\text{EeV}) \text{ g/cm}^2\end{aligned}\quad (5.22)$$

The uncertainty in R is significant for energies below  $10^{18.5} \text{ eV}$  but compatible

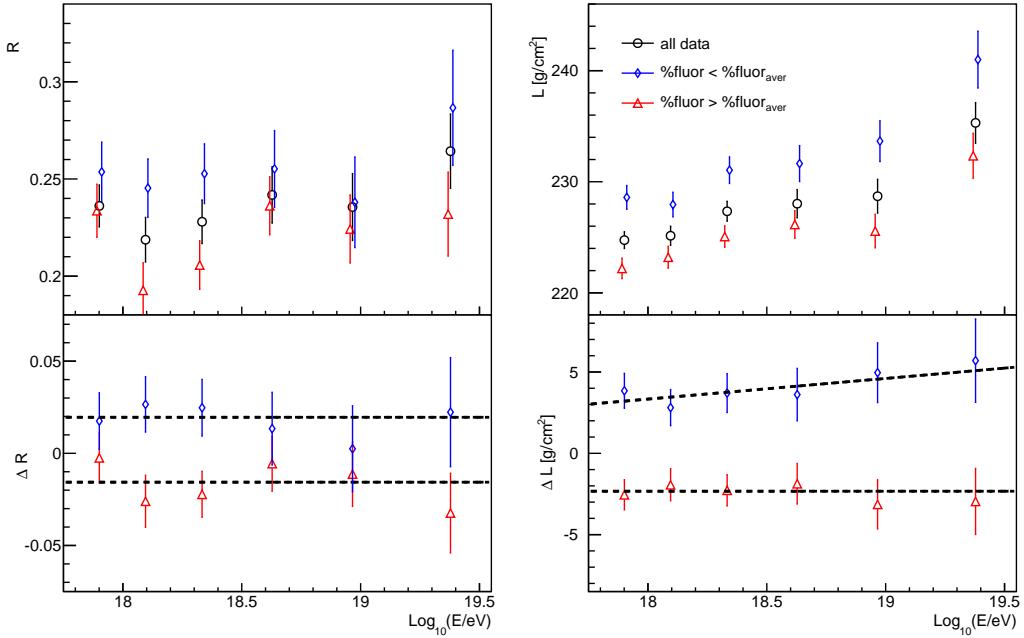


Figure 5.12: Top: R (left) and L (right) as a function of energy for all data (circles), and high (triangles) and low (rhombus) fluorescence samples. Bottom: Residuals of the fluorescence separated samples w.r.t. the total dataset. Fitted lines dashed

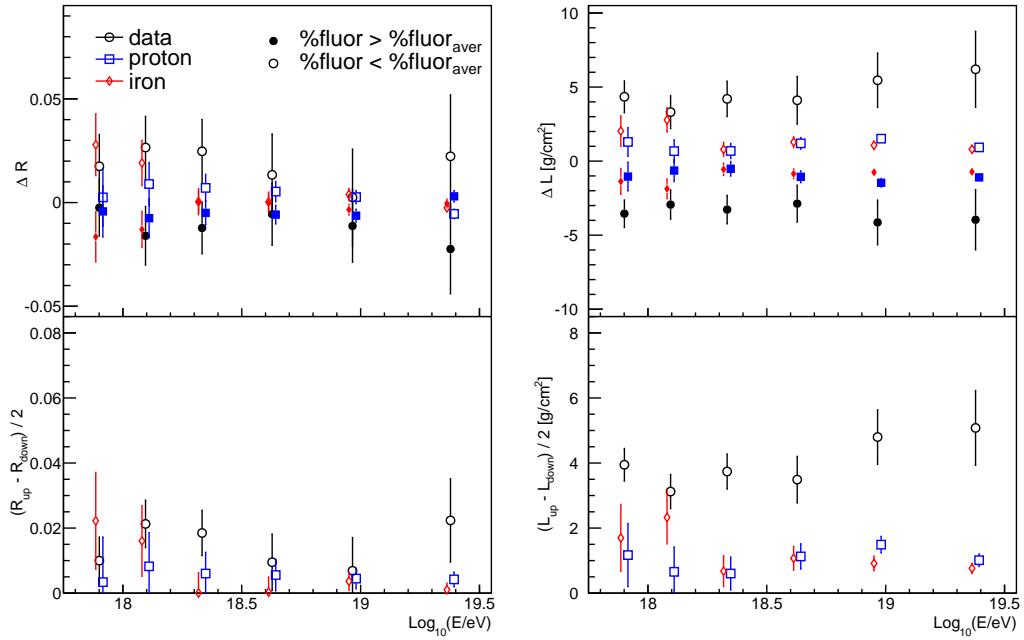


Figure 5.13: Top: Residuals of the fluorescence separated data samples w.r.t. the total dataset for data (black), proton (blue) and iron (red). Bottom: Half of the absolute difference in the above plot, so the uncertainty on data and simulation can be more easily compared.

with 0 at the 3 highest energy bins. However, as no clear trend can be seen or is expected, it is fitted to a straight line:

$$\begin{aligned}\Delta R_{\text{const}}^- &= -0.015 \\ \Delta R_{\text{const}}^+ &= 0.020\end{aligned}\tag{5.23}$$

## 5.6 Atmosphere

The description of atmosphere plays a fundamental role in the light propagation from production to the detector. It is non-uniform: the aerosol distribution is very concentrated near the ground and uncertainties affect the end of profiles, while clouds change the beginning. We will study the effect of both, plus the difference observed in different seasons of the year, when the atmospheric temperature (and therefore density) changes significantly.

### 5.6.1 Aerosols

The atmosphere is monitored due to its importance to the longitudinal profile reconstruction. Perhaps the hardest component to monitor is the aerosol content, as it is neither uniform in time or in space. In this section the impact of the uncertainty in the aerosol measurement on the average shape will be discussed.

#### 5.6.1.1 VAOD uncertainty

The atmosphere is constantly monitored during data taking nights for its aerosol content. As explained in the previous chapter, events in which this database is not available are discarded, so all data analyzed in this work has a measurement of the VAOD. However, the uncertainty in this value can be quite large, and for certain events it can affect significantly the measured profile. This effect can be quantified by altering for each event the VAOD value used in reconstruction by  $\pm 1\sigma$ .

The results in figure 5.14 go qualitatively according to predictions. When a larger VAOD is used for reconstruction, the light at aperture is considered to have been more attenuated and thus a larger energy deposit in the atmosphere is reconstructed. This leads to a larger width. Since the aerosols are mostly concentrated near the ground, the latter part is going to be more affected upwards, which translates into a larger asymmetry. The uncertainty in L can be parametrized by

$$\begin{aligned}\Delta L_{\text{const}}^+ &= 2.0 \text{g/cm}^2 \\ \Delta L_{\text{const}}^- &= -1.2 \text{g/cm}^2\end{aligned}\tag{5.24}$$

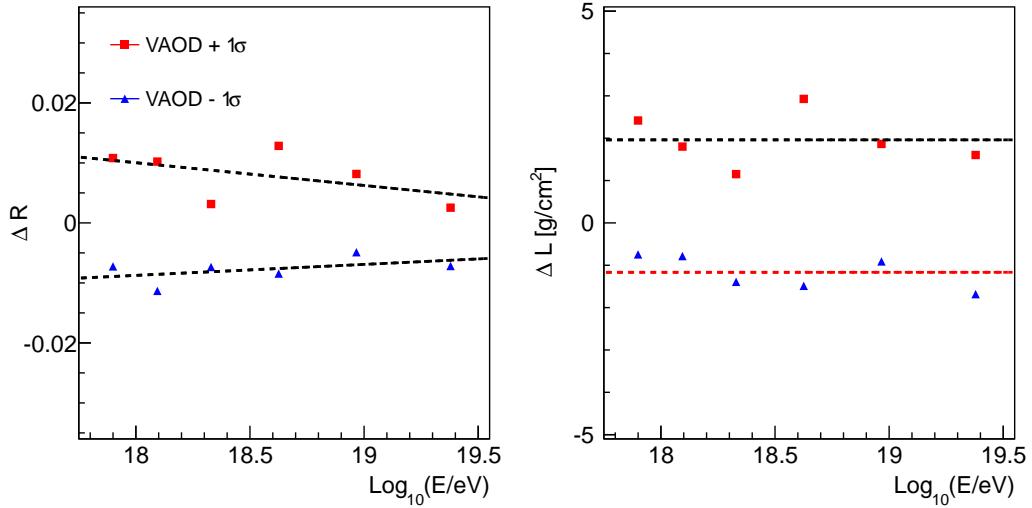


Figure 5.14: Difference in R (left) and L (right) when the measured VAOD is changed by  $\pm 1\sigma$  in reconstruction

and in R

$$\begin{aligned}\Delta R_{\text{const}}^+ &= 0.010 - 0.004 \log_{10}(E/\text{EeV}) \\ \Delta R_{\text{const}}^- &= -0.008 + 0.002 \log_{10}(E/\text{EeV})\end{aligned}\quad (5.25)$$

### 5.6.1.2 Aerosol distribution

The atmospheric attenuation due to aerosol can be inferred measuring the amount of light collected by the FD telescopes of CLF vertical laser tracks, whose energy and geometry are known. The standard method used in Auger to measure the VAOD is the Data Normalized (DN) Method [129].

Assuming that the horizontal uniformity for the atmosphere, it can be shown that the aerosol optical depth is [119]:

$$VAOD(h) = -\frac{\sin \phi_1 \sin \phi_2}{\sin \phi_1 + \sin \phi_2} \left[ \ln \left( \frac{N_{\text{obs}}(h)}{N_{\text{mol}}(h)} \right) - \ln \left( 1 + \frac{P_a(\theta)}{P_m(\theta)} \right) \right] \quad (5.26)$$

where  $N_{\text{mol}}(h)$  is the number of photons expected on a completely clear night (the so called Rayleigh night, when molecular attenuation only is present) and  $P_a(\theta)$  and  $P_m(\theta)$  are the fractions of photons scattered by aerosols and molecules, respectively. For nearly all scattering angles except for very forward directions,  $P_a(\theta)$  is small compared to  $P_m(\theta)$ , so the second term in eq. 5.26 can be neglected. Therefore, for vertical laser shots equation reduces to:

$$VAOD(h) = -(\ln [N_{\text{obs}}(h)] - \ln [N_{\text{mol}}(h)]) \left[ 1 + \frac{\sin \phi_2}{1 + \sin \phi_2} \right] \quad (5.27)$$

With these simplifications, the CLF optical depth measurement depends only on the elevation angle of each laser track segment and on the Rayleigh night reference profile  $N_{mol}(h)$ . Using measurements recorded on Rayleigh nights, the hourly CLF observations can be properly normalized without the need for absolute photometric calibrations of the FD or laser, and the aerosol optical depth may be calculated directly from 5.27. The big caveat with this analysis is the assumption of Rayleigh nights, in which no aerosols are present. This assumption has come into question recently, e.g. [132], as it may happen that the aerosol content of these reference nights might be non-negligible.

An alternative method developed within the Pierre Auger Collaboration is the Laser Simulation (LS) method. This method is based on the comparison between measured and simulated events. Simulated laser events are generated at a fixed energy using a parametric description of the aerosol attenuation. The vertical aerosol optical depth is in the upper part of the atmosphere described with a function of two parameters

$$VAOD(h_2 - h_1) = -\frac{H_{aer}}{L_{aer}} \left[ \exp\left(-\frac{h_2}{H_{aer}}\right) - \exp\left(-\frac{h_1}{H_{aer}}\right) \right] \quad (5.28)$$

where  $L_{aer}$  is the aerosol horizontal attenuation length and  $H_{aer}$  the aerosol scale height. In the lowest part of the atmosphere, the planetary boundary layer (PBL), the aerosol content is directly influenced by the contact with a planetary surface. Is can be parametrized by a constant factor for heights below the PBL height  $H_{PBL}$ , i.e,

$$VAOD(h - H_g) = \begin{cases} \frac{1}{L_{aer}} (h - H_g) & \text{if } h < H_{PBL} \\ -\frac{H_{aer}}{L_{aer}} \left[ \exp\left(-\frac{h - H_{PBL}}{H_{aer}}\right) - 1 \right] + \frac{1}{L_{aer}} (H_{PBL} - H_g) & \text{otherwise} \end{cases} \quad (5.29)$$

where  $H_g$  is the height at which the ground is.

The Laser Simulation (LS) Analysis is divided in three main phases:

- production of a grid of simulated laser events;
- reconstruction of measured laser events;
- research of the simulation that best fit the measured event to evaluate the aerosol optical depth profile

This analysis returns a preferred shape for the VAOD as a function of the height. The normalization of that shape can come from two souces. The first is also assuming

a reference Rayleigh-night, i.e., setting the value of the VAOD at 3 km in the simulation of that day as the zero of the analysis. This will still yield different results from the DN method as the shape as a function of the height will differ. We call this the LS (3 par.) RN model. The second choice is assuming that the simulation is working fine and all absolute calibrations (CLF, FD) are well understood, in which case there is no need for a reference Rayleigh night. This is the LS (3 par.) no-RN model.

The main difference between the LS and DN method is that the parametrization of the VAOD function on the first gives an exponential rise on the aerosol content as we approach the planetary boundary layer, while the latter does not assume any function for the aerosol shape with atmospheric height. This leads to distortions in the average profiles, as the number of photons at aperture is translated into similar energy deposits at earlier depths (high altitudes) in both models but into different ones closer to the ground.

In figure 5.15, R and L are plotted as a function of energy for the three cases studied: DN and LS (3 par.) with RN and no-RN. The shift in L is significant, but is well within the previously studied systematic uncertainties. The uncertainty on the aerosol distribution was expected to distort the profile, and in fact the shift in R for the non-normalized LS analysis w.r.t. the standard one is the largest studied in this work. Not only that, it changes R down, making it more compatible with data, although not fully within model predictions.

The uncertainty in L can be parametrized by:

$$\begin{aligned}\Delta L_{\text{const}}^+ &= 2.4 + 2.5 \log_{10}(E/\text{EeV}) - 1.3 \log_{10}^2(E/\text{EeV}) \text{ g/cm}^2 \\ \Delta L_{\text{const}}^- &= \exp(-0.4 - 2.4 \log_{10}(E/\text{EeV})) \text{ g/cm}^2\end{aligned}\quad (5.30)$$

and in R by:

$$\begin{aligned}\Delta R_{\text{const}}^+ &= 0.010 \\ \Delta R_{\text{const}}^- &= -0.012 \log_{10}^2(E/\text{EeV})\end{aligned}\quad (5.31)$$

### 5.6.2 Clouds

Clouds can distort the profile. The most extreme cases, like for example figure 5.16, are discarded by the  $\chi^2$  selection, but some distorted profiles may still pass.

One of the main cloud cuts in our analysis is requiring that if there is a LIDAR measurement, the cloud coverage is below 25%. However, only around 74% of the

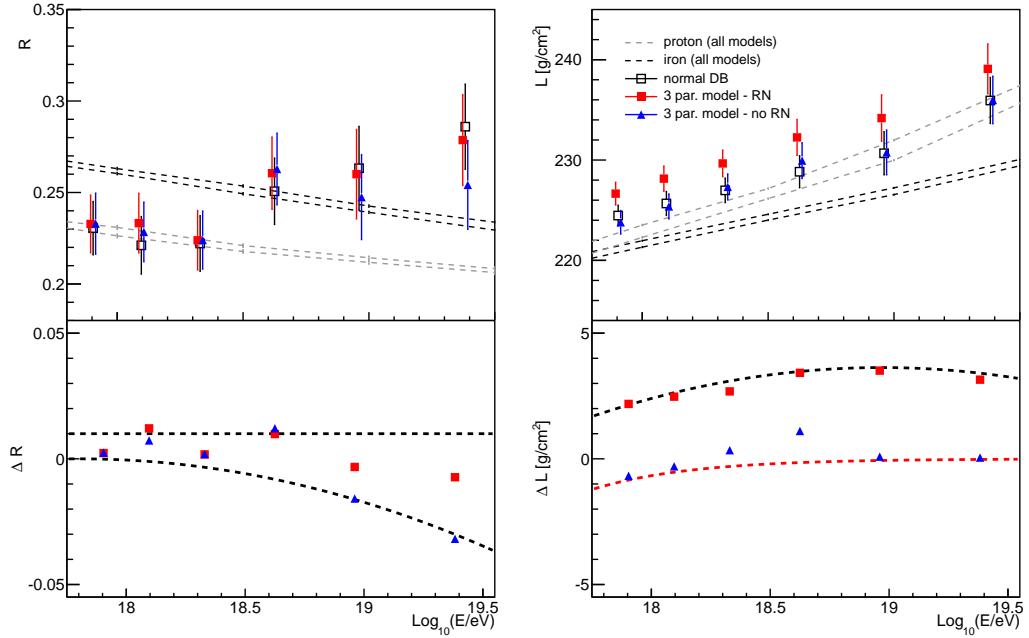


Figure 5.15: Difference in  $R$  (left) and  $L$  (right) using the different databases shown in the legend

events have LIDAR working, so there could still be some cloudy nights contamination in our sample. Looking at the events for which there is a measured cloud coverage, we see that only 27% of them are above 25%. This means that the expected contamination of events with cloud coverage above 25% is  $f = 0.26 \times 0.27 = 7\%$ .

In figure 5.17 we plot the difference between profiles for only cloudy nights (which we cut) and our selected data. For energies below  $10^{19.2}$  eV, profiles with and without clouds seem to be compatible. However, at higher energies they differ significantly. To estimate our systematic uncertainty we multiply the difference at the last energy bin by the factor  $f$ , which gives us an upper limit on our expected cloud contamination. The values are -0.012 for  $R$  and  $-1.1 \text{ g cm}^{-2}$  for  $L$  at the last energy bins. For all the other bins the uncertainty is considered to be 0.

### 5.6.3 Season of the year

As the seasons change, so does the atmospheric profile. This is taken into account in our reconstruction with month dependent models, but we can test whether the changes are described correctly in our `Offline` framework by checking whether the profiles are constant all year round. There is a small difference in profile width, with  $L$  being around  $3 \text{ g}/\text{cm}^2$  larger in spring and summer than autumn and winter (see

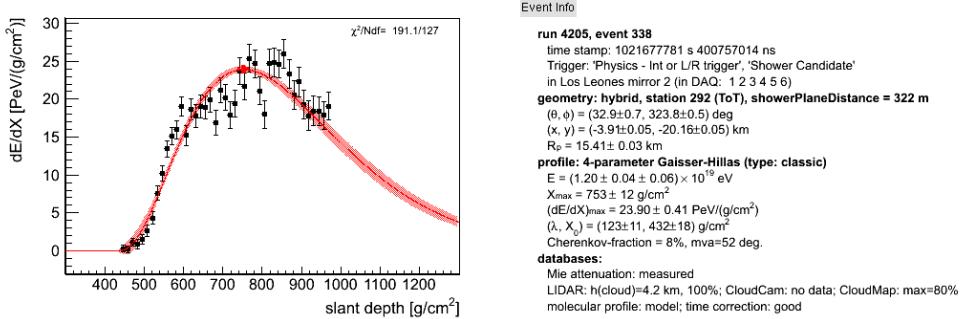


Figure 5.16: High energy event with measured cloud coverage of 100%

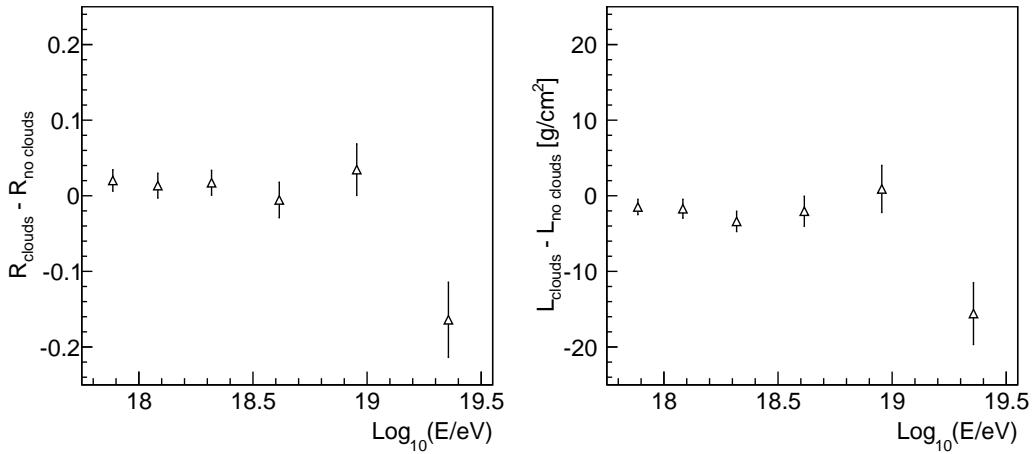


Figure 5.17: Difference in R (left) and L (right) between profiles measured when a cloud coverage above 25 % was detected and events measured with clear skies

figure 5.18). In *realMC* the atmosphere used to simulate and reconstruct the proton and iron profiles is the same as in data. Looking at the differences between season in the same figure it is clear that no effect is seen in simulations, so we consider as a systematic uncertainty half the difference observed between the first and last two seasons of the year.

$$\begin{aligned}\Delta R_{\text{season}} &= 0.005 \\ \Delta R_{\text{season}} &= 1.5 \text{ g/cm}^2\end{aligned}\tag{5.32}$$

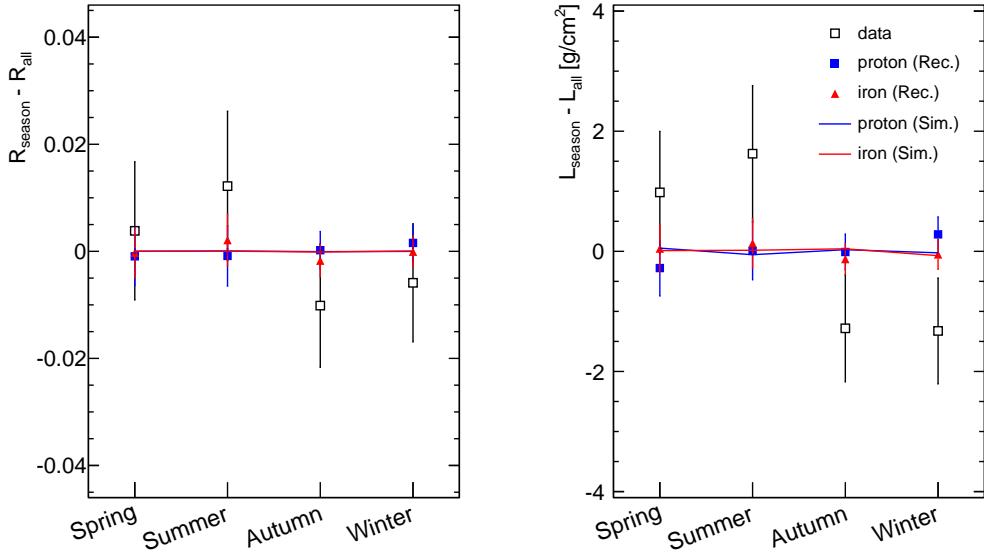


Figure 5.18: Difference in  $R$  (left) and  $L$  (right) as a function of the season

## 5.7 Light Integration

The profile reconstruction is treated as a one-dimensional procedure (in time or depth), since after the geometry determination everything is calculated within the shower axis. However, light arriving at the camera has a lateral distribution as well, so we define an angle ( $\zeta$ ) in the camera w.r.t. the shower plane within which we integrate the signal in the pixels. In this section the systematic uncertainties related to variations in the  $\zeta$  size and the definition of the light integration area are described.

### 5.7.1 Pixel Integrated vs Circle

It is known that the light integration angle,  $\zeta$ , resultant from the minimization procedure implemented in Auger does not integrate the entire lateral light distribution from the shower. Since it comes from a S/N minimization, the shower light which extends into background dominated pixels is discarded. In the Auger reconstruction, all this light is assumed to be lost, and this loss is corrected a posteriori. However, the whole picture is slightly more complicated.

In the standard reconstruction (Circle)  $\zeta$  is a circle in the camera around the

maximum signal point for a given time bin. This means that in general the circle borders are going to pass somewhere within 2 or 3 pixels, which collected light that is fully integrated, but only use a partial area for the lost light correction. The PixelIntegrated method was designed to deal more accurately with this. There  $\zeta$  is calculated not related to a collection area, but to pixels it integrates. Then, the correction is made over the more complicated area of the pixels which are considered as signal.

Note that while this might look like a huge difference in integration area, the corrections and pixel efficiencies in the Circle method were calibrated by data, so the difference in integration area is largely absorbed on average. The measured energy, for example, is around 3% smaller with the Circle method for low  $\zeta$  angles, but equal for  $\zeta$  above 2°.

The dataset was reconstructed with the two methods and the average profile variables calculated for each case (figure 5.19). The differences are small when compared to other studied effects. Since it is a global effect that influences (mostly) evenly the whole shower it is larger in L than in R, as expected. The parametrization of the difference between the two methods is given by

$$\begin{aligned}\Delta R_{\text{PixInt}}^- &= -0.003 + 0.002 \log_{10}(E/\text{EeV}) \\ \Delta L_{\text{PixInt}}^+ &= 1.0 - 0.3 \log_{10}(E/\text{EeV}) \text{ g/cm}^2\end{aligned}\tag{5.33}$$

### 5.7.2 Lateral Width Correction

As explained previously, there is still light from the air shower reaching the telescope camera at angles larger than  $\zeta$ . This was simulated by models of the lateral light distribution in the camera. It is however very hard to model, as several effects can increase the signal's lateral width, mainly the corrector lenses and mirror which are both subject to dust. In Auger it was found [85] that the models severely underestimated the amount of light outside the calculated  $\zeta_{opt}$  - around 50% but distance and age dependent. Thus, a data based correction for the light outside  $\zeta_{opt}$  was derived, called LW (Lateral Width) correction.

To estimate a profile shape uncertainty due to this correction, data was reconstructed with and without it and the difference measured. As can be seen in figure 5.19, the difference in R is small, since this is a global effect on the profile. In L however, particularly at low energies, it is significant. The fact that showers are in

average thinner without the correction is also expected, as less light is attributed to the shower given the same number of photons in aperture.

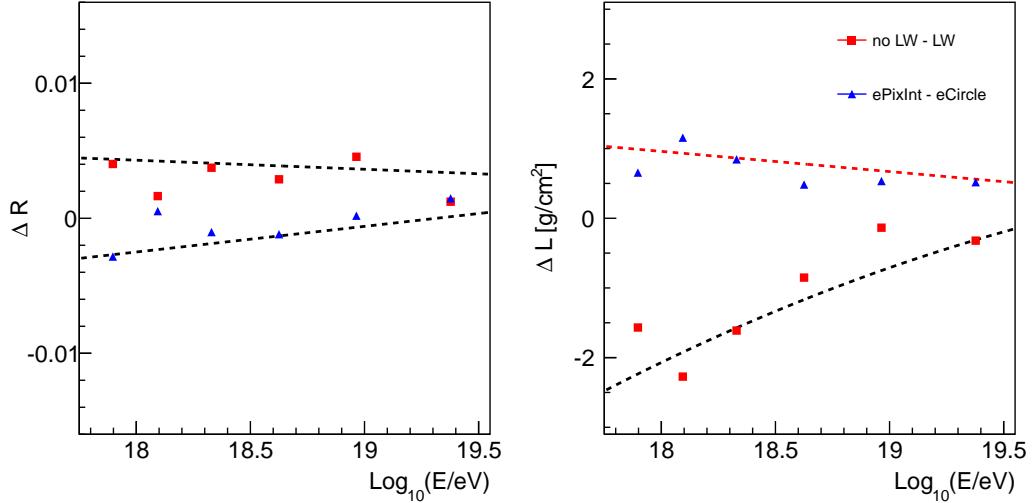


Figure 5.19: Half of the difference in R (left) and L (right) between events reconstructed with the eCircle (LW-correction on or off) or ePixelIntegrated algorithm. Since we are using the same events for both no statistical error was considered. The dashed lines are the upper and lower fitted uncertainties.

The uncertainty can be parametrized as:

$$\begin{aligned}\Delta R_{\text{LW}}^+ &= 0.004 + 0.001 \log_{10}(E/\text{EeV}) \\ \Delta L_{\text{LW}}^- &= -2.1 + 1.6 \log_{10}(E/\text{EeV}) - 0.2 \log_{10}^2(E/\text{EeV}) \text{ g/cm}^2\end{aligned}\quad (5.34)$$

### 5.7.3 Variations in $\zeta_{\text{safety}}$

The S/N minimization is not the be-all and end-all procedure for the  $\zeta_{\text{opt}}$  calculation, i.e., our results should not be robust against slight changes in the light integration area. A way to estimate the systematic due to this is not obvious. However, we should at least have a feeling for whether an over- or under-estimation of the area leads to a large difference in the shape. If this is true, it means we are relying strongly on the outside light correction.  $\zeta_{\text{opt}}$  is of the order of  $1.2^\circ$ , although slightly energy dependent. We decided to manually change  $\zeta_{\text{opt}}$  by  $\pm 0.2^\circ$  and  $+0.5^\circ$ , as we start getting problems if  $\zeta$  is much smaller than 1 (the minimum  $\zeta$  usually accepted in the S/N minimization is  $0.9^\circ$ ).

Since our reconstruction (and the LLW correction in particular) was tuned for  $\zeta_{\text{opt}}$ , this is almost certainly going to over-estimate our uncertainty. The results are shown in figure 5.20. The shower width increases as we increase the light collection

area and vice versa. This is expected as we are integrating over more background and multiple scattered light, which are not correcting for correctly. As with the previous two uncertainties studies, this is much smaller in R than in L.

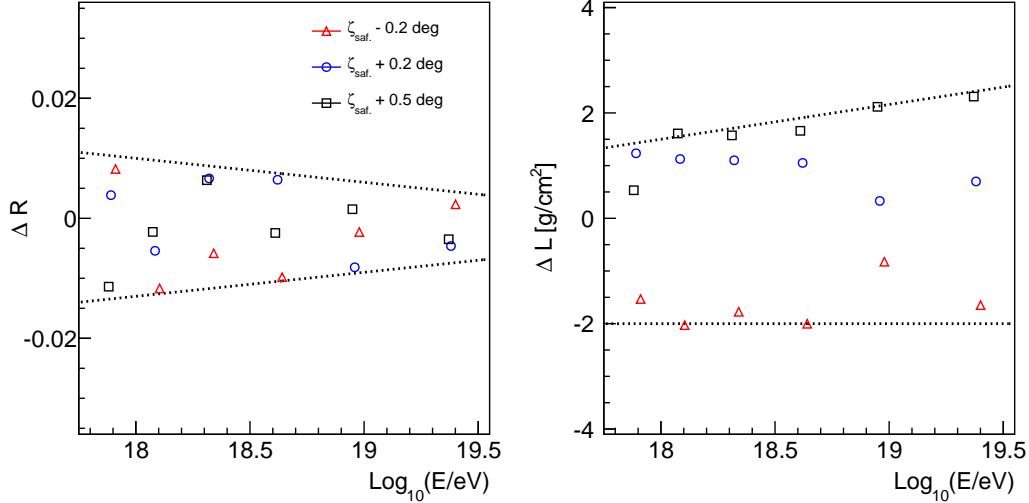


Figure 5.20: Difference in R (left) and L (right) between events reconstructed with the shown  $\zeta$  correction w.r.t. the one calculated by the S/N minimization. The dashed lines are the upper and lower fitted uncertainties.

The considered uncertainty, which encompasses all points, is for R

$$\begin{aligned}\Delta R_{\text{axis}}^- &= -0.013 + 0.004 \log_{10}(E/\text{EeV}) \\ \Delta R_{\text{axis}}^+ &= 0.010 - 0.004 \log_{10}(E/\text{EeV})\end{aligned}\quad (5.35)$$

and for the width

$$\begin{aligned}\Delta L_{\text{axis}}^- &= -2.0 \text{ g/cm}^2 \\ \Delta L_{\text{axis}}^+ &= 1.5 + 0.66 \log_{10}(E/\text{EeV}) \text{ g/cm}^2\end{aligned}\quad (5.36)$$

## 5.8 Geometry

### 5.8.1 SDP

The first step in the FD shower geometry reconstruction is defining the plane that contains both the shower axis and the detector, as all light propagation is going to be calculated in 2D within this plane. This plane is defined by two angles:  $\phi$ , which measures the direction of its projection on the ground with respect to the FD, and  $\theta$  which is its inclination with respect to vertical plane. The SDP is fitted by a S/N

minimization on the integrated light in the candidate shower pixels. We can take the results of this minimization and shift the calculated angles,  $\phi$  and  $\theta$ , by their correspondent uncertainties. The results for R and L as are in figure 5.21. No clear

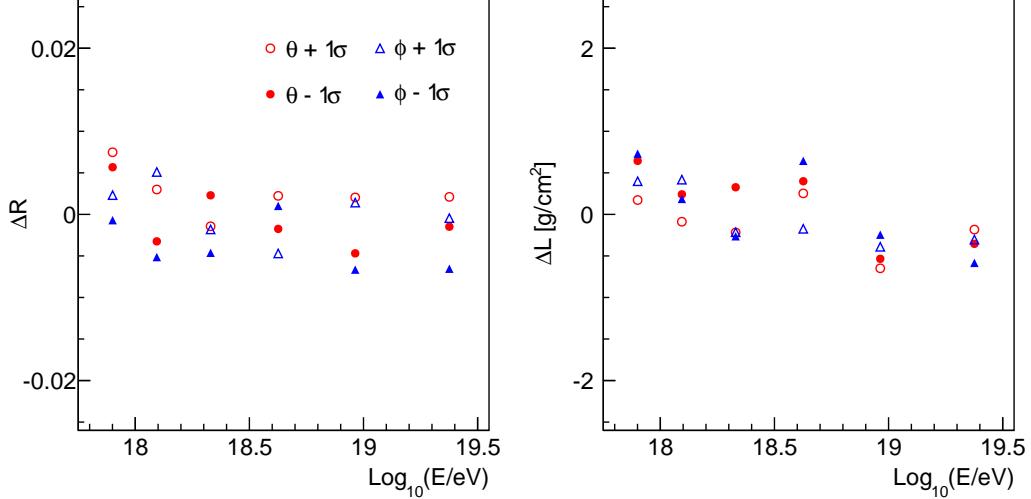


Figure 5.21:

trend is seen in either variable, so we chose a constant value that encompasses all points as our uncertainty:

$$\begin{aligned}\Delta R_{\text{SDP}} &\approx 0.005 \\ \Delta L_{\text{SDP}} &\approx 0.75\end{aligned}\tag{5.37}$$

### 5.8.2 Shower Axis

Fixing the SDP leaves three parameters to be reconstructed:  $R_p$ ,  $\chi_0$  and  $T_0$ . A change in the shower axis direction has the effect of either extending or diminishing the reconstructed track. In case of a FD-pointing shower, for example, a higher  $\chi_0$  implies a longer track. This, in turn, leads to a larger width in the reconstructed profile and a smaller asymmetry. However, in case it is pointing the other way, the same change in  $\chi_0$  brings the opposite effect. Since we are measuring the average profile, and the distribution in  $\phi$  is flat, the cumulative effect on the average profile is unclear, but not expected to be large. So, events are reconstructed using exactly the same SDP and axis algorithm, but changing each of the three parameters ( $R_p$ ,  $\chi_0$  and  $T_0$ ) in turns by  $\pm 1\sigma$  of their fitted error.

The results are in figure 5.22. The most clear result is that, as expected, the

uncertainty diminishes with energy, as geometry reconstruction is more accurate.

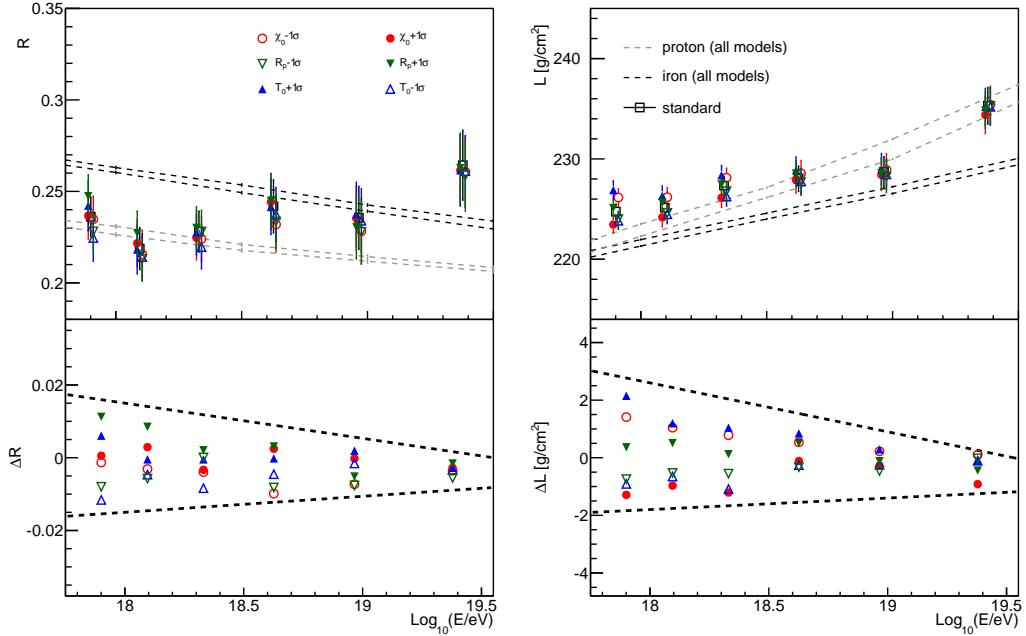


Figure 5.22:

The fitted line in this case are to the (quadratically) summed contributions, as we thought it is more clear than counting them separately

$$\begin{aligned}\Delta R_{\text{axis}}^- &= -0.015 + 0.004 \log_{10}(E/\text{EeV}) \\ \Delta R_{\text{axis}}^+ &= 0.015 - 0.010 \log_{10}(E/\text{EeV})\end{aligned}\quad (5.38)$$

and for the width

$$\begin{aligned}\Delta L_{\text{axis}}^- &= -1.8 + 0.4 \log_{10}(E/\text{EeV}) \text{ g/cm}^2 \\ \Delta L_{\text{axis}}^+ &= 2.6 - 1.7 \log_{10}(E/\text{EeV}) \text{ g/cm}^2\end{aligned}\quad (5.39)$$

## 5.9 Summary

The list of contributions to the systematic uncertainties of  $R$  and  $L$  studied in this chapter are summarized in table 5.2. The value indicated is a round number around  $10^{18.5}$  eV and gives a rough idea of the systematic size. However, most are energy dependent, and can be one-sided. The energy evolution of the uncertainties, separated in four main categories, are shown in figures 5.23 and 5.24.

At lower energies, the uncertainty in  $L$  is dominated by the light integration and fluorescence determination. For  $R$  the picture is more complicated, as the

uncertainty is well distributed among all four considered categories of systematics, but the larger contribution comes from the same sources as the one in  $L$ .

At higher energies the picture is the same in  $L$ , as particularly the fluorescence ratio systematic grows to be the largest considered in this study, at around  $4 \text{ g/cm}^2$ . The  $R$  measurement is clearly dominated by a downwards systematic uncertainty related to the aerosol measurement. This is due to the fact that  $R$  is reconstructed smaller when we use one of the alternate aerosol databases.

Also, the depth dependent systematic uncertainty from all these effects was estimated, and the resultant profiles are shown in figure 5.25. It is interesting to see that the systematic uncertainty increases as the tails are approached, especially in the upper tail where the uncertainty in the contribution due to atmospheric attenuation of emitted shower fluorescence light is larger.

Type	Description	Uncertainty	
		R	L [g/cm <sup>2</sup> ]
Reconstruction	Bias Correction	0.005	0.7
Energy Scale	$\pm 14\%$	0.010	1.0
Detector	SD-FD Time Offset	0.001	1.0
	Telescope Alignment	0.005	0.5
	Detector age	0.005	1.0
Shape Constraints	$\lambda = -121 \pm 171$ , $X_0 = 61 \pm 13$	0.015	3.0
Light Component determination	Fluorescence Yield	0.003	0.7
	Light propagation model	0.003	0.4
	Rayleigh $\rho_N$ parameter	0.001	0.2
	Cherenkov Model	0.002	0.2
	Multiple Scattering	0.005	1.0
	Fluorescence/Cherenkov ratio	0.015	4.0
Atmosphere	Aerosols	0.020	3.0
	Clouds (only above $10^{19}$ eV)	-0.010	-1.0
	Season of the year	0.005	1.0
Light Integration	$\zeta$ variations	0.010	2.0
	Integration method	-0.003	1.0
	Lateral Width Correction	0.005	-2.0
Geometry	SDP	0.008	0.8
	Shower Axis	0.010	2.0
Total		0.040	6.0

Table 5.2: Table of systematic uncertainties

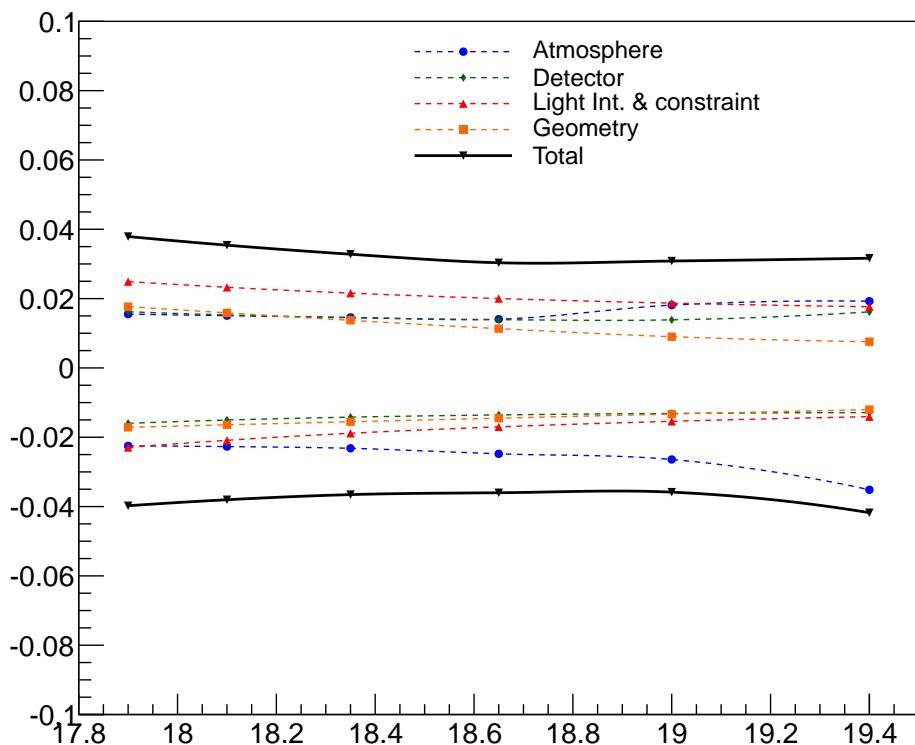


Figure 5.23: R systematic uncertainty separated by component

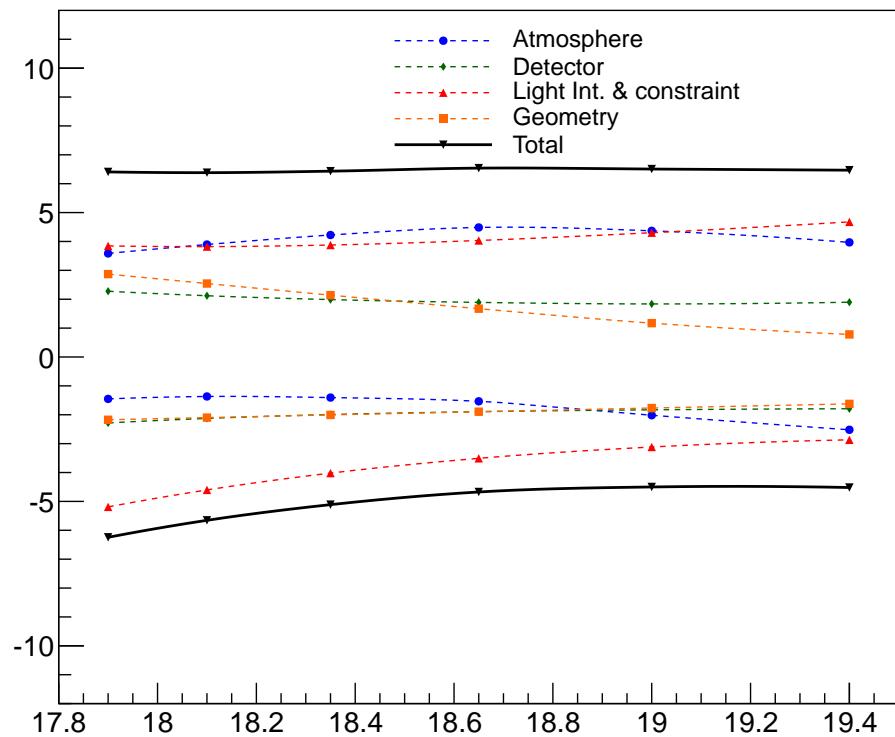


Figure 5.24: L systematic uncertainty separated by component

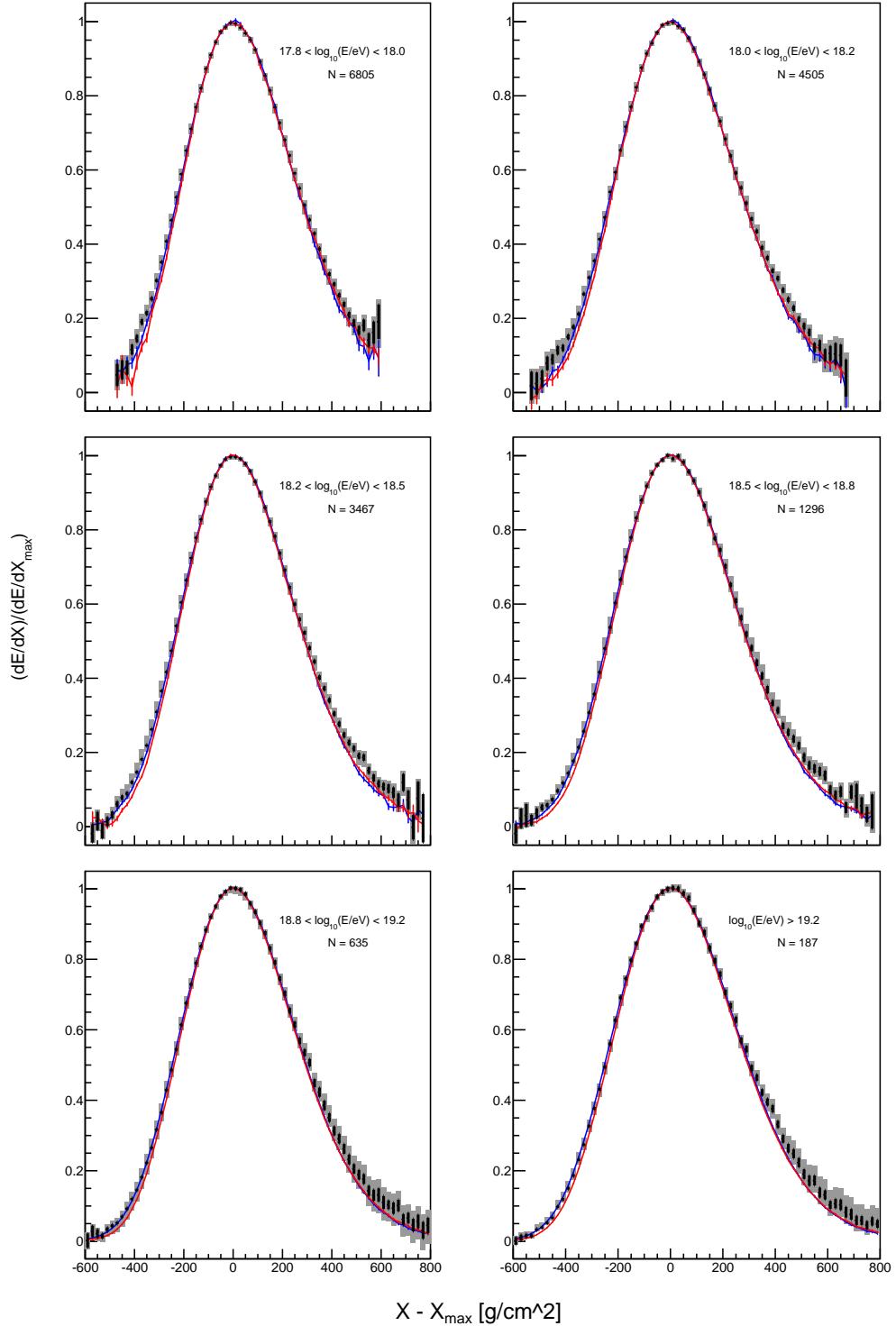


Figure 5.25: Average profiles for all the energy bins used in this work. The statistical error is shown as a black line and an estimation of the systematic uncertainty bin-by-bin as a gray area. Reconstructed profiles for MC are show in blue (proton) and red (iron). The high energy model used is QGSJETII03.



# CHAPTER 6

## Results and interpretation

---

In this chapter the values of the  $R$  and  $L$  variables as a function of energy with the systematic uncertainties will be presented, and interpreted in conjunction with independent Auger measurements and alternative interaction models.

In section 6.1, the fit results and their systematic uncertainties, estimated in the previous chapter, are shown.

If a given hadronic interaction model gives a consistent description of shower behaviour, all measured variables have to be simultaneously predicted by it for a given primary particle. Therefore, the approach in the first section will be to compute the absolute value and energy evolution from  $X_{\max}$ ,  $X_{\max}^\mu$ ,  $\ln R_\mu$ ,  $\sec \theta_{\max}$ ,  $R$  and  $L$  and compare it to model predictions. Since there is now a considerable number of variables that are sensitive to the shower development properties, we can try to categorize them according to the shower component they are most sensitive to. This will be presented in section 2, which is mostly based on an internal Auger note [127], which analyzed the publicly available measurements. Here, I decided to keep the same format, first describing  $X_{\max}$ ,  $X_{\max}^\mu$ ,  $\ln R_\mu$  and  $\sec \theta_{\max}$ , and then expand it by adding a detailed interpretation of the  $R$  and  $L$  measurements.

Recently, and mostly as an answer to the muon deficit problem in standard hadronic models, some alternative interaction models have been proposed. They generally increase the number of muons at the ground by altering substantially the high energy interaction properties, where there are few constraints from accelerator measurements. The process by which this change is made, however, is disparate among them, resulting in substantially different profile shapes for each of the models studied. This will be the subject of section 2.

Finally, in the third section, the dependence of the electromagnetic profile shape on some macroscopic multiparticle production parameters will be studied. For this, a fast Monte Carlo shower simulation based on the Heitler-Matthews model was developed. Despite being a very simplified model compared to full shower simulations, e.g., CONEX and CORSIKA, it can reproduce well the energy evolution of the main shower observables. Since it depends on just a few initial interaction parameters,

the sensitivity of the shower profiles to changes in these parameters can be studied.

## 6.1 Results

The analysis described in the previous chapters has the purpose of allowing the measurement of the two parameters that describe the average longitudinal profile shape,  $R$  and  $L$ , together with the energy dependent systematic uncertainty associated with their measurement.

The results are shown in table 6.1: the values of  $R$  and  $L$  for the average longitudinal profile at each energy bin, along with their statistical and estimated systematic uncertainty. The average energy in each bin and the number of events are also shown for reference.

Energy [eV]	$\langle \log_{10}[E/eV] \rangle$	N	R			L		
			$\langle value \rangle$	stat.	syst.	$\langle value \rangle$	stat.	syst.
$10^{17.8} - 10^{18}$	17.89	6996	0.229	0.009	$\pm 0.038$ $\pm 0.039$	224.8	0.6	$\pm 6.4$ $\pm 6.2$
$10^{18} - 10^{18.2}$	18.09	5132	0.208	0.010	$\pm 0.035$ $\pm 0.038$	224.9	0.8	$\pm 6.4$ $\pm 5.6$
$10^{18.2} - 10^{18.5}$	18.33	4412	0.221	0.010	$\pm 0.033$ $\pm 0.037$	226.9	0.8	$\pm 6.4$ $\pm 5.1$
$10^{18.5} - 10^{18.8}$	18.62	1780	0.240	0.013	$\pm 0.030$ $\pm 0.036$	227.9	1.1	$\pm 6.5$ $\pm 4.7$
$10^{18.8} - 10^{19.2}$	18.96	921	0.245	0.015	$\pm 0.031$ $\pm 0.036$	229.5	1.3	$\pm 6.5$ $\pm 4.5$
$> 10^{19.2}$	19.37	306	0.276	0.017	$\pm 0.032$ $\pm 0.042$	235.6	1.7	$\pm 6.6$ $\pm 4.5$

Table 6.1: R and L values obtained per energy bin, along with their statistical and systematic uncertainty

The values of both variables,  $R$  and  $L$ , as a function of  $\log_{10}[E/eV]$ , is shown in figures 6.1 and 6.2, with all uncertainties considered. The solid black vertical lines represent the statistical uncertainty, while the brackets depict the systematic one. In colored lines are the model expectations for proton (full lines) and iron showers (dashed lines).

The width,  $L$ , in data agrees well with the predictions of all models for the entire energy range considered, as said before. The energy evolution is similar to the predictions of model for a light component, however the systematic uncertainty is very large, and a mixed composition cannot be ruled out. The asymmetry,  $R$ , has an increase with energy not predicted by models. This is significant within the statistical uncertainty. However, the estimated systematic uncertainties are large, and the strength of this effect decreases when they are taking into consideration. A

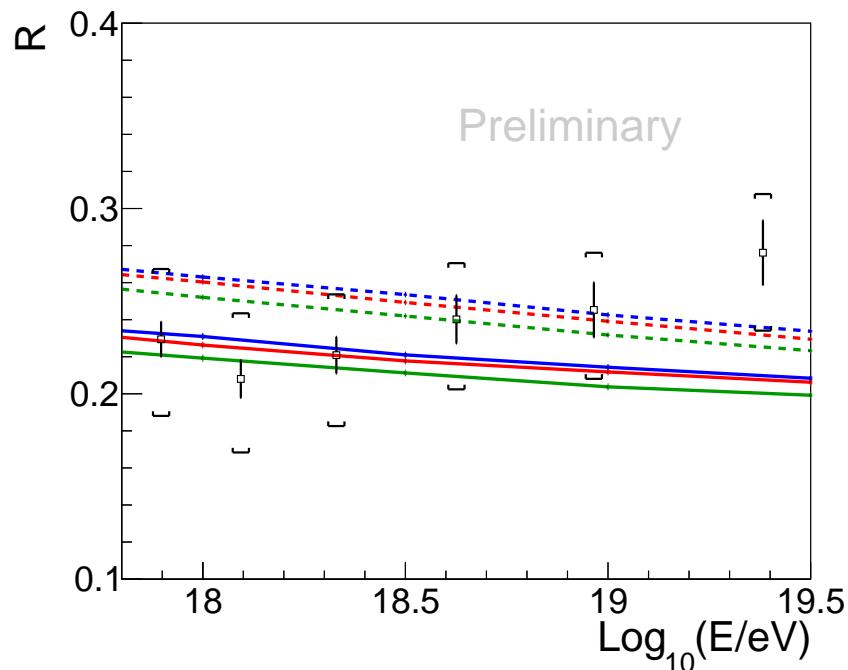


Figure 6.1:  $R$  as a function of energy Brackets represent systematic uncertainties

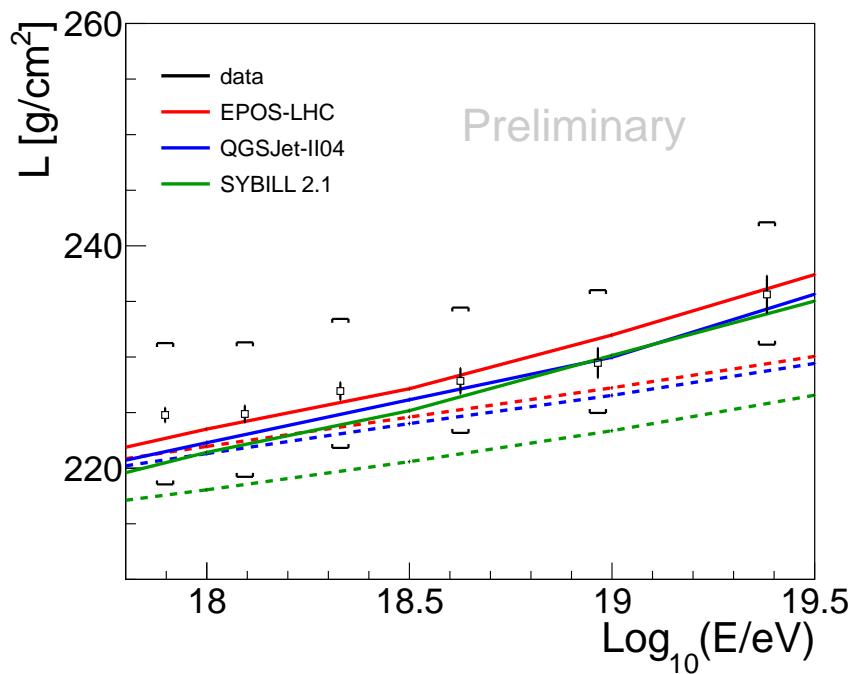
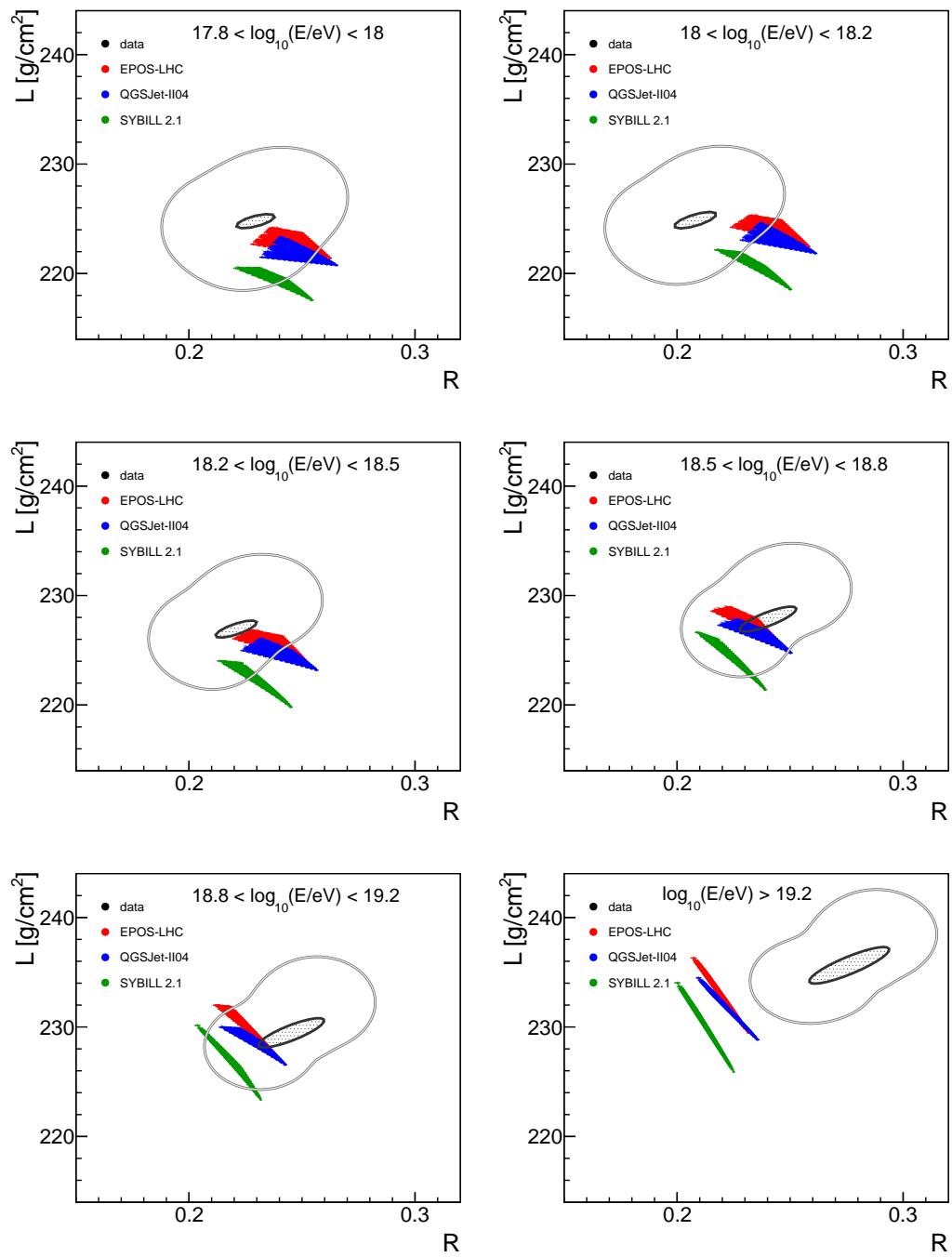


Figure 6.2:  $L$  as a function of energy Brackets represent systematic uncertainties

Figure 6.3:  $R$  vs  $L$  for all energies

more detailed analysis of this effect will be given ahead, in section 6.2.

The two variables come from the same fit and are not independent. In addition, at the highest energies  $L$  is within the predictions of models for proton primaries, while a much heavier composition is inferred from  $R$ . It is thus interesting to see the results on the (R-L) plane for a fixed energy. This makes data analysis easier in two ways:

- We can plot all combinations of primaries for each hadronic interaction model, which defines an area that is much smaller than the multiplication of its projection in each axis
- the error ellipsis constructed using the error matrix information gives a better estimation of the uncertainty distribution than the 1-dimensional error band.

In figure 6.3, the two dimensional plots for the measurements of  $R$  and  $L$  are drawn along with all the p, He, N and Fe combinations in the three main high energy hadronic models. Since p has a high  $L$  and low asymmetry, it is the top left point within a given model, while iron has a low  $L$  and high asymmetry, thus being on the bottom right. The statistical and systematic uncertainty ellipses are drawn on top. For the three lowest energy bins, the measured values are located more to the left, which corresponds to the lighter composition area. As the energy increases, it moves towards the heavier primaries, particularly at the last energy bin where it differs more from the models. Although it is compatible with data within  $1.5\sigma$ , this is a very interesting result, which will be analyzed further in the next sections.

## 6.2 Energy evolution of shower variables

The Pierre Auger Observatory currently measures several showers observables, which are sensitive to different shower components. These include variables from the electromagnetic longitudinal profile, from the LDF measured at ground and from the Muon Production Depth (MPD) profile. The usual way to extract the physical interpretation of the measurement is to compare it to simulated showers using different high-energy hadronic interaction models. This method has a drawback: it results in several independent analysis, and the full picture does not become immediately clear.

The question then becomes: how to extract the most information from all these variables simultaneously, since they can be correlated to each other in different de-

grees, both experimentally and in terms of the shower component they are most sensitive to? One of the ways to test the hadronic models w.r.t. their predicting accuracy is to use the lines given by their predictions for proton and iron primaries on each variable, and transform the measurement into the corresponding mass according to those lines. For this, the only thing we need is that the variables are approximately linear with  $\langle \ln A \rangle$ :

$$\langle \ln A \rangle = \frac{p - p_p}{p_{Fe} - p_p} \ln 56 \quad (6.1)$$

where  $p_p$  and  $p_{Fe}$  are the values of the observable  $p$  for proton and iron primaries, and we use the fact the for proton  $A = 1$ , so  $\ln A = 0$ . Then, we can use equation 6.1 to calculate  $\langle \ln A \rangle_i$  for each variable  $i$ .

Whenever such transformation from two observables  $p_1$  and  $p_2$  result in a different  $\langle \ln A \rangle_1 \neq \langle \ln A \rangle_2$ , it necessarily means that the phase-space of observables  $p_1$  and  $p_2$ , and in particular, some of the values  $p_{1p}$ ,  $p_{1Fe}$ ,  $p_{2p}$  or  $p_{2Fe}$  is wrongly predicted by the hadronic model under consideration. Figure 6.4 displays  $\langle \ln A \rangle$  calculated from  $\langle X_{\max} \rangle$ ,  $\langle X_{\max}^\mu \rangle$  and  $\langle \ln R_\mu \rangle$  as a function of the energy. There is an energy region around  $10^{19.4}$  eV ( $pp$  equivalent center of mass energy  $\sqrt{s} = 217$  TeV) where all shower quantities are measured, and which will be used all through this section as the reference point.

Table 6.2 displays the value of  $\langle \ln A \rangle$  at  $10^{19.4}$  eV (interpolated from a linear fit) for the published results of:  $\langle X_{\max} \rangle$  [90, 93],  $\sec \theta_{\max}$  [94],  $\langle X_{\max}^\mu \rangle$  [95],  $\langle \ln R_\mu \rangle$  [96].

While all these results are published by the Pierre Auger Collaboration, their joint interpretation in terms of  $\langle \ln A \rangle$  has never been made. These variables are, as we said, sensitive to different shower components.  $\langle X_{\max} \rangle$ , as the maximum of the longitudinal profile, is the de facto electromagnetic variable. It is controlled by the photon sub-showers that come from the  $\pi^0$  production in the first highest energy collisions. This sub-showers have virtually no input from the muonic component.  $\langle X_{\max}^\mu \rangle$  and  $\langle \ln R_\mu \rangle$ , on the other hand, can be considered almost purely "hadronic". Unlike the electromagnetic profile, they are constantly being influenced by the hadronic backbone of the air shower, and are driven by the muon production coming from hadronic interactions at different energy thresholds. They also are virtually un-affected by the electromagnetic sub-shower.

On the other hand,  $\sec \theta_{\max}$  is a hybrid variable, as the total signal measured by the tanks on the ground is sensitive to both e.m. particles and muons. Its measurement has been performed in two regions: one closer to the shower core  $r \in$

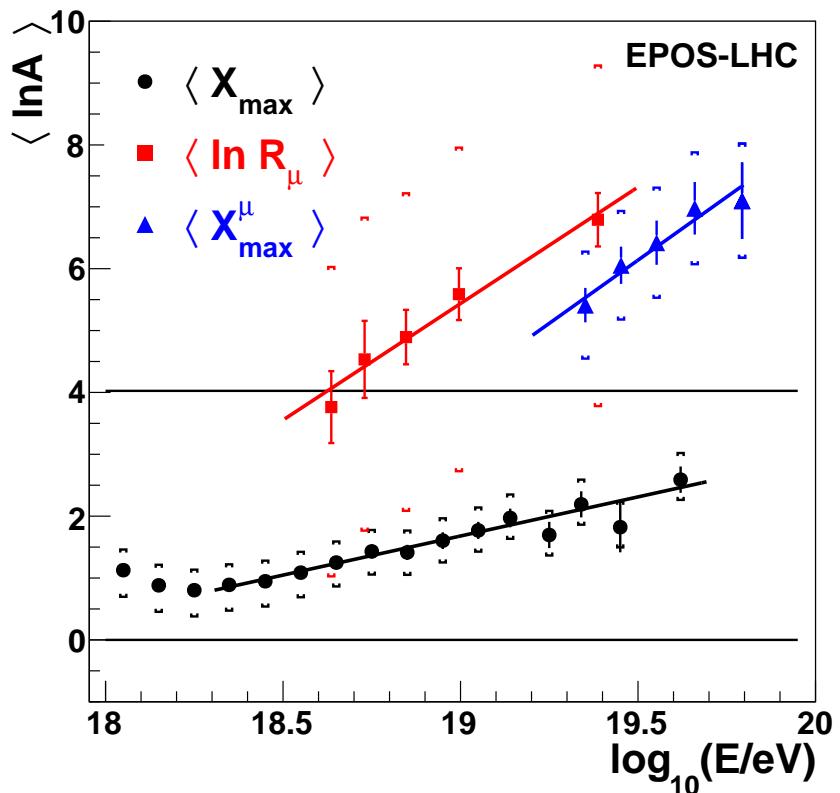
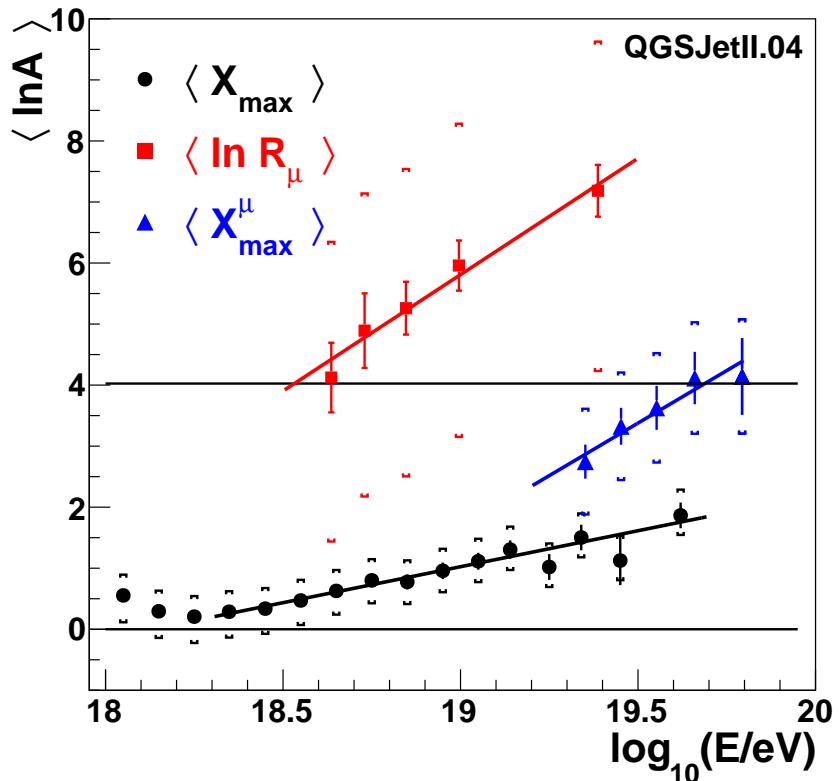


Figure 6.4:  $\langle \ln A \rangle$  vs  $\log_{10} E$  plot of the published results of  $\langle X_{\max} \rangle$  [90, 93],  $\langle X_{\max}^\mu \rangle$  [95],  $\langle \ln R_\mu \rangle$  [96] for the two post-LHC hadronic models, as labeled.

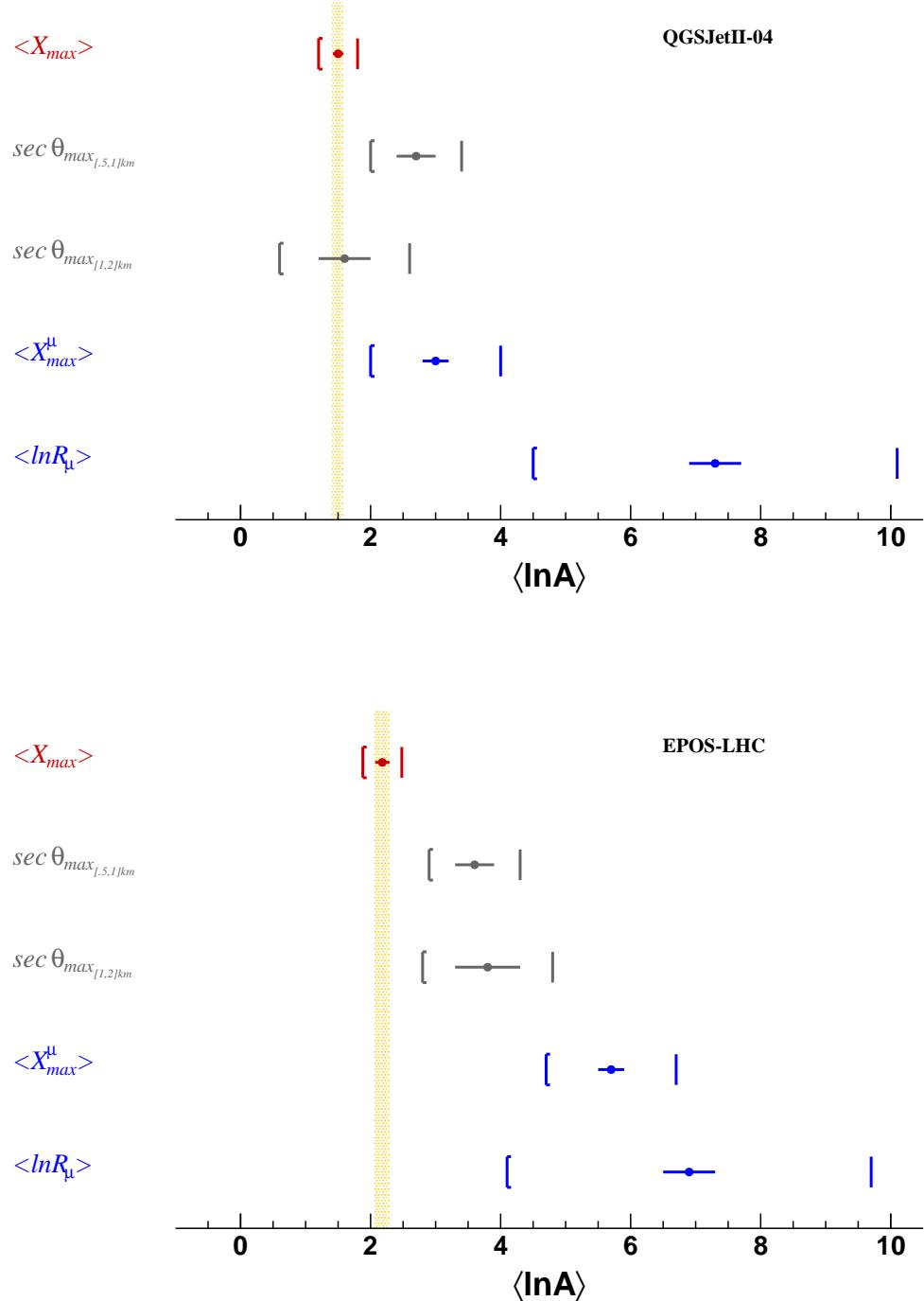


Figure 6.5: Interpolated values of  $\langle \ln A \rangle$  at  $10^{19.4}$  eV for QGSJET-II.04 (top) and EPOS-LHC (bottom). Calculations are based on the published results of  $\langle X_{max} \rangle$  [90, 93],  $\sec \theta_{max}$  [94],  $\langle X_{max}^\mu \rangle$  [95] and  $\langle \ln R_\mu \rangle$  [96]. The vertical yellow line represents the mass value inferred by each model within statistical uncertainty.

observable	QGSJET-II.04			EPOS-LHC		
	$\langle \ln A \rangle$	stat	syst	$\langle \ln A \rangle$	stat	syst
$\langle X_{\max} \rangle$	1.50	$\pm 0.08$	$\pm 0.3$	2.18	$\pm 0.11$	$\pm 0.3$
$r \in [0.5, 1.0] \text{ km}$ $\sec \theta_{\max}$	2.7	$\pm 0.3$	$\pm 0.7$	3.6	$\pm 0.3$	$\pm 0.7$
$r \in [1.0, 2.0] \text{ km}$ $\sec \theta_{\max}$	1.6	$\pm 0.4$	$\pm 1.0$	3.8	$\pm 0.5$	$\pm 1.0$
$\langle X_{\max}^{\mu} \rangle$	3.0	$\pm 0.2$	$\pm 1.0$	5.7	$\pm 0.2$	$\pm 1.0$
$\langle \ln R_{\mu} \rangle$	7.3	$\pm 0.4$	$\pm 2.8$	6.9	$\pm 0.4$	$\pm 2.8$

Table 6.2: Results of  $\langle \ln A \rangle$  interpolated from linear fit at  $10^{19.4}$  eV. Calculations are based on the published results of  $\langle X_{\max} \rangle$  [90, 93],  $\sec \theta_{\max}$  [94],  $\langle X_{\max}^{\mu} \rangle$  [95],  $\langle \ln R_{\mu} \rangle$  [96].

[500, 1000]m, and another far away  $r \in [1000, 2000]$ m, where the relative importance of the muonic component with respect to the electromagnetic one is larger.

Figure 6.5 presents the result of the transformation of these variables into  $\langle \ln A \rangle$  evaluated at  $10^{19.4}$  eV, where statistical and systematic uncertainties have been summed in quadrature. Observables linked to the hadronic/muonic cascade are coloured in blue and the one linked to the electromagnetic cascade is in red. Red and blue observables are independent of each other, while  $\sec \theta_{\max}$  depends on the overall muon content of the shower ( $\langle \ln R_{\mu} \rangle$ ), the arrival time-distribution of the muon content (which depends on  $\langle X_{\max}^{\mu} \rangle$ ) and the arrival time distribution of the electromagnetic content (which depends on  $\langle X_{\max} \rangle$ ), and is thus colored grey.

### 6.2.1 The $\langle \ln A \rangle$ derivative, $D_{10}\langle \ln A \rangle$

Not only is the mass inferred from shower observables through the hadronic models an interesting variable, but also its energy evolution. As discussed in the case of  $\langle \ln A \rangle$ , the value of all the moments derivatives should be consistently predicted by models. Derivatives contain additional independent information with respect to the absolute values of the moments themselves. They are also interesting from an experimental viewpoint: some of the systematic uncertainties cancel when the energy evolution is being computed, as the measurements in each energy bin are not independent from each other.

Following the same line of arguments used in the previous section, it is possible to calculate the derivative  $D_{10}\langle \ln A \rangle \equiv \frac{d\langle \ln A \rangle}{d\log_{10} E}$  from each one of the observable moments  $p$ . Note that a linear evolution  $p = a + b \log_{10} E$  does not necessarily transform into a linear evolution in  $\langle \ln A \rangle$  vs  $\log_{10} E$ , but:

$$\langle \ln A \rangle = \frac{(a - a_p) \ln 56}{(a_{Fe} - a_p) + (b_{Fe} - b_p) \log_{10} E} + \frac{(b - b_p) \ln 56}{(a_{Fe} - a_p) + (b_{Fe} - b_p) \log_{10} E} \log_{10} E. \quad (6.2)$$

In fact, only when  $b_{Fe} = b_p$ , straight lines in the  $(p, \log_{10} E)$ -space transform into straight lines in  $(\langle \ln A \rangle, \log_{10} E)$ -space. In general we must speak of  $\langle D_{10}\langle \ln A \rangle \rangle$  within a given range of energies, although we will maintain the simpler  $D_{10}\langle \ln A \rangle$ .

notation in general. Figure 6.6 displays a plot generated by a toy-MC where this effect has been exaggerated.

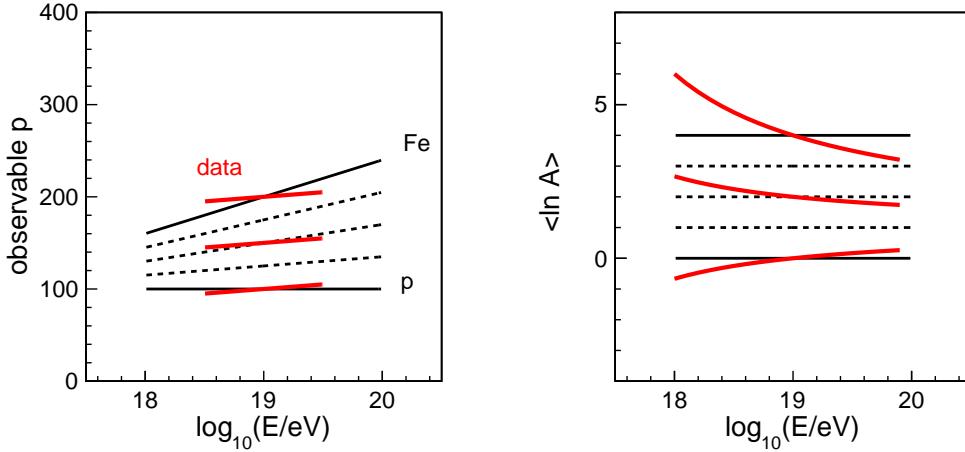


Figure 6.6: Toy example of the transformation of an observable that evolves as  $p = a + b \log_{10} E$  (left plot) into  $\langle \ln A \rangle (\log_{10} E)$  (right plot). It can be seen that a constant shift  $k$  as  $p' = a + b \log_{10} E \pm k$  introduces a non-linear change in  $\langle \ln A \rangle$  vs  $\log_{10} E$  evolution and a change in the overall average derivative  $\langle D_{10} \langle \ln A \rangle \rangle$ .

Note also that if  $b_{Fe} \neq b_p$ , a constant shift  $p'(\log_{10} E) = p(\log_{10} E) + k$  introduces a change in  $\langle D_{10} \langle \ln A \rangle \rangle$ . Some observables, like  $\langle X_{\max}^\mu \rangle$  and  $\langle \ln R_\mu \rangle$ , are not within the p-Fe reference lines. Since the distance between primaries is, for some variables, slightly energy dependent, a constant change in the overall normalization of the variable  $p$  at energies  $E_1$  and  $E_2$  can lead to different changes in the calculated  $\langle \ln A \rangle$  for the two energies. More specifically, for a perfectly parallel proton and iron reference line, if a variable  $p(E_1)$  translates to mass  $\ln A(E_1)$  and  $p(E_1) + \Delta_p$  to  $\ln A(E_1) + \Delta_A$ , then  $p(E_2) + \Delta_p$  must translate to  $\ln A(E_2) + \Delta_A$ . If they are not parallel, the same  $\Delta_p$  will correspond to different  $\Delta_A$  values at the two energies. So, we have calculated an additional contribution to the systematic uncertainty of  $D_{10} \langle \ln A \rangle$  labelled "tr" (from translation), corresponding to the changes introduced when a constant shift of the observables is applied over the energy range under consideration.

observable	QGSJET-II.04				EPOS-LHC			
	$D_{10} \langle \ln A \rangle$	stat	syst	tr	$D_{10} \langle \ln A \rangle$	stat	syst	tr
$\langle X_{\max} \rangle$	1.18	$\pm 0.10$	$^{+0.24}_{-0.08}$	$\pm 0.02$	1.26	$\pm 0.11$	$^{+0.24}_{-0.08}$	$\pm 0.02$
$r \in [0.5, 1.0] \text{ km sec } \theta_{\max}$	1.0	$\pm 0.5$	$\pm 0.7$	$\pm 0.1$	0.9	$\pm 0.5$	$\pm 0.7$	$\pm 0.7$
$r \in [1.0, 2.0] \text{ km sec } \theta_{\max}$	1.6	$\pm 0.8$	$\pm 0.8$	$\pm 0.5$	2.4	$\pm 0.9$	$\pm 1.0$	$\pm 0.7$
$\langle X_{\max}^\mu \rangle$	3.4	$\pm 1.0$	$\pm 1.0$	$\pm 0.6$	4.1	$\pm 1.0$	$\pm 1.0$	$\pm 0.6$
$\langle \ln R_\mu \rangle$	3.8	$\pm 0.8$	$\pm 0.8$	$\pm 0.6$	3.8	$\pm 0.9$	$\pm 0.8$	$\pm 0.6$

Table 6.3: Results of  $D_{10} \langle \ln A \rangle$ . Calculations are based on the published results of  $\langle X_{\max} \rangle$  [90, 93],  $\text{sec } \theta_{\max}$  [94],  $\langle X_{\max}^\mu \rangle$  [95],  $\langle \ln R_\mu \rangle$  [96].

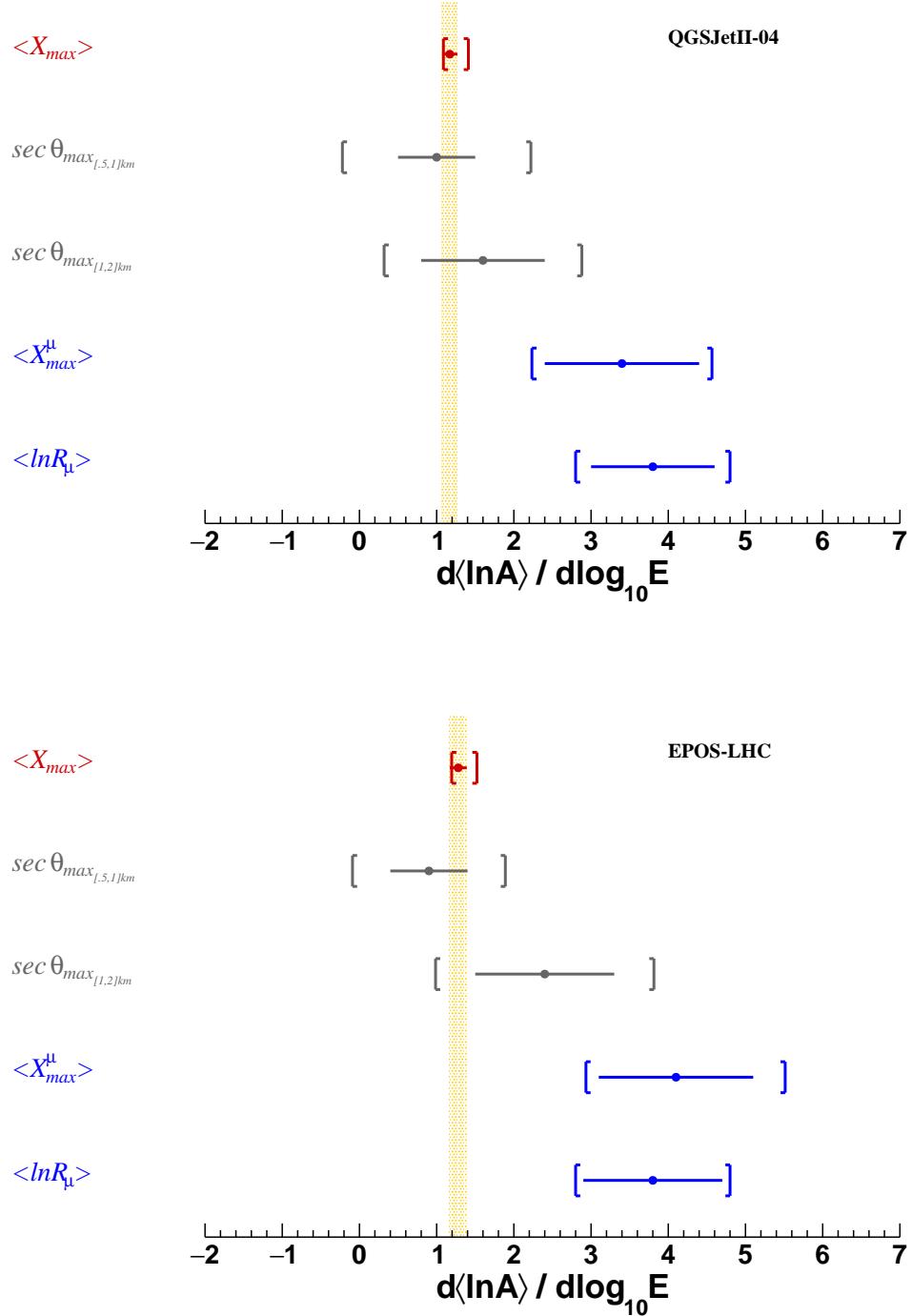


Figure 6.7:  $D_{10}\langle \ln A \rangle$  for QGSJET-II.04 (top) and EPOS-LHC (bottom). Calculations are based on the published results of  $\langle X_{max} \rangle$  [90, 93],  $\sec \theta_{max}$  [94],  $\langle X_{max}^\mu \rangle$  [95] and  $\langle \ln R_\mu \rangle$  [96]. The vertical yellow line in represents the  $D_{10}\langle \ln A \rangle$  value inferred by each model within statistical uncertainty.

### 6.2.1.1 Systematic uncertainties

The correct calculation of the systematic uncertainties of  $D_{10}\langle \ln A \rangle$  is of utmost importance in order to extract valid conclusions. We have followed a conservative approach, and calculated separately two different contributions.

For the first contribution, labeled as "*syst*", we have used the systematics which were available in the Auger publications. In the case of  $\langle X_{\max} \rangle$  and  $\langle \ln R_\mu \rangle$  we have used the systematics for  $D_{10}\langle \ln A \rangle$  and  $D_{10}\langle \ln R_\mu \rangle$  reported in [90] and [96] respectively. In these publications it was accounted that some contributions are correlated bin-to-bin, for instance the energy calibration systematics. For the remaining variables, systematics were considered to be independent bin-to-bin and thus having the worst possible scenario for the  $D_{10}\langle \ln A \rangle$  uncertainty. These calculations should be revisited in future studies, resulting in a likely reduction of the systematics.

Under the label *th* we have calculated a systematic uncertainty resulting from overall constant shift variations  $p'(\log_{10} E) = p(\log_{10} E) \pm k$ . The parameter  $k$  was chosen to be the p-Fe separation at  $10^{19.4}$  eV, ( $k = p_{Fe}(19.4) - p_p(19.4)$ ). The results of  $D_{10}\langle \ln A \rangle$  for all our observables moments are shown in table 6.3 and displayed in figure 6.7. The first immediate observation is that none of the hadronic interaction models provide a consistent picture, similarly to what happens with  $\langle \ln A \rangle$ . Moreover, it can be observed that observables related to the muonic component have a significant larger derivative.

### 6.2.2 Average shape: $\langle \ln A \rangle$ at $10^{19.4}$ eV and $D_{10}\langle \ln A \rangle$

Previously in this section, we decided to work only with variables already published by the Pierre Auger Collaboration for the calculation of  $\langle \ln A \rangle$ . This thesis, however, is mainly concerned with two new variables,  $R$  and  $L$ , which also carry information about shower development. The interpretation of these two variables within both models will be discussed, and compared with the expectations given their sensitivity to the respective shower component. Only bins above  $10^{18.33}$  eV will be used for  $R$  and  $L$ , as this is the energy for which a break is seen in the evolution of  $X_{\max}$ .

In figures 6.8 and 6.9, the transformation of the  $R$  and  $L$  measurements into  $\langle \ln A \rangle$  is shown as a function of energy.

Despite the large uncertainties, mainly at lower energies in  $L$ , as the proton and iron values predicted by models are very close, two things can be seen: the value is larger in  $R$  at around  $10^{19.4}$  than in  $L$ , and the energy evolution is also much steeper in  $R$ . Also, the transformed data points are compatible with a straight line, so this was chosen as the fitting function. The value of  $\langle \ln A \rangle$  at  $10^{19.4}$  eV, as well as its statistical uncertainty, is the result of this fit at the aforementioned energy. The systematic uncertainty in  $\langle \ln A \rangle$  is given by the transformation of the total systematic uncertainty in  $R$  or  $L$  in the last energy bin to  $\langle \ln A \rangle$ . The values for both variables and hadronic models are compiled in table 6.4.

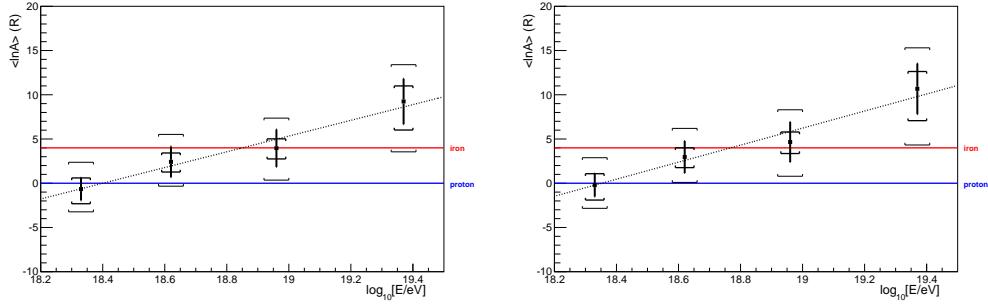


Figure 6.8: Transformation of  $R$  into  $\langle \ln A \rangle$  for the four highest energy bins for the QGSJetII-04 (left) and EPOS-LHC (right) model. The coloured lines represent the mass values of protons (blue) and iron (red). Data points are shown along with the statistical uncertainties (black lines), total systematic uncertainties (thin brackets) and energy dependent systematic uncertainties, calculated w.r.t. the  $10^{18.75}$  eV reference (thick brackets).

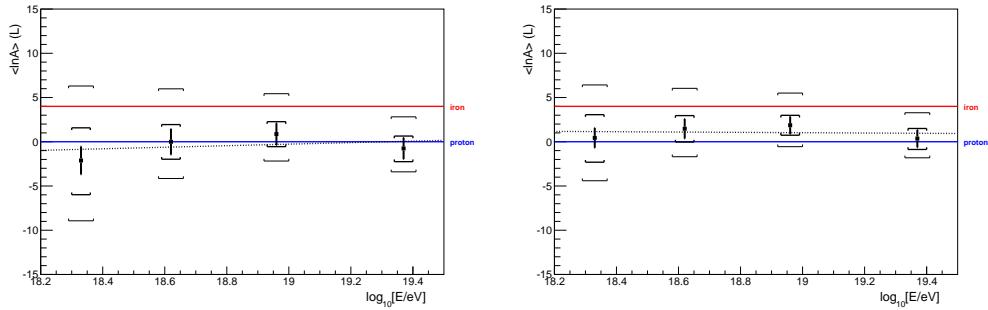


Figure 6.9: Transformation of  $L$  into  $\langle \ln A \rangle$  for the four highest energy bins for the QGSJetII-04 (left) and EPOS-LHC (right) model. The coloured lines represent the mass values of protons (blue) and iron (red). Data points are shown along with the statistical uncertainties (black lines), total systematic uncertainties (thin brackets) and energy dependent systematic uncertainties, calculated w.r.t. the  $10^{18.75}$  eV reference (thick brackets).

The transformation of the total systematic uncertainty in each variable into  $\langle \ln A \rangle$  is straightforward - the values in table 6.1 (data+syst. and data-syst.) are all transformed according to 6.1. However, the absolute value of the systematic uncertainty in each energy is very correlated with the uncertainty on the contiguous ones. So, the energy dependent systematic uncertainty in each point has to be calculated. In chapter 5 the different contributions to the uncertainty on the measurement of the average shape were calculated. For each of these contributions the reference is taken as  $E = 10^{18.75}$  eV, and the systematic uncertainty in each point is calculated with respect to that point, i.e.  $\text{syst}_E(E) = |\text{syst}(E) - \text{syst}(10^{18.75}\text{eV})|$ . So, the energy dependent systematic uncertainty,  $\text{syst}_E$ , is the maximal difference in the systematic uncertainty of a measurement at two different energies. It is minimal at the reference point, and increases as we get farther way from it. For two of the components, detector age and season of the year, no energy dependence could be

observable	QGSJET-II.04			EPOS-LHC		
	$\langle \ln A \rangle$	stat	syst	$\langle \ln A \rangle$	stat	syst
$R$	8.9	$\pm 2.1$	$^{+4.1}_{-5.7}$	10.1	$\pm 2.3$	$^{+4.6}_{-6.3}$
$L$	0.1	$\pm 1.1$	$^{+2.9}_{-2.2}$	1.0	$\pm 0.9$	$^{+2.5}_{-1.9}$

Table 6.4: Results of  $\langle \ln A \rangle$  interpolated from linear fit at  $10^{19.4}$  eV for the  $R$  and  $L$  variables.

observable	QGSJET-II.04				EPOS-LHC			
	$D_{10}\langle \ln A \rangle$	stat	syst	tr	$D_{10}\langle \ln A \rangle$	stat	syst	tr
$R$	8.9	$\pm 2.5$	$^{+2.3}_{-3.4}$	$\pm 0.7$	9.6	$\pm 2.7$	$^{+2.4}_{-3.7}$	$\pm 1.1$
$L$	0.9	$\pm 1.8$	$^{+2.7}_{-3.9}$	$\pm 3.9$	-0.2	$\pm 1.4$	$^{+2.1}_{-2.9}$	$\pm 3.3$

Table 6.5: Results of  $D_{10}\langle \ln A \rangle$  for the  $R$  and  $L$  variables

calculated due to lack of statistics, so these are added at all energies. From the  $\langle \ln A \rangle$  values with their respective energy dependent systematic uncertainty, the energy evolution corresponding to each variable is calculated. This is represented by the thick error brackets in figures 6.9 and 6.8.

Finally, the error coming from the fact that the slope in terms of  $\langle \ln A \rangle$  depends on the absolute value at some reference point for the variables is calculated. For that end, we shift data by  $1\sigma$  of the total uncertainty and fit it to a straight line again. In case of  $R$  the effect is small, as proton and iron lines are almost parallel, but in  $L$  it is the largest uncertainty in this analysis. This results from the fact that values of proton and iron are virtually the same at  $10^{18}$  eV, which means any small change in measured  $L$  leads to a vastly different composition interpretation.

### 6.2.3 Discussion

It is long known that simulations do not reproduce well the muon number and muon production depth expected from the average composition  $\langle \ln A \rangle$  as suggested by  $\langle X_{\max} \rangle$  data. Moreover, the new results presented in figure 6.7 for the derivative  $D_{10}\langle \ln A \rangle$  unveils two separate groups of shower observables, where the evolution with the energy tend to diverge from each other. Hence, they cannot be explained in terms of a simple constant shift in  $\langle \ln A \rangle$  by any of the post-LHC hadronic interaction models.

According to the nature of the observables, one can make the following considerations:

- “Pure” observables:

The depth of the shower maximum,  $\langle X_{\max} \rangle$  is sensitive to the electromagnetic shower component and measured with FD, while  $\langle \ln R_\mu \rangle$  and  $\langle X_{\max}^\mu \rangle$  are measured by SD, and are sensitive to the hadronic component of the shower through muons. The evolution of the shower observables related to the electromagnetic component is significantly different with respect to the ones associated to the hadronic cascade. The latter present a faster evolution with energy towards - what seems like - heavier elements (or "heavy-like" behaviour), i.e. a larger  $D_{10}\langle \ln A \rangle$ .

- "Mixed" observables:

The observable  $\sec \theta_{\max}$  (SD-based) is sensitive to both the electromagnetic and muonic shower components. The values of  $\langle \ln A \rangle$  and  $D_{10}\langle \ln A \rangle$  deduced from it should be somewhere between the electromagnetic quantities and the hadronic ones. Although the error bars are large,  $\langle \ln A \rangle$  and  $D_{10}\langle \ln A \rangle$  confirm these expectations independently of the chosen hadronic interaction model. The obtained result in terms of  $\langle \ln A \rangle$  for both radius cuts, are well within the uncertainties. For  $D_{10}\langle \ln A \rangle$ , the  $\sec \theta_{\max}$  observable with the highest mean value is the one with more muons, at  $r \in [1000, 2000]\text{m}$ .

- Independent FD observables:

Finally, in this work we present the first measurement of the average electromagnetic shower profile. The shape variables of this profile,  $L$  and  $R$ , are known to be sensitive to the primary mass composition and to have a nearly linear relation with  $\langle \ln A \rangle$  [98]. Although uncertainties are large, results are tantalizing, with larger values for  $\langle \ln A \rangle$  and  $D_{10}\langle \ln A \rangle$  deduced from  $R$  when compared with those from  $L$ . The connection of these two variables with the shower components is not trivial, and will be explained in more detail later, but it is known that the width  $L$  is related to the shower energy carried by the electromagnetic component. On the other hand,  $R$  traces the start-up phase of the electromagnetic shower, it is thus natural to be more sensitive to the hadronic interactions before the full decoupling of the hadronic and electromagnetic cascade. In fact, CONEX simulations show that  $R$  has a stronger correlation with the number of muons than  $L$ . The results obtained with the shower shape variables favours (in an independent manner) a faster evolution with energy for hadronic related quantities  $R$ , with respect to the electromagnetic ones,  $L$ .

It was found that the observables related with the hadronic shower component evolve faster with energy than those related with the electromagnetic shower component. A naive direct interpretation would suggest that either models are predicting a growth of the hadronic cascade which is too slow, or that the electromagnetic component is predicted to evolve too fast. To restore the consistency it might be necessary to reduce the rate of energy transfer to the electromagnetic channel in the start-up of the shower (for instance, though  $\pi^0$  suppression in the high energy interactions[101]). The ultimate reasons behind these changes span from just a simple reassessments of the multi-particle production characteristics to new phenomena at the highest energies.

Since the "muon deficit" problem was discovered in Auger, a group of hypotheses were introduced under the form of a constant scaling that could explain the observed mismatch. These constant scalings, which we will call generically *k-hypotheses*, could be explained by changes on the shower physics, but also by a shift in the FD energy scale or some unknown source of systematic related with the SD.

At  $10^{19} \text{ eV}$ , the ratio of measured  $\langle R_\mu \rangle$  with respect to the  $\langle X_{\max} \rangle$ -inferred

value  $\langle R_\mu \rangle$  is  $k=1.3$  according to EPOS-LHC. This value is compatible with the excess found in  $\sigma[R_\mu]$  [97]. Note that in principle, the fluctuations in the number of muons do not have necessarily to be increased by the same factor as the number of muons itself. Among the phenomena that could increase the muon signal and its fluctuations with the same constant factor, is for instance, a miss-calibration of the Cherenkov-tank response by a factor  $k$ . Nevertheless, the  $k$ -hypothesis  $R'_\mu = kR_\mu$  **cannot explain the derivative mismatch** since

$$D_{10}\langle \ln kR_\mu \rangle = D_{10}\langle \ln k + \ln R_\mu \rangle = D_{10}\langle \ln R_\mu \rangle \quad (6.3)$$

provided that  $k$  is constant with energy,  $D_{10}k = 0$ .

At the same time, the  $k$ -hypothesis  $E' = kE$  (miss-calibration of the FD energy scale by a constant factor) does not fix the problem of the diverging derivatives either, due to;

$$d \log_{10} kE = d(\log_{10} k + \log_{10} E) = d \log_{10} E \quad (6.4)$$

leaving all log-derivatives  $D_{10}p$  and  $D_{10}\langle \ln A \rangle$  unaltered. It should be noted here that, although it is not the subject of this work, the RMS of  $X_{\max}$  also has a tension for some models with the  $\langle X_{\max} \rangle$  value, which also cannot be fixed by this  $k$ -hypothesis.

#### 6.2.4 Summary

In this section we have collected and computed  $\langle \ln A \rangle$  and  $D_{10}\langle \ln A \rangle$  at  $E = 10^{19.4}$  eV ( $\sqrt{s} = 217$  TeV for a proton primary) for shower quantities **already published by Auger**. Besides the well known differences in  $\langle \ln A \rangle$ , we found that observables linked to the electromagnetic component have a significant smaller derivative than those linked to the hadronic component, which cannot be fixed by a simple overall shift of the observables or the energy scale.

The values, averaged with the two post-LHC hadronic models are:

- **Pure electromagnetic cascade** (Result of  $\langle X_{\max} \rangle$ ):

$$D_{10}\langle \ln A \rangle = 1.22 \pm 0.11 \text{ (stat)} \pm 0.24 \text{ (syst}_{total}\text{)} \pm 0.04 \text{ (models)}$$

- **Pure hadronic cascade** (Weighted average of  $\langle X_{\max}^\mu \rangle$  and  $\langle \ln R_\mu \rangle$  results):

$$D_{10}\langle \ln A \rangle = 3.7 \pm 0.7 \text{ (stat)} \pm 0.8 \text{ (syst}_{total}\text{)} \pm 0.2 \text{ (models)}$$

This mismatch could indicate an unbalance in models of the energy flow between the electromagnetic and hadronic cascade at the start-up of the shower.

The results from the average shape also support these claims, with the derivative for  $L$  being around 0-1 depending on the model, so more "electromagnetic", and  $D_{10}\langle \ln A \rangle \approx 9$  for  $R$ , so more "hadronic". For the  $R$  derivative, the downward uncertainty (systematic + theoretical) is around 4 for both models, so there is tension between the energy evolution and the one inferred from  $X_{\max}$ . This supports the findings of the other variables, and is further indication of a mismatch in the description of both shower components in current high-energy hadronic models.

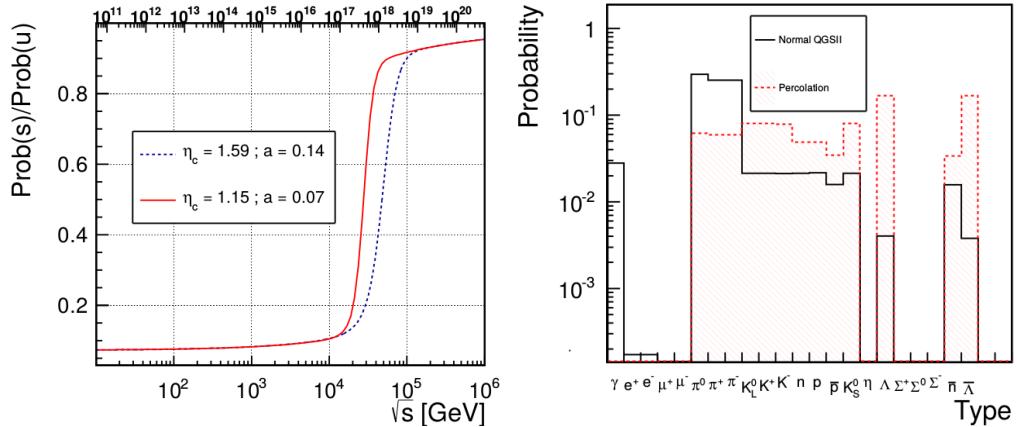


Figure 6.10: Left: Probability of producing a  $s\bar{s}$  relatively to  $u\bar{u}$  pair as a function of  $s$  in the QGSJet-II model accounting for string percolation effects. The red line is for a proton with an uniform density profile while the blue line is assuming a Gaussian profile. The upper scale is the equivalent energy of a proton in the laboratory frame of the Earth. Right: Probability of particle production as a function of the type of particle as predicted by the QGSJet-II.03 model without percolation and after enhancing the probability of strange quark production.

In fact, looking at the  $\langle \ln A \rangle$  values as a function of the energy (figure 6.4), it can be seen that the mismatch at  $10^{19.4}$  eV is correlated with the largest  $D_{10}\langle \ln A \rangle$  value: the extrapolation of the lines seems to indicate that at an energy around  $10^{18}$  eV, variables were compatible and "light", and the large e.m.-muon component mismatch increases from around  $10^{18}$  to  $10^{19.4}$  eV.

## 6.3 Alternative Interaction Models and New Physics

Mainly motivated by the muon deficit in current hadronic models w.r.t. Auger data, several changes to interaction properties at the highest energies have been proposed to increase the total muon content predicted. In this section we will describe some of these models, and especially the processes by which they alter the hadronic component of the cascade. Note that this list is not comprehensive, but illustrates the different ways strong changes in interaction properties can alter the profile shape, as well as giving a rough idea on the amplitude we can expect from these changes.

### 6.3.1 String Percolation Model

In [87], the hypothesis of a sudden increase in parton density at high energies (around  $10^{18.5}$  eV) and its effects on air shower development were investigated. Here, we will give a brief description of this work and show its result for the average shape.

In the string model [89], color strings are formed in high energy collisions. In the transverse plane of a collision (with an interaction area defined by the effective radius  $R$ ) the strings appear as disks with a characteristic transverse radius  $r_0$ . The

relevant parameter in the String Percolation Model (SPM) is the so-called transverse string density  $\eta \equiv (r_0/R)^2 N_s$ , where  $N_s$  is the average number of strings formed in the collision. Large values of  $\eta$  means strings largely overlap, while  $\eta \ll 1$  corresponds to a system with mostly isolated strings. For sufficiently large  $\eta$  values strings may percolate (fuse). The probability for this to happen is controlled by the percolation function [88]

$$f(\eta) = \frac{1}{(1 + e^{-(\eta - \eta_c)/a})} \quad (6.5)$$

where  $\eta_c$  is the critical transverse string density, and  $a$  the parameter controlling the slope of  $f(\eta)$  at the transition point.

There are two main effects of string percolation in air shower development:

- when string percolation is effective the string tension increases due to the presence of higher color charges. As a consequence the probability of producing a  $s\bar{s}$  pair relative to that of producing a  $u\bar{u}$  pair ( $P_{s/u}$ ) increases. In the usual hadronic interaction models this quantity is of the order of 10%. With string percolation it can rise up to almost one, with its energy behavior dependent on the model parameters.
- as strings are stretched in rapidity, energy conservation implies a natural reduction of the hadronic interaction multiplicity. Around  $10^{19}$  eV, string percolation predicts a multiplicity less than half of that in normal models, and with correspondingly larger average particle energy.

These changes were implemented in CONEX shower simulations using QGSJetII-03 as a baseline model. As this process only kicks in above the LHC energy, there is some freedom in the choice of parameters in equation 6.5, although the rise in  $P_{s/u}$  is constrained up to 7 TeV. In [87], two extreme cases were considered: a uniform proton density profile ( $\eta_c = 1.15$  and  $a = 0.07$ ) and a Gaussian profile ( $\eta_c = 1.59$  and  $a = 0.14$ ).

String percolation increases the number of strange mesons being produced, in particular kaons, as well as the production of baryons. Particles directly responsible for the development of electromagnetic shower development ( $\gamma$  and  $\pi^0$ ) are suppressed. In standard models, an high energy hadronic interaction yields around 90% pions (evenly divided, giving 60%  $\pi^\pm$  and 30%  $\pi^0$ ). With string percolation, the pool of particles to evenly divide the energy is larger ( $p$ ,  $n$ ,  $\Lambda$  and their antiparticles,  $K$ ,  $\pi$ ) so we end up with slightly less than 10% of each type of pions (20%  $\pi^\pm$  and 10%  $\pi^0$ ).

### 6.3.1.1 Impact on air shower variables

The largest difference w.r.t. standard models is an increase in the number of muons reaching the ground of around 40% (slightly larger in the uniform proton density profile model than in the gaussian one). This is a direct consequence of the decrease in the feeding of the electromagnetic shower in the first interactions. This effect

is of the order of the muon excess observed in Auger compared to those predicted by standard models for an intermediate composition. Also, both the electromagnetic and muon production maximum depth increase. The first is the result of the smaller percentage of energy that goes into the electromagnetic shower in the first interactions, and the second is also expected from the larger energy of the produced muons. As both maximums are measured in Auger between proton and iron predictions, increasing the maximum of proton is not going to make this model incompatible with data. However, assuming this model would result in a heavier composition estimation at the highest energies.

Finally, we predict how string percolation affects the average shape. The decrease in multiplicity at the highest energies which leads to deeper showers should also result in wider profiles, i.e., larger L. The asymmetry, R, is more difficult to predict. From the argument presented above, it is clear the rise in the profile is going to be slower. However, the asymmetry depends on the interplay between the speed of shower development before and after the maximum. In figure 6.11, the difference in normalized energy deposit between the showers with and without string percolation is plotted. The profiles falls more slowly around the maximum in the string percolation model, but do so almost symmetrically - very small change in R. The difference in values of fitted L and R in the string percolation model compared to the standard QGSJetII-03 model is shown in figure 6.12. The asymmetry is indeed unchanged, but the width is much larger, between 7-11 g/cm<sup>2</sup> depending on the model parameters. This is an interesting motivation for measuring the average profiles: standard models can be tweaked in ways that make them compatible with data in some variables (small changes in  $X_{max}$  for example) while having measurable changes in shower width.

### 6.3.2 Chiral Symmetry Restoration (CSR) Model

Here we will describe the CSR model [102] and its implication in proton interactions and subsequent shower development at the highest energies.

High energy interactions dictate the energy partition between the electromagnetic and hadronic components of any given atmospheric shower, with the energy in the hadronic shower governing the number of muons at ground. So, increasing the total number of muons implies increasing the fraction of energy in the hadronic component of the shower. Defining  $N_{gen}$  to be the number of generations required for most pions to have energies below 100 GeV (approximately the energy below which pions decay before interacting), and  $f_{EM}$  to be the average energy fraction in EM particles, the hadronic shower carries a fraction  $f_{had} \approx (1 - f_{EM})^{N_{gen}}$  of the total primary UHECR energy. Evidently, reducing either  $N_{gen}$  or  $f_{EM}$  increases  $f_{had}$  and hence the number of muons.  $N_{gen}$  is hard to reduce without making  $X_{max}$  incompatible with data, as it implies a strong reduction in multiplicity. This leaves the option of reducing the energy fraction carried by electromagnetically decaying particles, which is also non-trivial from a theoretical point of view.

At accelerator experiments, e.g., central region of high energy collisions, as well

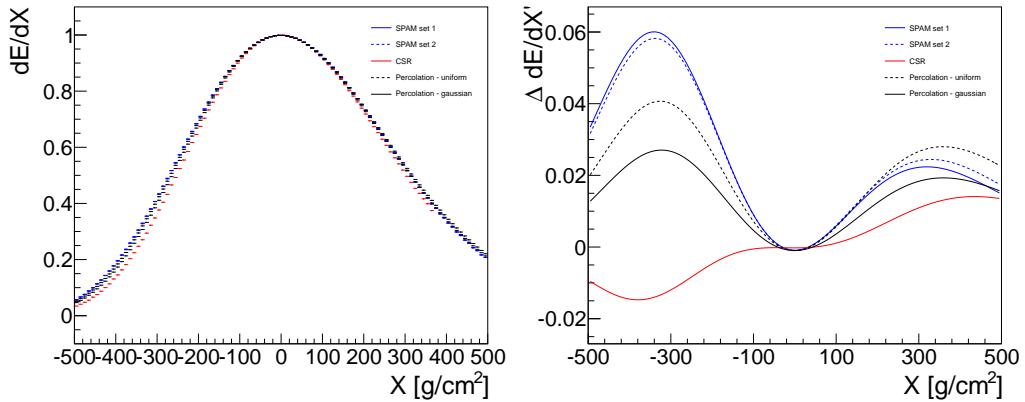


Figure 6.11: Left: Longitudinal profiles after translation by  $X_{\max}$  and normalization for all alternative models described in the text. Right: Residuals in normalized energy deposit between the alterntive hadronic interaction model shown in legend and QGSJetII-04. Note that the residuals, by definition, go to zero at the origin as all profiles are normalized to 1 and at the tails where all profiles go to 0

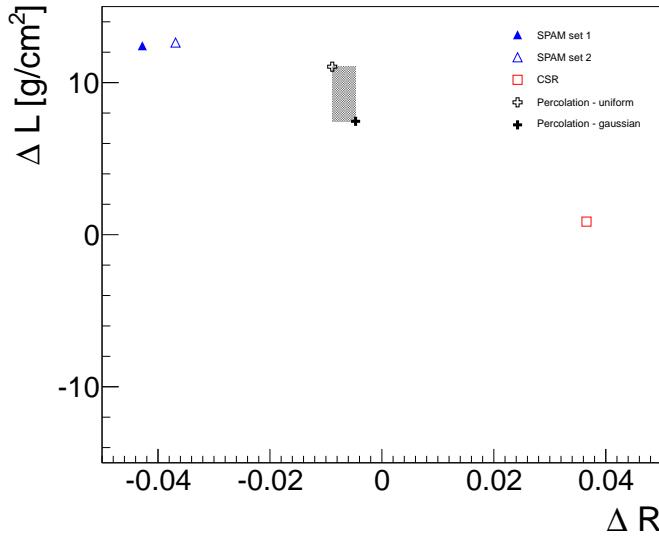


Figure 6.12: Difference in R and L for  $E = 10^{19.5}$  eV between the alternative models described and the ones they used as baseline. This means, for the X axis: SPAM model - QGSJetII-04, CSR - EPOS1.99 and Percolation - QGSJetII-03, and corresponding for L in the Y axis.

as the final states of quark-gluon-plasma probed at ALICE, particle ratios for non-leading particle types are to a good approximation universal. The average  $f(\pi^0)$  in  $\pi^\pm$ -air collisions above  $10^{16}$  eV is around 0.25 for QGSJet03-II and 0.20 for EPOS-1.99 and Sibyll 2.1, with an additional 0.04 contribution to  $f_{EM}$  coming from  $\eta$  and  $\eta^0$ 's. The lower  $f_{EM}$  of EPOS compared to QGSJet is the reason of the muon discrepancy between the models.

Chiral Symmetry is a near-exact symmetry of the QCD Lagrangian, because the quark masses ( $m_u < 5$ ,  $m_d < 10$  MeV) are very small compared to the QCD scale ( $\approx 100\text{-}200$  MeV). Chiral symmetry is broken spontaneously in the normal, low temperature vacuum, due to QCD interactions which reduce the vacuum energy in the presence of a light quark-antiquark condensate:  $\langle q\bar{q} \rangle \approx 200$  MeV. When Chiral Symmetry is spontaneously broken, Goldstone's theorem requires the existence of a light "Goldstone boson", which is the reason for the anomalously low pion mass of 135 MeV. This is why pions are much more abundant than other hadrons. Above the Chiral Symmetry Restoration (CSR) temperature, Chiral Symmetry can be realized either by massless fermions or parity doublets, and mesons have no reason to be light - or even to be bound. There are several possible mechanisms by which CSR may suppress pion production relative to baryon-antibaryon production, as theoretical studies to date have addressed static properties of the CSR phase rather than the highly non-equilibrium conditions produced in a UHE collision. So, a phenomenological approach was used and a model of CSR in UHE collisions based on analogies with other systems was developed.

The CSR phase is probably more readily created in central than in peripheral collisions, so an event generator (EPOS) was used to simulate a realistic distribution of centralities, for which the elasticity of the generated event was taken to be a proxy. When EPOS creates an event whose elasticity is above some threshold value that event is simply used as-is, while if the elasticity is below the threshold, an event in the CSR phase is created as discussed below. The threshold for CSR production was taken to be  $10^{17}$  eV and the elasticity threshold chosen so that 50% of the events are in the CSR phase at  $10^{17}$  eV, increasing linearly in  $\log_{10}[E]$  to 100% at  $10^{19.5}$  eV. The final states were produced by the EPOS model, as the CSR final states are expected to be similar to those in ordinary non-diffractive events. However, as meson content is suppressed and overall multiplicity can be modified, a fraction of mesons is converted to baryons and anti-baryons and the average multiplicity is controlled by imposing a cap on the elasticity of CSR-phase events: a lower elasticity cap results in a higher average multiplicity and vice versa. Finally, since the cross-section is unconstrained except by the UHECR data that is being fitted, a more rapid cross-section increase than the default in EPOS is allowed. This is not a consequence of the CSR effect but an independent parameter on the model. For simplicity, the toy model implemented uses a very crude scheme to decide whether a particular event will be in the normal or CSR phase, based on a hard elasticity cut. It is important to note the simulations discussed here uses a set of parameters that fit well Auger data, but doesn't exhaust all possibilities for the impact of a possible chiral symmetry restoration at high energy on air shower variables.

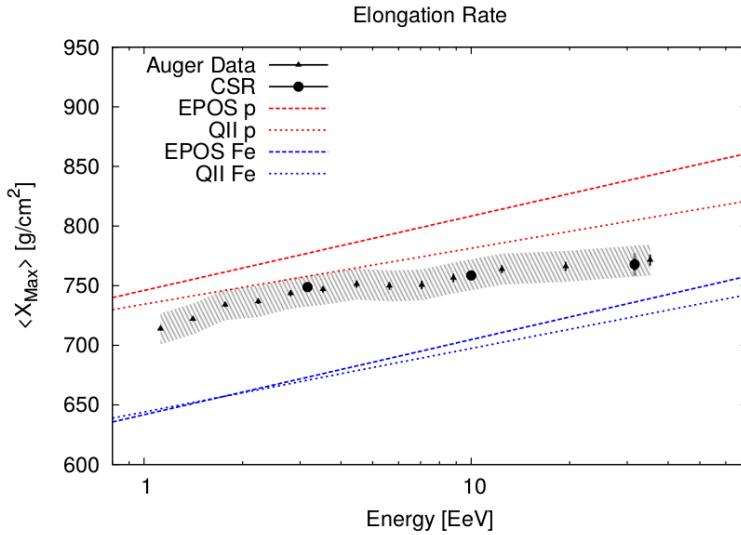


Figure 6.13: From [102].  $\langle X_{\max} \rangle$  versus energy. The systematic uncertainty in the  $X_{\max}$  measurement is around  $11 \text{ g cm}^{-2}$ .

### 6.3.2.1 Impact on air shower variables

The main impact of this model is a significant increase of the number of muons (almost two-fold at the highest energies) and decrease of  $\langle X_{\max} \rangle$  (and its variance) with increasing energy. The larger muon number comes, as previously explained, from the decreased production of pions, thus not transferring so much energy so rapidly to the electromagnetic channel. This in turn means that the hadronic channel carries more energy to the later generations, leading to more pions at lower energies, where they have a higher likelihood of decaying to muons.

$\langle X_{\max} \rangle$  is dominated by the showers coming from the first high energy photons, which are also the deepest. As energy is taken away from the EM component at the first generations, the maximum gets more prominent contributions from the subsequent sub-showers which have shallower maximums. As the fraction of events to which CSR was applied increases with energy,  $\langle X_{\max} \rangle$  decreases with it compared to the reference model, as shown in figure 6.13. The same thing happens with  $\sigma[X_{\max}]$ , as showers become more homogeneous at higher energy. Also, some of the decrease in  $X_{\max}$  comes from the increase in the cross-section, making the first interaction occur upper in the atmosphere.

In this case, predicting the average shape is very difficult, as none of the main shower parameters (inelasticity, multiplicity) are directly altered, except for the cross-section which does not affect the shape. Moving energy from pions to kaons causes the shower to fall more slowly after the maximum, as it takes more generations to move the energy to the electromagnetic component. Thinking about changes in  $R$  as a "rotation" of the electromagnetic profile, with increasing  $R$  meaning less energy deposit before and more after the maximum, it is expected that  $R$  should be larger for CSR. This is indeed seen in simulations, as is shown in figure 6.12.

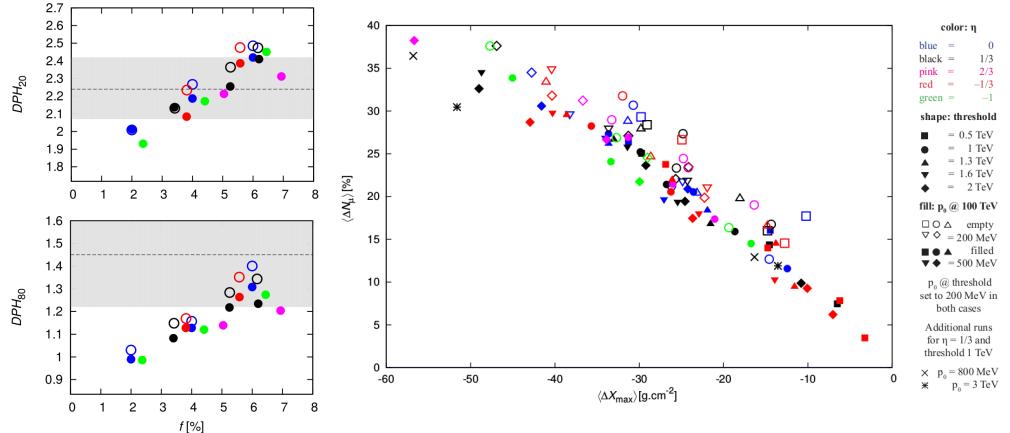


Figure 6.14: Left:  $DPH_{20}$  (top) and  $DPH_{80}$  (bottom) as a function of the converted energy fraction  $f$ . The horizontal line shows the value measured by DELPHI with a 1-sigma band. Right:  $\langle \Delta N_\mu \rangle$  as a function of  $\langle \Delta X_{\max} \rangle$  for proton showers at  $10^{18.5}$  eV. The plot includes all available simulations at once.

The increase in  $R$  is very sizable, making this parameter even larger in protons after CSR than the iron prediction in standard models. To analyze the effect in  $L$ , which is harder to predict, we look at the resulting profile in figure 6.11. There, we can see that  $L$  remains mostly unaltered, as the lack of energy that is deposited before the maximum from the first pions is compensated by a component after the maximum from later generations, and they cancel out.

### 6.3.3 Soft Particle Addition Model

An excess of muons with respect to simulations has been found not only at the Pierre Auger Observatory but also at lower energies: the DELPHI collaboration at cosmoLEP studied multi-muon events from cosmic ray showers and reported an excess of muons in data compared to Monte Carlo at the highest multiplicity events [128]. For low multiplicities both data and Monte Carlo were in fair agreement. In this section we will briefly describe the work in [103], where a mechanism to simultaneously increase the muon count at LEP and Auger energies was developed by adding soft particles to hadronic interaction secondaries.

The main cosmic ray result from DELPHI is the measurement of bundles of muons from extensive air showers. At high muon multiplicities, the observed flux of events is in excess with respect to simulations even for a pure iron primary beam. For observed multiplicities larger than 20, there are  $(2.24 \pm 0.17)$  times more events than in pure proton simulations and, for multiplicities larger than 80, there are  $(1.45 \pm 0.23)$  times more events than in pure iron simulations. Ideally, we would like to find which model reproduces the  $DPH_{20}$  and  $DPH_{80}$  the best for a realistic mixture of primary particles, according to KASCADE and KASCADE-Grande data. This muon excess in DELPHI occurs at  $P_{lab} \approx 10^{15}$  to  $10^{17}$  eV, and the lower part of this range corresponds to LHC energies, where no such effect has been observed. One of the ways to reconcile these results is to produce the extra soft particles

in a very narrow cone along the interaction axis - remnant jets. In the case of high energy cosmic rays, the increased production of pions in the forward direction would increase the number of muons at ground. The idea of adding soft particles in the forward-backward direction while leaving all other interactions for the standard hadronic model by itself leads to a large number of degrees of freedom, so some simplifications have been made.

The first choice is deciding which fraction  $f$  of the center-of-mass system energy is transformed into the extra soft particles. While  $f$  could be energy dependent, a sharp energy threshold above which  $f$  is constant was set. The efficiency of the process was also varied by a factor  $\eta$ , depending on the size of the interacting system: the  $f$  constant is, for each interaction, multiplied by the number of wounded nucleons to the power of  $\eta$ . So, the energy given to the extra soft particles is

$$E_{soft} = \sqrt{s} \cdot f \cdot N_{\text{wounded}}^{\eta} \quad (6.6)$$

To make room for the extra soft particles, energy has to be taken from the standard ones predicted by hadronic models. It was decided to remove a fraction of energy only from particles with c.m. energy above 10 GeV, with the fraction taken proportional to the logarithm of its energy. Also, only particles along the forward cone in which the extra soft component is being added are altered. The next step is the addition of the particles. Only pions, kaons and protons are used, with the ratio of production for  $\pi$  and  $K$  being always 2/3 charged to 1/3 neutral. The opening angle has to be very small from LHC constraints, so it was set to  $1^\circ$ , with tweaks to this value showing very little sensitivity. Finally, the momentum distribution was chosen to be  $p \exp(-p/p_0)$ , where  $p_0$  is a (possibly energy dependent) model parameter. For pions, the best approximation for the fit of this function is around  $p_0 \approx 200$  MeV, value which is used here as default. Also shown are simulation for which  $p_0$  starts at 200 MeV but grows logarithmically to 500 MeV at 100 TeV c.m. energy.

#### 6.3.3.1 Impact on air shower variables

The addition of soft particles to high-energy interactions has the potential to increase the number of muons at ground. But since a larger number of parameters were left free compared to the other models described, and their interplay is not trivial, the analysis is not as straightforward. It was found that the DELPHI excess could be explained by using approximately  $>5\%$  of the energy of each interaction for the soft particle production (3-8% were simulated), so we only discuss this range from now on. The main constraint from cosmic ray observations is the value of  $X_{max}$ , which within this model always decreases: the model increases multiplicity, producing lower energy particles, and thus showers develop faster. Data, however, is only  $\approx 30 \text{ g cm}^{-2}$  below the proton line, so this is the upper value by which  $X_{max}$  can be shifted downward. In figure 6.14, the correlation of  $\Delta X_{max}$  with  $\Delta N_{\mu}$  is shown. If we accept the shift of  $30 \text{ g cm}^{-2}$  as maximal, we can obtain roughly 25% more muons on the ground for primary protons.

For the average shape, we are interested in higher energies (above  $10^{19}$  eV) than were used in [103], so we do not have access to simulations of the full parameter range of this model. Two sets of parameters representative of the favored range, where DELPHI data is fitted and  $\Delta N_\mu$  is around 25%, were chosen and simulated at  $10^{19}$  and  $10^{19.5}$  eV. The two sets share some common parameter values: the energy threshold is set at 1 TeV and the mean momentum is 200 MeV at threshold and 500 MeV at  $10^{19}$ . In set 1,  $f$  was set to 6% of  $\sqrt{s}$  of the interaction and  $\eta = 1/3$ , so the total fraction of energy used in each interaction to the production of soft particles is given by

$$E_{\text{soft}} = \sqrt{s} \cdot 6\% \cdot N_{\text{wounded}}^{1/3} \quad (6.7)$$

where  $N_{\text{wounded}}$  is the total number of wounded nucleons in the interaction (target+projectile). In set 2,  $f$  was set to 22% and the  $\eta$  factor to  $-1$ .

Figure 6.15 shows the model predictions for both sets of values implemented on proton and iron primaries. Protons behave very similarly between both sets, while iron is very different. This can be explained by looking at equation 6.7: the number of wounded particles is 2 in proton collisions, while for iron it is generally around one third of the total number of nucleons in the interaction, 70 (56 from iron and 14 from nitrogen)<sup>1</sup>. Therefore, for proton the percentage of energy used for soft particle production is around  $0.06 \cdot 2^{1/3} = 8\%$  for set 1 and  $0.22 \cdot 2^{-1} = 11\%$ , so there isn't a large difference between both sets. For iron, the percentages are  $0.06 \cdot 26^{1/3} = 17\%$  for set 1 and  $0.22 \cdot 26^{-1} = 1\%$ . This explains both the huge difference between iron values in both data sets and the fact that in set 2 the value of this model is much closer to the one from standard models: only a very small fraction of the energy is used for soft particle production.

#### 6.3.4 Comparison with Auger data

Although it might be an oversimplification, it can be said that all models described in this section aim to increase the muon number, while being constrained by accelerator measurements and  $X_{\text{max}}$ . The way they do this is, however, very different from each other, which leads to a large variation in the shape parameters predicted by them. In figure 6.15,  $R$  and  $L$  for data and all models discussed in this work are shown.

The average shape measured in Auger tends to disfavor both the SPAM and percolation model, as they lead to a larger width and smaller asymmetry than normal models, while in data precisely the opposite is measured, i.e., we observe a larger  $R$  than predicted by all hadronic models. The CSR model, however, is compatible with data as it predicts a larger asymmetry while maintaining the width virtually unchanged. We reiterate that this is not a comprehensive study on the possibilities of new physics, as the models described represent a very small choice of the possible parameter space. What we can conclude is that the shape has currently, even with its large systematic uncertainties, some power to constrain hadronic models, as the parameter space spanned by them is larger than our current uncertainty.

---

<sup>1</sup>The integration of the nucleons in the interaction region for all values of the impact parameter  $b$  yields an average of 26 interacting nucleons for an iron hitting a nitrogen nucleus

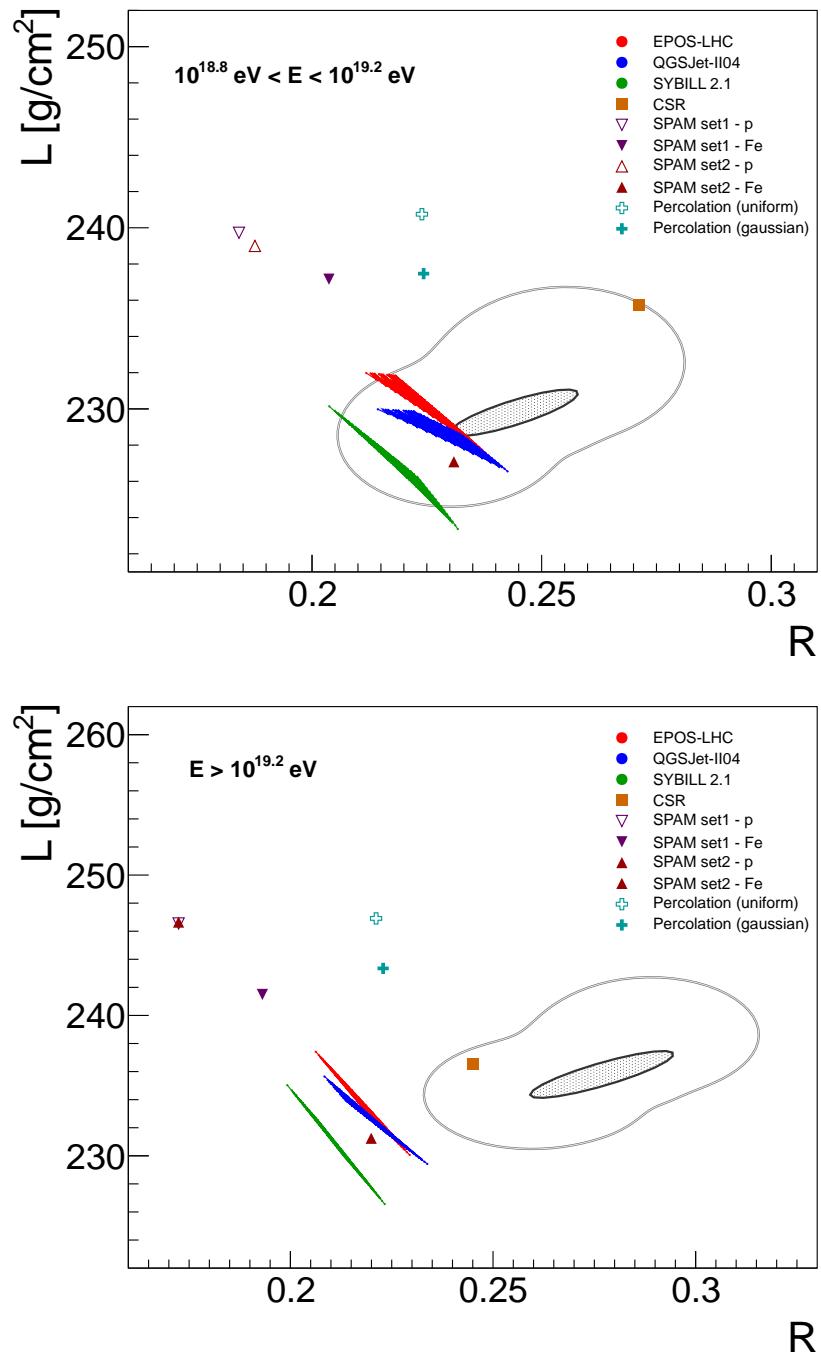


Figure 6.15:  $R$  and  $L$  for Auger data (filled ellipse is statistical uncertainty, its larger contour the systematic one) along with the predictions of all the alternative models considered in this chapter, for the energy ranges  $10^{18.8}$  to  $10^{19.2}$  eV (top) and  $E > 10^{19.2}$  eV (bottom).

## 6.4 Impact of hadronic parameters on R and L

In the previous section some alternative models of high energy interactions were described. These run through a full shower simulator and have a very detailed and self-consistent description of all interaction secondaries. However, and partly due to this, their results can be difficult to analyze: each model alters a few parameters simultaneously and the changes in the average shape cannot be pinpointed to a specific factor. In this section we will develop a simplified Monte Carlo shower simulation, which does not fully describe the whole shower development intricacies but can shed light on which are the main parameters that can influence the average shape.

The Heitler model, more specifically its hadronic extension, the Heitler-Matthews model [46], provides a framework from which the main shower variables can be derived using just a few initial parameters. This approach, however, has limitations: it does not take into account the finer details of the interactions (particles created besides pions, spectrum, interaction-to-interaction variance, etc) so we cannot derive a full description of all shower components. In this section a Heitler-Matthews based Monte Carlo simulation of showers will be implemented to study the impact of hadronic parameters on it, focusing on the shape variables.

### 6.4.1 Introduction

The Heitler model is a description of electromagnetic showers. They develop from a initial photon of energy  $E_0$ , which produces a  $e^+e^-$  pair, and each of this particles in turn radiates a photon with half its energy after a radiation length  $\lambda_r$ . Particles stop multiplying when they reach the critical energy,  $\xi_c$ . The resultant profile is described by the Greisen function [45]:

$$N_e(X) = \frac{0.31}{\sqrt{y_c}} \exp \left[ \frac{X}{\lambda_r} (1 - 1.5 \ln[s]) \right] \quad (6.8)$$

where  $y_c = \ln[E_0/\xi_c]$  and  $s$  is the age parameter.

This model can be extended to hadronic primaries by considering, instead of  $e^\pm$  and  $\gamma$ , a cascade of pions in air [46]. A charged pion interacts creating  $c \times m$  charged pions and  $(1 - c) \times m$  neutral ones.  $m$  is the (energy dependent) interaction multiplicity, and  $c$  is the ratio of charged to neutral pions created, which is typically defined as 2/3, as equal amounts of  $\pi^0$ ,  $\pi^+$  and  $\pi^-$  are created. The  $\pi^0$  decay instantly in two  $\gamma$  which developed a electromagnetic shower, while the charged pions continue this process until they reach their critical energy,  $\xi_\pi$ .

This original description has been improved to included some second-order effects. Perhaps the most prominent simplification is that the energy is equally divided between all secondary particles. This excludes the leading particle effect, where one secondary particle takes a very significant fraction ( 0.3 to 0.4) of the original energy. This effect can be incorporated by introducing a elasticity parameter:

$$\kappa_{el} = E_{leading}/E_0 \quad (6.9)$$

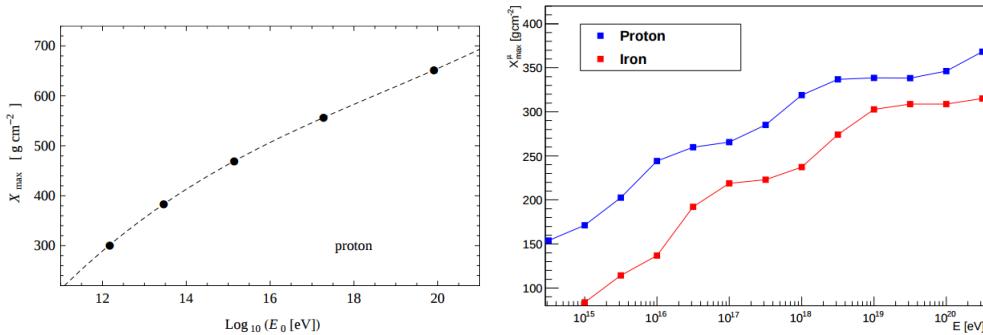


Figure 6.16: Left:  $X_{\max}$  from a Heitler-Matthews implementation [104]. For reference, the  $X_{\max}$  predicted by hadronic models at  $10^{19}$  eV is around  $800 \text{ g cm}^{-2}$ . Right:  $X_{\max}^\mu$  from a Heitler-Matthews implementation [105]. For reference, the  $X_{\max}^\mu$  predicted by hadronic models at  $10^{19}$  eV is around  $550 \text{ g cm}^{-2}$

The remaining fraction,  $\kappa_{inel}$ , is available for the production of secondary pions, and is usually divided equally between them.

So far, the approach described here yields analytical expressions for the main shower variables. This comes at the cost of two major simplifications: all particles interact after the same interaction length, and there are no pion decays to  $\mu$  and its neutrino. This problem can be solved by a MC approach [105]: probabilities of decay ( $p_d$ ), depending on the pion's relativistic half-life, as well as of interacting ( $p_i$ ), from the energy dependent interaction length, are computed. The probability of either happening is evaluated in depth steps through the atmosphere for each particle. So at depth  $X_n$ , where there are  $N_n$  pions with energy  $E_n$ ,  $p_d \times N_n$  muons of  $E_n/2$  and  $p_i \times N_n \times \text{mult}_n$  of energy  $E_n/\text{mult}_n$  pions are created. The remaining pions,  $(1 - p_n - p_i) \times N_n$ , are propagated unscathed to the next depth step. This is a more realistic implementation of the idea behind the Heitler model, but also more computationally intensive and harder to interpret, as our final variables, like  $X_{\max}$ , cannot be written anymore as a function of the initial parameters.

This approach has been applied to predict the muon production depth given by the Heitler-Matthews model [105]. It was found that the depth of the maximum and the width of the profile were underestimated compared to predictions from standard hadronic interaction models. Also, the  $X_{\max}$  predicted by the hadronic Heitler model for the usual initial parameters is slightly smaller than the one predicted by hadronic models [104]. A possible cause of this problem is the even energy division between secondaries in the hadronic interactions.

Here, we decided to add one more feature to the Heitler-Matthews inspired shower simulation: a realistic pion energy spectrum for the high energy interactions. It only takes one parameter, which was varied between its limits to study the effect on the shower development.

### 6.4.2 Implementation

An incoming particle with energy  $E_0$  has a probability of interacting in a given depth slab  $\Delta X$  given by  $1 - \exp[-\Delta X/\lambda_p]$ , where  $\lambda_p$  is the interacting length and was parametrized from the post-LHC hadronic models as

$$\lambda_p = 60 - 10 \times \log_{10}[E_0/EeV] \text{ g cm}^{-2} \quad (6.10)$$

At each depth step it is evaluated stochastically whether the particle interacts or not. When it does, a proton with energy  $(1 - \kappa_{inel}) \times E_0$  is propagated to the next level, while the remaining energy is divided by  $M$  pions. The default value of  $\kappa_{inel}$  used here is 0.7, which is around the one given in [104], and is the average of the value in hadronic models (0.69 to 0.72 at the highest energies). The value of  $N_{mult}$  is strongly energy dependent, and is given by [104]:

$$M = 0.15 \cdot E_0^{0.18} \quad (6.11)$$

It is also assumed as default that the ratio of charged pions to neutral ones,  $r_{ch}$ , is 2:1 in all interactions. So, the multiplicity of charged pions is related to the total one by  $M_{\pi^\pm} = \frac{2}{3}M$ . At this point, there is a very energetic proton, for which exactly the same process is repeated until it reaches 10% of its initial energy, and  $M$  pions. The neutral pions have a very short lifetime - of the order of  $10^{-17}$  s while charged ones are around  $10^{-8}$  s - so in our implementation they decay instantly. The main decay mode, with a branching ratio of around 99%, is  $\pi^0 \rightarrow \gamma\gamma$ , which go and feed the electromagnetic cascade. These particles are no longer propagated in our program, as the resultant profile for each  $\gamma$  is given by a Greisen function, and therefore the whole electromagnetic profile is the sum of each individual contribution.

The core of our simulation is then the propagation of charged pions. In the original Heitler-Matthews model, pions produce a cascade until they reach the critical energy and then all decay simultaneously. On a Monte Carlo approach, this assumption is not necessary. As explained before, the results of a pion crossing a slab of atmosphere can be three-fold:

- it interacts, producing  $M(E)$  pions, for which the process is repeated: neutral ones give Greisen functions, charged ones are propagated to the next depth. This process occurs with probability  $p_{int} = 1 - \exp[-\Delta X/\lambda_\pi]$ , where the interaction length is given by  $\lambda_\pi = 200 - 3.3 \log[E/eV] \text{ g cm}^{-2}$  [104]
- it decays, yielding a muon and a neutrino. This process occurs with probability  $1 - \exp[-\Delta h/(t_0 \cdot c \cdot \gamma_{fac})]$ , where  $t_0$  is the pion half life and  $\gamma_{fac} = E/m_\pi$  the relativistic dilation factor.
- it does neither decay nor interact, and is propagated to the next depth.

In these calculations, an approximate relation between the depth in the atmosphere  $X$  and height  $h$  is used:

$$X(h)[\text{g cm}^{-2}] = 1030 \cdot \exp\left[-\frac{h[\text{km}]}{8}\right] \quad (6.12)$$

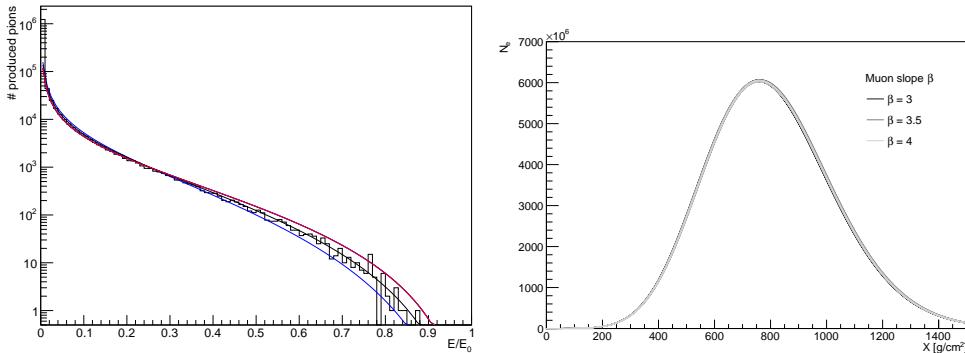


Figure 6.17: Left: Pion production energy fraction spectrum. The histogram is the result of our simulation (with  $\beta = 3.5$ ) and the lines are function 6.13 with parameters  $\beta = 3$  (red),  $\beta = 3.5$  (black) and  $\beta = 4$  (blue). Right: Average profile using the parameter  $\beta$  between 3 (black) and 4 (light grey) in 0.1 steps.

This model has been shown to work in predicting qualitatively the behaviour with energy of the main shower variables. It underestimates, however, the length of the development of both shower components:  $X_{\max}$  and  $X_{\max}^\mu$  are underestimated by around  $150 \text{ g cm}^{-2}$  each [104, 105] (see figure 6.16).

The most important simplification made in the Heitler-Matthews model is that all secondary pions in an interaction have the same energy, i.e.,  $E_\pi = E_0/M$ . However, we know this not to be true. The pion energy spectrum can be described by the function (cite lipari)

$$f_\pi(E) = \frac{(1 - E/E_0)^\beta}{E/E_0} \quad (6.13)$$

where  $\beta$  is the slope parameter.  $\beta$  is not strictly universal in energy, but can take values between 3 and 4. In our implementation, the default value of 3.5 was chosen. This choice however, carries virtually no systematic uncertainty into our shower variable analysis, as can be seen in figure 6.17. The muon energy distribution does not depend strongly on  $\beta$  within this range, and the longitudinal profile dependence is even smaller - no difference can be seen by eye between a shower simulated with  $\beta = 3$  and  $\beta = 4$ . The only measurable difference is a few ( $< 5$ )  $\text{g cm}^{-2}$  in  $X_{\max}$ , and after translating the profiles no sizable change is measured in either  $R$  or  $L$ .

Also, the two main muon energy loss mechanisms ( $e^+e^-$  pair production and bremsstrahlung) were implemented (from [106]). Their effect on the electromagnetic profile are also very small and only really visible at its upper tail: around 1% of the shower energy is brought back to the electromagnetic component above  $X' \approx 300 \text{ g cm}^{-2}$  (both values are slightly energy dependent). So, it affects neither our fitting region ( $X' < 200 \text{ g cm}^{-2}$ ) for the average shape nor the  $X_{\max}$  fitted value.

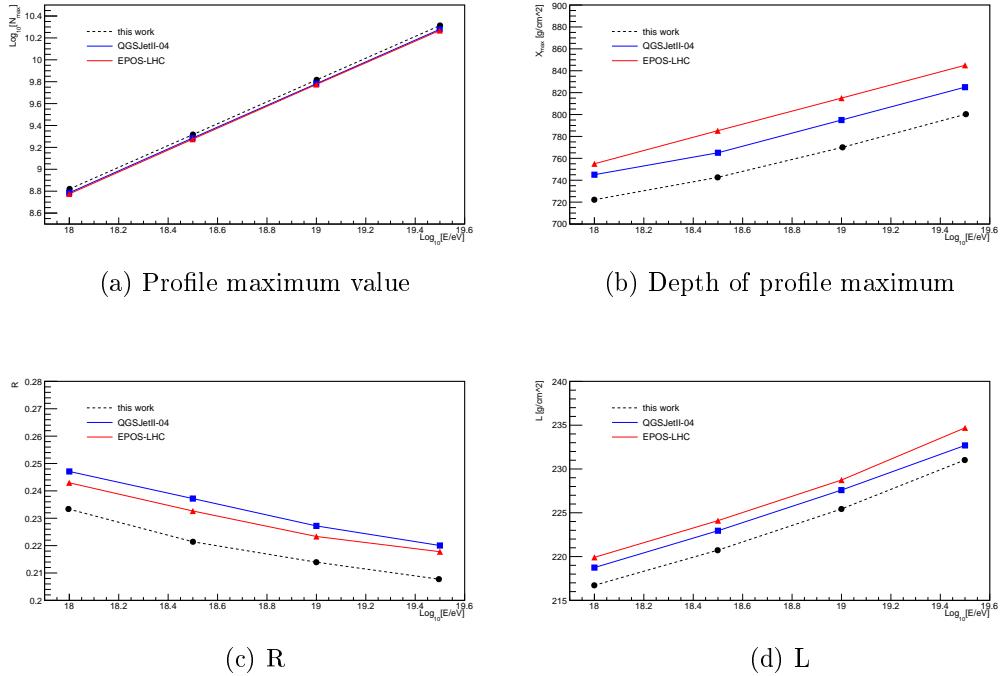


Figure 6.18: The four electromagnetic profile variables as indicated in each sub-caption as a function of energy. Red and blue lines indicate predictions

### 6.4.3 Results

The four parameters of the electromagnetic profile ( $X_{\max}$ ,  $N_{\max}$ ,  $R$  and  $L$ ) predicted by our model, as well as by the post-LHC models, are shown in figure 6.18<sup>2</sup>. In general, the agreement is very good. Most notably, there is an improvement in the prediction of the depth of shower maximum, which increases from around  $600 \text{ g cm}^{-2}$  [104] to  $750 \text{ g cm}^{-2}$  - more in line with standard model predictions. However, this is still slightly below all models, as is the width of the shower,  $L$ . An explanation for this might be the fact that only pions are being produced, while in real showers more than 10% of produced particles are other hadrons. More specifically, within our model, in each interaction 1/3 of all secondaries are  $\pi^0$ , which decay instantly feeding the electromagnetic cascade. In standard models, however, only 27-30% (energy and model dependent) of produced particles are neutral pions, with the rest being mostly other hadrons, like protons and kaons. In the alternative models explained in the previous section, this value can be as low as 10% (see figure 6.10). The simulation of kaons and other hadrons is beyond the scope of this work, as full, well tested, shower simulators already exist and we are aiming for a simplified model which can explain the shower main features. It should be noted, however, that varying the charged pion ratio,  $r_{ch}$ , which is by default set to 2/3, can partially

<sup>2</sup>This is a Monte Carlo simulation, so values shown here are averages over several simulated showers, 500 usually - this was found to be a large enough number of events such that the average of all showers tends to the true average one.

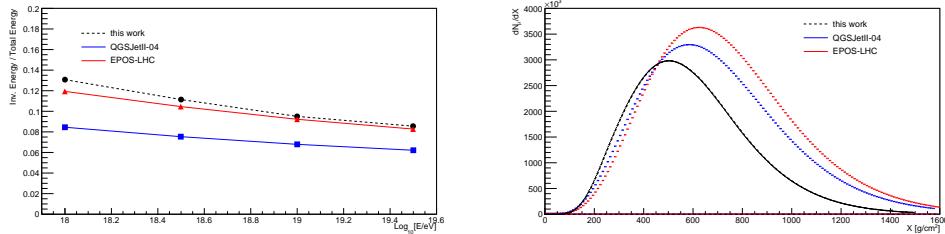


Figure 6.19: Left: Missing Energy fraction as a function of energy for this work and the hadronic models shown in the legend. Right: Muon Production Depth (MPD) profile for this work and the hadronic models shown in the legend for simulations at  $10^{19.5}$  eV. For reference also the profile for the case in which the energy is evenly distributed between all secondary pions in interactions is shown [105].

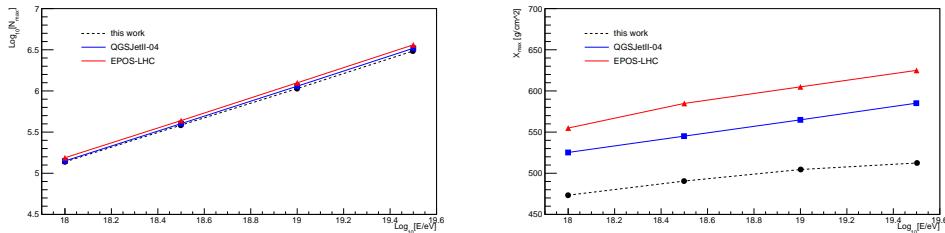


Figure 6.20: Value at maximum (left) and depth of maximum (right) of the muon production depth profile.

account for this effect as protons and kaons behave more like charged pions than neutral ones.

Even with these simplifications made, and with only a few parameters - multiplicity, elasticity, charged pion ratio, pion distribution  $\beta$  parameter, interaction length - the energy evolution of all electromagnetic profile variables, as well as the missing energy (see figure 6.19), is very accurate when compared to hadronic models.

Another interesting result to look at is the muon production depth (MPD) profile. The profile for  $10^{19}$  eV showers is shown in figure 6.19, while the energy evolution of its two main parameters, the depth of the maximum ( $X_{\max}^{\mu}$ ) and the number of muons produced there, is shown in figure 6.20. The results are very encouraging. While it suffers from the same caveat we pointed when discussing the e.m. profile, i.e., it develops too fast, at least partially due to the fact not enough hadrons (besides pions) are being created, the energy evolution is well reproduced and the normalization is very much improved with respect to previous implementations of the model where no pion production spectra was considered [105].

We have demonstrated in this section that a simple Monte Carlo implementation of the Heitler-Matthews model with an added pion production energy spectrum can reproduce the energy evolution of both the electromagnetic and muon production profiles. The absolute value does not perfectly match the one in hadronic models, but it is improved with respect to the standard Heitler-Matthews model prediction, i.e., the addition of the pion production energy spectrum deepens the maximum of both the e.m. and MPD profiles, and increases the maximum of the MPD, bringing

these variables closer to those predicted by hadronic models. The main process we are missing, which is the production of other hadrons besides pions, would result in further deepening both profiles and increasing the number of muons, as less energy would be put directly into photons through the creation of  $\pi^0$ .

#### 6.4.4 Impact of hadronic parameters on the longitudinal profile

Full shower simulations have several free parameters. However, not all of those parameters are equally important. It is commonly considered that the two most important ones for each interaction are the number of particles produced and the fraction of energy carried by the most energetic particle. The distribution of particles produced can not be described by one parameter, as it is much more complex, but the percentage of particles that are neutral pions can somewhat function as a proxy to it. This is due to the fact that charged pions are much closer to protons and other hadrons, in the sense that at high energies they usually interact, while neutral pions decay virtually immediately to photons.

So, in this work, these three important parameters are considered:

- the elasticity,  $\kappa_{el} = E_{leading}/E_0$ , which is by default 0.3.
- the multiplicity,  $M$ , which is defined as the total number of pions created in each interaction and is by default  $0.15E^{0.18}$ .
- the pion neutral ratio,  $c = 1 - r_{ch} = M_{\pi^0}/M$ , and is by default 1/3.

By shifting these parameters individually up and down we can test the repercussions of the uncertainty in their value in the main electromagnetic shower variables, like  $X_{max}$ , or  $R$  and  $L$  for example.

The idea of altering the hadronic parameters to investigate the resultant uncertainty in the shower variables predictions is not new - see for example [108] for the effect of directly shifting  $\kappa_{el}$  or [109] where a more holistic approach was preferred, and several sets of initial hadronic model parameters were tested. The most comprehensive work in this area, and the one we will use as a benchmark, is [107]. There, predictions of existing hadronic event generators were used and the output of these models modified in a suitable way to probe the phase-space of interaction characteristics exhaustively. These modifications are ad-hoc and explicitly not based on an underlying fundamental theory or phenomenology, and therefore the influence of the modification of one parameter on the shower evolution can be studied almost independently of the other parameters (apart from the correlation imposed by the energy and charge conservation constraints).

The hadronic parameters are measured at low energies, and extrapolated to the highest ones. Therefore, it makes sense that the model parameters are altered by an energy dependent factor, which was chosen to be of the form [107]

$$f(E, f_{19}) = 1 + (f_{19} - 1)F(E) \quad (6.14)$$

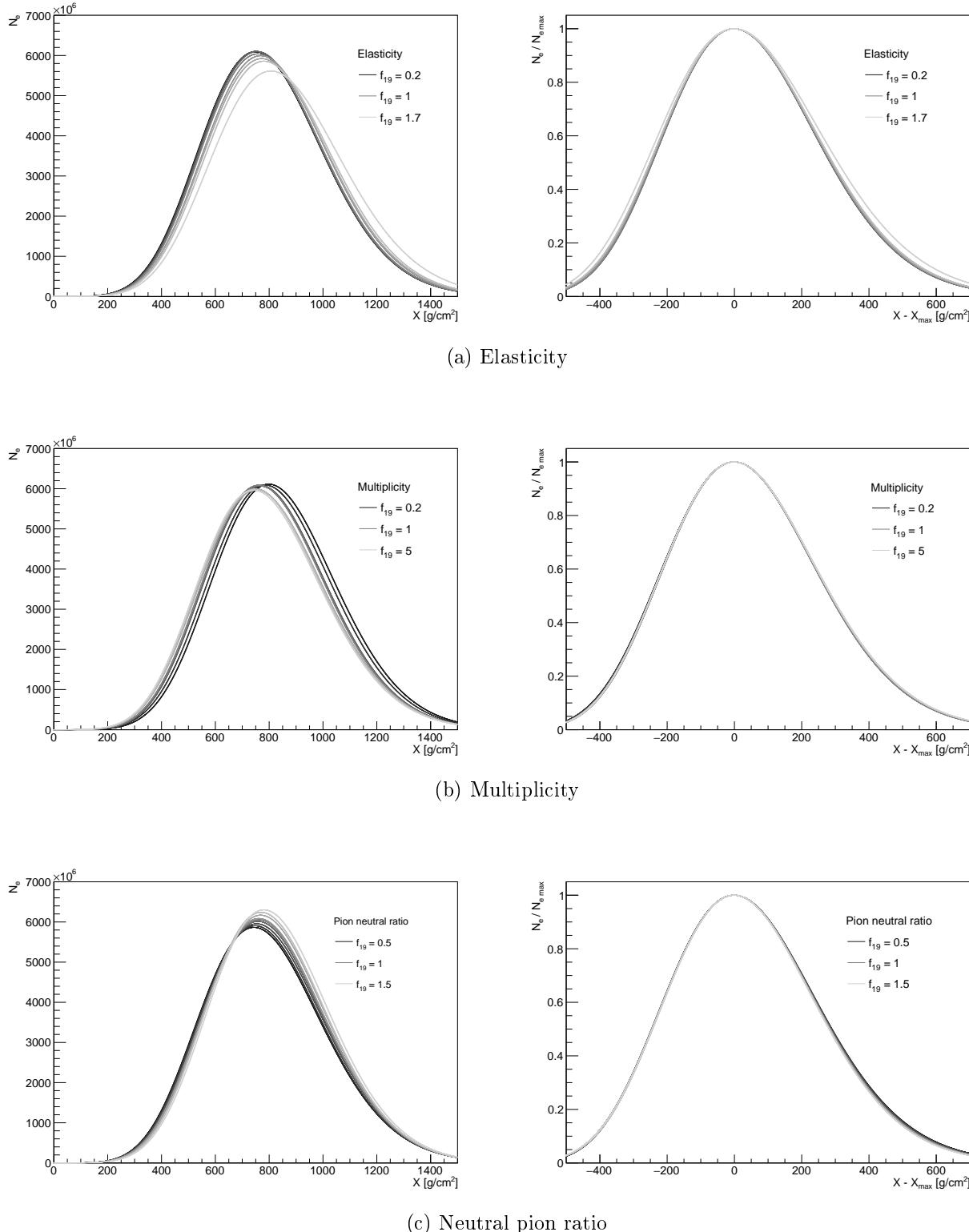


Figure 6.21: Electromagnetic profile (left) and the result of its normalization and translation by its maximum (right) for different values of the  $f_{19}$  factor affecting each of the captioned parameters. The color ranges from black (low  $f_{19}$ ) to light grey (high  $f_{19}$ ).

with

$$F(E) = \begin{cases} \frac{\log_{10}[E/1PeV]}{\log_{10}[10EeV/1PeV]}, & \text{if } E > 1PeV \\ 0, & \text{otherwise} \end{cases} \quad (6.15)$$

We chose to keep the threshold energy of  $f(E, f_{19})$  at  $10^{15}$  eV, so we can compare the results with our benchmark. However, it should be noted that there are already post-LHC models, which are constrained up to  $10^{16}$ - $10^{17}$  eV. Above  $10^{15}$  eV, a logarithmic increase in the deviation of the parameter from 1 is allowed, until it reaches the reference value,  $f_{19}$ , at 10 EeV. Each interaction depends on a set of hadronic parameters. For each realization, we simulate showers where in each interaction one of these parameters is multiplied by the corresponding  $f(E, f_{19})$  factor. The resultant longitudinal profiles for  $10^{19.5}$  eV showers are shown in figure 6.21. The main features are qualitatively as expected: larger elasticity leads to deeper showers, while large multiplicity has the opposite effect - makes proton initiated showers more similar to iron initiated ones. Increasing the fraction of neutral pions has two clear effects: larger  $X_{max}$  and a larger number of produced electrons, as more energy goes into the electromagnetic component.

In [107], the variation of shower variables w.r.t. different multiplicity, elasticity and neutral pion ratio values was made. Two of these are average longitudinal profile variables -  $X_{max}$  and  $N_e(1000)$ , the number of electrons at a depth of 1000 g cm<sup>-2</sup> - and the other two are more related to the muonic sector of the cascade - the invisible energy and the number of muons,  $N_\mu$ . This will be used as a cross-check of the accuracy of our simplified model to reproduce the main features of the more complex modifications to the CONEX output in [107]. The results are shown in figure 6.22. The agreement is qualitatively very good, as the behaviour of both  $X_{max}$  and  $N_e(1000)$  with respect to variations in all the hadronic parameters agrees between this work and [107]. The number of muons and the invisible energy are quite correlated variables, and the behaviour with the neutral pion ratio is well modelled - both increase with multiplicity as expected. Elasticity is the only variable where the qualitative behaviour of both models does not agree, with modified CONEX predicting a small decrease of the number of muons for higher values of this variable, while for our model almost no change is seen. The difference, however, is manageable for our objective, within 5% for almost the entire elasticity range. Quantitatively, the agreement is also very good. Our model underestimates the changes brought by increasing the multiplicity (by  $\lesssim 50\%$ ), but accurately predicts those resultant of decreasing it. Otherwise, and apart from the highest point in elasticity (70%), which increases a bit more in our model, both works are compatible.

#### 6.4.5 Impact on profile shape - R and L

Finally, having validated our model, we can start to tackle the question that is at the core of this section: what is the impact that changing each of the hadronic parameters independently on the average profile shape? The answer to this question is shown in figure 6.23.

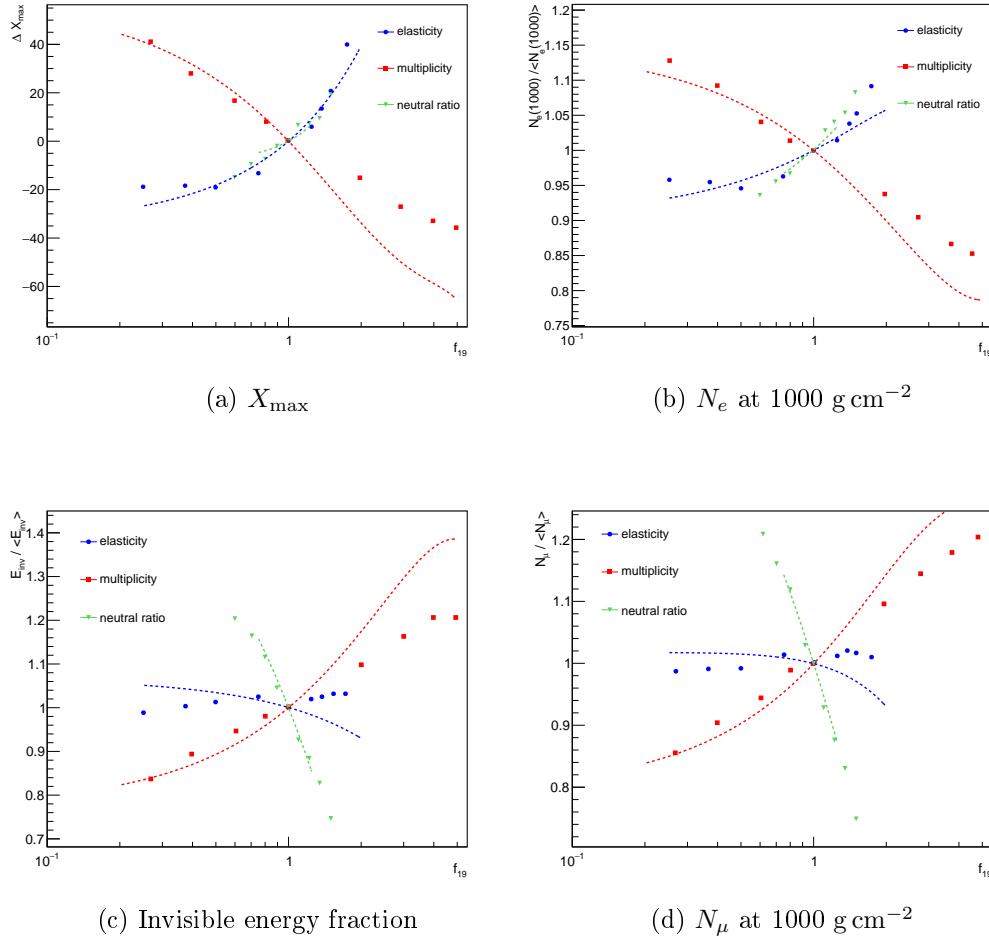


Figure 6.22: Impact of hadronic interaction features on the shower maximum,  $X_{\max}$  (top left), number of electrons at  $1000 \text{ g cm}^{-2}$ ,  $N_e(1000)$  (top right), invisible energy fraction (bottom left) and number of muon at  $1000 \text{ g cm}^{-2}$ ,  $N_\mu(1000)$  (bottom right). Points are our results, dashed lines are the fits from [107]

$L$  behaves exactly as expected with elasticity, as it grows larger with increasing leading primary energy. The effect is also highly non-linear, again as expected - if the leading particles always takes all the energy,  $L \rightarrow \infty$ . Knowing that a Greisen function has a maximum approximately given by

$$N_{emax} \approx 0.37 \frac{E_0/E_c}{\sqrt{\log_2[E_0/E_c]}} \quad (6.16)$$

which occurs at depth  $X_{\max} = \log[E_0/E_c]$ , we can perform a simple exercise to study the behaviour of  $X_{\max}$  with elasticity: for each interaction, we take the inelastic portion and calculate the maximum value and its depth, and then advance the elastic portion by one interaction length and repeat the process until 99% has been utilized. Then, the weighed  $X_{\max}$  is calculated for different elasticity values. What

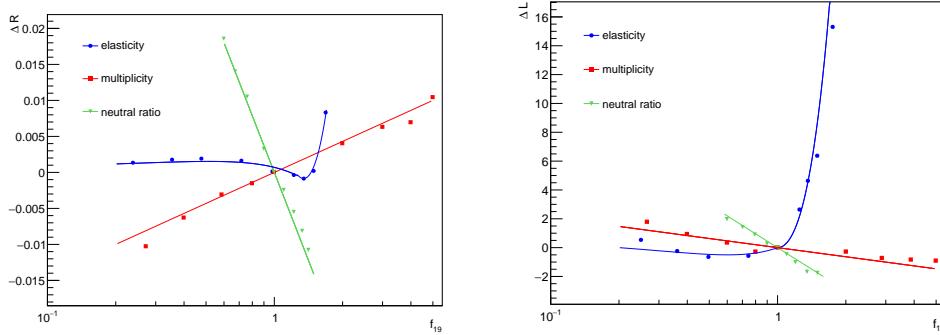


Figure 6.23: Impact of hadronic interaction features on  $R$  (left) and  $L$  (right).

we see is that changes in elasticity up to 50% are not relevant, since the proton interaction length (around  $60 \text{ g cm}^{-2}$ ) is much shorter than the width of the first Greisen (around  $200 \text{ g cm}^{-2}$ ), and by the time we get to older generations no energy is left. However, above  $\kappa_{el} = 60\%$  ( $f_{19} \approx 1.5$ ),  $\Delta X_{\max}$  rises very steeply, until it diverges at 1. A fit to this upper part yields the approximate dependence:

$$\Delta X_{\max}^{\kappa_{el}}(f_{19}) = 120 \cdot (f_{19} - 1)^2 \text{ g cm}^{-2} \quad (6.17)$$

The behaviour of the width is more complicated, but in general the rule that  $X_{\max}$  is approximately 3 to 3.5 times larger than  $L$  holds. The curve in figure 6.23 is

$$\Delta L_{\kappa_{el}}(f_{19}) = 35 \cdot (f_{19} - 1)^2 \quad (6.18)$$

which is exactly what was predicted by the simple model explained. So, the behaviour of the width with elasticity is well understood: it is mostly insensitive to it up to values around 50 to 60%, and then  $L$  increases very rapidly as the elasticity approaches 1.

For  $R$ , we can not predict the behaviour from the Heitler model.  $R$  is a deviation of the profile from symmetry, so its increase can be made in two ways: a faster rise to the maximum or a slower decay after it. What we observe is something similar to the effect in  $L$ : only large values of elasticity influence the shape in a significant way, and they contribute for making showers more asymmetric. Having said that, the change in elasticity necessary to affect profiles is very large:  $f_{19} = 1.75$  corresponds to  $\kappa_{el} \approx 0.6$ , while most models have this parameter at around 0.3 to 0.35 depending on the energy.

For multiplicity, we expect a clear behaviour from both variables.  $L$  should decrease with increasing multiplicity, as the shower develops faster and there is less variance in the energy and decay length of its highest energy particles.  $R$  should increase with multiplicity for the same reason: the shower develops faster (more iron-like). Both are observed in figure 6.23, and both can be parametrized as linear variations with the logarithm of  $f_{19}$

$$\Delta L_m(f_{19}) = -0.9 \log(f_{19}) \text{ g cm}^{-2} \quad (6.19)$$

$$\Delta R_m(f_{19}) = 6 \times 10^{-3} \log(f_{19}) \quad (6.20)$$

The neutral pion ratio is the hardest variable to predict, as an analysis of the Heitler-Matthews model does not yield its behaviour in an obvious way. The width  $L$ , like  $X_{max}$ , is dominated by the Greisen function resultant from the first very high energy photons (with width around  $200 \text{ g cm}^{-2}$ , energy dependent). The rest comes from the next generations of sub-showers, which have smaller  $\Delta \equiv X_{max} - X_1$  and  $L$ , but start later in the atmosphere. Typically the  $X_{max}$  value of the first generation pions is larger than the total one, i.e., even though further generations start later, they still reach their maximum in average before the first one (this is dependent on the elasticity value, as explained before). That is why  $X_{max}$  increases when more energy is given to  $\pi^0$ s (figure 6.22) - the relative importance of the first generation increases as more energy is directly injected into the electromagnetic component in it. From this description we can see that although  $L$  is usually directly correlated with  $X_{max}$ , in this case it is not: having only the first generation ( $c = 1$ ) corresponds to the thinnest shower possible, with only one component, while adding energy to other components with maxima at different depths widens the final profile. The variations in  $L$  are very small though, only a few  $\text{g cm}^{-2}$  in the entire range studied. They can be parametrized by

$$\Delta L_c(f_{19}) = -4 \log(f_{19}) \text{ g cm}^{-2} \quad (6.21)$$

which, since  $f_{19}(c)$  only goes from 0.5 to 1.5 corresponds to a variation of  $\pm 2 \text{ g cm}^{-2}$ .

In  $R$ , the effect can also not be predicted by the Heitler model but is intuitively simpler: as the generations after the first have on average shallower maximums than it, giving more energy to sub-showers of older generations means increasing their value at maximum, and thus increasing their relative importance. This brings  $X_{max}$  down by producing showers which rise faster, and that translates into a larger  $R$ , as observed in figure 6.23. The variation can be described by:

$$\Delta R_c(f_{19}) = -0.035 \log(f_{19}) \quad (6.22)$$

The main conclusions from this analysis are:

- there is no easy way to significantly ( $\gtrsim 5 \text{ g cm}^{-2}$ ) alter  $L$  - the only way to do so is increasing elasticity by a factor of almost 2, which is wildly outside any model prediction, and, to our knowledge, not even predicted by any alternative model based on new physics.
- $R$  is very sensitive to the fraction of energy going to neutral pions in each interaction,  $c$ . In fact, a large increase in  $R$  - let us say half the proton-iron difference, 0.01, for reference - implies either bringing the elasticity to 0.6, having a multiplicity 4 or 5 times larger than the standard one, or bringing the  $\pi^0/\pi$  production ratio to 70% of its standard value. The first two seem very unlikely, while the last one is not only reasonable but also predicted by some of the alternative models described in the previous section.

This last point is particularly important as a larger  $R$  is seen at highest energies, and, more significantly, an increase of  $R$  with energy not predicted by models.

#### 6.4.6 Comparison with Auger data

The objective of this section was a general estimation of the effects that hadronic model changes can produce on shower variables. However, we already know some sources of tension between models and data, so we will address them here. We chose two plots to study: one is the main plot of this work, the  $R$  vs  $L$  at the highest energy bin,  $10^{19.4}$  eV, and the other is the  $X_{max}$  vs  $R_\mu$ , that has been shown previously at  $10^{19}$  eV [96], but that we produced at our reference energy,  $10^{19.4}$  eV. We reproduced the plots of the variable changes with model parameters at this energy, and no visible differences were found with respect to the ones at  $10^{19.5}$  eV, so the values in figures 6.22 and 6.23 will be used.

From data, it is known that simulations are predicting a smaller number of muons and  $R$  than expected, while being compatible with  $X_{max}$  and  $L$ . Looking for parameters that can produce sizable differences in the first two variables without significantly altering the latter two, only one seems to fit the bill: the neutral pion ratio,  $c$ . Multiplicity also increases  $R$  but  $R_\mu$ , while the inelasticity can increase the number of muons but at the cost of an increase in  $X_{max}$  and  $L$  that worsens the compatibility of models with data.

The plots for the  $R : L$  and  $X_{max} : R_\mu$  plane at  $10^{19.4}$  eV are shown in figure 6.24. Superimposed are the arrows corresponding to 5 different "changes" to the reference, which is proton simulated with QGSJetII-04: changing model (to EPOS-LHC), changing composition (to iron) or changing the three hadronic parameters in study by the values that most shift upwards  $R_\mu$  and  $R$  while shifting  $X_{max}$  less than  $30 \text{ g cm}^{-2}$  down (data is only around  $30 \text{ g cm}^{-2}$  below the proton predictions) and  $L$  less than  $5 \text{ g cm}^{-2}$  up (systematic uncertainty on the measurement). The values are a pion neutral ratio of 60%, elasticity 50% larger and double the multiplicity of the nominal values. Although the values might be somewhat arbitrary, the purpose of this work is not to look at the exact magnitude of the effect, but to analyze the general behaviour of the shower with each parameter.

Figure 6.24 shows that for both plots a combination of having a mixed composition and a smaller pion neutral ratio brings the model predictions closer to data. In the case of the  $X_{max} : R_\mu$  plot, the agreement can be slightly improved also if EPOS-LHC is used and by an increase in the elasticity.

It should be noted that the disagreement with data in each plot, especially in the case of the shape, is less than  $2\sigma$  and is therefore not conclusive, so this exercise might seem premature. However, taking together all of the variables (as explained in the first section of this chapter) something is clearly missing in the shower description of current models. The results presented here show the potential of a change in the model parameters, especially the neutral pion ratio to bring the models towards data, although it is not the only one (for example, a combination of parameter changes could be employed).

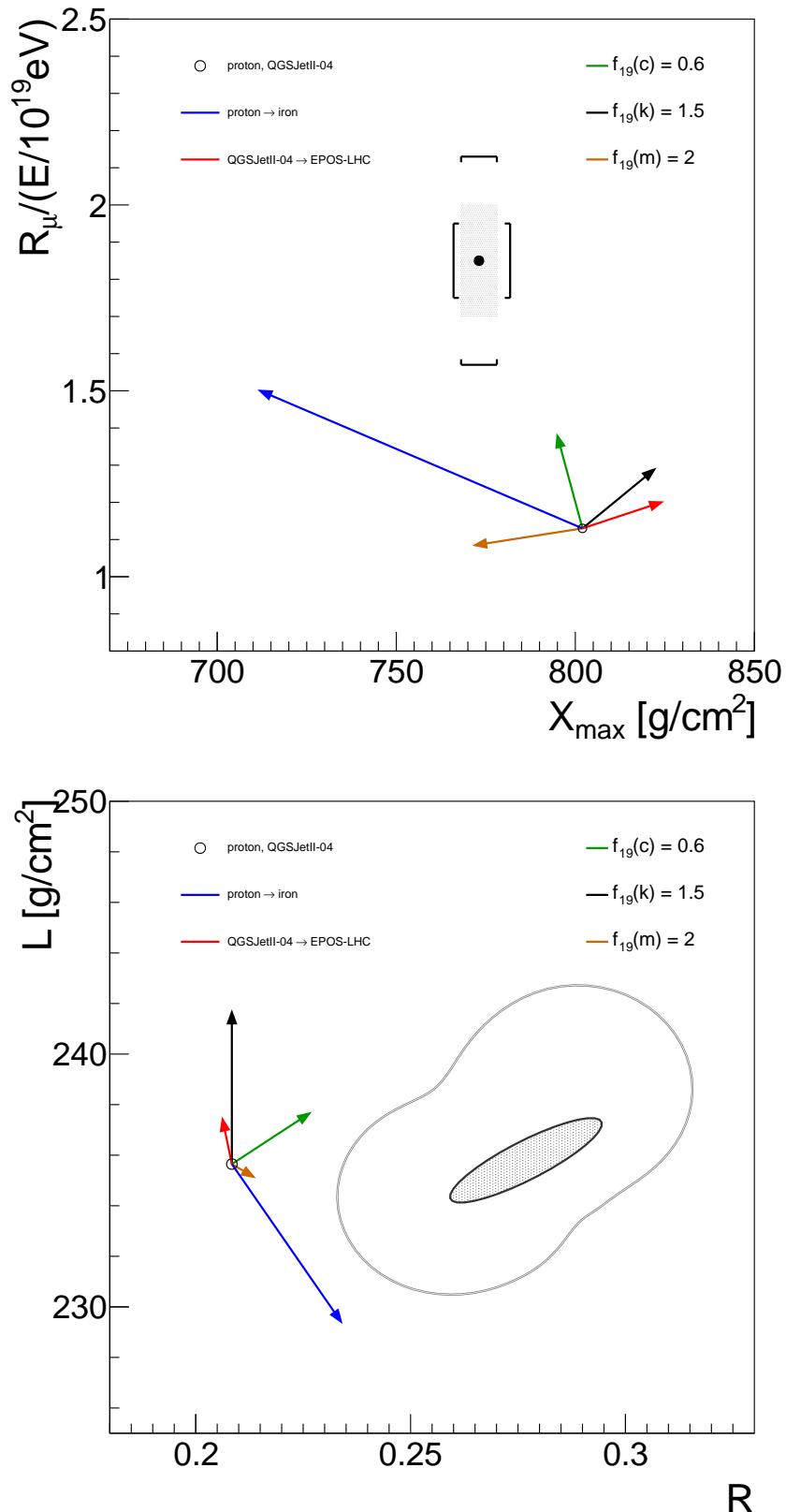


Figure 6.24:  $X_{\max} : R_\mu$  (top) and  $R : L$  (bottom) planes at  $10^{19.4}$  eV. Shaded area corresponds to statistical uncertainty on the measurement at the Pierre Auger Observatory, and brackets to the systematic one. The arrows represent the changes shown in the legend from a baseline of proton simulated with QGSJetII-04.

## 6.5 Discussion

In this chapter we have provided a joint interpretation of several variables measured by the Pierre Auger Observatory, with emphasis on the shape variables measured in this work,  $R$  and  $L$ . Taken at once, they portray a clear picture: air showers at the highest energies are neither accurately nor even consistently described by current shower simulations, regardless of the high-energy hadronic interaction model used.

In the first section the source of this incompatibility was identified - the variables sensitive to the hadronic component, like  $X_{\max}^\mu$ ,  $R_\mu$  and  $R$ , have a larger absolute value and faster increase with energy than what is predicted by hadronic models for the variables sensitive to the electromagnetic component, like  $X_{\max}$  and  $L$ . "Mixed" variables (with respect to their component sensitivity), the two measurements of  $\sec \theta_{\max}$ , have values that lie between them. Versions of these claims had already been done before (for example  $X_{\max}$  vs  $R_\mu$  and  $\frac{d\langle \ln R_\mu \rangle}{d \ln E}$  (data) vs  $\frac{d\langle \ln R_\mu \rangle}{d \ln E}$  ( $\langle \ln A \rangle$ ) [110],  $X_{\max}$  vs  $X_{\max}^\mu$  [111]) but here several variables were all put together and analyzed according to their shower component sensitivity.

The number of muons has been, for several years now, known to be incompatible with data. To fill this gap in the shower description left by standard hadronic models, a few alternative high-energy interaction models have been proposed. Since the main objective of this work is the study of the average shape, we focused on the effect these alternative interaction descriptions produce on the parameters  $R$  and  $L$ . The constraining power of these variables is clear: two of the models are incompatible with our measurement at the highest energy bin. It shows that even this first measurement of the average shape, with relatively large systematics, can be a factor in our understanding of high-energy interactions.

Even having these alternative models in addition to the standard ones does not produce a full phase space span of the possible descriptions of the shower development. Also, the cause of the changes we see in the variables is not always clear: the models adjust several parameters at the same time, and shower-to-shower fluctuations in which these parameters are correlated can obfuscate the contribution of each individual one. To shed some light on the importance of the main model parameters - multiplicity, elasticity and pion neutral ratio - we developed a Monte Carlo simulation based on the Heitler-Matthews model with a few extra features, most notably muon decay and energy loss, and a pion production spectrum. The results are very encouraging. The energy evolution of all shower variables studied is well reproduced, and the absolute values, while not always compatible with full shower simulation predictions, are close and much improved in comparison to previous Heitler based simulations. The main improvement with respect to these is the addition of the pion production spectrum, which pushes the shower development deeper into the atmosphere. Equipped with this model we studied the impact of the hadronic parameters on the shape. We compared the results of our model for  $X_{\max}$  and  $N_e(1000)$  to a previous similar work on CONEX, and our model held very consistent results.

Finally, the average shape was analyzed, and the main dependences of both its

parameters identified. The width  $L$  is very sturdy w.r.t. changes on all parameters around their central value, i.e., it is very hard to significantly alter  $L$  without going to extreme values of elasticity or multiplicity. It is not sensitive to the neutral pion ratio. The asymmetry,  $R$ , on the contrary, is very sensitive to the fraction of energy going to  $\pi^0$ s, while it changes little for small deviations of the other two parameters from the reference value. This is an interesting result as this analysis points to different "knobs" that can control independent variables that describe the shower behaviour.

# Summary

High energy cosmic rays are the most energetic particles in the universe. The behaviour of the air showers they produce when interacting with the Earth's atmosphere are our only current insight into interactions at energies above those achieved at the LHC. Large uncertainties beset their study: particles are accelerated in unknown sources, of uncharted temporal and spatial distribution, by uncertain acceleration mechanisms. This results in an unknown chemical composition and injection energy spectrum. The fact that they travel through intergalactic space, deflected by poorly known galactic and intergalactic magnetic fields and interacting with background photons with unclear energy density, further obfuscates the question.

There are now, however, some solid results. The cosmic ray energy spectrum is an approximately constant power law (a small deviation from it occurs at the ankle) until there is a cutoff above 40 EeV, and they arrive from nearly isotropic directions. The highest energy cosmic ray flux does not contain a significant percentage of either photons or neutrinos, so they are not produced by the decay of supermassive particles. In fact they are light nuclei up to the highest energies, where the composition is still unknown. But the questions still outweigh the answers.

The only means by which a clearer picture can be achieved is studying, in increasingly finer detail, the only objects measurable at Earth: extensive air showers produced by high energy cosmic rays. This thesis is focused on a new observable never before measured by Auger or any other experiment, the shape of the longitudinal profile of high energy showers.

Firstly, a study of the behaviour of the two shape parameters in simulations is done. It is shown that their values are sensitive to the properties of hadronic interactions at the highest energies and to the primary composition - showers from iron nuclei develop faster, and have thus a larger R, than those from proton. It is also shown that while two methods of building average profiles, in shower age and in traversed depth, may seem sensible, the former is mostly dominated by values of  $X_{\max}$  and thus not independent from that previously measured variable.

Following, in chapter 4, the longitudinal profile shape of air showers measured in Auger is presented. For each energy interval, the individual profiles are summed after being normalized to their respective maximum and translated by  $X_{\max}$ . In this way, the information from the two shape parameters is isolated in the average longitudinal profile shape. The values of L and R obtained are, in general, compatible with hadronic model predictions: L is within the values expected for proton at all energies, while at low energies R is bracketed by the values predicted for proton and iron by all models. At the highest energies, however, R is above the values expected for any primary particle.

These variables are affected by several systematic uncertainties, which are detailed in chapter 5. The relative importance of each contribution varies significantly between both variables. The largest uncertainty in L comes from its variation as

a function of the estimated percentage of scattered light in the shower, while R is very sensitive to the aerosol distribution in the atmosphere.

In chapter 6, the interpretation of these results, together with other measured Auger variables, is done. The average shape for models which implement alternative interaction properties at the highest energy, mainly in order to increase the number of muons, is studied. Data can already constrain the parameters for some of these models, which predict steeper changes with energy than standard hadronic models. Also, the sensitivity of the average shape to changes in the main macroscopic hadronic interaction parameters - multiplicity, elasticity, charged pion ratio - was estimated. A Heitler-Matthews inspired model was implemented, in which the three above mentioned quantities are the defining variables. Two further ingredients were added: the decay of charged pions, a process which competes with interaction, and a pion energy spectrum. For nominal values of the parameters, the model has a good agreement in the main variables with standard hadronic models. Variations in elasticity and multiplicity around this central value lead to subtle changes in the profile shape, while modest variations in the charged pion ratio can lead to sizable differences in the shower asymmetry parameter, R.

To improve the longitudinal profile measurement in the future, two factors are important. The first is increasing the average measured track length. The way to do this is dependent on the shower energy. At lower energies, the upper threshold of the telescope field-of-view is too low for close, not very penetrating showers. The HEAT extension, increasing the FOV up to  $60^\circ$  greatly improves the measurement of longitudinal profiles below  $10^{18}$  eV. As some details in its calibration and profile fit were not fully understood while this thesis was being prepared, we chose not to analyze this data, but its results in the future should be a great addition to this work. At higher energies, showers are mostly farther away and the main limitations are low statistics at the profile tails and the field-of-view cutoff near ground level for deeply penetrating showers. The latter is hard to circumvent, the only option being to go to lower altitudes. The former, and most important since the beginning of the shower carries more information, can only be solved by either more statistics, allowing stricter event selection, or, more likely, by better photomultiplier cameras and overall apparatus, allowing us to go to lower thresholds. The other factor that limits this analysis is the overall largest systematic uncertainty, the understanding of light scattering in the atmosphere. Light traveling from the shower to the telescope has to cross a medium which has a quickly changing temporal and spatial distribution of particles and particulates. While this is not very important for the measurements of the surface detector, which tracks comparatively high energy muons and electromagnetic particles, it is fundamental for the longitudinal profile, which is built from detected fluorescence photons which are considerably attenuated in their journey. In particular, a better understanding of the aerosol spatial and temporal distribution is needed, since the fact that this vastly changes with altitude can skewer the longitudinal profile, altering its measured asymmetry.

There is now a complete description of the electromagnetic longitudinal profile of UHECR in terms of its four parameters. The two new variables introduced -

R and L - can be used to test the consistency of our shower description, helping us move one step closer to understanding the chemical composition and hadronic interaction properties of high energy cosmic rays.



---

## APPENDIX A

# Event-by-event contribution to average shape and outliers

We cannot measure the shape event-by-event. However, we can build the average profile with all events except one, and do this for all events in each energy bin. Doing this we can get an idea of the single event contribution to the average shape. This is not a measurement of the standard deviation of the R and L distribution, and we do not want to take a physical variable out of this study. Here we are particularly interested in checking whether the shower-to-shower fluctuation decreases with energy as expected and, particularly in the highest energy bins, where R was measured above MC predictions, understanding whether this is caused by a few events with a very strange shape that distort the profile, or whether the showers are all on average consistent with the measured value. So, first we take the standard average profile for each energy bin and take its R and L value. Then, we construct N profiles, where for each of these one event (different in each) is not used. Then we compute the difference between each of these profiles and the standard one. The resultant distributions are shown in figure A.1.

We can see data and MC actually agree very well, as the width of the distributions around 0 follows closely the one in simulations. However, we can also see some events which change the average profile by a large amount. These type of events are not present in simulation, so we decide to look at them to see if something strange happens during reconstruction. At the highest energy, there are two events clearly out of the R distribution, one upwards at 0.005 and one downwards at -0.007. Also, we decide to look at the previous energy bin, and select the event further from 0 at R. Curiously, all of them are from the Coihueco site. So, we look carefully at the following three events:

- 7436826 - stereo event, profile in Coihueco is the outlier. Energy is  $10^{19.8}$  eV and singlehandedly changes R by 0.005 and L by 0.63
- 12018427 - seen by 3 eyes, profile in Coihueco is the outlier. Energy is  $10^{19.7}$  eV and singlehandedly changes R by -0.007 and L by -0.31
- seen only by Coihueco, energy is  $10^{19}$  eV. Changes R by 0.003 and L by 0.2

The first event, 7436826, is shown in figures A.2 and A.3. On the first figure we can see the profiles seen by Coihueco and Los Leones superimposed, and the reconstructed energies and directions and the right. Although the measured energy is the same in both sites this is deceptive, as this is achieved by multiplying two

## Appendix A. Event-by-event contribution to average shape and outliers

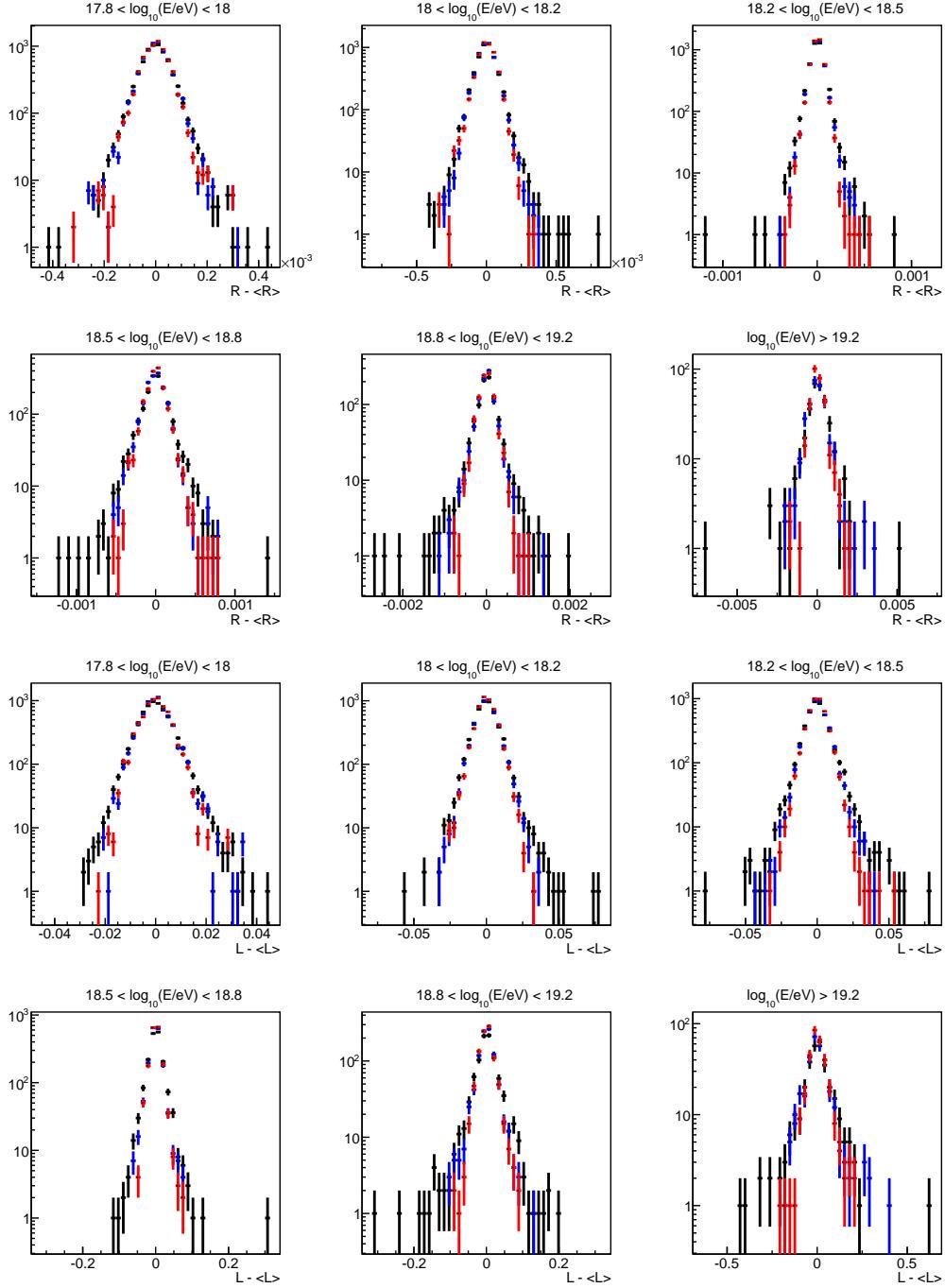


Figure A.1: Top 6:  $R - \langle R \rangle$  for the 6 energy bins. Bottom 6:  $L - \langle L \rangle$  for the 6 energy bins. Data in black, proton in blue, iron in red

incompatible maximums by two incompatible shape integrals, which in the end evens out. Looking in more detail at the Coihueco reconstruction in figure A.3 we can see there are indeed some strange features. There are some pixels along the track that were removed during the geometry fit, which seems to indicate some trouble with their timing. Also the profile does not seem very continuous: it starts rising very fast in the first bins, then more slowly for the next ones, then has a large jump around 600, especially given the small errors, and it is spiky near the maximum. The reconstructed shape is also strange. With  $\lambda = 39 \pm 4 \text{ g/cm}^2$  it is much thinner than usual and has a  $X_0$  of  $-515 \pm 112$ , which is much smaller than any other measured profile in data, or any seen in simulations. But, no clear reason to cut this event can be found.

Next we look at the other high energy event, 12018427 (figure A.4). This is almost certainly a shower measured in bad atmospheric conditions. A lot of the pixels had to be discarded during geometry fit, and can be seen in gray color. Also, the profile is very non-continuous by eye, although it ends up passing the  $\chi^2$  cut somehow. But, since we do have a cloud map and Lidar measurement for this event, and they are both 0, and the VAOD is also small (0.01), we do not really understand what could cause the strange energy deposit shape and timing problems with the pixels.

Finally, let's look at the lower energy event, 7923232 (figure A.5). Again, this is almost certainly measured during poor atmospheric conditions, as spikes can be seen throughout the profile. In this case actually only a cloud map measurement is available, and it shows there were clouds over the array, just not in the specific place where this shower crossed. So, it is possible the cloud coverage in FOV was mis-estimated.

These events have indeed some strange features, and by eye look like they were measured in poor atmospheric conditions. However, no clear cut criteria was found. If we take them away from our sample we can see two things:

- The distributions in data and MC become almost identical - this is another validation of our analysis, as a large difference between data and simulations would show that there are things we are not accounting for
- removing the events far from 0 the average value of R and L do not change significantly (within statistical error). This is clear just looking at the distributions, as they are not skewed. Here we were mainly worried about the last energy bin where we see a larger R than predictions, which we thought might be caused by a few very asymmetric events distorting the sample. However, as we can see this is not the cause

## Appendix A. Event-by-event contribution to average shape and outliers

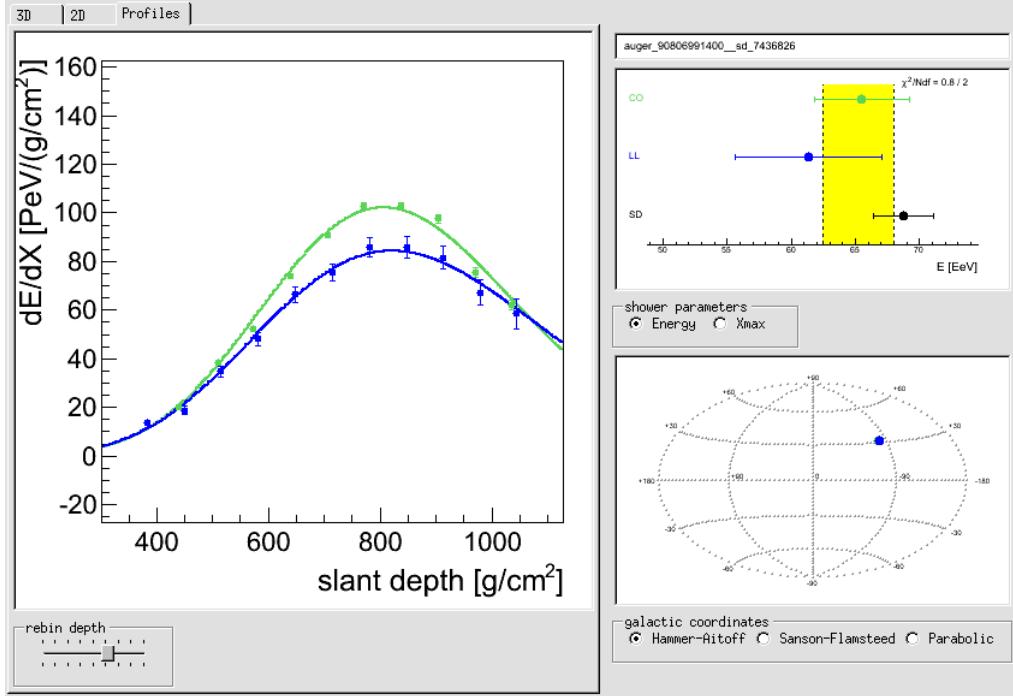


Figure A.2: SD event 7436826

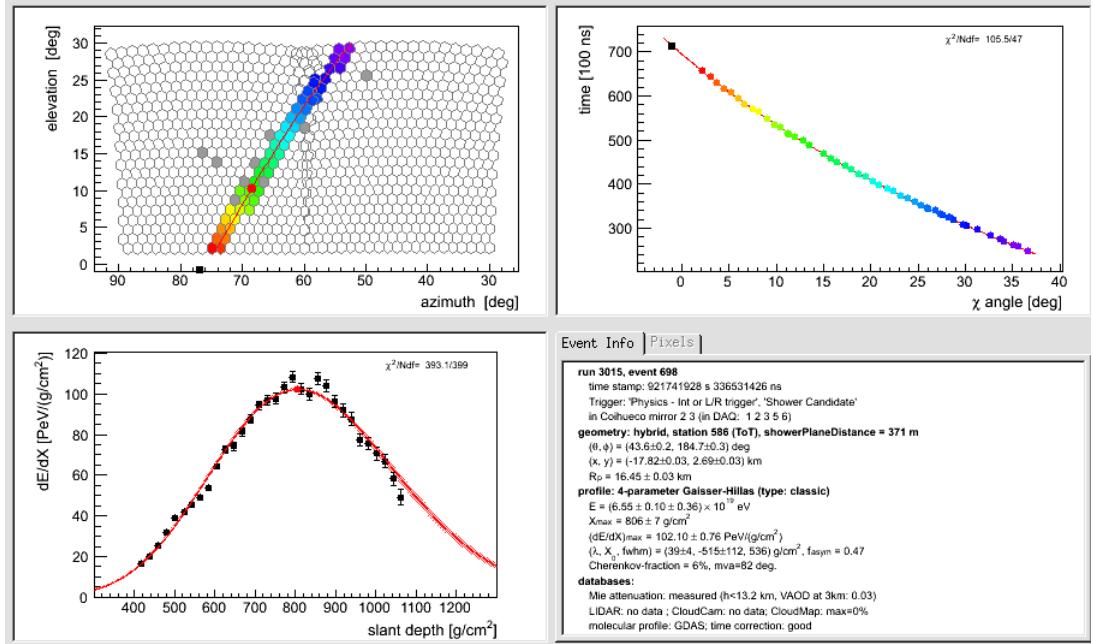


Figure A.3: SD event 7436826

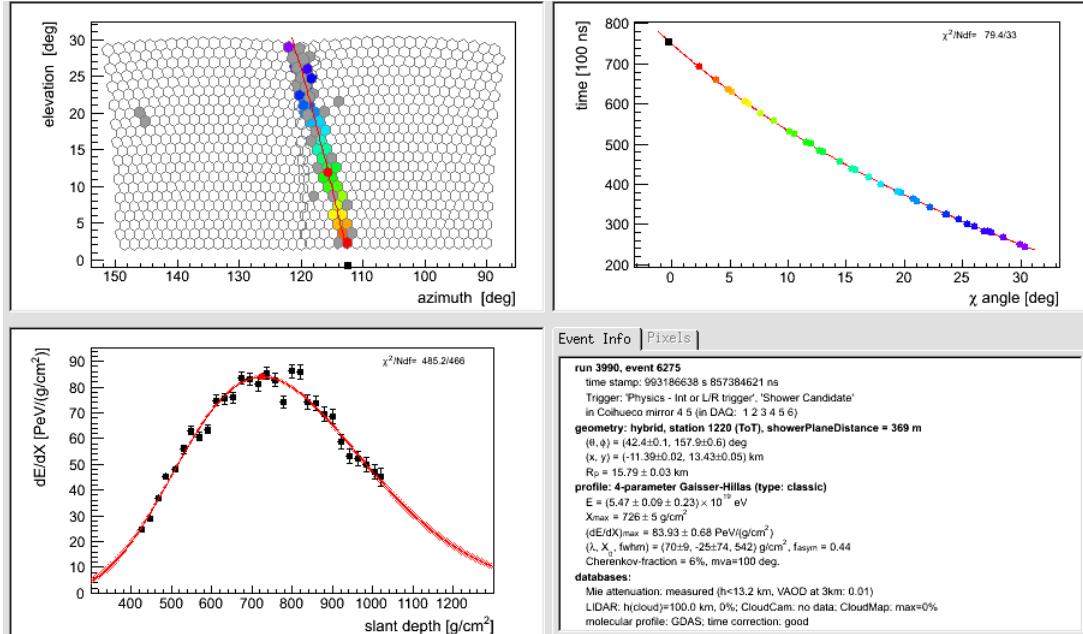


Figure A.4: SD event 12018427

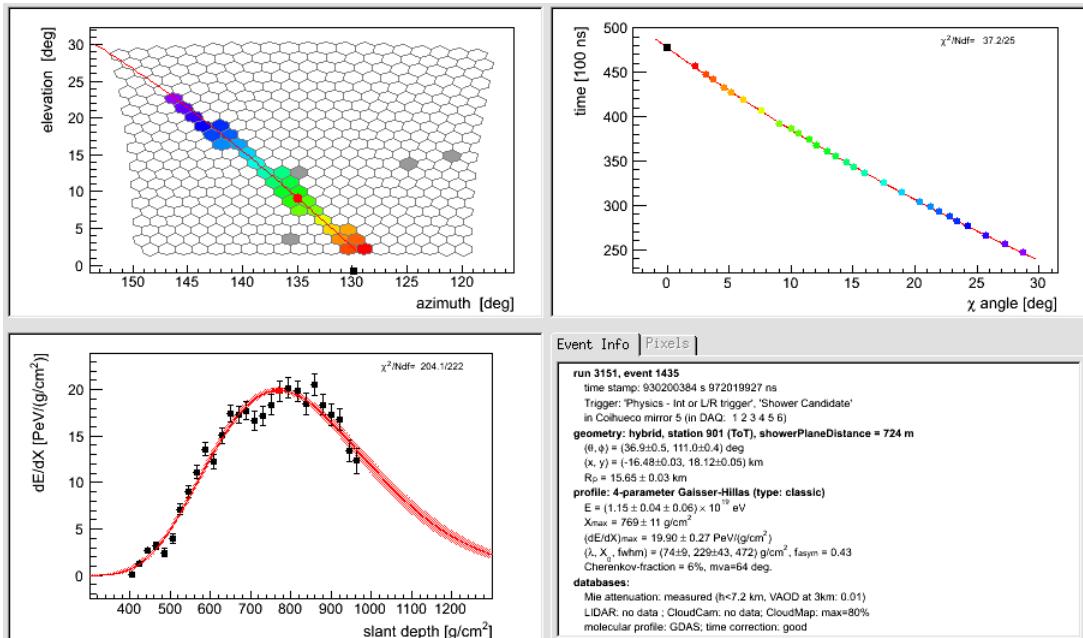


Figure A.5: SD event 7923232



# APPENDIX B

## Events with large Time-Fit $\chi^2$ - alternative geometry reconstruction

---

While looking at some control plots, we came across two events with a much larger time-fit  $\chi^2$  than all the rest. This indicates a very sporadic reconstruction problem, as it only affects less than 1 out of each 10000 events.

The algorithm used to calculate the shower detector plane (SDP) works in the following way:

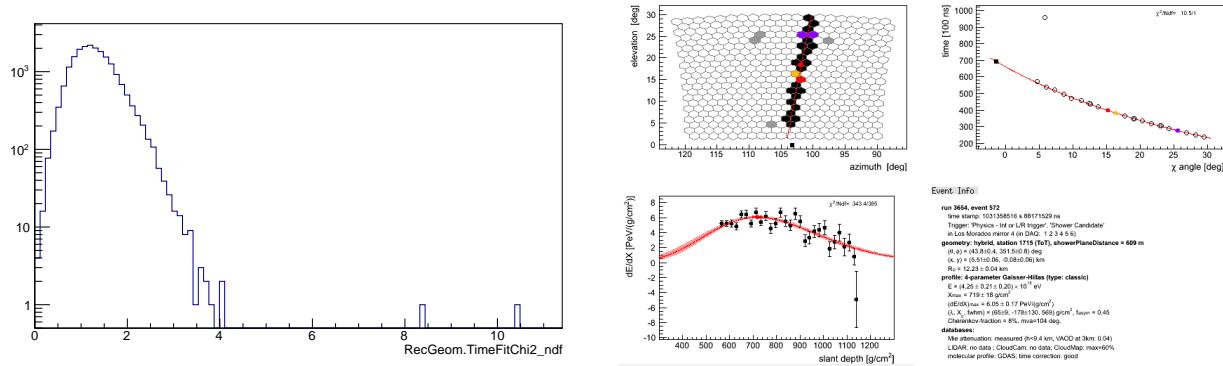
- a rough first guess of the optimal SDP is performed: all planes defined by pairs of triggered pixels are calculated, and the one which contains more pixels in it is taken as the first guess. All pixels outside of the track bed - a  $1.5^\circ$  band around this line (in the camera) - are not considered as part of the reconstruction chain from now on, i.e., they are not added to the SDP pixels list.
- The angle in the SDP is calculated for each point from the pixel pointing direction and the measured arrival time.
- The SDP is fitted: the difference between the measured time (weighed by the uncertainty in the centroid of the pulse signal) of all time bins in the SDP pixels list and the one expected for a given shower-detector plane (considering particles travel at the speed of light) is minimized, and a new SDP is found.
- this means some pixels that were incompatible with the previous SDP may now be compatible. So, all pixels outside the current SDP pixel list are tested for their time and spatial consistency with the last fitted plane.
- if it is found that one of them is now compatible, it is added to the SDP pixel list and the fit is repeated. This step and the previous one are repeated iteratively.
- when no single pixel can be readmitted in the previous, the final shower-detector plane and its constituting pixels are saved.

The time fit  $\chi^2$ , shown in figure ??, is the result of the last fit. The distribution, as expected, has a peak around 1 and the vast majority of events are below 3.

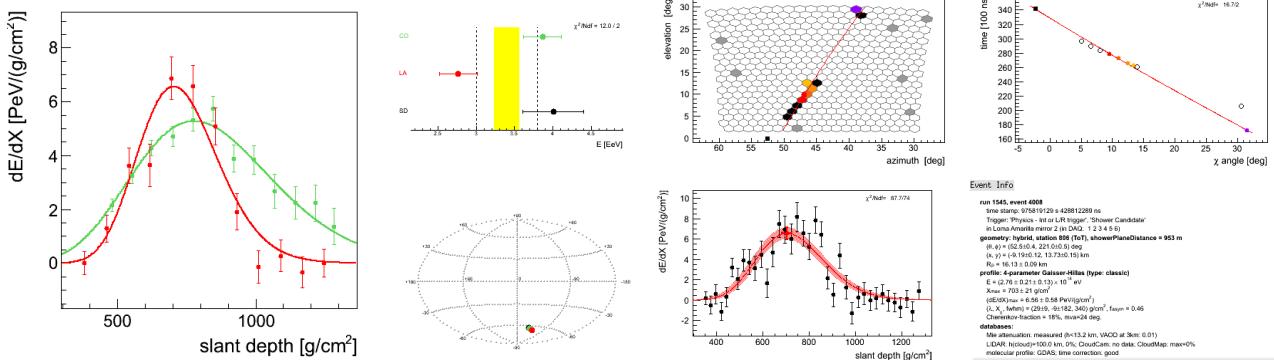
## Appendix B. Events with large Time-Fit $\chi^2$ - alternative geometry reconstruction

---

168



(a) Left: Time Fit  $\chi^2$  distribution for all events in the  $X_{\max}$  analysis. Right: Image in camera (top left), time fit (top right), and reconstructed profile and its results (bottom) for the Los Morados site in event 20075748.



(b) Event browser for event 10732320, with the profile reconstructed in Coihueco and Loma Amarilla sites (left), and in detail the Loma Amarilla reconstruction (right).

However, two standout, with  $\chi^2$  values of 8.5 and 10.5. The two events are shown in detail in figure ?? . In event 20075748, the reason for the poor fit is clear: there is a point at 95000 ns that is outside the fit line and was not cut. The reason for this is that one of the pixels around it, but outside the track-bed has a temporal correspondence with it. So, in this case the problem is that the pixel in question does not have a temporal correspondence with any of the SDP pixels, only with one outside this list. Our solution was to change the procedure to readmit pixels only if there is a temporal correspondence with a pixel within the shower-detector plane. The results are shown in figure B.2. The reconstruction seems much improved, with both the profile and the time fit  $\chi^2$  around 1, and almost all pixels in the track-bed compatible with the fit. The profile parameters now also have standard values,  $\lambda = 60 \text{ g cm}^{-2}$  and  $X_0 = -140 \text{ g cm}^{-2}$ . So, with the new algorithm this event is not anymore an outlier according to any of our control parameters.

The other event with a large time-fit  $\chi^2$  is 10732320, shown in the bottom of figure ?? . This case is very similar to the other one, in the sense that to include a point in the fit (the one with  $t = 18000$  ns in the figure), other five pixels were rendered incompatible, which is not the intended outcome of the algorithm. The reason for this is also very similar: there is a pixel outside the SDP pixel list that corresponds in time to this one, so it passes that condition. Applying the same algorithm to this event yields the results in figure B.2. Again, there seems to be an improvement: all pixels are now compatible with the fit and the  $\chi^2$  is around one. The profile, which before was very thin ( $\lambda = 30 \text{ g cm}^{-2}$ ), now has a standard value of  $\lambda = 64 \text{ g cm}^{-2}$ .

**Appendix B. Events with large Time-Fit  $\chi^2$  - alternative geometry reconstruction**  
**170**

---

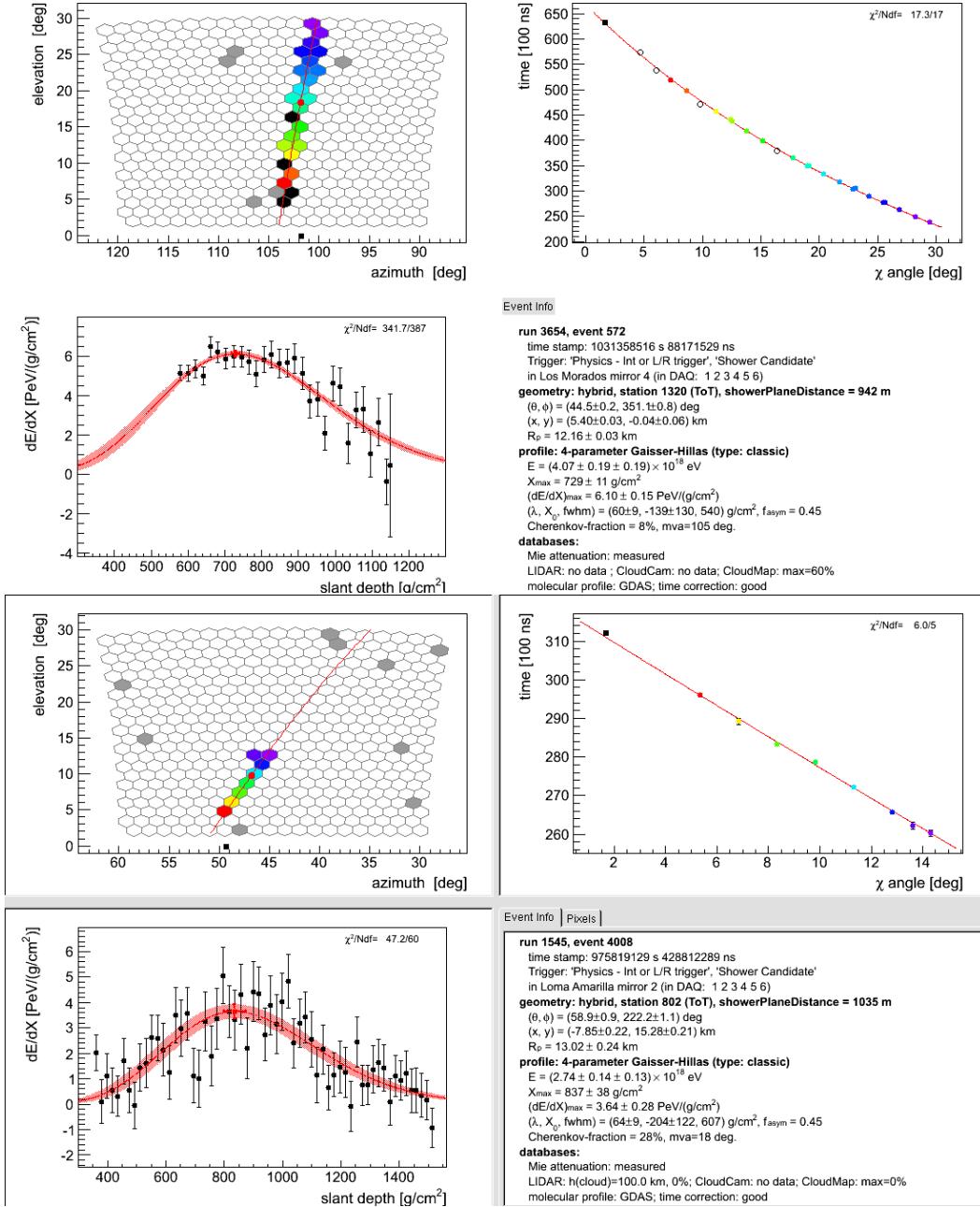


Figure B.2: Reconstruction with altered procedure for event 20075748 (top) and 10732320 (bottom)

# Bibliography

- [1] V.F. Hess, "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten", Phys. Zeitschr., XIII:1084, 1912 (Cited on page 5.)
- [2] C.D. Anderson, "The Positive Electron", Phys.Rev., 43:491-494, 1933. (Cited on page 5.)
- [3] S.H. Neddermeyer and C.D. Anderson, "Note on the Nature of Cosmic Ray Particles", Phys.Rev., 51:884-886, 1937 (Cited on page 5.)
- [4] P. Auger, P. Ehrenfest, R. Maze, J. Daudin, and A.F. Robley, "Extensive cosmic ray showers" Rev.Mod.Phys., 11:288-291, 1939 (Cited on page 5.)
- [5] B. Rossi, "Misure sulla distribuzione angolare di intensità della radiazione penetrante all'Asmara", Ric. Sci. Suppl., 1:288, 1934 (Cited on page 5.)
- [6] J. Linsley, "Measurement of multiply charged cosmic rays by a new technique", Phys.Rev., 97:1292-1302, 1955 (Cited on page 6.)
- [7] J. Linsley, "Evidence for a primary cosmic-ray particle with energy  $10^{20}$  eV", Phys.Rev.Lett., 10:146-148, 1963 (Cited on page 6.)
- [8] K. Suga, Proc. 5th Interamerican Seminar on Cosmic Ray, La Paz, Bolivia, vol.II:49, 1962 (Cited on page 6.)
- [9] A.E. Chudakov, Proc. 5th Interamerican Seminar on Cosmic Ray, La Paz, Bolivia, vol.II:44, 196 (Cited on page 6.)
- [10] T. Abu-Zayyad et al, "The prototype high-resolution Fly's Eye cosmic ray detector", Nucl.Instrum.Meth., A450:253-269, 2000. (Cited on page 6.)
- [11] J. Abraham et al, "Properties and performance of the prototype instrument for the Pierre Auger Observatory", Nucl.Instrum.Meth., A523:50-95, 2004 (Cited on page 6.)
- [12] S. Ogio et al, Proceedings of the 30th International Cosmic Ray Conference (ICRC2007), Merida, Mexico, 2007 (Cited on page 6.)
- [13] C. Baus, "Study of hadronic interactions using longitudinal shower profiles of the Pierre Auger Observatory" Master thesis, 2011 (Cited on page 7.)
- [14] KASKADE-Grande Collaboration, "KASKADE-Grande measurements of energy spectra for elemental groups of cosmic rays", Astropart. Phys. 47 (2013) 54 (Cited on page 7.)
- [15] J. Abraham et al, "Measurement of the energy spectrum of cosmic rays above  $10^{18}$  eV using the Pierre Auger Observatory", Phys. Lett. B, 685:239-246, 2010 (Cited on pages 7 and 8.)
- [16] A. Aab et al, "The Pierre Auger Observatory: Contributions to the 33rd International Cosmic Ray Conference (ICRC 2013)", In: eprint (2013). arXiv:1307.5059 (Cited on page 8.)
- [17] R.U. Abbasi et al, "Observation of the ankle and evidence for a high-energy break in the cosmic ray spectrum", Phys. Lett. B, 619:271-280, 2005 (Cited on page 8.)
- [18] T. Abu-Zayyad et al, "Energy Spectrum of Ultra-High Energy Cosmic Rays observed with the Telescope Array using a hybrid technique", Astrop. Phys., 61:93-101, 2015 (Cited on page 8.)
- [19] D. Allard, "Propagation of extragalactic ultra-high energy cosmic-ray nuclei: implications for the observed spectrum and composition", Proc. of the Rencontres de Moriond 2009, La Thuile, Italy, arXiv:0906.3156 (Cited on page 9.)
- [20] A.A. Penzias and R.W. Wilson, "A Measurement of excess antenna temperature at 4080 Mc/s", Astrophys.J., 142:419-421, 1965. (Cited on page 8.)
- [21] K. Greisen, "End to the cosmic ray spectrum?", Phys.Rev.Lett., 16:748-750, 1966. (Cited on page 8.)
- [22] G.T. Zatsepin and V.A. Kuzmin, "Upper limit of the spectrum of cosmic rays", JETP Lett., 4:78-80, 1966. (Cited on page 8.)

- [23] M. Kachelriess, "The rise and fall of top-down models as main UHECR sources", Proc. of the XXth Rencontres de Blois, Blois, France, arXiv:0810.3017 [astro-ph]. (Cited on page 10.)
- [24] E. Fermi, "On the origin of cosmic radiation", Phys. Rev. 75: 1169-1174, 1949 (Cited on page 10.)
- [25] R. D. Blandford and J. P. Ostriker, "Particle acceleration by astrophysical shocks", Astrophys. J. 221 (1978) L299-L32 (Cited on page 10.)
- [26] H. Blumer and K.H. Kampert, "Die Suche nach den Quellen der kosmischen Strahlung", Physikalische Blätter, vol. 56: 39-45, 2000 (Cited on page 11.)
- [27] A.M. Hillas, "The Origin of Ultrahigh-Energy", Ann.Rev.Astron.Astrophys., 22:425-444, 1984 (Cited on page 11.)
- [28] E. Waxman, "Cosmological Gamma-Ray Bursts and the Highest Energy Cosmic Rays", Phys.Rev. Lett. 75 (1995) 386. (Cited on page 11.)
- [29] P. Blasi, R. Epstein, A. V. Olinto, "Ultra-High-Energy Cosmic Rays from Young Neutron Star Winds", ApJ 533 (2000) L123 (Cited on page 11.)
- [30] G. R. Farrar, A. A. Berlind, D. V. Hogg, "Foreground and Source of a Cluster of Ultra-High-energy Cosmic Rays", ApJ 642 (2006) L89. (Cited on page 12.)
- [31] 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. (Cited on page 12.)
- [32] V. Verzi for the Pierre Auger Collaboration, "Cosmic Rays: air showers from low to high energies", In PoS(ICRC2015), 2015. (Cited on page 12.)
- [33] Pierre Auger Collaboration, C. Di Giulio, "The Pierre Auger Observatory Upgrade Program", presentation at CRIS 2015, Gallipoli, Italy (2015). (Cited on page 13.)
- [34] J. Abraham, et al, "Correlation of the highest-energy cosmic rays with the positions of nearby active galactic nuclei", Astropart. Phys., 29:188-204, April 2008. (Cited on page 13.)
- [35] 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. (Cited on page 13.)
- [36] Telescope Array Collaboration, K. Kawata, "Ultra-High-Energy Cosmic-Ray Hotspot Observed with the Telescope Array Surface Detectors", Proc. 34th ICRC (2015) 276. (Cited on page 13.)
- [37] Pierre Auger Collaboration, J. Aublin, "Arrival directions of the highest-energy cosmic rays detected with the Pierre Auger Observatory", Proc. 34th ICRC (2015) 310, [arXiv:1509.03732]. (Cited on page 13.)
- [38] E. Barcikowski et al, "Mass Composition Working Group Report at UHECR-2012", In: EPJ Web Conf. 53 (2013), p. 01006. doi: 10.1051/epjconf/20135301006. arXiv: 1306.4430 [astro-ph.HE]. (Cited on page 14.)
- [39] M. Risse, "Properties of extensive air showers", Acta Phys. Polon., B35:pp.1787-1797, 2004, astro-ph/0402300 (Cited on page 17.)
- [40] P.M.S. Blackett and G.P.S. Occhialini, "Some Photographs of the Tracks of Penetrating Radiation", Proc. of the Royal Society of London, Ser. A, Vol. 139, no. 839, 1933, pp:699-720+722+724+726 (Cited on page 16.)
- [41] T.K. Gaisser, "Cosmic rays and particle physics", Cambridge University Press, 1990 (Cited on page 17.)
- [42] W. Heitler, "The Quantum Theory of Radiation", Oxford University Press, 1954 (Cited on page 17.)
- [43] A. B. Migdal, "Bremsstrahlung and pair production in condensed media at high-energies", Phys. Rev., 103:pp. 1811-1820, 1956 (Cited on page 19.)
- [44] A. N. Cillis et al, "Influence of the LPM effect and dielectric suppression on particle air showers", Phys. Rev. D, 59(11):p. 113012, 1999 (Cited on page 19.)

- [45] K. Greisen, "The Extensive Air Showers", *Prog. in Cosmic Ray Physics*, III, 1 (1956) (Cited on pages 19, 36, 43 and 141.)
- [46] J. Matthews, "A Heitler model of extensive air showers", *Astroparticle Physics* 22 (2005) 387-397 (Cited on pages 19 and 141.)
- [47] J. Alvarez Muniz et al, "Hybrid simulations of extensive air showers", *Phys. Rev. D* 66:p. 033011, 2002, astro-ph/0205302 (Cited on page 20.)
- [48] A. M. Hillas, "Shower simulation: Lessons from MOCCA", "Nucl. Phys. Proc. Suppl.", 52B,:pp. 29-42, 1997 (Cited on page 22.)
- [49] S. Sciutto, "AIRES: A System for air shower simulations. User's guide and reference manual. Version 2.2.0", 1999, <http://www.fisica.unlp.edu.ar/auger/aires>, astro-ph/9911331 (Cited on page 22.)
- [50] D. Heck et al, "CORSIKA: A Monte Carlo Code to simulate extensive air showers", 1998 (Cited on page 22.)
- [51] T. Pierog et al, "First results of fast one-dimensional hybrid simulation of EAS using CONEX", *Nucl. Phys. Proc. Suppl.*, 151:pp. 159-162, 2006, astro-ph/0411260 (Cited on page 22.)
- [52] H. J. Drescher and G. R. Farrar, "Air shower simulations in a hybrid approach using cascade equations", *Phys. Rev. D* 67:p., 116001, 2003, astro-ph/0212018 (Cited on page 22.)
- [53] R. Conceição, "Sensitivity to Cosmic Ray composition and Hadronic Models from the Fluorescence Detector Data of the Pierre Auger Observatory", Doctoral Thesis (Cited on pages 37 and 38.)
- [54] S. Ostapchenko, *Non-linear screening effects in high energy hadronic interactions*, *Phys. Rev.*, D74:p. 014026, 2006, doi:10.1103/PhysRevD.74.014026, hep-ph/0505259 (Cited on page 23.)
- [55] S. Ostapchenko, *On the re-summation of enhanced Pomeron diagrams*, *Phys. Lett.*, B636:pp. 40–45, 2006, doi:10.1016/j.physletb.2006.03.026, hep-ph/0602139 (Cited on page 23.)
- [56] K. Werner, F.-M. Liu and T. Pierog, *Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron-gold collisions at RHIC*, *Phys. Rev.*, C74:p. 044902, 2006, doi:10.1103/PhysRevC.74.044902, hep-ph/0506232 (Cited on page 23.)
- [57] T. Pierog and K. Werner, *The hadronic interaction model EPOS and air shower simulations: New results on muon production*, *Proc. of the 30th Int. Cosmic Ray Conf.*, 2007 (Cited on page 23.)
- [58] J. Engel, T. K. Gaisser, T. Stanev and P. Lipari, *Nucleus-nucleus collisions and interpretation of cosmic ray cascades*, *Phys. Rev.*, D46:pp. 5013–5025, 1992, doi:10.1103/PhysRevD.46.5013 (Cited on page 23.)
- [59] R. Engel, T. K. Gaisser, T. Stanev and P. Lipari, *Air shower calculations with the new version of SIBYLL*, prepared for 26th International Cosmic Ray Conference (ICRC 99), Salt Lake City, Utah, 17–25 Aug 1999 (Cited on page 23.)
- [60] R. S. Fletcher, T. K. Gaisser, P. Lipari and T. Stanev, *SIBYLL: An Event generator for simulation of high-energy cosmic ray cascades*, *Phys. Rev.*, D50:pp. 5710–5731, 1994, doi:10.1103/PhysRevD.50.5710 (Cited on page 23.)
- [61] L. Durand and P. Hong, *Phys. Rev. Lett.*, 58, 303, 1987 (Cited on page 23.)
- [62] V. N. Gribov, *Sov. Phys. JETP*, 26, 414, 1968. (Cited on page 23.)
- [63] S. Ostapchenko, "Monte Carlo treatment of hadronic interactions in enhanced pomeron scheme: QGSJet-II model", *Phys. Rev.*, vol. D83 (014018), 2011 (Cited on page 24.)
- [64] T. Pierog et al, "EPOS-LHC: test of collective hadronization with LHC data", Prepared for submission to JHEP (Cited on page 24.)
- [65] E.-J. Ahn et al, "Cosmic ray interaction event generator SIBYLL 2.1", *Phys. Rev.*, vol. D80 (094003), 2009 (Cited on page 24.)
- [66] R. Ulrich, "Measurement of the proton-air cross section using hybrid data of the Pierre

- Auger Observatory", Ph.D. Thesis, 2007 (Cited on page 23.)
- [67] T. Pierog, "Air Shower Simulation after LHC", QCD Cosmic Woorkshop, Paris, France, May 2013 (Cited on page 25.)
- [68] R. Ulrich for the Pierre Collaboration, "Extension of the measurement of the proton-air cross section with the Pierre Auger Observatory", ICRC2015, arXiv:1509.03732 [astro-ph.HE]. (Cited on page 26.)
- [69] M. Bohacova (AIRFLY), "Study of the air fluorescence by AIRFLY", Nucl. Phys., Proc. Suppl., 190:pp. 266-271, 2009, doi:10.1016/j.nuclphysbps.2009.03.098 (Cited on page 28.)
- [70] T. Waldenmaier, J. Blumer, D. M. Gonzalez and H. Klages, "Measurement of the air fluorescence yield with the AirLight experiment", Nucl. Instrum. Meth., A597:pp. 67-74, 2008, doi:10.1016/j.nima.2008.08.047 (Cited on page 28.)
- [71] M. Nagano, K. Kobayakawa, N. Sakaki and K. Ando, "New measurement on photon yields from air and the application to the energy estimation of primary cosmic rays", Astropart. Phys., 22:pp. 235-248, 2004, doi:10.1016/j.astropartphys.2004.08.002, astro-ph/0406474 (Cited on page 28.)
- [72] R. M. Baltrusaitis et al., "The UTAH FLY's Eye Detector", Nucl. Instrum. Meth., A240:pp. 410-428, 1985, doi:10.1016/0168-9002(85)90658-8 (Cited on pages 29 and 31.)
- [73] M. Ave et al, "Precise measurement of the absolute fluorescence yield of the 337 nm band in atmospheric gases", Astr. Phys. 42, pp.90-102, astro-ph/1210.6734 (Cited on page 29.)
- [74] A. Bucholtz, "Rayleigh scattering calculations for the terrestrial atmosphere", Applied Optics 34, 2765 (1995) (Cited on page 93.)
- [75] Pierre Auger Collaboration, A. Aab *et al.*, *The Pierre Auger Cosmic Ray Observatory, accepted for publication in Nucl. Instrum. Meth. A* (2015), [arXiv: 1502.01323]. (Not cited.)
- [76] T.K. Gaisser and A.M. Hillas, *Proc. 15th ICRC, Plovdiv, Bulgaria* **8** (1977) 353. (Cited on page 36.)
- [77] M. Unger *et al.*, *Nucl. Instrum. Meth. A* **588** (2008) 433, [arXiv: 0801.4309]. (Not cited.)
- [78] The HiRes/MIA Collaboration, T. Abu-Zayyad *et al.*, *Astropart. Phys.* **16** (2001) 1-11. (Cited on pages 35, 36, 42 and 44.)
- [79] G. Hughes for the High Resolution Fly's Eye Collaboration, *Proc. 30th ICRC, Merida, Mexico* **4** (2007) 405-408. (Cited on pages 35, 36, 37, 42, 43 and 46.)
- [80] S. Andringa, R. Conceição, and M. Pimenta, *Astropart. Phys.* **34** (2011) 360-367. (Cited on pages 35, 49, 69 and 70.)
- [81] R. Conceição *et al.*, *24th European Cosmic Ray Symposium* (2014) (to appear in JPCS). (Not cited.)
- [82] J.A.J. Matthews *et al.*, *J. Phys. G* **37** (2010) 025202, [arXiv: 0909.4014]. (Cited on page 39.)
- [83] Pierre Auger Collaboration, A. Aab *et al.*, *Phys. Rev. D* **90** (2014) 122005. (Cited on pages 55, 56 and 91.)
- [84] Alessio Porcelli, "Measurement of the depth of shower maximum in the transition region between galactic and extragalactic cosmic rays with the Pierre Auger Observatory", Doctoral Thesis (Cited on page 56.)
- [85] V. Verzi for the Pierre Auger Collaboration, *Proc. 33rd ICRC, Rio de Janeiro, Brazil* (2013), [arXiv: 1307.5059]. (Cited on page 106.)
- [86] F. Meyer, F. Vernotte, *Time tagging tests in Besancon Observatory*, GAP 2001-050 (Cited on page 56.)
- [87] J. Alvarez-Muñiz *et al.*, Muon production and string percolation effects in cosmic rays at the highest energies, [arXiv: 1209.6474v1] (Cited on pages 131 and 132.)
- [88] A. Rodrigues, R. Ugoccioni, J. Dias de Deus, Percolation approach to phase transitions in high-energy nuclear collisions, *Phys.Lett. B*458 (1999) 402-406 [arXiv: 9812364] (Cited on page 132.)

- [89] A. Capella, U. Sukhatme, C.I. Tan, J. Tran Thanh Van, Dual parton model, Phys.Rept. 236 (1994) 225-329. doi:10.1016/0370-1573(94)90064-7. (Cited on page 131.)
- [90] A. Aab *et al.* [Pierre Auger Collaboration], "Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above  $10^{17.8}$  eV", Phys. Rev. D **90** (2014) 12, 122005 (Cited on pages 14, 59, 62, 120, 121, 122, 123, 124, 125 and 126.)
- [91] R. Abbasi et al, "Study of Ultra-High Energy Cosmic Ray Composition Using Telescope Array's Middle Drum Detector and Surface Array in Hybrid Mode", In: Astropart. Phys. 64 (2014), p. 49. doi:10.1016/j.astropartphys.2014.11.004. arXiv: 1408.1726 [astro-ph.HE] . (Cited on page 14.)
- [92] T. Abu-Zayyad et al, "Pierre Auger Observatory and Telescope Array: Joint Contributions to the 33rd International Cosmic Ray Conference (ICRC 2013)", In: e-print (2013). arXiv: 1310.0647 [astro-ph.HE] . (Cited on page 15.)
- [93] A. Porcelli for the Pierre Auger Collaboration, "Measurements of Xmax above  $10^{17}$  eV with the fluorescence detector of the Pierre Auger Observatory", ICRC2015, arXiv:1509.03732 [astro-ph.HE]. (Cited on pages 120, 121, 122, 123, 124 and 125.)
- [94] A. Aab *et al.* [Pierre Auger Collaboration], "Azimuthal asymmetry in the risetime of the surface detector signals of the Pierre Auger Observatory", arXiv:1604.00978 [astro-ph.HE]. (Cited on pages 120, 122, 123, 124 and 125.)
- [95] A. Aab *et al.* [Pierre Auger Collaboration], "Muons in air showers at the Pierre Auger Observatory: Measurement of atmospheric production depth", Phys. Rev. D **90** (2014) no.1, 012012, [arXiv:1407.5919 [hep-ex]]. Addendum: [Phys. Rev. D **90** (2014) no.3, 039904] Erratum: [Phys. Rev. D **92** (2015) no.1, 019903] (Cited on pages 120, 121, 122, 123, 124 and 125.)
- [96] A. Aab *et al.* [Pierre Auger Collaboration], "Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events", Phys. Rev. D **91** (2015) 3, 032003 (Cited on pages 120, 121, 122, 123, 124, 125, 126 and 153.)
- [97] L. Cazon, M. Pimenta, G. Torralba Elipe, I. Valiño and E. Zas, "Measurement of the  $N_\mu$  fluctuations using inclined showers", Auger Internal Note - GAP2015-056 (Cited on page 130.)
- [98] R. Conceição, S. Andringa, F. Diogo, and M. Pimenta, "The average longitudinal air shower profile: exploring the shape information", J. Phys. Conf. Ser. **632**, 1 (2015) 012087 (Cited on pages 48 and 129.)
- [99] S. Andringa, R. Conceição, F. Diogo, and M. Pimenta, "Sensitivity to primary composition and hadronic models from average shape of high energy cosmic ray shower profiles", arXiv:1209.6011 [hep-ph] (Cited on page 35.)
- [100] F. Diogo for the Pierre Auger Collaboration, "Measurement of the average electromagnetic longitudinal shower profile at the Pierre Auger Observatory.", ICRC2015, arXiv:1509.03732 [astro-ph.HE]. (Cited on page 64.)
- [101] L. Cazon, "Extensive Air Showers: from the muonic smoking guns to the hadronic backbone,", EPJ Web Conf. **52** (2013) 03003, [arXiv:1301.3340 [astro-ph.HE]]. (Cited on page 129.)
- [102] G. R. Farrar and J. D. Allen, "A new physical phenomenon in ultra-high energy collisions", arXiv:1307.2322v1 [hep-ph]. (Cited on pages 133 and 136.)
- [103] J. Ebr, P. Nescal, J. Ridky, "Soft Particle Production in Very High Energy Hadron Interactions", arXiv:1606.03288v1 [astro-ph.HE]. (Cited on pages 137 and 139.)
- [104] J. M. C. Montanus, "An extended Heitler-Matthews model for the full hadronic cascade in cosmic air showers", Astropart. Phys. 59, 4 (2014) [astro-ph.HE/1311.0642v3]. (Cited on pages 142, 143, 144 and 145.)
- [105] V. Novotny, D. Nosek, J. Ebr, "A branching model for hadronic air showers", Proceedings for the 34th ICRC, 2015, arXiv:1509.00364v1 [astro-ph.HE] (Cited on pages 142, 144 and 146.)

- [106] D.E. Groom, N.V. Mokhov, and S.I. Striganov, "Muon stopping-power and range tables: 10 MeV-100 TeV," *Atomic Data and Nuclear Data Tables* 78, 183-356 (2001). (Cited on page 144.)
- [107] R. Ulrich, R. Engel and M. Unger, "Hadronic Multiparticle Production at Ultra-High Energies and Extensive Air Showers", *Phys. Rev. D* 83, 054026 (2011) (Cited on pages 147, 149 and 150.)
- [108] S. Ostapchenko, *Czech. J. Phys.* 56, A149 (2006), arXiv:hep-ph/0601230 (Cited on page 147.)
- [109] R.D. Parsons et al, "Systematic uncertainties in air shower measurements from high-energy hadronic interaction models", *Astropart.Phys.*34 (2011) 832-839 [arxiv:1102.4603] (Cited on page 147.)
- [110] A. Aab for The Pierre Auger Collaboration, "Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events", *Phys. Rev. D* as doi:10.1103/PhysRevD.91.032003, arXiv:1408.1421v6 [astro-ph.HE] (Cited on page 155.)
- [111] A. Aab for The Pierre Auger Collaboration, "Muons in air showers at the Pierre Auger Observatory: Measurement of atmospheric production depth", *Phys. Rev. D* 90, 012012 (2014); *Phys. Rev. D* 92, 019903(E) (2015), arXiv:1407.5919v2 [hep-ex] (Cited on page 155.)
- [112] T. Harrison, B. Dawson, J. Bellido, "Alignment of camera 6 of Coihueco", GAP2010-104 (Cited on page 65.)
- [113] E. J. Ahn, J. Bellido, S. BenZvi, R. Engel, F. Schussler, R. Ulrich, M. Unger, "Measurement of the Depth of Shower Maximum of Cosmic Rays above 1EeV", GAP2009-078 (Cited on pages 59, 60 and 72.)
- [114] C. de Donato et al, "Using star tracks to determine the absolute pointing of the Fluorescence Detector telescopes", *Pierre Auger GAP Note*, GAP-2005-008 (Cited on page 87.)
- [115] M. Prouza et al, "Star tracking using background data of FD telescopes-whole star track optimization method", *Pierre Auger GAP Note*, GAP-2005-041 (Cited on page 87.)
- [116] P. Younk and B. Fick, "FD Telescope alignment using CLF laser shots", *Pierre Auger GAP Note*, GAP-2006-087 (Cited on page 87.)
- [117] M. Scuderi et al, "Alignment method of the Pierre Auger Observatory fluorescence telescopes based on the central laser facilities", *Pierre Auger GAP Note*, GAP-2006-032 (Cited on page 87.)
- [118] S. Falk et al, "Telescope Alignment Studies", *Pierre Auger GAP Note*, GAP-2011-123 (Cited on pages 86 and 87.)
- [119] S. Saffi and B. Dawson, "Determining the pointing directions of the Pierre Auger fluorescence telescopes using observed cosmic ray extensive air showers", *Pierre Auger GAP Note*, GAP-2012-106 (Cited on page 86.)
- [120] P. Sokolsky, "Introduction to ultrahigh energy cosmic ray physics", *Frontiers in Physics*, Vol. 76 (1989) (Cited on page 93.)
- [121] B. Dawson, "New Rayleigh Scattering Implementation in the Offline", *Pierre Auger GAP Note*, GAP-2006-007 (Cited on page 94.)
- [122] D. Kumpel et al, "Geometry Reconstruction of Fluorescence Detectors Revisited", Contribution to the 30th ICRC 07, arxiv:0708.3976v1 (Cited on page 94.)
- [123] F. Nerling et al, "Universality of electron distributions in high-energy air showers - description of Cherenkov light production", *Astropart.Phys.*24:421-437,2006, 10.1016/j.astropartphys.2005.09.002, arXiv:astro-ph/0506729 (Cited on page 95.)
- [124] A. Hillas, "Cerenkov light images of EAS produced by primary gamma", *Proc. 19nd I.C.R.C. (La Jolla)*, Vol 3, 445 (1985) (Cited on page 95.)
- [125] M. D. Roberts, "The role of atmospheric multiple scattering in the transmission of fluorescence light from extensive air showers", 2005, *J. Phys. G: Nucl. Phys.* 31, 1291 (Cited

- on page 96.)
- [126] P. Pekala et al, "Atmospheric multiple scattering of fluorescence and Cherenkov light emitted by extensive air showers", Nucl.Instrum.Meth.A605:388-398,2009, doi:10.1016/j.nima.2009.03.244, arXiv:0904.3230 [astro-ph.HE] (Cited on pages 48 and 96.)
  - [127] L. Cazon et al, "dlnA/dlogE from hadronic-type and EM-type cascade observables", Pierre Auger GAP Note, GAP-2016-022 (Cited on page 115.)
  - [128] J. Abdallah et al. [DELPHI Collaboration], Astropart. Phys. 28 (2007) 273 [arXiv:0706.2561 [astro-ph]]. (Cited on page 137.)
  - [129] L. Valore for the Pierre Auger Collaboration, "Measuring Atmospheric Aerosol Attenuation at the Pierre Auger Observatory", *Proc. 33rd ICRC, Rio de Janeiro, Brazil* (2013), [arXiv: 1307.5059]. (Cited on page 100.)
  - [130] J. Matthews, R. Mesler, B. Becker, M. Gold, J. Hague, "A Parametrization of Cosmic Ray Shower Profiles Based on Shower Width", J.Phys.G G37 (2010) 02 5202, [arXiv: 0909.4014], doi:10.1088/0954-3899/37/2/025202. (Cited on page 35.)
  - [131] M. Giller, A. Kacperczyk, J. Malinowski, W. Tkaczuk, G. Wieczorek, "Similarity of extensive air showers with respect to the shower age", J.Phys.G G31 (2005) 947-958, doi:10.1088/0954-3899/31/8/023. (Cited on pages 35 and 40.)
  - [132] M. Malacari et al, "Testing the Aerosol Databases using the Stereo Energy Balance Technique", Pierre Auger GAP Note, GAP-2016-035. (Cited on page 101.)
  - [133] S. Andringa, R. Conceição, F. Diogo, M. Pimenta, "Measurement of the shape parameter of FD average profiles in age", Pierre Auger GAP Note, GAP-2014-104. (Cited on page 35.)
  - [134] M. Giller et al, "Energy spectra of electrons in the extensive air showers of ultra-high energy", J. Phys. G30 (2004) 97. (Cited on page 35.)
  - [135] S. Lafebre et al, "Universality of electron-positron distribution in extensive air showers", Astropart. Phys. 31 (2009) 243. (Cited on page 35.)
  - [136] S. Müller for the Pierre Auger Collaboration, "Energy scale derived from fluorescence telescopes using Cherenkov light and shower universality", Contribution to the 31st ICRC 2009 (Cited on pages 35 and 41.)
  - [137] R.M. Baltrusaitis et al, "The Utah Fly's Eye Detector", Nucl. Instr. Meth. A240 410 (1985) (Cited on page 36.)
  - [138] V. Scherini et al, "Simulation study of shower profiles from ultra high energy cosmic rays", prepared for the 30th ICRC, Merida, Yucatan, Mexico, 2007 (Cited on page 37.)
  - [139] Bruno Rossi and Kenneth Greisen, Rev. Mod. Phys. 13, 240 (1941). (Cited on page 40.)
  - [140] Keilhauer B, Bohacova M, Fraga M, Matthews J, Sakaki N. et al., 2013, EPJ Web Conf. 53 01010 (Preprint 1210.1319) (Cited on page 48.)
  - [141] Pierog T. et al., 2006, Nucl. Phys. Proc. Suppl. 151-159 (Preprint astro-ph/0411260) (Cited on page 49.)
  - [142] Bergmann T. et al., 2007, Astropart. Phys. 26 420-432 (Preprint astro-ph/0606564) (Cited on page 49.)
  - [143] Ahn E J, Engel R, Gaisser T K, Lipari P and Stanev T 2009, Phys.Rev. D80 094003 (Preprint 0906.4113) (Cited on page 49.)
  - [144] Ostapchenko S, 2011 Phys.Rev. D83 014018 (Preprint 1010.1869) (Cited on page 49.)
  - [145] Werner K, Liu F M and Pierog T 2006 Phys.Rev. C74 044902 (Preprint hep-ph/0506232) (Cited on page 49.)
  - [146] Pierog T, Karpenko I, Katzy J, Yatsenko E and Werner K 2013 (Preprint 1306.0121) (Cited on page 49.)
  - [147] Andringa S, Cazon L, Conceição R and Pimenta M, 2012, Astropart.Phys. 35 821-827 (Preprint 1111.1424) (Cited on page 49.)
  - [148] P. Lipari, "The Concepts of "Age" and "Universality" in Cosmic Ray Showers",

10.1103/PhysRevD.79.063001, arXiv:0809.0190 [astro-ph] (Cited on page 44.)