



---

GRAN SASSO SCIENCE INSTITUTE  
ASTROPARTICLE PHYSICS PROGRAM  
XXXI CYCLE

RECONSTRUCTION OF EVENTS FROM THE SURFACE DETECTOR  
OF THE PIERRE AUGER OBSERVATORY  
USING AIR SHOWER UNIVERSALITY

---

PhD CANDIDATE  
ANASTASI GIOACCHINO ALEX

---

SUPERVISOR:  
PROF. PETRERA SERGIO

Co-SUPERVISOR:  
DR. SALAMIDA FRANCESCO

---



## Abstract

---

The chemical composition of ultra-high energy cosmic rays (UHECRs) is a key element to solve several open questions about the origin and propagation of these rare particles. When reaching the Earth, UHECRs interact with the atmosphere generating showers of secondary particles. The atmospheric depth of the maximum in the electromagnetic part of the shower ( $X_{\max}$ ) is a quantity commonly employed to infer information about the nature of primary cosmic rays. It is measured with good accuracy through fluorescence detection, but the related duty cycle is insufficient to extend our knowledge about composition at the highest energies, namely above  $10^{19.5}$  eV with the current experimental setups. Therefore many efforts have been put in the development of a composition analysis which makes use only of data recorded with ground arrays of surface detectors, operational nearly 100% of the time.

With this aim, in the framework of the Pierre Auger Observatory a new reconstruction procedure has been developed, based on the concept of air shower Universality. According to the Universality paradigm, extensive air showers created by ultra-high energy cosmic rays can be described, to a remarkable degree of precision, in terms of a reduced set of macroscopic parameters: shower geometry, energy,  $X_{\max}$  and muon content. The so-called Universality reconstruction consist in a best-fit of the values of these macroscopic parameters to describe the data collected by the water Cherenkov detectors of the ground array, comparing measurements and expectations from specific parameterizations of the integrated and time-dependent signals.

The work I carried out during my PhD led to two main outcomes.

The first one is an improvement of the fitting procedure already implemented in the Offline software of the Collaboration, to achieve a higher accuracy in the reconstructed quantities, in particular of the  $X_{\max}$  determination. The adjustments were guided by the comparison between the results of the Universality and the standard reconstruction for a selected set of high quality hybrid events, i.e. events measured simultaneously by the surface and fluorescence detectors. In this way an improved likelihood function and a better description of uncertainties in the time-dependent signal are achieved, allowing the definition of an optimized fitting procedure.

The second outcome is the reconstruction of the dataset collected by the surface detector between 2004 and 2016, using the optimized version of the Universality fit. The resulting values of the shower maximum are used to calculate the first two moments of the  $X_{\max}$  distributions above  $10^{19}$  eV, from which the detector independent mean and variance of the logarithmic mass,  $\ln A$ , of the primary cosmic rays reaching the Earth are obtained. Finally, the  $X_{\max}$  distributions in their entirety are used to estimate the contribution of different groups of masses (light, intermediate, heavy) and to attempt a study of the energy spectrum separated in a lighter and a heavier component.



# Contents

<b>Introduction</b>	<b>1</b>
<b>1 The cosmic rays</b>	<b>3</b>
1.1 The cosmic ray energy spectrum . . . . .	4
1.2 Production and acceleration of cosmic rays . . . . .	6
1.2.1 The Fermi model and the Diffusive Shock Acceleration . . . . .	6
1.2.2 The Hillas plot and the source candidates . . . . .	8
1.3 The propagation of ultra-high energy cosmic rays . . . . .	11
1.3.1 The GZK cutoff . . . . .	12
1.4 The detection of cosmic rays . . . . .	14
<b>2 The detection of Ultra-High Energy Cosmic Rays</b>	<b>19</b>
2.1 The Extensive Air Showers . . . . .	20
2.1.1 The Heitler model for electromagnetic showers . . . . .	21
2.1.2 A generalized Heitler model for hadronic showers . . . . .	23
2.1.3 The superposition model . . . . .	25
2.2 The detection techniques for EASs . . . . .	26
2.2.1 The fluorescence technique . . . . .	27
2.2.2 The surface array detection . . . . .	30
2.3 Mass composition analysis . . . . .	33
2.3.1 The interpretation of the first and second $X_{\max}$ moments . . . . .	35
<b>3 The Pierre Auger Observatory</b>	<b>37</b>
3.1 The surface detector . . . . .	37
3.1.1 SD reconstruction . . . . .	39
3.2 The fluorescence detector . . . . .	41
3.2.1 FD reconstruction . . . . .	44
3.2.2 SD energy calibration . . . . .	46
3.3 The Observatory enhancements . . . . .	47
3.3.1 HEAT . . . . .	48
3.3.2 AMIGA . . . . .	48
3.3.3 AERA . . . . .	49

3.4	The upgrade of the Observatory: <i>AugerPrime</i>	50
3.5	Selected results	52
3.5.1	Energy Spectrum	52
3.5.2	Mass composition studies	53
3.5.3	Anisotropy searches	55
<b>4</b>	<b>Air shower Universality</b>	<b>59</b>
4.1	Universality of the shower components	59
4.2	Universality of the ground signals	64
4.3	Universality of the total integrated signal	68
4.4	Universality of the time-dependent signal	75
<b>5</b>	<b>The Universality reconstruction</b>	<b>85</b>
5.1	The Universality fitting procedure	86
5.2	The selected set of golden hybrid events	88
5.3	The original Universality fit	91
5.4	A first change in the Universality fit: deviance minimisation	93
5.4.1	The results for the modified Universality reconstruction	97
5.5	Study of the bin signal fluctuations	99
5.5.1	“Quasi poissonian” case	99
5.5.2	Toy “cluster model”	101
5.5.3	The bin signal uncertainties from the toy cluster model	105
5.6	A second change in the Universality fit: reduced bin signal uncertainties	107
5.7	A different paradigm for the Universality reconstruction	109
5.8	Constrained Universality reconstruction	113
5.8.1	The Universality fitted energy	117
<b>6</b>	<b>The Universality reconstruction of the SD-1500 dataset</b>	<b>125</b>
6.1	A simplified Universality fit	125
6.2	The SD-1500 dataset	127
6.2.1	Event selection	127
6.2.2	Quality cuts	130
6.3	Mass composition	132
6.3.1	The first two moments of the $X_{\max}$ distribution	134
6.3.2	Estimation of the first two moments of the $\ln A$ distribution	136
6.3.3	Fit of the mass fractions	139
6.4	Future prospects	143
<b>Conclusions</b>		<b>146</b>

<b>A Some additional details about the Universality fit</b>	<b>149</b>
A.1 The deviance for the axis constraint . . . . .	149
A.2 Study of mock traces generated through the cluster model . . . . .	150
A.2.1 The expected mean of a set of mock traces . . . . .	150
A.2.2 The expected variance of a set of mock traces . . . . .	151
A.2.3 Effect of a different choice for the single cluster signal . . . . .	153
<b>B Azimuthal asymmetry, core position and core time resolutions</b>	<b>155</b>
B.1 The azimuthal asymmetry . . . . .	155
B.2 The shower coordinate system . . . . .	158
B.3 Comparison of different core positions . . . . .	158
B.4 Evaluating the effect of azimuthal asymmetry on the standard SD core position . . . . .	159
B.5 Standard SD and hybrid core times comparison . . . . .	161
<b>Bibliography</b>	<b>162</b>
<b>Acknowledgements</b>	<b>170</b>



# Introduction

More than 100 year ago, it was discovered that our planet is continuously hit by high-energy particles coming from outer space, later called cosmic rays. There are still many uncertainties about the origin and nature of these particles, in particular for those with energies above  $10^{18}$  eV, usually referred to as ultra-high energy cosmic rays (UHECRs). Due to the very low flux (in the order of a few tens of particles per square kilometre per steradian per year), UHECRs can only be indirectly detected through cascades of secondary particles generated in their interaction with the atmosphere, known as extensive air showers (EASs). Then, for the purpose of measuring EASs, arrays of particle detectors covering vast areas have been deployed over the years, sometimes accompanied by fluorescence telescopes, which collect the radiation emitted by the air nitrogen molecules excited by the shower particles. The largest of such experiments, the Pierre Auger Observatory, has been operating since 2004 in the Argentinian *Pampa Amarilla*. It comprises 1,660 water-Cherenkov detectors spread over a  $3,000 \text{ km}^2$  area, overlooked by 27 fluorescence telescopes distributed in 5 sites.

There is now a general belief that almost all the UHECRs are charged particles (protons and/or heavier nuclei) mostly coming from outside our galaxy, even if neither the positions and type of these sources, nor the acceleration mechanisms are established. To answer these question, it is critical to better determine the mass composition of cosmic rays at the highest energies. The most commonly used mass-sensitive observable extracted from EAS measurements is  $X_{\max}$ , the atmospheric depth corresponding to the maximum development of the electromagnetic part of the shower. This quantity is measured with good quality in hybrid events, i.e. showers detected by both fluorescence and surface detectors. However, the duty cycle of fluorescence telescopes is about 15% and the measurements that can be actually used for physics studies are only a few %, insufficient to extend our knowledge about the mass of primaries at the highest energies (namely above  $10^{19.5}$  eV). Therefore many efforts have been put in the development of a composition analysis which makes use only of the data recorded with the surface detector of the Pierre Auger Observatory, operational nearly 100% of the time.

With this aim, a new reconstruction procedure has been proposed in the framework of the Pierre Auger Collaboration, based on the concept of air shower Universality. According to the Universality paradigm, extensive air showers created by UHECRs can be described, to a remarkable degree of precision, in terms of a reduced set of macroscopic

parameters: shower geometry, energy,  $X_{\max}$  and muon content. Besides its implications on astroparticle physics, shower Universality allowed the development of models for the expected size and time structure of signals collected by water-Cherenkov detectors at observational level. In this sense, the Universality-based reconstruction can be considered as a very advanced multivariate analysis of the measurements, that employs parameterized physical relations to combine and interpret the shower data. The Universality method is extendable to other detector types and it is of essential importance for the *AugerPrime* upgrade of the Pierre Auger Observatory and for future analyses.

During my PhD, I worked in the optimization of the existing Universality reconstruction, starting from the current implementation in the `Offline` software of the Collaboration. The goal was to improve the accuracy in the reconstructed physical parameters of the shower, in particular in the determination of  $X_{\max}$ . The adjustments in the fitting procedure were guided by the comparison between Universality and standard reconstruction results, using as benchmark a selected set of high quality hybrid events. Then I employed the improved Universality fit to reconstruct the events measured by the surface detector of the Pierre Auger Observatory in the period between 2004 and 2016 and to finally accomplish a study of the mass composition, even above  $10^{19.5}$  eV.

The thesis is organized as follows. Chapter 1 deals with a general outline on cosmic rays, from the energy spectrum to the acceleration and propagation processes. Chapter 2 concerns the physics of EAS and the main techniques employed in their detection. Chapter 3 describes the Pierre Auger Observatory, including the main procedure for the reconstruction of detected showers and a few recent results on UHECRs. In Chapter 4 an overview on air shower Universality is given, starting from the application to pure electromagnetic showers, the extension to hadronic ones, and concluding with the parameterizations implemented to reproduce the signals collected by the surface detector of the Pierre Auger observatory. In Chapter 5, the characteristics of the Universality reconstruction are introduced; then the improvements of the fitting procedure elaborated during my research work are described. Finally, in Chapter 6 the results obtained from the optimized Universality reconstruction of the surface detector dataset are presented, showing the implication on the mass composition of primaries with energies above  $10^{19}$  eV.

# Chapter 1

## The cosmic rays

The cosmic rays are particles coming from space which continuously hit the Earth, bringing unique information about the Universe as the only messengers from outside the Solar System (or even from outside the Milky Way) that we are able to directly observe or indirectly measure. At the top of the atmosphere, the charged component is mainly made by electrons and ionized nuclei (about 90% of protons, 9% of alpha particles, the rest heavier nuclei) with relativistic energies. A small fraction of these nuclei present energies comparable to macroscopic objects (even more than 1 Joule per nucleon, much higher than any value achievable by man-made experiments), providing a natural probe for the study of particle interactions in extreme conditions.

The discovery of cosmic rays is strictly related with the first studies in radioactivity, but their existence solved a question which dates back to the 19<sup>th</sup> century. When the first experiments using a leaf electroscope were performed, it was noticed that the electroscope spontaneously discharges overtime, even if the system was isolated. This process was later explained through the ionization of the air surrounding the leaves, but the cause of such ionization remained unknown. Only at the beginning of the 20<sup>th</sup> century, the discovery of spontaneous radioactivity allowed to find a source of air ionization in the natural decay processes. However, an obvious consequence of this hypothesis is that the ionization has to decrease far from the source of natural ionizing radiation, that was identified in the Earth's crust. Therefore, experiments were performed comparing the rate of discharge of an electroscope at ground level with the rate far from the surface or underground (and underwater). The result of these first experiences were unclear, although the expected decrease in ionization was not evident.

The final step of this journey was achieved independently by the italian Domenico Pacini and the austrian Victor Franz Hess, between 1907 and 1912. Hess highlighted an increase of air ionization with altitude [1], while Pacini measured a decrease with depth in water [2] due to absorption of the ionizing radiation. Both these results confirmed that a penetrating radiation must come from outside the Earth, producing a noticeable ionization in the atmosphere<sup>1</sup>.

---

<sup>1</sup>However, physicists were reluctant to abandon the hypothesis of a terrestrial origin for the mystery

Robert Millikan in the '20 named this radiation *cosmic rays* [3], believing it was composed mainly of  $\gamma$  rays. However, in the following years, it was understood that such radiation originates in the interaction of energetic particles with the atmosphere molecules: such particles are nowadays designated as cosmic rays. The origin of cosmic rays is still a mystery, even if likely connected with the most catastrophic events in the Universe, such as SuperNovae explosions in the Milky Way, or Active Galactic Nuclei and Gamma Ray Bursts in extragalactic space (since the most energetic cosmic rays are disfavoured to be produced in our galaxy, as demonstrated in the following).

The first chapter deals with a general description of cosmic ray observations, starting from the spectrum of primary cosmic rays<sup>2</sup> and then dealing with the journey of these particles from the possible production sites towards the Earth, with a final summary of the fundamental detection techniques.

## 1.1 The cosmic ray energy spectrum

A first step in the field of cosmic ray physics is usually a survey of the all-particle energy spectrum (Fig. 1.1), intended as the number of particles ( $N$ ) that reach the Earth per unit of energy, time, surface and solid angle. The behaviour of the differential flux as function of the energy  $E$  is a **power law**:

$$\frac{dN}{dE} \propto E^{-\gamma} \quad [\text{m}^{-2}\text{s}^{-1}\text{sr}^{-1}\text{eV}^{-1}] \quad (1.1)$$

where  $\gamma$  is the so-called *spectral index*. A value of  $\gamma \approx 2.8$  fairly describes the overall energetic spectrum, but in order to highlight features of the flux at high energies the differential flux is usually multiplied by a power of the energy<sup>3</sup> (see Fig. 1.2). It's necessary to point out that below  $\sim 10^9$  eV cosmic rays are modulated by the solar activity and deviated by the magnetic fields in the Solar System, so this energetic range is not considered in the following.

The energy spectrum spans over 12 orders of magnitude in energy (from  $10^9$  eV to  $10^{20} \div 10^{21}$  eV) and above 30 orders of magnitude in flux (from about 1 particle/ $\text{m}^2/\text{s}$  for low energies, to 1 particle/ $\text{km}^2/\text{century}$  for the highest), giving an overall idea of the enormous number of phenomena that contribute in the production, propagation and observation of these particles. The spectrum can be roughly divided into three parts depending on the  $\gamma$  value, which is supposed to change because of different origin, acceleration mechanisms and propagation of the primary particles:

---

penetrating radiation. Only in 1936 Hess was awarded with the Nobel Prize for this result, while Pacini has died in 1934.

<sup>2</sup>Actually, in the acceleration and propagation of cosmic rays across the Universe, it may occur that high energetic particles interact generating other particles that then reach the Earth. However, in the following the charged and neutral particles that reach the Earth atmosphere from space are indicated as *primary* cosmic rays, in contrast with the particles generated by the interaction of primaries with the Earth's atmosphere, called *secondaries*.

<sup>3</sup>The disadvantage is that the two axes become coupled, enhancing the systematic differences due to different energy estimation between the experiments.

**Cosmic Ray Spectra of Various Experiments**

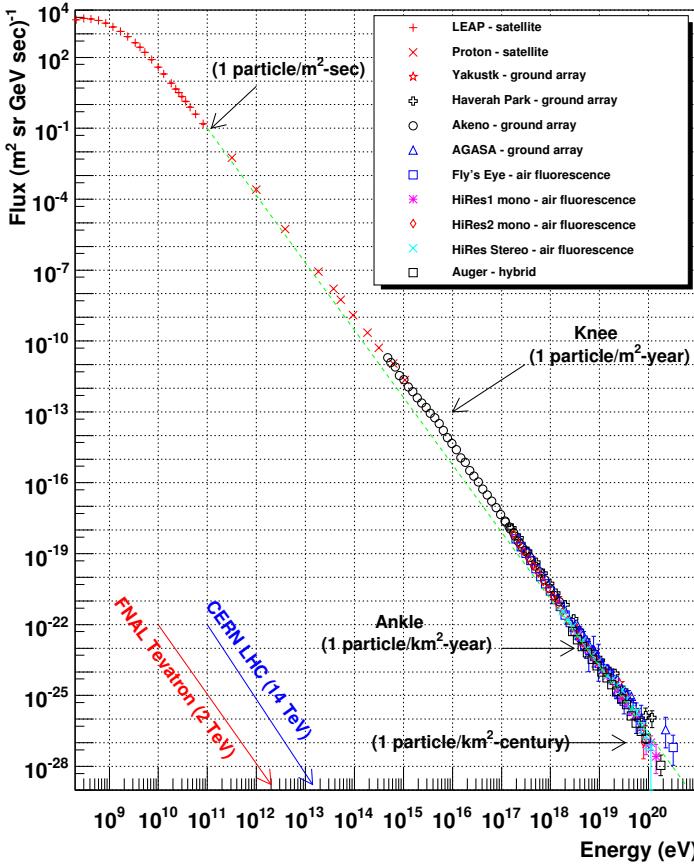


Figure 1.1: All-particle energy spectrum for primary cosmic rays. The two main changes of slope (knee and ankle) for the differential flux, with the related values, are highlighted.  
Image from <http://www.physics.utah.edu/whanlon/spectrum.html>.

- *low energy region*, from above  $10^9$  eV to the *knee* ( $E \approx 10^{15}$  eV): the spectral index is  $\gamma \sim 2.7$  and the primaries of this region are supposed to be galactic;
- *intermediate region*, from the *knee* to the *ankle* ( $E \approx 10^{18}$  eV): the slope grows being  $\gamma \sim 3.1$ ; in this region the transition from a galactic origin to an extragalactic one is expected;
- *high energy region*, above the *ankle*: the spectral index decreases ( $\gamma \sim 2.6$ ) approximately to the value before the *knee*; moreover, a flux suppression is observed above  $10^{19.5}$  eV (see sect. 1.3.1 for more details).

Observing the energy spectrum, it clearly appears that a single experimental set-up cannot measure the cosmic rays in all the different ranges of flux. For the lower energies (and therefore higher fluxes) a direct detection<sup>4</sup> is possible, while from energies near and above the knee several techniques of indirect detection (detailed in the next chapter) have

---

<sup>4</sup>For direct detection it is intended the observation of cosmic rays at the top of the atmosphere, which implies the identification and characterization of each measured particle. In sect. 1.4 is presented a more detailed discussion.

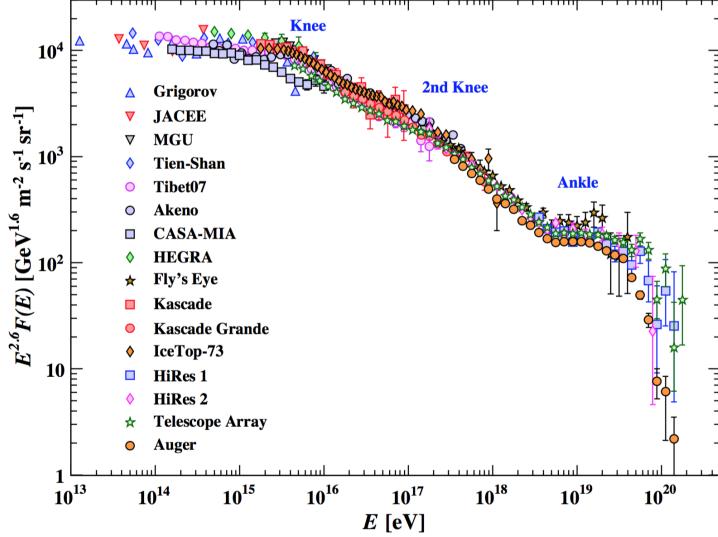


Figure 1.2: Upper part of the all-particle energy spectrum, where the differential flux is multiplied by  $E^{2.6}$  to emphasise the observed sub-structures. Image from [4].

been developed during the years, exploiting the interactions of high energy particles with the atmosphere. The cosmic ray detection methods, with a focus on nuclei, are presented in sect. 1.4; instead, the next sections deal with the production and propagation of cosmic rays from the (supposed) sources toward the Earth.

## 1.2 Production and acceleration of cosmic rays

In order to explain the observed energies, it is natural to suppose as production and acceleration sites the most energetic astrophysical objects inside (and outside) the Milky Way. Those objects are characterized by surrounding matter in the state of plasma and by the presence of intense, strongly variable magnetic fields. Such environments are perfect to accelerate charged particles through multiple interactions with a moving magnetised plasma, as proposed by Enrico Fermi in a famous paper [5], based on the concept of *collisionless magnetic diffusion*. Currently, the most widely accepted model for cosmic ray acceleration is an evolution of the Fermi model, based on *diffusive shock acceleration*. Both scenarios are illustrated in the following.

### 1.2.1 The Fermi model and the Diffusive Shock Acceleration

A first realistic acceleration mechanism was originally proposed by Fermi in 1949 (called after him **Fermi mechanism** or model) to explain the primary energies measured until that time. This mechanism is based on the interaction of charged particles, already relativistic ( $v \approx c$ ), with a magnetized molecular cloud moving at a velocity  $V$  with respect to an external observer. The primary particle enters into the cloud and is accelerated by the variable magnetic fields in the plasma, according to a collisionless statistical mechanism called magnetic diffusion.

Actually this original model, nowadays called *II order Fermi mechanism*, cannot explain the observation of very energetic particles. In fact, the energy gained per interaction ( $\Delta E$ ) is in the order of [6]:

$$\left\langle \frac{\Delta E}{E} \right\rangle = \frac{8}{3} \frac{V^2}{c^2} \sim 10^{-8} \quad (1.2)$$

where  $c$  is the light velocity. Then to reach the observed energies a huge number of interactions is needed, that is not compatible with the observations.

However, it is possible to have enough energy gain in a regime of diffusive shock acceleration (Fig. 1.3), where the charged particles interact with a shock wave front<sup>5</sup> of magnetized matter. This scenario is also called *I order Fermi mechanism* and can be found for instance in the environment surrounding a Supernova explosion.

In this model the energy gain per interaction, i.e. in one round trip of a particle across the shock front and back again, gives (as detailed in [6]):

$$\left\langle \frac{\Delta E}{E} \right\rangle = \frac{4}{3} \frac{V}{c} + O\left(\frac{V^2}{c^2}\right) \approx \frac{V_{\text{shock}}}{c} \quad (1.3)$$

where the last equality stands according with the relation  $V_{\text{shock}} = 4/3 V$  from fluid mechanics. As the shock front can reach sub-relativistic velocities (even  $V_{\text{shock}} \approx 0.1 c$ ) this result is acceptable to (at least partially) explain the observations.

The essence of the original Fermi idea is that accelerated particles reach high energies after many processes, that in the case of diffusive shock acceleration means after cyclic interactions in and out from the shock front. But the key feature of this model is that the acceleration of a particle is first order in the shock velocity, which results automatically

<sup>5</sup>A shock wave front is generated when the front speed of the wave is larger than the sound speed in the gas or plasma.

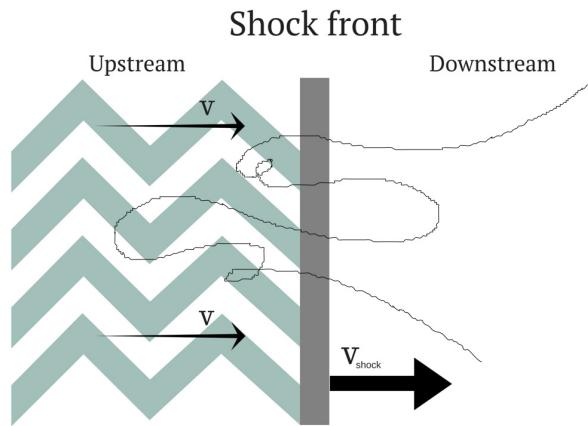


Figure 1.3: Sketch of the physical scenario of the *I order Fermi mechanism*, i.e. the regime of diffusive shock acceleration. The shock front, in motion at speed  $V_{\text{shock}}$  with regard to an external observer, divides the space in 2 regions: the *downstream*, at rest for the external observer and the *upstream*, where matter has a velocity  $V < V_{\text{shock}}$  and variable non-uniform magnetic fields (in light blue) are present. Charged particles are accelerated passing through the shock many times before leaving.

in a power-law spectrum compatible with the observations on Earth, as demonstrated in the following (referring to [6]).

If  $E_0$  is the initial energy of the particle, after one interaction it becomes  $E_1 = \beta E_0$ , i.e. increased of a factor  $\beta$ , that from eq. 1.3 results  $\beta = 1 + V_{\text{shock}}/c$ . Moreover, if  $N_0$  is the initial number of charged particles accelerated, after one interaction it reduces to  $N_1 = PN_0$ , where  $P$  is the probability of remaining in the shock region. After  $k$  interactions,  $E_k = \beta^k E_0$  and  $N_k = P^k N_0$ . Applying the logarithm and matching the resulting expressions for  $k$ , it results:

$$\frac{\ln(N_k/N_0)}{\ln(E_k/E_0)} = \frac{\ln P}{\ln \beta} \quad (1.4)$$

Then, dropping the  $k$  index and considering  $E$  and  $N$  the energy and remaining particles after a very large number of interactions:

$$\frac{N}{N_0} = \left( \frac{E}{E_0} \right)^{\ln P / \ln \beta} \implies \frac{dN}{dE} \propto E^{\left( \frac{\ln P}{\ln \beta} - 1 \right)} \quad (1.5)$$

that is the required power law behaviour (see sect. 1.1). In order to quantitatively determine the spectral index, one has to estimate the escape probability, encoded in the value of  $P$ . Using an argument derived by Bell from classical kinetic theory (therefore assuming the shock is non-relativistic), it is possible to demonstrate that only a small fraction of particles, equal to  $V_{\text{shock}}/c$ , is lost in each cycle. Thus it is obtained  $P = 1 - V_{\text{shock}}/c$  and consequently, at first order approximation:

$$\frac{\ln P}{\ln \beta} = \frac{\ln(1 - V_{\text{shock}}/c)}{\ln(1 + V_{\text{shock}}/c)} \sim -1 \quad (1.6)$$

So the energy spectrum at the sources results  $dN/dE \propto E^{-2}$ , in agreement with the expectations from measurements<sup>6</sup>.

### 1.2.2 The Hillas plot and the source candidates

The diffusive shock acceleration can explain the lower region of the cosmic ray spectrum but this mechanism has strong limits in the energies that can be effectively reached. In fact, the expression for the maximum energy of a cosmic ray with electric charge  $Z$ , accelerated in a region of size  $L$  where the average magnetic field intensity is  $B$ , can be written (see [6, 7] and references therein) as:

$$E_{\text{max}} \propto Z e B L \quad (1.7)$$

---

<sup>6</sup>The cosmic ray spectrum observed from Earth arises from the combination of source injection and of processes occurring during the propagation, that alter the original spectral index. Using the *Leaky box model*, a simple propagation model detailed in [7], the measured spectral index  $\gamma$  in the region before the knee can be written as  $\gamma = \alpha + \delta$  for particles that reach the Earth directly from the sources (i.e. not for cosmic rays obtained from the fragmentation of heavier nuclei), where  $\alpha$  is the spectral index at the sources and  $\delta$  is the hardening due to propagation effects. From measurements, for instance of the boron over carbon (B/C) ratio, it is obtained  $\delta \approx 0.6$ . Then, the energy spectrum expected at the sources for energies below the knee is  $Q(E) \propto E^{-2.1}$ , in full accordance with the result from the I order Fermi mechanism.

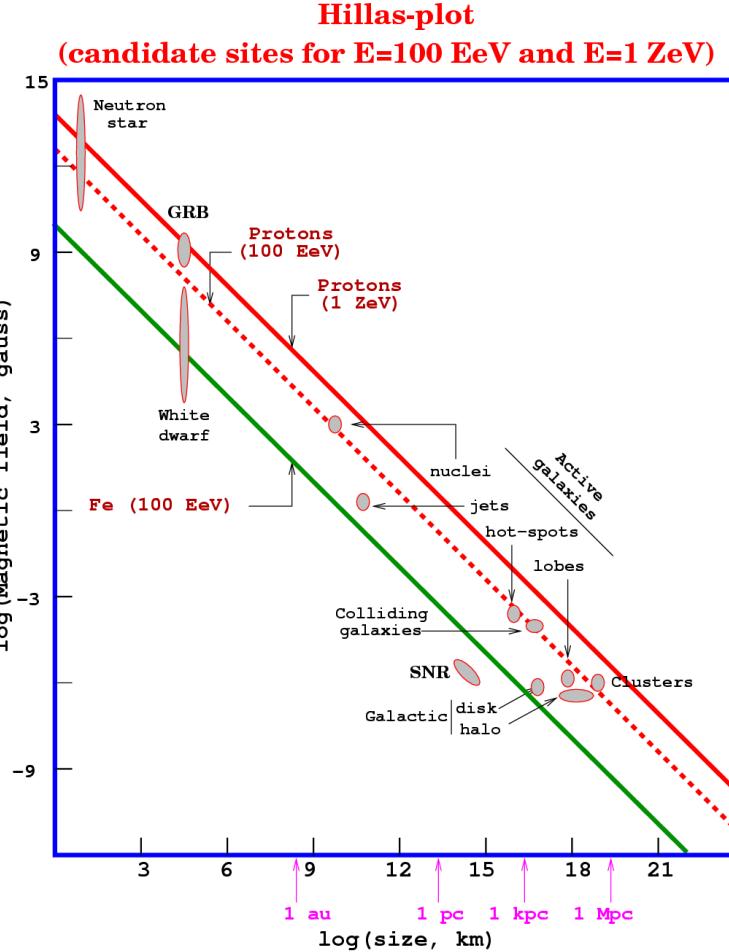


Figure 1.4: The Hillas plot in  $\log - \log$  scale, with the linear dimension of the acceleration region in the x-axis and the magnetic field intensity in the y-axis. According to the Hillas relation (eq. 1.7), it is possible to express through a line the minimal required conditions for different primary species and energies: only objects above this line are candidate sources.

Image from <http://inspirehep.net/record/796916/files/Hillas-plot2.png>

This equation is known as **Hillas relation** [8] and allows to predict (in first approximation) the magnetic field intensity required to accelerate a charged particle to a certain energy given the source size. This concept can be expressed in a graphical form through the Hillas plot (Fig. 1.4). However, it should be noticed that in eq. 1.7 as well as in Fig. 1.4 an efficiency factor for the particles accelerations is not present.

Below the knee ( $\sim 10^{15}$  eV) possible source candidates are the *Supernovae*. In fact, considering a magnetic field of intensity  $B \sim 3 \mu G$  in a region of dimension<sup>7</sup>  $L \sim 1$  pc, the maximal reachable energy (eq. 1.7) as function of  $Z$  is  $E_{\max} \approx Z \times 10^{15}$  eV.

Another argument supporting the scenario of a Galactic origin for cosmic rays below the knee is the classical comparison between the luminosity  $L_{SN}$  of the Supernovae in our galaxy and the luminosity  $L_{CM}$  of low energy cosmic rays, as emphasized long ago

<sup>7</sup>1pc  $\cong 3.1 \times 10^{16}$  m  $\cong 3.26$  ly  $\cong 203.8$  AU

The *parsec* (pc) is a unit of length used to measure large distances outside the Solar System; it is named from an abbreviation of “parallax of one arcsecond”.

by Ginzburg & Syrovatskii. The former is obtained considering an energetic emission of  $E_{SN} \sim 10^{51}$  erg per Supernova explosion and an average time between two events of  $\tau = 30$  y: the result is a luminosity  $L_{SN} \approx 10^{42}$  erg/s. The latter can be estimated as  $L_{CR} = V \rho_{CM} / \tau_{escape}$ , where  $V \sim 10^{66}$  cm<sup>3</sup> is the Milky Way volume,  $\rho_{CM} \sim 1$  eV/cm<sup>3</sup> is the average energy density for cosmic rays and  $\tau_{escape} \sim 10^6$  y is the mean escape time from the magnetic confinement of the galaxy: so the result is  $L_{CM} \approx 10^{40}$  erg/s. Even within such a rough evaluation, only  $\sim 1/100$  of the energy emitted by Supernovae is needed to accelerate all the cosmic rays produced inside the Milky Way, supporting the hypothesis that primaries below the knee are accelerated inside our galaxy.

In principle the described *I order Fermi model* can be effective also at higher energies: the candidate sources become galactic accelerators, with processes in a hugely higher energy scale and dimensions  $L \sim$  kpc. However, the production of UHECRs up to  $10^{20}$  eV is not trivial even in this scenario and possible sources are still debated, even if there are few known objects that present intense magnetic fields in regions of space enough large to accelerate charged particles to the highest energies.

*Active Galactic Nuclei* [9, 10] and *Gamma Ray Bursts* [11] are among the firstly proposed candidates. The former are galaxies with a very luminous nucleus (usually powered by a supermassive black hole) which can represent an ideal gigantic accelerator; the latter are among the most energetic events observed in the Universe, and in their formation sites also the acceleration of UHECR could occur. In addiction, the *lobes of giant radio galaxies* [12] have been proposed as possible sources. They are believed to be the endpoints of galactic jets and so an environment fitting the shock acceleration paradigm. In more recent times, *starburst galaxies* [13] are being re-evaluated. These objects presents an outstanding stellar activity (with associated supernova explosions and stellar winds) which generate galactic-sized superwinds. The shock generated by such winds has been reconsidered as potentially able to accelerate cosmic rays at the highest energies. Finally, *pulsars* [14, 15] have been proposed since their discovery, even if the possibility of hadron acceleration in these sites is still unclear. Besides, the acceleration around pulsars is not a stochastic process (like for diffusive shock acceleration) but is completely governed by magnetohydrodynamics.

The acceleration in all these sources constitute the so-called *Bottom-Up* scenario. Therein an additional element to take into account is the energy loss in the propagation of the generated UHECRs (detailed in the following sect. 1.3) since the mentioned astrophysical objects are typically far away (more than 100 Mpc).

Over the years an alternative scenario (called *Top-Down* in contrast with the *Bottom-Up* scenario) has been studied. In this case, UHECRs would be the decay products of topological defects<sup>8</sup> or super-heavy relic particles [16] (with masses above  $10^{20}$  eV) maybe produced at the beginning of the Universe. However, the requirements on such decaying

---

<sup>8</sup>Extended metastable objects formed in the first phases of the Universe.

objects to be considered the source of the observed UHECRs are hardly matched in observations<sup>9</sup>. Furthermore, the limits on the flux of photons and neutrinos in the ultra high energy region measured by the Pierre Auger Observatory [17, 18] severely constrain top-down models for the origin of UHE cosmic rays. The general indication is that top-down mechanisms cannot be the main source of the observed flux.

### 1.3 The propagation of ultra-high energy cosmic rays

The propagation of primary cosmic rays is intrinsically connected with the magnetic fields they pass throughout the journey to the Earth; in fact any charged particle suffers magnetic deflection, quantified by its Larmor radius:

$$R_L = \frac{p}{ZeB_\perp} \approx \frac{E}{ZeB_\perp c} \quad (1.8)$$

where  $p$  is the particle momentum,  $Ze$  the charge,  $E$  the energy,  $c$  the light speed and  $B_\perp$  the magnetic field strength perpendicular to the particle trajectory.

The galactic magnetic field presents a roughly uniform component with intensity  $10^{-6}$  G on a large scale. Then, for a proton at the knee energy, the Larmor radius inside the Milky Way is  $R_L \sim 1$  pc, which results of the same order of the galactic magnetic field irregularities [19]. Therefore the propagation of primaries with energies below  $10^{15}$  eV is a diffusion process and they can be considered confined inside the galaxy. At larger primary energies the Larmor radius grows, becoming comparable to the thickness of the galactic disk ( $\approx 300$  pc) around the ankle energy, meaning that the diffusion approximation is no longer appropriate and particles could escape more readily from the galaxy.

Instead, UHECRs are supposed to be produced outside the Milky Way and have to travel for much longer distances before reaching the Earth. In these conditions, the effect of small extragalactic magnetic fields is not negligible. The intergalactic space is modelled [20] with large regions (with a typical size of  $\sim 1$  Mpc) of almost uniform magnetic field of  $10^{-9}$  G, usually with random directionality. Simple simulations (Fig. 1.5), where energies losses are not applied, show that the propagation of UHE protons can be considered diffusive (in the extragalactic space, not inside our galaxy) at 1 EeV ( $= 10^{18}$  eV) and becomes more and more rectilinear approaching 100 EeV. Given the fact that protons are the nuclear primaries less influenced by magnetic fields, it is clear the difficulty in the task of correlate the cosmic rays with the respective sources.

However, a fundamental aspect in the study of UHECRs' propagation is the evaluation of the energy losses which occur during the long travel of the primaries, since in this case the distances are cosmological (hundreds of Mpc).

---

<sup>9</sup>For instance, they have to present a long lifetime on a cosmological scale, a sufficient abundance to explain the measurements and a small galactic anisotropy in the UHECRs arrival directions.

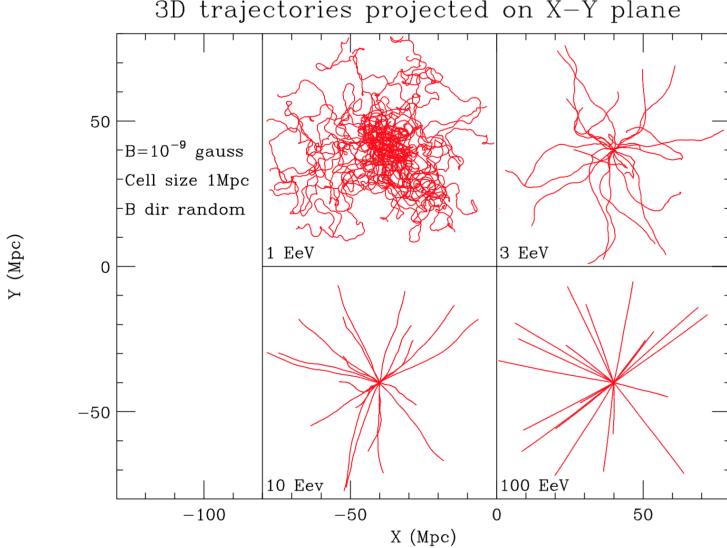


Figure 1.5: Projected view of 20 trajectories of proton primaries emanating from a point source for several energies. Trajectories are plotted until they reach a physical distance from the source of 40 Mpc. No energy loss was applied in these calculations, while the characteristics of the intergalactic magnetic field are reported in the picture. Image from [21].

### 1.3.1 The GZK cutoff

With the discovery [22] of the Cosmic Microwave Background radiation (CMB)<sup>10</sup>, the american Greisen and the russians Zatsepin and Kuz'min found out independently [23, 24] that the interaction between UHECRs and these background photons is not negligible, with important consequences on the propagation.

For an UHE proton ( $p$ ) the main process of interaction with a CMB photon ( $\gamma_{CMB}$ ) is the pion photoproduction through Delta resonance:

$$p + \gamma_{CMB} \longrightarrow \Delta^+ \longrightarrow N + \pi \quad (1.9)$$

where  $N$  is a nucleon (proton or neutron) and  $\pi$  a pion (charged or neutral). In order to calculate the minimum energy required for this reaction to occur, the process is studied in the Center of Mass System, where both nucleon and pion are produced at rest. For a head-on collision with a photon of energy  $\varepsilon_\gamma \approx 10^{-3}$  eV, the threshold energy  $E_{th}$  for the cosmic ray proton is:

$$E_{th} = \frac{m_\pi^2 + 2m_p m_\pi}{4 \varepsilon_\gamma} \approx 7.5 \cdot 10^{19} \text{ eV} \quad (1.10)$$

The proper calculation requires an integration over the CMB spectrum and over the collision angle, which result is a lower threshold  $E_{th} \sim 5 \times 10^{19}$  eV. A proton produced with

---

<sup>10</sup>Electromagnetic background radiation coming quasi-isotropically from the whole Universe; the spectrum is an almost perfect black body emission at a temperature of 2.7 K which presents a maximum in the microwaves. This radiation is a relic of the age of recombination, as it was originated in the recombination of electrons and nuclei. Since before that time the Cosmos was opaque to light, the CMB radiation also forms the so-called last scattering surface.

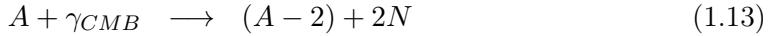
an energy above  $E_{th}$  is subjected to pion photoproduction many times in its propagation. Since the energy loss per interaction is about 10-20%, this process causes a substantial decrease of the initial energy, until a value below the threshold is reached.

A similar calculation [6] can be carried out for the electron-positron photo-pair production process:



The threshold energy for this process is 1.02 MeV, about 200 times less than that for the pion photoproduction mechanism. Moreover, the cross-section in the ultra-relativistic limit is about 40 times larger than that for the production of pions. Therefore, the photo-pair production becomes relevant for protons with energies of about  $10^{18}$  eV. However, each photo-pair production event removes only  $10^{-3}$  of the energy of the proton and so this process is less important for cosmic ray protons of energy  $E_{th} \approx 5 \times 10^{19}$  eV, but it could result in a distortion of the particle spectrum down to energies of about  $10^{18}$  eV (the so-called “dip” [25, 26]).

The described processes are effective only for UHE protons. In fact, also UHE nuclei propagating towards our galaxy interact with CMB photons, but in this case the dominant source of energy loss is the nuclear photo-disintegration<sup>11</sup>. In this reaction a photon is absorbed by the nucleus, which then decays emitting one or more nucleons. Indeed, the dominating decay modes are:



An accurate description of this process is more complicated than in the proton case, and it is beyond the scope of this chapter. However, the important results is a threshold energy for the photo-nuclear disintegration of the same order of magnitude (i.e.  $\sim 10^{19}$  eV) of  $E_{th}$  from eq. 1.10 and a suppression for the spectrum of UHE nuclei qualitatively similar to the proton one [7]. For this reason, the subnuclear reactions between UHECRs and the CMB photons, resulting in substantial energy losses during the propagation, are overall known under the name of **GZK effect**.

A first consequence of the GZK effect is that the Universe can be considered opaque to cosmic rays of extreme energies, an effect quantified studying the attenuation length  $\lambda$ , in a sense a mean free path for cosmic rays in their propagation. Such quantity is calculated as  $\lambda \approx (n_\gamma \sigma_{CR\gamma})^{-1}$ , where  $n_\gamma$  is the numerical density of the CMB photons (function of the photon momentum) and  $\sigma_{CR\gamma}$  is the cross section for the selected energy and primary (proton or nucleus). Since an UHECR statistically decreases its energy below the threshold after few interactions,  $\lambda$  defines a GZK horizon, intended as the maximum distance where the cosmic rays can be produced in order to reach the Earth without

---

<sup>11</sup>It is worth noticing that already Greisen, Zatsepin & Kuz'min pointed out that nuclei would also suffer suppression at high energies due to photo-disintegration processes, with about the same threshold of protons.

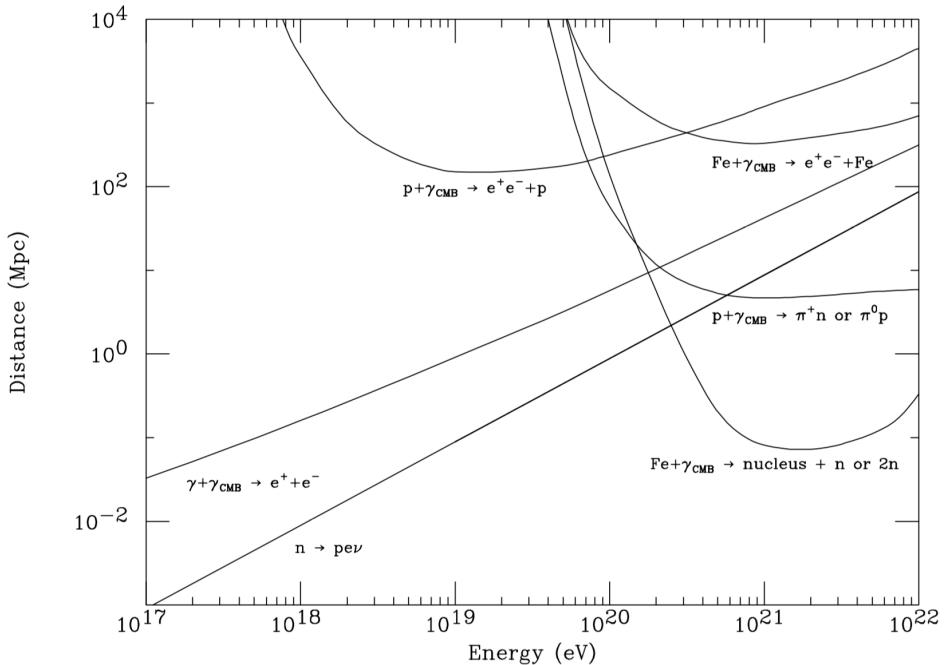


Figure 1.6: Summary of interactions lengths for the main reactions of cosmic primaries with the CMB photons. Curves correspondent to proton and Fe pair production are energy loss lengths (distance for which a proton or Fe nucleus loses  $1/e$  of its energy due to pair production). Instead, curves for pion photo-production from proton and photo-nuclear disintegration of Fe nuclei are mean free paths. Also the mean free path for photon pair production and the mean decay length for neutron are shown for reference. Image from [21].

attenuation. In Fig. 1.6 the attenuation lengths are shown for the main interaction processes between cosmic rays and CMB photons. It is evident that already for energies of about  $3 \times 10^{20}$  eV the mean free path of a proton is less than 10 Mpc, while for an iron nucleus is of the order of 1 Mpc. The result is that, if most of the sources are well beyond the defined GZK horizon, UHE particles cannot be detected at an energy higher than  $\sim 10^{20}$  eV, independently from the production energy.

Thus the GZK effect predicts observable consequences on the cosmic ray spectrum in its extremal part, in particular the *GZK cutoff*, i.e. a sudden drop of the flux above the threshold energies. Nowadays a cutoff in the energetic spectrum at extreme energies is experimentally observed (fig. 1.7), but it is not possible to assert that it is due to the GZK effect. For instance, it could be a suppression in the acceleration mechanisms that prevents reaching energies above  $10^{20}$  eV already at the source.

## 1.4 The detection of cosmic rays

As mentioned in sect. 1.1, the rapid decrease of the flux with energy requires the employment of different techniques for the detection of cosmic rays reaching the Earth, which are summarised in this section.

It should be preliminarily underlined that primary nuclei interact with the atmo-

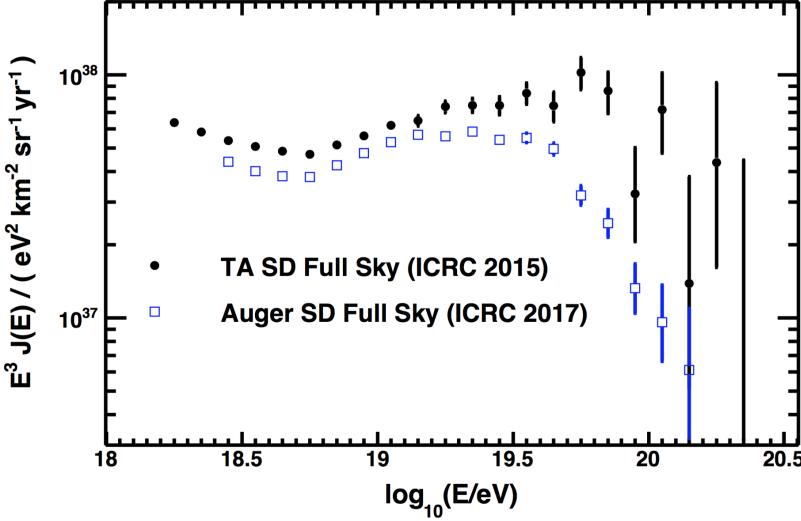


Figure 1.7: Energy spectra by the Pierre Auger Observatory and Telescope Array surface detectors. In both cases, a suppression in the flux is evident above  $10^{19}$  eV. The systematic uncertainty in the overall energy scale is 14% for Auger and 21% for TA. Image from [27].

sphere at altitudes between 10 and 20 km, which means far away from the Earth surface. Therefore, experiments for the direct measurement of these particles have to be performed on high-altitude balloons or in near space, putting the detectors on satellites or space stations. In this typology of observations, called *direct detection*, the goal is to identify each primary cosmic ray, obtaining its mass, charge and energy through a measurement of the particle speed or its calorimetric release. The direct detection is anyway limited by the rapid decrease of the cosmic rays flux. Already around 100 TeV the flux is so low that becomes difficult to collect a significant event statistics in a reasonable timescale, given the typical detector sizes feasible in such conditions. To understand this limit it is useful to introduce the concept of acceptance of an experiment.

In general, a particle has to entirely traverse an experimental apparatus, composed by several instruments, to be effectively measured. Thus it is clear that only a small fraction of the particles entering in the upper part of the detector can reach the bottom surface, i.e. only those under a certain, relatively small solid angle. This geometrical efficiency in the particle detection is quantified by the acceptance, measured in [ $\text{m}^2 \cdot \text{sr}$ ], which is a function of the dimensions and shape of each experimental setup. The typical acceptances of experiments for direct cosmic ray detection range from several tens of  $\text{cm}^2 \cdot \text{sr}$  to few  $\text{m}^2 \cdot \text{sr}$  in modern apparatuses.

The acceptance, together with the time of data taking, actually defines the range of energies that can be explored by a certain experiment. For instance, a detector with an acceptance of  $\sim 0.003 \text{ m}^2 \cdot \text{sr}$  can detect only few particles/year with energies larger than 100 TeV; instead experiments presenting an aperture of  $\sim 1 \text{ m}^2 \cdot \text{sr}$ , can observe several particles/year also for energies near 1000 TeV. Experiments on satellites are typically small but planned to take data for several years; experiments on balloons can be bigger

(they do not have to be brought in orbit) thus have a larger acceptance, but the time of flight is limited, going from few days to few months.

Furthermore, different detectors have to be employed at the same time, according with the focus of each experiment (see Fig. 1.8). The general paradigm is to couple particle tracking and calorimetric measurements with other techniques. In fact, when the energy per nucleon of a cosmic ray (neglecting for the moment electron/positrons and gammas) is larger than  $\sim 1$  GeV, the cascade of particles generated in the interaction cannot be fully contained, given the size and weight of experiments that can be sent in the atmosphere or near space. For instance, at energies below  $\sim 1$  TeV per nucleon ( $1 \text{ TeV} = 10^{12} \text{ eV}$ ), magnetic spectroscopy can be performed, even if only in the last years it becomes technologically possible to apply a strong magnetic field in such extreme experimental conditions. Above TeV energies, too intense magnetic fields would be required to bend at a measurable level the observed particles; so the identification is performed combining time-of-flight, transition-radiation, RING detectors and so on.

In any case, the acceptances of this type of experiments can hardly exceed values

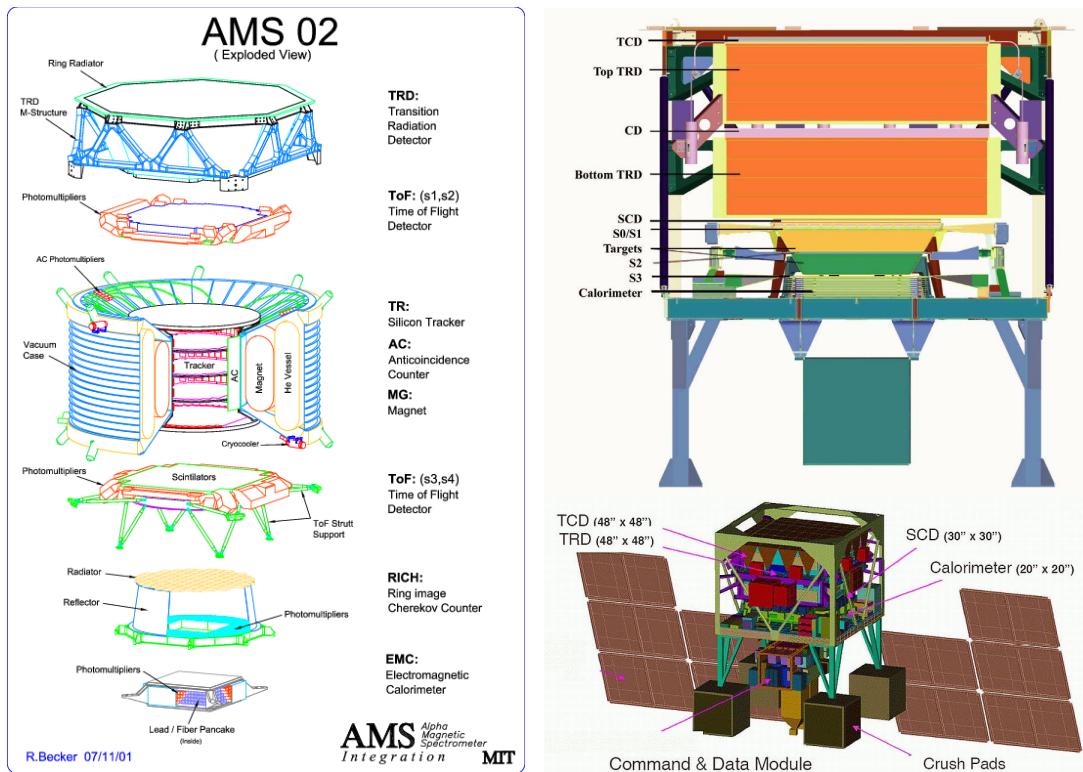


Figure 1.8: Schematic representations of two experiments for the direct detection of cosmic rays. (Left) Exploded view of the AMS-02 (*Alpha Magnetic Spectrometer*) experiment, installed in 2011 on the International Space Station. Image from [28]. (Right) Schematic view of the CREAM (*Cosmic Ray Energetics And Mass*) experiment and of the support structure, used to secure the detector to an high-altitude ( $\sim 40$  km) balloon flying over Antarctica. The acronyms indicate: TCD, Timing-based scintillator Charge Detector; SCD, Silicon Charge Detector; S0-S1-S2-S3, layers of scintillating fibers; other acronyms are explained in the AMS-02 picture. Image from <http://stratocat.com.ar/fichas-e/2009/MCM-20091201.htm>

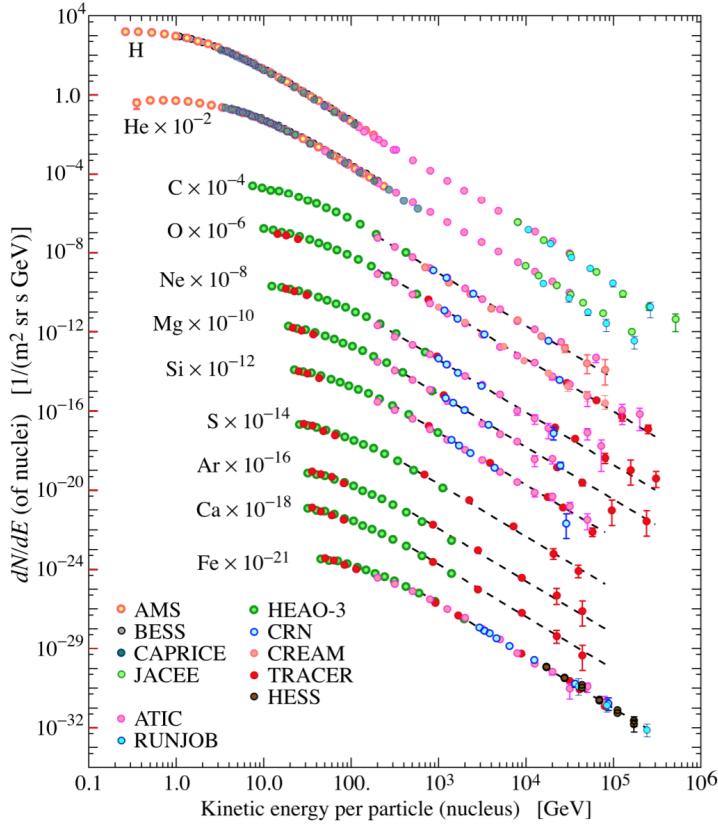


Figure 1.9: Summary of the results from multiple experiments for the energy spectrum of primary cosmic ray nuclei in the region of direct detection. The differential flux of each nucleus can be clearly isolated. Image from [4].

of few  $\text{m}^2 \cdot \text{sr}$ , since the dimensions of an object to be sent to high atmosphere or near space can not be larger than few meters, both for weight and integrity reasons. Therefore, techniques of indirect measurement are required to go beyond energies at about the knee, i.e. above  $10^{15}$  eV.

As detailed in the next chapter, the fundamental process that allows indirect detection (from energies around the knee to the extreme ones) is the production of an *Extensive Air Shower* (EAS, see sect. 2.1), i.e. a cascade of secondary particles generated in the interaction of the primary with the atmosphere molecules. Such EAS represents a footprint of the initial cosmic ray, and the measurements of its atmospheric development (for example through the emitted fluorescence radiation, sect. 2.2.1) and/or a sampling of the secondaries that reach the ground (sect. 2.2.2) lead to the “reconstruction” of relevant physical quantities for each event.

The great advantage of direct detection, compared to indirect one, is to provide mass composition information on an event-by-event basis (of course within certain efficiency and contamination), allowing the determination of different spectra for each mass group (or even for single elements) as shown in Fig. 1.9. Instead, the composition at higher energies has to be statistically inferred from other observables, since the primary is never

observed. This is possibly the most difficult task in air shower physics, as it relies on the comparison between data and simulations. For instance, hadronic interaction models need to be extrapolated towards energies much beyond those accessible in man-made particle accelerators, resulting in large systematic uncertainties. The issue about composition of UHECR and its implications are deeper discussed in sect. 2.3.

## Chapter 2

# The detection of Ultra-High Energy Cosmic Rays

In the early '30s, the Italian physicist Bruno Rossi noticed [29] that several separated particle counters put in a coincidence circuit<sup>1</sup> measure a much higher rate than expected for accidental events. Some years later, the French physicist Pierre Victor Auger conducted several high altitude experiments on the Alps, to study cosmic ray particles reaching the ground. In 1939, he found [32] that detectors located even hundreds of meters apart detect particles with exactly the same arrival times, confirming previous observations of Bruno Rossi [33]. These observations lead to the conclusion that many secondary particles are produced in the atmosphere by cosmic rays of high energy, that Auger initially estimated in  $10^{15}$  eV.

In the following decades, large arrays of detectors were built to better understand the physics of these cascades of particles, called **Extensive Air Showers** (EAS). In fact, it was soon realised that such phenomenon could allow the indirect detection of cosmic rays with extreme energies but a very low flux. This is especially true for ultra high energy cosmic rays (UHECR), that reach the Earth with a rate lower than one particle per year, per  $\text{km}^2$  and would be practically undetectable with other methods.

However, in order to actually reconstruct the characteristics of the primary particle, it is necessary to analyse the interaction processes that govern the atmospheric development of a cascade. The first part of this chapter deals with the description and modelization of EAS development. Then the experimental techniques for the detection of UHECRs will be addressed, focusing on the procedures used nowadays in the largest cosmic ray experiments based on the Earth's surface. Finally, in the third part, mass composition studies and their implications will be analysed.

---

<sup>1</sup>At the age of 24, Bruno Rossi read a paper [30] by Bothe and Kohlhörster about the first coincidence device for recording simultaneous events, based on a mechanical-photographic method. Inspired by this work, Rossi invented [31] an electronic coincident circuit (employing several triode vacuum tubes) that could register coincident pulses from any number of counters with a time resolution smaller than 1 millisecond. The Rossi coincidence circuit was much more practical than the previous design, therefore was rapidly adopted by particle physics experimentalists around the world.

## 2.1 The Extensive Air Showers

As discussed in the previous section, cosmic rays with energies above  $10^{14}$ - $10^{15}$  eV cannot be directly detected. However, the collision of the primary with the atmosphere molecules produces a cascade of particles, called Extensive Air Shower, that in this energy range becomes measurable through experiments on high mountains (and at lower altitudes for increasing primary energies).

An EAS initiated by a proton or a nucleus presents three components (fig. 2.1):

- *hadronic*: mainly composed by  $\pi$  and  $K$  mesons produced in the initial (and most energetic) phases of the shower, rapidly decays during the EAS development;
- *electromagnetic*: made of electrons, positrons and photons<sup>2</sup>, it becomes the dominant component after few interactions in the atmosphere, both from quantitative and energetic point of view (about 90% of the primary energy);
- *penetrating* or *hard*: composed by muons and neutrinos, it is mainly originated in the decay of pions and kaons and it is on average three order of magnitude

---

<sup>2</sup>In first approximation, the e.m. part of the shower is generated by the decay of neutral pions ( $\pi^0$ ) in two high-energy photons, which then produce small e.m. cascades (as described in sect. 2.1.1).

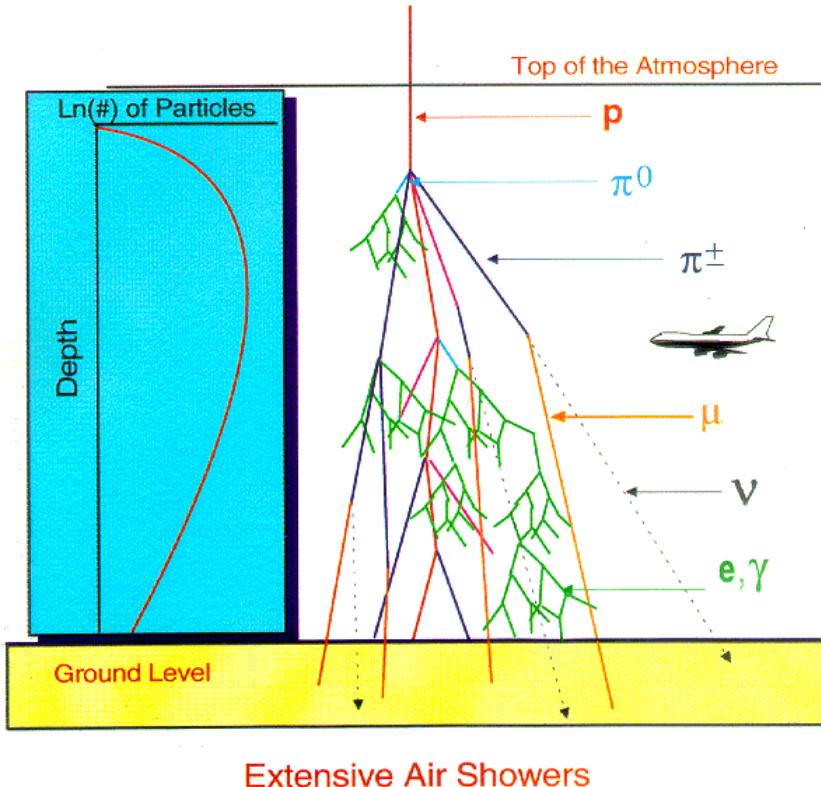


Figure 2.1: Sketch of EAS development in atmosphere: in blue the hadronic component (nucleons, mesons); in green the electromagnetic one ( $\gamma$ ,  $e^-$ ,  $e^+$ ); in orange muons and as dashed lines neutrinos. Neutrinos (and partially muons) do not interact after production. On the left, the evolution of the particle number as a function of the atmospheric depth. Image from <http://www.physics.adelaide.edu.au/astrophysics/hires/uhecr.html>

smaller than the electromagnetic component; it is called “penetrating” as muons and neutrinos reach the ground without further interacting in the atmosphere.

Intuitively the number of particles produced in the cascade grows with the path covered in atmosphere, while the average energy per particle lowers. In order to quantify the amount of matter traversed from the top of the atmosphere to a certain altitude  $h$ , the atmospheric depth  $X$  is defined:

$$X = X_v / \cos \theta \quad \text{with} \quad X_v = \int_h^\infty \rho(h) dh \quad (2.1)$$

where  $\theta$  is the zenith angle of the shower propagation axis with respect to the vertical direction and  $\rho(h)$  the density profile of the atmosphere ( $X_v$  is usually referred as vertical depth). The atmospheric depth is measured in  $g/cm^2$ .

The interaction processes continue until the average energy per particle reaches a certain critical value, a threshold below which particle creation stops. At this point, the maximum number of particle produced in the shower is reached: for this reason, the atmospheric depth corresponding to this stage of the cascade development is called  $X_{\max}$ . Afterwards, the showers weakens while the energy gets degraded in the atmosphere.

In the following the classical models for the description of EAS development in atmosphere are presented. The first step is to predict with good accuracy the behaviour of pure electromagnetic showers, since cross-sections and multiplicity of particles production in each reaction are well known. Subsequently, more complicated (and approximated) models which take into account the hadronic part of the shower can be discussed.

### 2.1.1 The Heitler model for electromagnetic showers

In order to model the EAS development, a starting point is the study of purely electromagnetic showers, which are dominated by two processes<sup>3</sup>: bremsstrahlung and pair production. These reactions are characterised by:

- *radiation length*  $\lambda_r$ , expressed as an atmospheric depth; after crossing a radiation length, an electron gets its energy reduced on average to  $1/e \approx 0.37$  of the initial value through radiative processes;
- *mean free path for pair production*  $\lambda_{pair}$ , approximately equal to  $9/7 \lambda_r$  ;
- *critical energy*  $\varepsilon_c$ , defined as the value when energy losses through radiation are equal to those in ionization processes; for electron in air  $\varepsilon_c \approx 86$  MeV.

A simple model was elaborated by Heitler [34] (and therefore today called **Heitler model**) in the '40s, taking into account the stochastic nature of the shower development. The model is based on the following assumptions:

1. cross sections of all processes are considered energy independent and the two characteristic interaction lengths are considered approximately equal ( $\lambda_{pair} \approx \lambda_r$ );

---

<sup>3</sup>Actually, to obtain the lateral spread of the particles in a shower, it is necessary to include not only pair production and bremsstrahlung, but also multiple Coulomb scattering. In fact, it is the latter that determines the characteristic size of the shower front. However, this aspect is not of interest in the following modelizations of the atmospheric development of the shower.

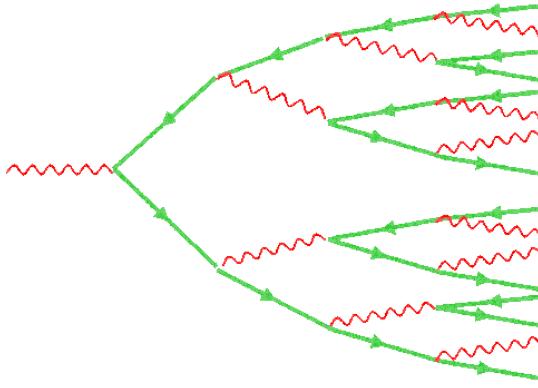


Figure 2.2: Sketch of the shower development according to the Heitler model. Each step corresponds to a radiation length  $\lambda_r \times \ln 2$ ; the wave lines are  $\gamma_s$ , the green lines oriented to the right are electrons, if to the left positrons. Image from <http://www.borborigmi.org/2012/02/28/rivelatori-di-particelle-a-lhc-quinta-parte-sciami-di-elettroni-e-foton>

2. in a bremsstrahlung process, the produced photon and electron share equally the energy of the initial electron;
3. in a pair production process, the produced electron and positron share equally the initial photon energy;
4. Compton scattering and multiple scattering are negligible;
5. the interaction step length is taken equal [6, 35] to  $d = \ln 2 \times \lambda_r$ .

In this picture, after each step (Fig. 2.2) the number of secondaries doubles and the primary energy is divided in a increasingly bigger number of particles. Therefore, after  $n$  steps, the average energy per secondary particle  $\langle E \rangle$  is:

$$\langle E \rangle = \frac{E_0}{2^n} \quad (2.2)$$

where  $E_0$  is the energy of the primary cosmic ray and  $2^n$  the number of  $\gamma/e^+/e^-$ .

By definition, the shower reaches its maximum when the average energy per particles is not enough to generate new secondaries, that is when  $\langle E \rangle$  approaches the critical energy  $\varepsilon_c$ . This condition corresponds to a number of steps  $n_{max} = (\ln 2)^{-1} \ln(E_0/\varepsilon_c)$  and consequently to an atmospheric depth at the shower maximum equal to:

$$X_{\max} = X_0 + \lambda_r \ln(E_0/\varepsilon_c) \quad (2.3)$$

where  $X_0$  is the atmospheric depth of the first interaction of the primary cosmic ray. From this expression, it can also be obtained the evolution rate of the depth of the maximum as a function of the primary energy. This quantity, called *elongation rate*, is usually expressed as the change per energy decade:

$$D_{10} \equiv \frac{dX_{\max}}{d \log_{10} E_0} \cong 2.3 \lambda_r \quad (2.4)$$

Given a value of the radiation length in the medium  $\lambda_r \approx 37 \text{ g/cm}^2$ , the elongation rate for pure e.m. showers results about  $85 \text{ g/cm}^2$  in air.

The Heitler model is only a first approach in the EAS analysis, as it does not take into account the hadronic part of the shower. Nevertheless, it correctly reproduces two fundamental features of an extensive air shower: the proportionality between shower size (intended as the maximum number of particle  $N_{max} = 2^{n_{max}}$ ) and primary energy  $E_0$ ; a logarithmic relation between  $X_{max}$  and  $E_0$ .

### 2.1.2 A generalized Heitler model for hadronic showers

Despite its simplistic nature, the Heitler model can be generalized to consider also the hadronic component of the shower (and therefore the penetrating one), to describe an EAS generated by a nucleon (see for example [36]).

The characteristic quantity for the hadronic shower development is the interaction length  $\lambda_I$  of strongly interacting particles, a quantity inversely proportional to the cross section of the processes and to the air density. In the energy range of interest,  $\lambda_I \sim 90$  g/cm<sup>2</sup> for nucleons and  $\lambda_I \sim 120$  g/cm<sup>2</sup> for pions [35]. The generalised model is based on the following assumptions:

1. a step in the atmosphere is a layer of thickness  $\lambda_I \times \ln 2$  ;
2. traversing one layer, each hadron interacts producing  $2N_\pi$  charged pions and  $N_\pi$  neutral ones; the energy is shared in equal parts;
3. neutral pions immediately decay into 2 photons, populating the electromagnetic part of the shower;
4. charged pions continue to interact, until the average energy per particle reaches a critical value ( $E_c^\pi$ ); after that  $\pi^\pm$  decay into muons and neutrinos.

A schematic representation of this model is shown in Fig. 2.3.

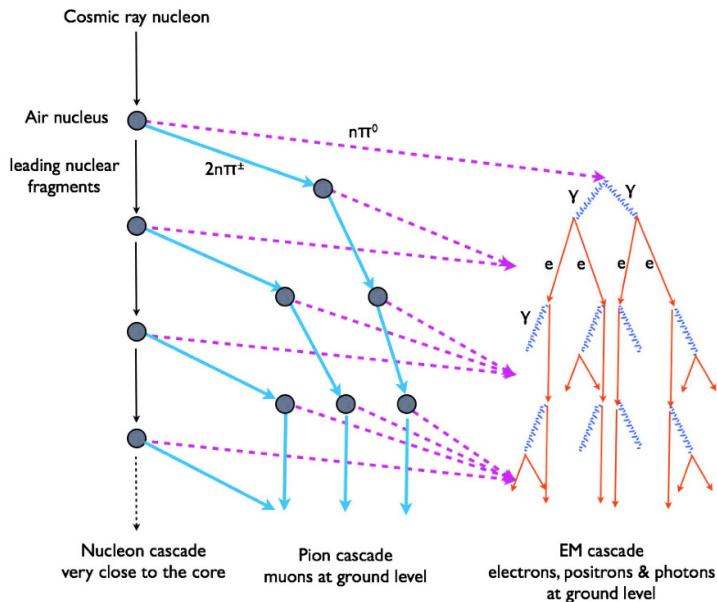


Figure 2.3: Sketch of the model for the development of a hadronic cascade. At each step about 1/3 of the energy is transferred to the electromagnetic component. Image from [35].

The first result of the generalized Heitler model is that, after each step, one-third of the energy is transferred to the e.m. shower component. After  $n$  interaction lengths, the energy is divided between the hadronic and e.m. parts according to:

$$E_{had} = (2/3)^n E_0 \quad E_{e.m.} = [1 - (2/3)^n] E_0 \quad (2.5)$$

Therefore, already for  $n = 6$  about 90% of the primary energy  $E_0$  is carried by the electromagnetic part of the shower.

Another interesting results is related with the total number of muons produced in a hadronic shower, in the assumption that all charged pions decay into muons when the critical energy  $E_c^\pi$  is reached. As a preliminary step, it is necessary to evaluate the average energy of charged pions after  $n$  interactions:

$$\langle E_{\pi^\pm} \rangle = \frac{(2/3)^n E_0}{(2N_\pi)^n} \quad (2.6)$$

where  $2N_\pi$  is the number of charged particles produced at each step, and the repartition of energy between hadronic and e.m. components of the shower is taken into account. Then the critical energy for the pion decay is reached after  $n_c$  interaction lengths, where:

$$n_c = \frac{\ln(E_0/E_c^\pi)}{\ln(3N_\pi)} \quad (2.7)$$

At this point, the total number of muons  $N_\mu$  can be written as:

$$N_\mu = (2N_\pi)^{n_c} = \left(\frac{E_0}{E_c^\pi}\right)^\alpha \quad \text{with} \quad \alpha = \frac{\ln(2N_\pi)}{\ln(3N_\pi)} \sim 0.9 \quad (2.8)$$

Therefore the muonic size of the shower increases with the primary energy, even if at a slower rate with respect to the electromagnetic one. Of course this result is valid only in first approximation: for instance, the multiplicity of the hadronic interaction (i.e. the effective number of charged and total particles produced in each interaction, which leads to the value of  $\alpha$ ) depends on the energy. As a consequence, systematics errors due to extrapolation of hadronic interactions at extreme energies (much higher than those reached in collider experiments) must be taken into account.

Finally, one can try to evaluate the atmospheric depth correspondent to the maximum number of particle produced in a proton-initiated shower. However, the cascade is numerically dominated by e.m. particles, thus  $X_{\max}$  is defined as the maximum of the electromagnetic part. In the simplifying assumption made in [36], the  $X_{\max}$  value is influenced only by the starting generation of e.m. sub-showers, generated by the decay of  $\pi^0$  produced by the primary particle in the first interaction length. Therefore:

$$X_{\max}^p(E_0) \approx X_0 + \lambda_I \ln 2 + X_{\max}^{e.m.} \left( \frac{(1/3)E_0}{2N_\pi} \right) \quad (2.9)$$

where  $X_{\max}^{e.m.}$  of each sub-shower is calculated for a cascade with an initial energy equal to 1/3 of the primary proton energy shared among  $(2N_\pi)$  particles, that is the number

of photons by neutral pions generated in the first step (i.e. after a travelled thickness of  $\lambda_I \ln 2$  for the primary particle). Substituting eq. 2.3 in the previous formula:

$$X_{\max}^p(E_0) \approx X_0 + \lambda_I \ln 2 + \lambda_r \ln \left( \frac{E_0}{6N_\pi \varepsilon_c} \right) \quad (2.10)$$

The determination of the actual depth of shower maximum is more complicated and requires a proper study of hadronic cross-sections, multiplicity and inelasticity of the interactions, together with detailed simulations. However, the expression in eq. 2.10 allows to evaluate the elongation rate for a hadronic shower:

$$D_{10}^p \equiv \frac{dX_{\max}^p}{d \log_{10} E_0} \approx D_{10}^{e.m.} \left[ 1 - \frac{d \log_{10}(6N_\pi)}{d \log_{10} E_0} \right] + \ln 2 \frac{d\lambda_I}{d \log_{10} E_0} \quad (2.11)$$

The particle multiplicity  $N_\pi$  slowly increases with energy, while the hadronic interaction length  $\lambda_I$  is naturally shorter for higher energies; therefore the elongation rate per energy decade of proton-initiated showers is always smaller than the correspondent value for e.m. cascades. This result is a direct consequence of the larger number of particle produced in hadronic EAS for each step, which gives a much higher conversion rate of the primary energy into secondary particles.

### 2.1.3 The superposition model

The generalized, even if simplistic, model for the description of proton-initiated (and thus hadronic) showers reported in the previous section has to be eventually extended to the case of a primary nucleus.

Since the binding energy per nucleon ( $\sim 5$  MeV) is much smaller than the interaction energies of interest for cosmic ray physics, one can consider a primary nucleus of mass  $A$  and energy  $E_0$  as a bunch of  $A$  primary nucleons with energy  $E_0/A$  which independently interact in the atmosphere (see Fig. 2.4). The behaviour of every sub-shower follows the generalized Heitler model but the complete shower development results as a **superposition** of the cascades generated by each nucleon.

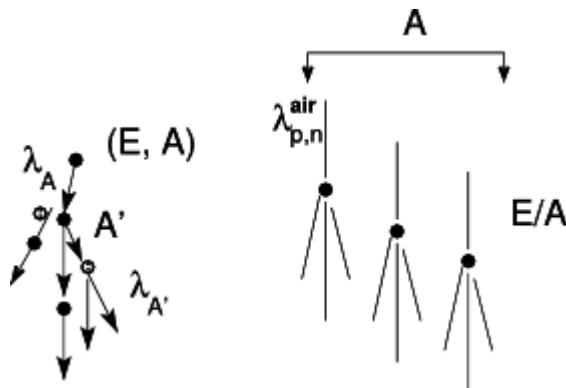


Figure 2.4: The basic assumption of the superposition model: a primary nucleus with mass number  $A$  and energy  $E$  behaves as  $A$  independent primary nucleons with energy  $E/A$ .

One of the important aspects of the superposition model is that, averaged over many showers, the distribution of nucleon interaction points in the atmosphere coincides with that of more realistic calculations [37, 38]. Therefore it is not surprising that the superposition model gives a good description of many features of air showers, such as the mean depth of shower maximum and the number of muons.

In particular, an interesting results for  $X_{\max}$  is immediately obtained from eq. 2.10 under the superposition assumption:

$$X_{\max}^A(E_0) = X_{\max}^p(E_0/A) = X_{\max}^p(E_0) - \lambda_r \ln A \quad (2.12)$$

which means that heavier nuclei develop upper in the atmosphere. Moreover, the shower-to-shower fluctuations of the  $X_{\max}$  position are intuitively reduced for heavy nuclei with respect to light ones, as result of the combination of several cascades.

The superposition model also predicts a faster development of the shower (since the initial energy of each nucleon is lower than the energy of the primary nucleus), meaning that the pions reach their critical energy sooner. Therefore the number of muons with respect to the e.m. component will be larger, as can be obtained from eq. 2.8:

$$N_\mu^A(E_0) = \left( \frac{E_0}{AE_c^\pi} \right)^\alpha = A^{1-\alpha} N_\mu^p(E_0) \quad (2.13)$$

For instance, an iron-initiated shower is supposed to contain  $\sim 50\%$  more muons than a proton-initiated one. Another result concerns the elongation rate for showers with the same mass  $A$ ; assuming that the evolution of cross section and multiplicity with energy for primary nuclei is the same as for protons:

$$D_{10}^A \equiv \frac{dX_{\max}^A}{d \log_{10} E_0} = \frac{d(X_{\max}^p - \lambda_r \ln A)}{d \log_{10} E_0} = D_{10}^p \quad (2.14)$$

which means that different nuclei present identical elongation rates. This is a very important prediction, since it implies that a change of the composition with primaries energy will be clearly reflected in the elongation rate.

Finally, it is necessary to underline that all the described trends for  $N_\mu$ ,  $X_{\max}$  and related evolutions with energy and mass are reproduced by detailed MC simulations, but with differences in quantitative terms. This is mainly related with our poor knowledge of the nuclear and subnuclear reactions at extreme energies, that reflects in hadronic interaction models which are not able to fully reproduce UHECR observations.

## 2.2 The detection techniques for EASs

As anticipated, the detection of high energy cosmic ray relies on the observation of extensive air showers. In particular EAS-related observables, such as the atmospheric development of the cascade, the density of particles at ground level and the muonic content, are employed to infer the fundamental characteristics of the primary particle, i.e. its energy and arrival direction.

Instead, the composition can be determined only with the statistical analysis of many EASs, since the shower-to-shower fluctuations nullify the direct correlation between the primary mass and the atmospheric development for a single event.

The detection of UHECR is extensively performed through two experimental techniques<sup>4</sup>: using **fluorescence telescopes**, which measure the UV light emitted by a cascade during its development; using **ground arrays** of surface detectors (usually scintillators or water-Cherenkov tanks), which sample the shower particles at ground level.

### 2.2.1 The fluorescence technique

The study of air fluorescence began in the Los Alamos Scientific Laboratories (early '60s) as a method for detecting the yield of nuclear explosions in atmospheric tests. Only at a later time it was understood that such method can be used in the field of cosmic ray physics. The experimental technique takes advantage of the radiation, called fluorescence light, emitted by the air nitrogen molecules excited by the passage of the (ionizing component of the) cascade. The physical process behind this emission is examined more in depth in sect. 2.2.1. Important features for fluorescence measurements are:

1. the fluorescence emission is detectable only for primary energies above  $10^{17}$  eV;
2. fluorescence radiation is emitted isotropically, its wavelength is in a narrow band (between 300 – 400 nm) and presents low absorption in air;
3. the fluorescence light yield is proportional to the energy loss per unit of length per charged particle (see next paragraph).

The basic idea is to use the atmosphere as an enormous calorimeter, which is possible thanks to the last listed characteristic of the fluorescence emission. In fact, collecting the fluorescence light emitted by secondary particles can be considered equivalent to a measurement of the energy dissipated during the shower development. Furthermore, isotropy and low absorption in air allow the realization of telescopes that can observe large regions of the sky, since the fluorescence light emitted by a shower can be observed by any direction and at large distances (several kilometers).

A typical fluorescence telescope presents one or more mirrors which focus the light into a matrix of photomultipliers (PMTs): each PMT independently observes a small area of the total field of view of the telescope (Fig. 2.5a). In this way the sky is seen as a collection of “tiles” (see Fig. 2.5b), usually referred to as pixels.

Each PMT measures a signal amplitude, that is the integrated value of the light collected during a characteristic time interval, and gives a timing information through a trigger time, i.e. the time when the collected photons reach a threshold value. Therefore the shower trajectory is observed as a time sequence of light pulses: from the arrival time at each pixel (correspondent to the passage of a cascade in the PMT field of view) it is

---

<sup>4</sup>In the last decade, also a technique based on the measurements of EAS radio emission has been conceived. However, being still innovative and under development it is not described in the following.

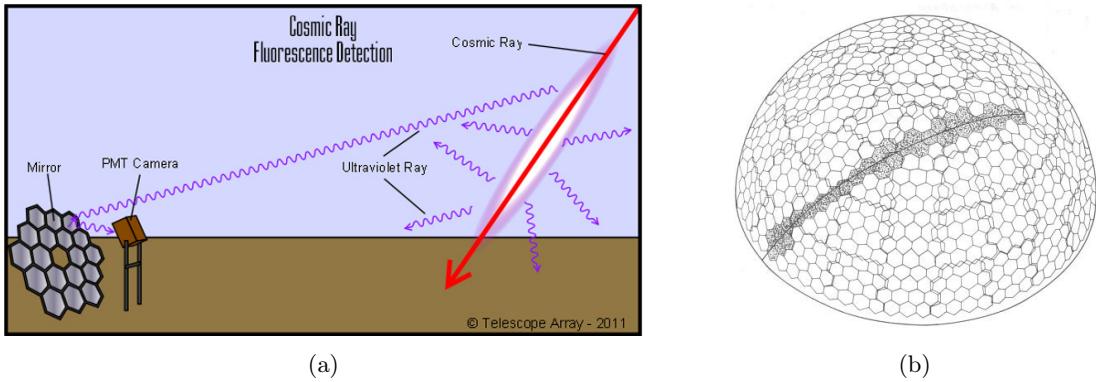


Figure 2.5: (a) Schematic representation of the detection of an EAS by means of a fluorescence telescope. Image from <https://www.livescience.com/46664-photos-telescope-array-images.html>. (b) A 3D composition of an event detected by the *Fly's Eye* experiment: the field of view is seen as a collection of “tiles” or pixels, each corresponding to a photomultiplier. Image from <http://www.cosmic-ray.org/reading/flyseye.html>.

possible to reconstruct the shower axis direction; then the detected signals are converted into an energy estimation<sup>5</sup>.

Nevertheless, the fluorescence emission is very faint, so this kind of detection is heavily affected by any light source (artificial or natural) other than the EAS<sup>6</sup>. For this reason, the environmental conditions suitable for data taking are matched only during clear and moonless nights (about one week before and one week after a new moon), which limits the acquisition time (duty-cycle) to the 10 – 15%. Furthermore, a continuous monitoring of the atmosphere is required to determine its optical properties and to perform a correct reconstruction of the EAS development.

The detection of an EAS from a single fluorescence telescope sometimes does not allow an accurate reconstruction of the shower geometry, in particular when the distance of the shower from the detector turns out to be evaluated with large uncertainties. In these cases, it is necessary to have another telescope observing the same event (*stereo observation*) or provide additional constraining information from a ground detector (*hybrid reconstruction*, a key feature in the case of the Pierre Auger Observatory).

The main advantage of the fluorescence detection technique is the direct experimental access to the position of  $X_{\max}$ , the most commonly used composition sensitive parameter. A deeper discussion on this topic is postponed to sect. 2.3.

### The fluorescence emission

The fluorescence light is emitted by the disexcitation of the air nitrogen molecules ( $N_2$ ) after the interaction with charged particles of the shower. The strongest emission is in

<sup>5</sup>A more complete description of shower reconstruction with fluorescence technique, as performed in the framework of the Pierre Auger Collaboration, is reported in sect. 3.2.1.

<sup>6</sup>Also the Cherenkov light emitted by the shower particles has to be evaluated and subtracted. In fact, even if this radiation is emitted in the direction of propagation of the cascade, it can be diffused by the atmosphere into the telescopes.

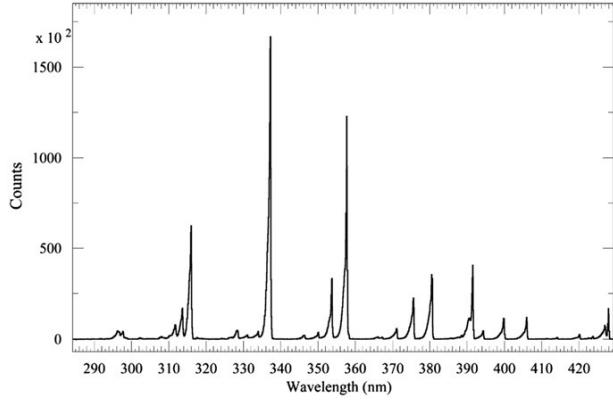


Figure 2.6: Spectrum of nitrogen emission in dry air at 800 hPa e 293 K. Image from [39].

the near UV region between  $300 \div 400$  nm, with two intense lines at 337.1 nm and 357.7 nm, as shown in the emission spectrum in Fig. 2.6.

The fluorescence yield  $Y_\gamma$ , that is the number of photons produced by a charged particles per unit of length crossed in air, results to be about 5 photon per MeV of released energy [39, 40]. However, it changes for different atmospheric conditions, in particular as a function of air density, pressure and temperature.

Once emitted, the fluorescence light faces attenuation processes in the atmosphere, which principal causes (see Fig. 2.7) are:

- *Rayleigh scattering*: the scattering centres are the air molecules; the intensity of this process is approximately constant in time, as the air composition presents only slight seasonal variations. However, it is usually evaluated from the atmospheric state variables (temperature, pressure, humidity) which are continuously measured. The cross section goes as  $\lambda^{-4}$  (with  $\lambda$  the wavelength of the emitted radiation) and the distribution of the scattering angles is proportional to  $(1 + \cos^2 \theta)$ ;
- *Mie scattering*: the scattering centres are the aerosol particles; the monitoring of aerosols is one of the “calibration challenges” in a fluorescence cosmic ray experiment, since their characteristics (distribution in dimension, composition, density profile with altitude/atmospheric depth, etc.) can vary in short space and time scales. The cross section weakly depends on the wavelength ( $\sim \lambda^{-1}$ ), and the angular distribution is strongly peaked forward.

Given the initial intensity  $I_0$  of the emitted radiation, the light intensity  $I$  at a distance  $r$  from the source is:

$$I(r) = I_0 \exp\left(-\frac{r}{N\sigma}\right) \quad (2.15)$$

where  $N$  is the numeric density of the scattering centres and  $\sigma$  the cross section. Often it is preferred to express the result of the attenuation as a extinction length  $r_{\text{ext}} = N\sigma$ , which is evaluated continuously by the atmosphere monitoring: approximately  $r_{\text{ext}} = 10$  km in the wavelength interval  $300 \div 400$  nm for clean and moonless nights.

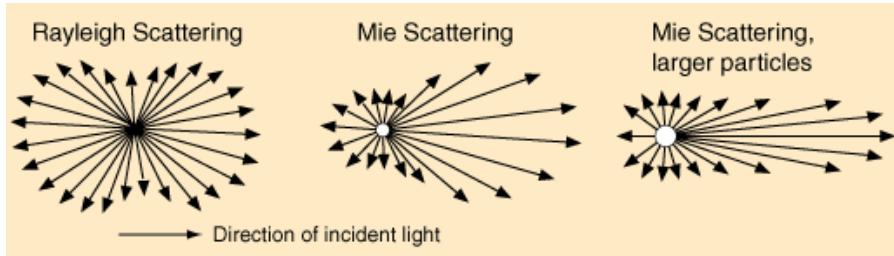


Figure 2.7: Sketches of the light scattering processes which are the main cause of the attenuation of the EAS fluorescence radiation in atmosphere. (*Left*) *Rayleigh scattering*, due to the air molecules; (*right*) *Mie scattering* due to the aerosols, with a forward emission even more directional for larger particles. Image from <http://hyperphysics.phy-astr.gsu.edu/hbase/atmos/blusky.html>.

### 2.2.2 The surface array detection

After the end of the WWII, many pioneering experiments demonstrated that cosmic rays can be actually studied through the measurement of Extensive Air Showers. Since it is impossible to build an extended but continuous device, the optimal solution is a network of ground-based particle detectors which sample the shower properties at ground level.

Nowadays the arrays of surface detectors are employed mainly for energies around and above the knee, that is when the flux becomes too low for direct detection. The main features of these instruments are:

- the covered surface, which is related with the flux in the energy range of interest for the experiment; for instance, few  $\text{km}^2$  are enough for measurements around the knee ( $10^{15}$  eV) while several hundreds of  $\text{km}^2$  are required to observe events above the ankle and approaching the cut-off (more than  $10^{19}$  eV);
- the spacing among the detectors, which is also dependent on the energies one is aiming to measure; denser arrays are necessary for lower energies, since the dimension of the EAS footprint at ground level grows with the primary energy. For UHECR detection, a spacing of 1 km or larger is suitable;
- the altitude, which is related to  $X_{\max}$ ; in fact a sampling measurement has to be accomplished near the maximum of the shower, where the number of secondary particles is larger and the fluctuations are smaller. As lower energy cosmic rays develop higher in the atmosphere, the experiments for detection in the knee region are built on the highest reachable altitudes, while quotes between 1000-2000 meters are sufficient for UHECR observations.

Another important characteristic is the continuous acquisition, also necessary to collect a significant set of observations in a reasonable interval of time.

The detectors commonly used are *scintillators* or *water-Cherenkov* detectors, with a sensitive surface of 1-20  $\text{m}^2$ , designed for a large acceptance in solid angle and 24/7 data acquisition. Therefore the actual physical quantity measured by a surface array is the energy released by secondary particles in each detector as a function of time. This kind of observation is by definition a sampling measurement of an EAS, and only at a certain

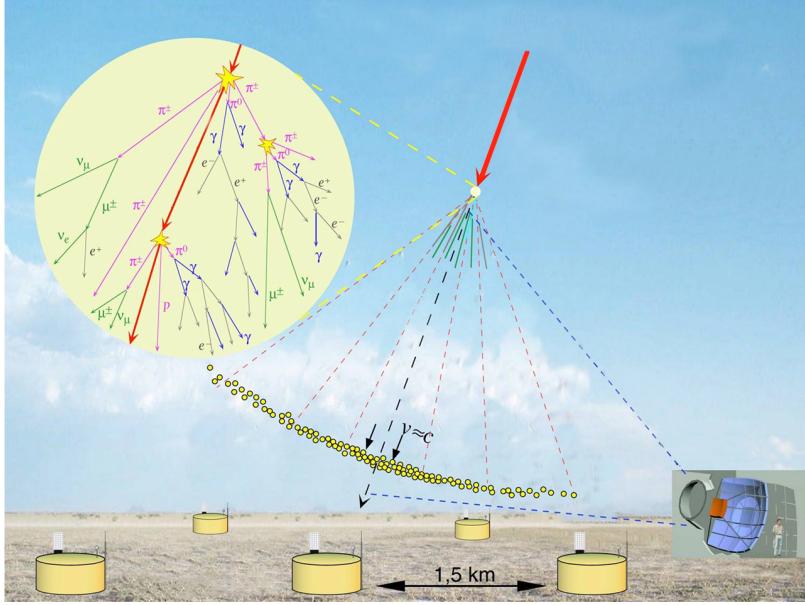


Figure 2.8: Sketch of the detection of an EAS by a ground array, with detectors and spacing as for the Pierre Auger Observatory. Image from <http://apcauger.in2p3.fr/Public/Presentation/>.

stage of its development, function of the shower characteristics and of the traversed atmospheric depth  $X$  (see eq. 2.1).

There are important differences between the above-mentioned two typologies of detectors. Scintillators are equally sensitive to all charged particles thus measure mostly the e.m. component of the cascade, which exceeds of several orders of magnitude the muonic one. Moreover, the aperture of scintillator arrays rapidly decrease with zenith angle, because of the e.m. particle absorption and the quick drop in the effective area for detection, given the fact that scintillators are usually flat. Therefore, measurements of scintillator arrays are often restricted to zenith angles below  $45^\circ$ . Instead, since the amount of emitted Cherenkov light is proportional to the pathlength of a particle inside matter, water tanks are more sensitive to the muonic component of the shower: in fact, a muon entirely crosses the detector, while  $e^+ / e^-$  are absorbed in a few centimetres/tens of centimetres. Furthermore, water-Cherenkov tanks present a height around 1 m, thus offering a non-zero effective surface for horizontal showers.

An EAS candidate is identified by the delayed coincidences among several neighbouring detectors (see Fig. 2.9a). From the relative time delay and position of the triggered detectors, together with the intensity of the signal (proportional to the particle density) recorded in each of them, one can reconstruct the *shower axis* and hence the primary direction, with typical resolutions of  $1\text{-}3^\circ$ .

At this point, the signal densities are analysed as a function of the distance from the *shower core*, identified as the intersection between the direction of propagation and the ground surface (and also the point where the particle density is maximal). Then the **lateral distribution function** (LDF) is evaluated comparing the measured signals

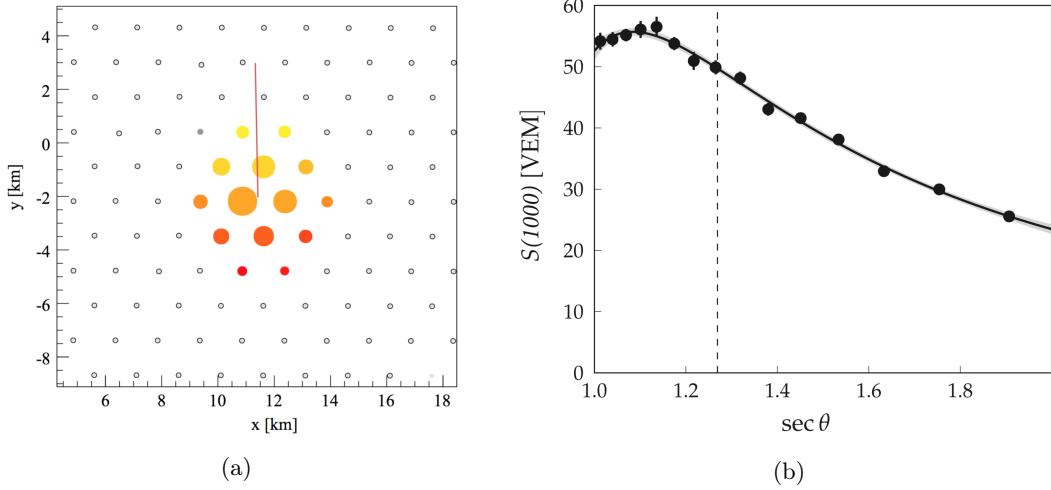


Figure 2.9: (a) An example of signals induced by an air shower in the stations of the surface detector of the Pierre Auger Observatory (spacing of 1.5 km). Colors represent the arrival time of the shower front, from yellow (early) to red (late), while the markers' size is proportional to the logarithm of the signal. The line represents the shower arrival direction. (b) Signal attenuation with increasing zenith angle for showers of approximately equal energy, according with the CIC procedure. In particular, it is represented the signal evaluated at 1000 m from the core (reference distance  $r_{opt}$  for the Pierre Auger Observatory) for showers considered equivalent to an event with a measured value of 50 VEM at the reference zenith angle  $\theta_{opt} = 38^\circ$  (corresponding to about 10.5 EeV). The line is a third degree polynomial used to fit the data. Images from [45].

with well-established empirical parameterizations<sup>7</sup>.

In order to reconstruct the primary energy, it is necessary to define an *estimator* to be extracted from the shower measurements, usually the particle density (or directly the signal) evaluated at an optimal distance  $r_{opt}$  from the shower core [41, 42]. Such  $r_{opt}$  is identified as the distance where the fluctuations, both shower-to-shower and statistical (from particle counting), are minimal; thus it changes with each specific experimental setup, depending on the energy range of interest and on the array spacing.

Furthermore, the attenuation due to the zenith angle  $\theta$  have to be taken into account. An air shower with  $\theta \sim 60^\circ$  traverses about twice the amount of atmosphere compared to a vertical shower (see eq. 2.1), a relevant difference in the absorption of electromagnetic particles, which are the dominant component of an EAS. Therefore showers initiated by identical primary cosmic rays present decreasing signals, as measured by the detectors of a surface array, for increasing zenith angle (at least up to around  $60^\circ$ ). The correction for such attenuation is accomplished through *Constant Intensity Cut* (CIC) methods (see for example [43, 44]), empirical techniques in which the shower size is normalized to a particular zenith angle  $\theta_{opt}$  (see Fig. 2.9b) in the assumption of an isotropic flux of primaries at the top of the atmosphere.

---

<sup>7</sup>One of the usual assumption is that an EAS presents axial symmetry. This assumption is valid in first approximation for zenith angles up to about  $60^\circ$ . For larger values, the e.m. part of the cascade starts to be heavily absorbed and muons to be bent by the geomagnetic field. However, even below  $60^\circ$ , the differences in the paths covered by e.m. particles from opposite sides of the cascade already produces a measurable effect on the signals at ground, the so-called *azimuthal asymmetry*.

Finally, the energy is parameterized as a function of the estimator evaluated at  $r_{opt}$  and  $\theta_{opt}$ . In the case of the Pierre Auger Observatory, it takes the form:

$$E = k \ S(r_{opt}, \theta_{opt})^\beta \quad (2.16)$$

where  $S(r_{opt}, \theta_{opt})$  is the signal evaluated at  $r_{opt}$  using the LDF parameterization and renormalized at  $\theta_{opt}$  through the CIC technique, while  $k$  and  $\beta$  are two parameters to be obtained by independent methods. In fact, an array of surface detectors must be always calibrated in order to obtain the energy of each event. The calibration is performed using simulations or (if available) by means of hybrid observations, i.e. studying events detected at the same time by the ground array and by fluorescence telescopes, since the last ones can directly estimate the shower energy. Then  $k$  and  $\beta$  are fitted and the energy is calculated for each event with a certain systematic uncertainty: this is the origin of the so-called “energy scale” differences among giant-array experiments.

One of the main limitations of this technique is the difficulty of accessing mass sensitive parameters; for instance, information about the atmospheric depth of the shower maximum cannot be directly obtained using arrays of ground-based detectors. A more in-depth discussion about the evaluation of the cosmic rays composition at high energies is addressed in the next section.

## 2.3 Mass composition analysis

The composition of cosmic rays is a fundamental element for the interpretation in a physics-wise picture of the energy spectrum features, such as the ankle and the flux suppression, and in the identification of the transition from galactic to extragalactic sources. Unfortunately, as discussed in sect. 1.4, in the energy range where indirect measurements becomes necessary (namely above the knee) the composition of cosmic rays cannot be determined on an event-by-event basis. For this reason, it is mandatory to establish which observable shower parameters are more sensitive to the primary mass and then extract the average behaviours as a function of the energy (in the hypothesis that cosmic ray composition does not abruptly changes with energy) from a statistically significant set of well-reconstructed events.

One of the first strategies invented in this field is based on the comparison of the total number of electromagnetic particles (usually identified with the shower size and mostly dependent on the primary energy) with the total number muons (that is correlated with the primary mass as described in eq. 2.13) measured at ground, as illustrated in Fig. 2.10a. For such type of analysis it is required to separately measure the muonic content of the shower, mostly using underground muon detectors.

At later times, with the development of the fluorescence technique, it became possible to evaluate with sufficient accuracy the  $X_{\max}$  value that, as exposed in sect. 2.1.3 (in particular with eq. 2.12), carries information about the nature of the primary particle.

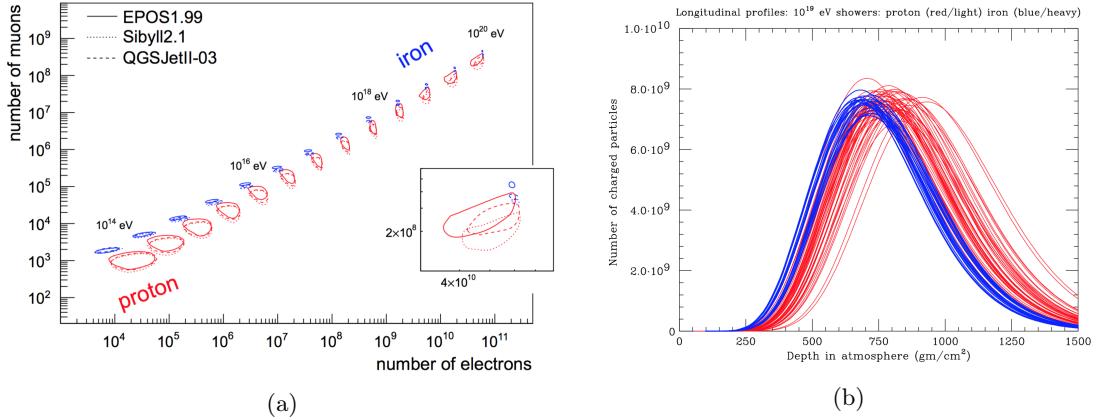


Figure 2.10: (a) Plots of the number of muons vs. electrons at ground level ( $800 \text{ g/cm}^2$ ) for simulated vertical shower. Contour lines indicate the regions which include 90% of the showers; the inset shows a detailed view at  $10^{20}$  eV. Image from [46]. (b) Simulated shower profiles for proton (red) and iron (blue) primaries at  $10^{19}$  eV, which represent the extreme cases for nuclear cosmic rays. It is evident that the  $X_{\max}$  of iron-initiated showers is less deep and the shower-to-shower fluctuations are smaller with respect to the proton-initiated ones. Image from [21].

However, the fluctuations in the atmospheric depth of the first interaction and the intrinsically stochastic behaviour of each EAS cause large differences in the development from shower to shower, also for a fixed primary, as illustrated in Fig. 2.10b. Then it remains impossible to determine the primary mass of a cosmic ray only from the  $X_{\max}$  measured from the induced cascade.

Moreover, this method presents strong limitations: the fluorescence emission is effective only for primary energies above  $10^{17}$  eV and the duty cycle of fluorescence telescopes is considerably short (around 10-15%). Consequently, for energies above  $\sim 10^{19.5}$  eV the number of events that can be collected in the lifetime of a typical ground-based experiment for UHECR detection is not sufficient for conclusive results. To overcome this problematic, independent methods based on observables available through surface arrays (with their approximately 100% duty cycle) have been studied over the years.

An example is represented by analysis that exploit the time evolution of the signals recorded by each surface detector, which was demonstrated to carry information on the position of the shower maximum (mainly in its earliest part)<sup>8</sup>.

In the framework of the Pierre Auger Collaboration, it has been recently proposed an innovative method based on the paradigm of air shower Universality. The idea is to develop an algorithm which reconstructs the macroscopic characteristics of the atmospheric development of an EAS using only the data collected by the surface detector of the Pierre Auger Observatory, in particular the total and time-dependent signals. The improvement of this reconstruction method is actually the subject of this thesis, so it will be deeply addressed in the next chapters.

Such methods are naturally less sensitive than a direct measurement of  $X_{\max}$  as done

<sup>8</sup>In the Pierre Auger Observatory, an employed parameter is the *risetime* [47], defined as the time of increase from 10% to 50% of the total integrated signal for a water Cherenkov detector.

by means of the fluorescence technique, and also heavily rely on simulations. Therefore, the depth of the shower maximum remains the most commonly used shower observable for the study of the composition of UHECR. For this reason, the next section deals with the interpretations of the first ( $\langle X_{\max} \rangle$ ) and second ( $\sigma(X_{\max})$ ) moments of the  $X_{\max}$  distribution as functions of the energy.

### 2.3.1 The interpretation of the first and second $X_{\max}$ moments

The widespread strategy, employed to extract hints about the mass composition from the measurements of  $X_{\max}$ , is based on the analysis of the mean and standard deviation of the  $X_{\max}$  distributions in small energy intervals. These quantities are often simplistically interpreted as two points of view over the same phenomenon and independently compared with the values expected for proton and iron primaries. Instead, the two moments of the  $X_{\max}$  distributions reflect different aspects of the underlying distribution of primary masses and require a more careful examination (see [48]).

As reported in sect. 2.1.3, according to the superposition model  $\langle X_{\max} \rangle$  is linear in the logarithm of  $A$ , the primary mass number (see eq. 2.12). Therefore, for nuclei of the same mass  $A$ , the shower maximum is expected to be on average:

$$\langle X_{\max} \rangle = \langle X_{\max} \rangle_p + f_E \ln A \quad (2.17)$$

where  $\langle X_{\max} \rangle_p$  denotes the mean depth at maximum of proton showers (for instance obtained from eq. 2.12 or more advanced parameterizations) and the energy-dependent parameter  $f_E$  encloses the specific hadronic interaction model features.

In the case of a mixed composition, eq. 2.17 can be simply rewritten as

$$\langle X_{\max} \rangle = \langle X_{\max} \rangle_p + f_E \langle \ln A \rangle \quad (2.18)$$

where  $\langle \ln A \rangle$  is the mean of the  $\ln A$  distribution at the top of the atmosphere.

The behaviour of  $\sigma(X_{\max})$  is more complex, as only in the case of pure composition there is a one-to-one correspondence between its value and a given logarithmic mass:

$$\sigma^2(X_{\max}) = \sigma_{sh}^2(\ln A) \quad (2.19)$$

where  $\sigma_{sh}^2(\ln A)$  is the  $X_{\max}$  variance for the mass  $A$ , due to the intrinsic shower-to-shower fluctuations. Instead, in the case of a mixed composition there is another independent source of fluctuations: the  $\ln A$  dispersion arising from the mass distribution. This contribution can be written as:

$$\left( \frac{d \langle X_{\max} \rangle}{d \ln A} \right) \sigma_{\ln A}^2 = f_E^2 \sigma_{\ln A}^2 \quad (2.20)$$

where  $\sigma_{\ln A}^2$  is the variance of the  $\ln A$  distribution. So the value of the  $X_{\max}$  variance in the case of a mixed composition results:

$$\sigma^2(X_{\max}) = \langle \sigma_{sh}^2 \rangle + f_E^2 \sigma_{\ln A}^2 \quad (2.21)$$

where the first term  $\langle \sigma_{sh}^2 \rangle$ , related with the shower-to-shower fluctuations, is the average variance of  $X_{\max}$  weighted according to the  $\ln A$  distribution.

The equations 2.18 and 2.21 can be inverted to get the first two moments of the  $\ln A$  distribution using the measurements of  $\langle X_{\max} \rangle$  and  $\sigma^2(X_{\max})$ . For the mean, one obtains:

$$\langle \ln A \rangle = \frac{\langle X_{\max} \rangle - \langle X_{\max} \rangle_p}{f_E} \quad (2.22)$$

Instead, for the variance, an explicit expression of  $\langle \sigma_{sh}^2 \rangle$  is needed. Using the parameterization given in [48], one obtains:

$$\sigma_{\ln A}^2 = \frac{\sigma^2(X_{\max}) - \sigma_{sh}^2(\langle \ln A \rangle)}{b \sigma_p^2 + f_E^2} \quad (2.23)$$

where  $\sigma_p^2$  is the  $X_{\max}$  variance for proton showers,  $b$  is a parameter depending on the hadronic interaction model and  $\sigma_{sh}^2(\langle \ln A \rangle)$  is the value of the shower fluctuation calculated at the average mass (obtained from eq. 2.22).

The equations 2.22 and 2.23 are the key expressions to interpret the  $X_{\max}$  data gathered by experiments that measure the shower development using fluorescence light profiles<sup>9</sup>. The power of this method is in the combined analysis of the mean and the variance of  $\ln A$ , that can heavily constraint astrophysical scenarios for the UHECR nuclei acceleration and propagation. In fact the parameter  $\sigma_{\ln A}^2$  represents the dispersion of the masses as they hit the atmosphere, thus reflecting not only the spread of nuclear masses at the sources but also the modifications that occur during their propagation to the Earth. Coupling this information with the average composition obtained from  $\langle \ln A \rangle$ , there are predictably only a few source models which are able to reproduce the observations.

---

<sup>9</sup>Furthermore, the results of such analysis can in principle be used to assess the validity of the available hadronic interaction models, if the uncertainties (statistical and above-all systematic) are enough low.

## Chapter 3

# The Pierre Auger Observatory

The Pierre Auger Observatory is the world’s largest cosmic ray experiment, designed for the detection of the highest energy cosmic rays (namely above  $10^{17}$  eV). Located near the town of Malargüe in the Province of Mendoza, Argentina, the Observatory is a hybrid detector, where the measurements of a surface array of 1660 ground detectors and of 24 fluorescence telescopes are combined to achieve an accurate reconstruction of EAS.

The surface detector (SD) occupies an area of  $\sim 3000$  km $^2$  in a vast high plain in the Andes region (*Pampa Amarilla*) at a mean altitude of 1400 m a.s.l. (corresponding to an atmospheric depth of  $\sim 875$  g/cm $^2$ ). This area is observed, during clear moonless nights, by the fluorescence detector (FD) composed by 4 buildings with 6 telescopes each, situated at opposite sides of the array (see Fig. 3.1).

The hybrid design is a key feature of the Pierre Auger Observatory, since the two detection techniques (fluorescence and surface array, described respectively in sect. 2.2.1 and sect. 2.2.2) allow to measure air showers in complementary ways, providing important cross-checks, data redundancy and substantially reducing the need of simulations.

This chapter deals with the description of the Observatory and the procedures of EAS reconstruction. Then some relevant physics results are shortly presented.

### 3.1 The surface detector

The surface detector (SD) is composed by 1660 water-Cherenkov detectors (WCD) placed in a triangular grid with side of 1500 m. This spacing is an optimal compromise between covering an extended area and assuring a sufficient sampling of the shower properties at ground level<sup>1</sup>. In fact, for energies above  $3 \times 10^{18}$  eV, an event falling on any part of the SD array is detected with 100% efficiency independently of primary particle mass. A smaller array (called *infill*) with a shorter spacing of 750 m is also present, constituting a extension of the standard SD for the measurement of lower energy showers.

Each WCD is a cylindrical tank of 1.2 m height and 3.6 m diameter, made of polyethy-

---

<sup>1</sup>Above 10 EeV the footprint of an EAS on the ground can extend over 25 km $^2$  (depending on its inclination), assuring the trigger of several stations also for a nearly vertical shower.

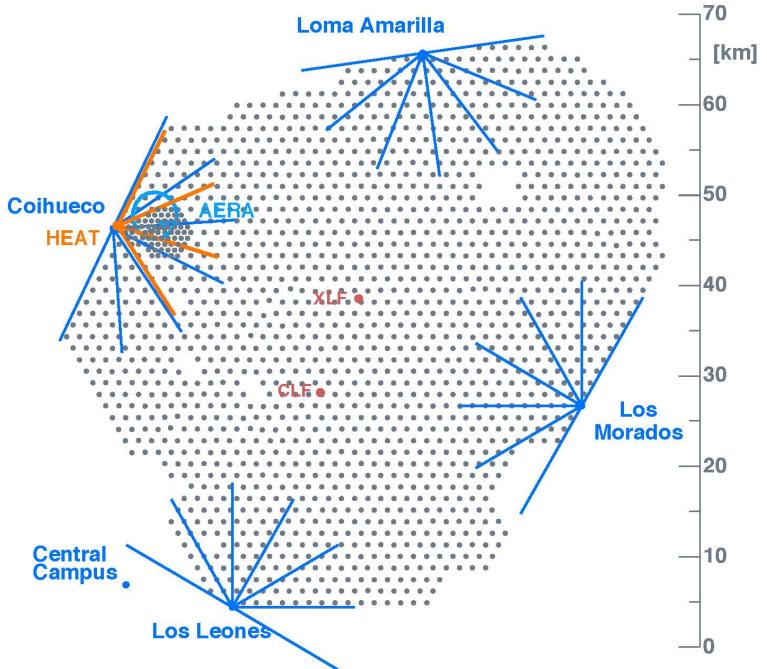


Figure 3.1: View of the Pierre Auger Observatory. The location of each SD station is indicated with a grey dot, while the 4 FD buildings (Los Leones, Coihueco, Loma Amarilla, Los Morados) are shown with blue dots, together with the field of view of each telescope (6 per building) delimited by the blue lines. Also the 3 *High Elevation Auger Telescopes* (HEAT) are shown (in orange). A denser array of WCDs (known as *infill*) is located in the HEAT FoV; in the same area also the *Auger Engineering Radio Array* (AERA) detectors have been set. The red dots indicate the *Central Laser Facility* (CLF) and the *eXtreme Laser Facility* (XLF), essential for atmospheric monitoring. Finally, the Central Campus identifies the position of Malargüe. Image from <https://pc.auger.unam.mx/sites/default/files/>.

lene with an average thickness of 1.3 cm, containing a sealed plastic liner with a reflective inner surface filled with ultra-pure water. Each station is self-powered by means of two solar panels, providing an average of 10 W for the PMTs and electronics (processor, GPS receiver, radio transmitter/receiver and power controller, as shown in Fig. 3.2).

Three photomultiplier tubes (PMTs) are placed in symmetrically distributed positions on the liner surface (as shown in the right panel of Fig. 3.2) and look downward collecting the Cherenkov photons produced by the passage of relativistic charged particles in water<sup>2</sup>. The PMTs have a 9 inch diameter photocathode, 8 dynodes, and are operated at a nominal gain of  $2 \times 10^5$ . The fast analog-to-digital conversion (FADC) electronics samples and digitizes the signal produced by each PMT in intervals of 25 ns (40 MH/s).

The Cherenkov light recorded by a tank is measured in units of the signal produced by a muon traversing the tank on a vertical trajectory, a unit called *vertical equivalent*

---

<sup>2</sup>As underlined in sect. 2.2.2, the energy release in a WCD is proportional to the length traversed inside water by each particle, and depends on the particle type and energy. For this reason, it is difficult to estimate the number of particles producing the observed signals, or their energy spectrum.

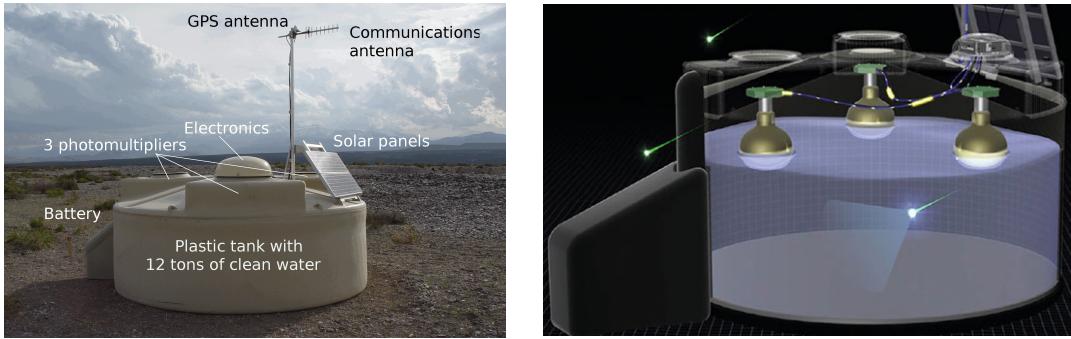


Figure 3.2: (Left) A water-Cherenkov detector in the field of the Pierre Auger Observatory; its main components are indicated. (Right) Representation of the inner part of a WCD, where the positions of the three PMTs, observing the water from above, are clearly visible. Image from <https://cerncourier.com/augerprime-looks-to-the-highest-energies>.

*muon* (VEM). The usage of VEM units provides a common reference between tanks, which is necessary considering that the calibration of each SD station is performed by the local electronics<sup>3</sup>. In short, the gain of each PMT is tuned to match a reference rate of background events (more details in [49, 45]), meaning that even the PMTs in a single tank can have different gains. The calibration is performed every 60 s but the (most recently determined) parameters are returned to the Central Data Acquisition System (CDAS) only when a shower is observed. In this way each measured event contains information about the state of each SD station in the minute preceding the trigger, allowing for a very accurate data calibration.

### 3.1.1 SD reconstruction

The reconstruction of the events measured by the surface detector is performed exploiting the integrated signals and the timing information from the WCDs hit by the shower particles. In particular two key characteristics of the detected EAS are obtained: the geometry, i.e. the shower core<sup>4</sup> and the arrival direction of the primary particle; the shower size, a quantity employed as energy estimator (see sect. 2.2.2). In this section only the reconstruction of events with zenith angle  $\theta < 60^\circ$  (vertical dataset) measured by the standard array (1.5 km spacing) is described.

A preliminary step is the selection of data actually corresponding to EAS events, that is performed by means of sequential triggers organized in a hierarchical structure. The SD acquisition, from local WCD thresholds to multi-station event triggers, result in the recording of 768 samples (i.e. time-dependent signals of  $19.2 \mu\text{s}$  duration) from each PMT of each involved WCD. Then physics triggers, based on the time correlation between adjacent detectors according with a simplified propagation of the shower front,

<sup>3</sup>This is required also because the total bandwidth available for communication with the Central Data Acquisition System (CDAS) is only 1200 bit/s. Furthermore, given the scale of the surface detector, the WCD calibration procedure has to be robust, allowing for failures of individual PMTs.

<sup>4</sup>Defined (see sect. 2.2.2) as the intersection between the shower axis and the ground surface, experimentally corresponds to the point with the higher density of secondary particles.

guarantee the selection of good quality data for physics analysis<sup>5</sup>.

The shower geometry is reconstructed by fitting the start times of the signals ( $t_i$ ) of each WCD. For events with enough triggered stations, a concentric-spherical model (which describes the shower front evolution as a sphere inflating at the speed of light, see Fig. 3.3a) can reproduce these times according with the equation:

$$c(t_i - t_0) = |\vec{x}_{sh} - \vec{x}_i| \quad (3.1)$$

where  $t_0$  and  $\vec{x}_{sh}$  identify a virtual origin (in space and time) of the shower development, while  $t_i$  and  $\vec{x}_i$  are the position and trigger time of each tank.

The next step is the evaluation of the shower core, accomplished with a fit of the lateral distribution function (LDF) of the signals measured by each station involved in the event. A modified Nishimura–Kamata–Greisen function is employed:

$$S(r) = S(r_{\text{opt}}) \left( \frac{r}{r_{\text{opt}}} \right)^\beta \left( \frac{r + r_1}{r_{\text{opt}} + r_1} \right)^{\beta + \gamma} \quad (3.2)$$

where  $r_1 = 700$  m,  $\beta$  and  $\gamma$  are slope parameters functions of the zenith angle<sup>6</sup>,  $r_{\text{opt}}$  is the optimal distance (introduced in sect. 2.2.2) and  $S(r_{\text{opt}})$  the corresponding signal, used as estimator of the shower size. For the surface detector of the Pierre Auger Observatory, the optimum distance was estimated [42] to be  $r_{\text{opt}} = 1000$  m.

This LDF fit also includes not-triggered stations and those, close to the shower axis, with saturated signals<sup>7</sup> as shown in Fig. 3.3b. The main outcome is the determination of the shower impact point at ground ( $\vec{x}_{gr}$ ) and of the shower size  $S(1000)$ .

At this point the shower axis  $\hat{a}$  can be calculated using the virtual shower origin  $\vec{x}_{sh}$  (from the geometrical reconstruction) and  $\vec{x}_{gr}$  (from the LDF reconstruction):

$$\hat{a} = \frac{\vec{x}_{sh} - \vec{x}_{gr}}{|\vec{x}_{sh} - \vec{x}_{gr}|} \quad (3.3)$$

The angular resolution [51] achieved with the whole procedure is lower than  $1.6^\circ$  for events with more than three stations, and improves reaching less than  $0.9^\circ$  for events with more than six stations (see Fig. 5.18a).

Finally, the value of  $S(1000)$  has to be converted in an energy estimator, taking into account the shower attenuation with the zenith angle and other geometrical effects.

---

<sup>5</sup>Furthermore, a fiducial cut (called 6T5 trigger) is applied to select only events where the station with the highest signal is surrounded by 6 neighbors all operating. This condition assures a very accurate reconstruction of the core position, and at the same time allows a simple geometrical calculation of the aperture/exposure, a quantity necessary for spectrum studies. For arrival-direction analysis less strict cuts can be used.

<sup>6</sup>The parameters  $\beta$  and  $\gamma$  are introduced to have a more flexible LDF fit, which takes into account the different contributions from the muonic and the electromagnetic components at ground. In fact, these two contributions change with the distance from the shower core and with the shower age, that depends on the zenith angle (less inclined showers are measured at an earlier stage at ground level).

<sup>7</sup>The saturation of the time-dependent signal verifies when a large number of particles reach the detector in a short amount of time, causing an overflow of the FADC read-out electronics and the transition of the PMTs response to a non-linear behavior. However, in the majority of cases, the missing part of the signals can be recovered using the procedure described in [50].

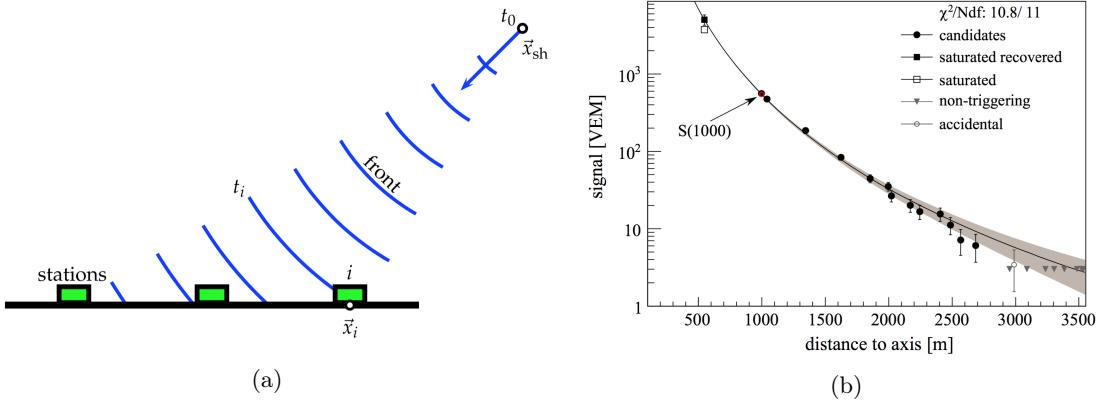


Figure 3.3: (a) Schematic representation of the shower geometry. (b) Example of an LDF fit, where the estimated value of the shower size  $S(1000)$  is indicated. Images from [45].

Such conversion is performed using the Constant Intensity Cut method (introduced in sect. 2.2.2). The attenuation curve  $f_{\text{CIC}}(\theta)$  (shown in Fig. 2.9b) employed in the standard SD reconstruction is a third degree polynomial in  $x = \cos^2 \theta - \cos^2 \bar{\theta}$ :

$$f_{\text{CIC}}(\theta) = 1 + ax + bx^2 + cx^3 \quad (3.4)$$

where  $a = 0.96$ ,  $b = -1.66$ ,  $c = -1.19$  (from the ICRC-2017 analysis) and  $\bar{\theta} = 38^\circ$  (median of the zenith distribution for the events measured by the surface detector) is chosen as reference angle. The energy estimator is

$$S_{38} \equiv \frac{S(1000)}{f_{\text{CIC}}(\theta)} \quad (3.5)$$

which can be considered as the signal that a shower with a certain size  $S(1000)$  would have produced arriving with an inclination of  $38^\circ$  instead of  $\theta$ .

The final energy value for each vertical ( $\theta < 60^\circ$ ) EAS recorded with the SD-1500<sup>8</sup> is obtained from a calibration with the (independent) results of the FD reconstruction, following the procedure exposed in sect. 3.2.2.

## 3.2 The fluorescence detector

The fluorescence detector (FD) of the Pierre Auger Observatory is formed by four sites located on small hills at the boundaries of the SD array (see Fig. 3.1): Los Leones (south), Los Morados (east), Loma Amarilla (north) and Coihueco (west). Each building hosts six independent telescopes with field of view of approximately  $30^\circ \times 30^\circ$  in azimuth and elevation, providing a  $180^\circ$  coverage in azimuth. The FD observes the atmosphere above the array and it was designed to ensure that every event above  $10^{19}$  eV arriving within the FD on-time is recorded by at least one fluorescence telescope [45].

<sup>8</sup>Surface detector with 1.5 km spacing.

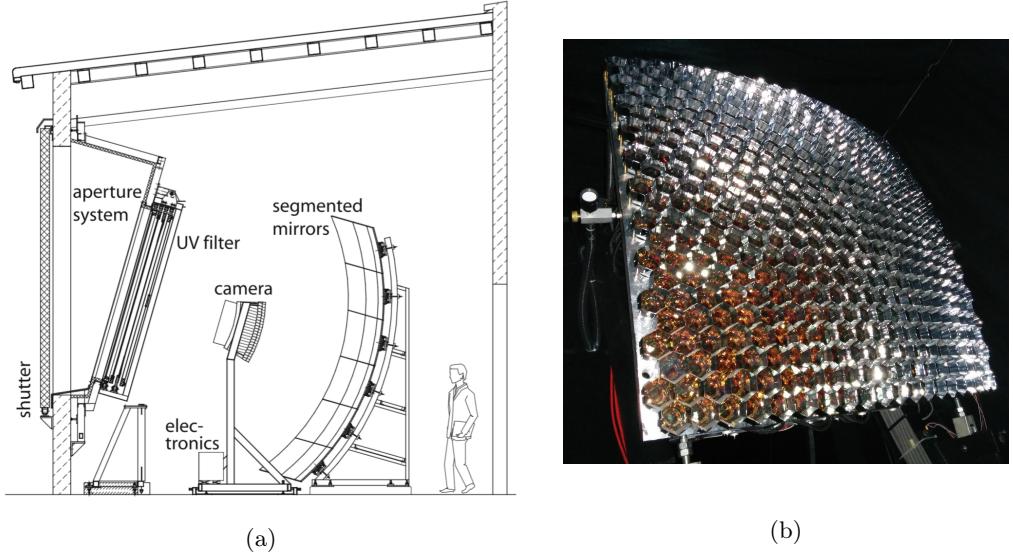


Figure 3.4: (a) Schematic view of a fluorescence telescope, where the main components are indicated. Image from [45]. (b) Photo of the PMT matrix from a telescope at the Coihueco site.

The measure of air showers via fluorescence light is possible only in clear (low cloud coverage) and moonless (low amount of direct or scattered moonlight) nights<sup>9</sup>. In fact, the night sky brightness is the main source of background for EAS events. The mean length of the dark observation period is 17 nights per month, which reduces to an average FD on-time of  $\sim 15\%$  because of adverse weather conditions.

The optical system in each FD telescope (Fig. 3.4a) is composed by: a filter at the entrance window, a corrector ring at the circular aperture, a mirror and a camera of photomultipliers. The filter is needed to reduce the background light flux<sup>10</sup>, improving the signal-to-noise ratio for the measurement of EAS, and as a window over the telescope aperture, to keep the inner environment clean and climate-controlled. The corrector ring allows to almost double the aperture area of the telescope with respect to the optical system without any correcting element. The mirror is a spherical surface of  $\sim 3400$  mm radius of curvature<sup>11</sup> focusing the light onto a surface with radius  $\sim 1700$  mm occupied by the camera (Fig. 3.4b), a matrix of 440 hexagonal photomultiplier tubes (PMTs) arranged in 22 rows and 20 columns. More detailed information about the optical system can be found in [52, 45].

The camera is the sensitive element of a telescope, observing the sky as collection of pixels (see sect. 2.2.1) each one corresponding to the field of view of one PMT. The selection of candidate events is performed by a three-stage trigger system. The first level

---

<sup>9</sup>In particular, the sun must be lower than  $18^\circ$  below the horizon and the illuminated fraction of the moon must be less than 70%.

<sup>10</sup>The filter transmission in the UV range (of interest for the fluoresce detector, see sect. 2.2.1) is above 80% in the range [330, 380] nm (50% in [310, 390] nm).

<sup>11</sup>Due to the large surface ( $\sim 13 \text{ m}^2$ ), each mirror is segmented to reduce the costs. The average reflectivity of a clean mirror tile is above 90% at a wavelength  $\lambda = 370$  nm.

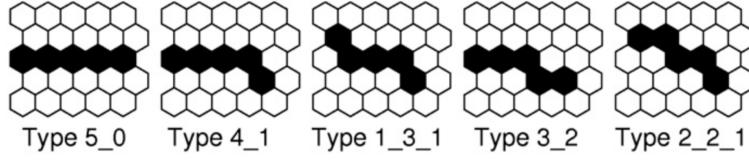


Figure 3.5: Fundamental patterns for the second level FD trigger. Image from [52].

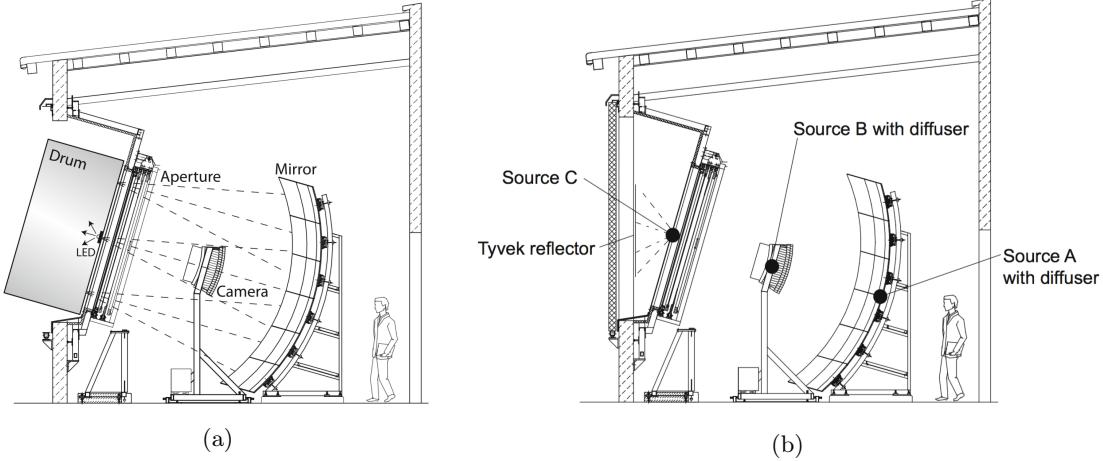


Figure 3.6: Schematic view of a fluorescence telescope, showing (a) the absolute calibration through the *drum* light source, (b) the position of the 3 permanent sources for the relative calibration. Images from [52].

selects the PMT signals over a certain threshold value, that is adjusted to keep the the trigger rate of each pixel in a camera around 100 Hz. At the second level, the trigger searches for track segments (within a camera) long at least 4/5 pixels (see Fig. 3.5), resulting in a typical rate per camera between 0.1 and 10 Hz. The third level trigger is a more complicate software algorithm, designed to reject noise events (for instance produced by lightning or by cosmic ray muons impacting on the camera) which appears to be uncorrelated in time and/or space.

After the selection of good quality data, the reconstruction of the shower longitudinal profiles requires a conversion from the signal measured by each pixel to a light flux relative to the portion of the cascade observed by the pixel itself. Therefore an end-to-end calibration of the detector response is essential. The so-called *absolute* calibration is accomplished using a drum-shaped light source of 2.5 m diameter that is periodically put at the telescopes aperture, releasing a uniform light flux over the camera as illustrated in Fig. 3.6a. Knowing accurately this flux and the response of the acquisition system, the calibration parameters for each pixel are obtained. Furthermore, a *relative* calibration is performed before and after each night of operation, to monitor both short and long term changes (for instance between two absolute calibrations) in the telescopes. Three permanent light sources are installed in different positions of each telescope (Fig. 3.6b) to monitor different groups of elements of the optical system.

Finally, during each night of FD measurements, thousands of collimated UV laser pulses are shot into the atmosphere from two facilities (the CLF and the XLF, indicated in Fig. 3.1) located near the center of the SD array. Such laser pulses are recorded by the fluorescence telescopes in the same way of tracks generated by air showers, but their characteristics are well known, thus allowing a test of the FD performances and a continuous monitoring of the atmosphere.

### 3.2.1 FD reconstruction

The FD reconstruction of an EAS candidate (Fig. 3.7) counts two steps: a *geometrical* reconstruction, from which distance and direction of propagation of the shower are obtained; a reconstruction of the longitudinal development, to evaluate the atmospheric depth of the maximum ( $X_{\max}$ ) and the shower energy.

The geometrical reconstruction begins with the determination of the *Shower-Detector Plane* (SDP), that is the plane containing the fluorescence telescope and the observed EAS trace. The SDP is established using [45] the pointing direction of each pixel involved in the event, weighted with the respective integrated signals.

With this information, the shower axis can be determined through the time sequence of the triggered pixels. In fact each pixel can be associated with an angle  $\chi_i$  on the SDP, as illustrated in Fig. 3.8, and the arrival time of the shower photons  $t(\chi_i)$  should follow the equation:

$$t(\chi_i) = t_0 + \frac{R_p}{c} \tan\left(\frac{\chi_0 - \chi_i}{2}\right) \quad (3.6)$$

where  $R_p$  is the minimum distance between the shower axis and the detector,  $t_0$  the time when the shower front reaches the point at this minimum distance, and  $\chi_0$  the angle between the shower axis and the intersection of the ground surface with the SPD.

These three quantities are evaluated minimizing a likelihood which is a function of the pixels trigger times [45] based on eq. 3.6. However, the fit of an event detected by a single fluorescence detector (*monocular* reconstruction) may present a degeneracy between  $R_p$

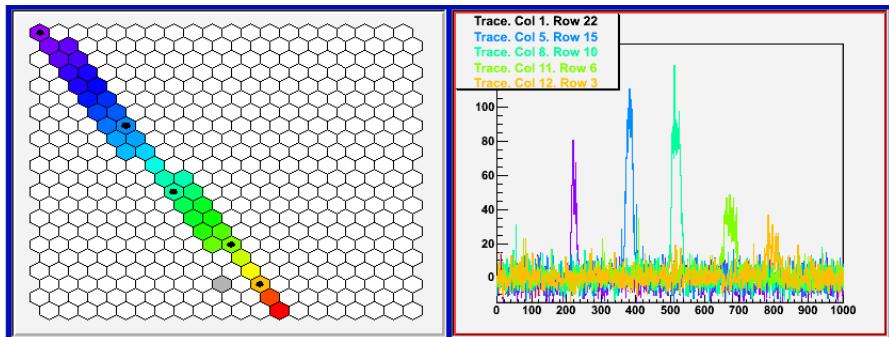


Figure 3.7: Example of an event recorded by a fluorescence telescope. On the left, the track of triggered pixels where the different colours suggest the time sequence. On the right, the histograms of PMT signals (corresponding to the pixels with black dots of the left panel) in steps of 100 ns. Image from <http://people.roma2.infn.it/auger/cosmic3.html>.

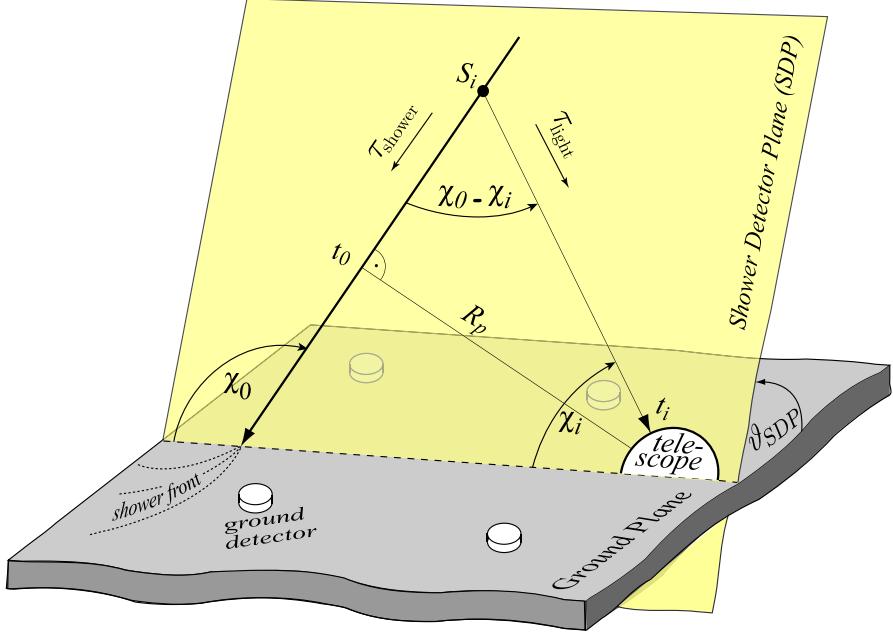


Figure 3.8: Geometrical reconstruction of a shower with a fluorescence telescope. Image from [52].

and  $\chi_0$ , leading to an inaccurate determination of the shower parameters. To break this degeneracy, the timing information provided from at least one SD station is exploited in combination with the fluorescence measurement, performing a *hybrid* reconstruction.

Once the geometry has been established, the reconstruction of the shower development can be accomplished. In particular, the signals recorded by the PMTs are converted into an energy deposit as a function of atmospheric slant depth ( $X$ ), taking into account the fluorescence yield (defined in sect. 2.2.1), the calibrated response of each pixel and the light attenuation from the shower to the detector<sup>12</sup>. Furthermore, the contributions of direct and scattered Cherenkov light and of multiply scattered light (left panel of Fig. 3.9) must be subtracted.

The final longitudinal energy deposit profile is fitted with a Gaisser-Hillas function:

$$f_{\text{GH}} = \left( \frac{dE}{dX} \right)_{\max} \left( \frac{X - X_0}{X_{\max} - X_0} \right)^{(X_{\max} - X_0)/\lambda} \exp \left( \frac{X_{\max} - X_0}{\lambda} \right) \quad (3.7)$$

where  $X_0$  and  $\lambda$  are two shape parameters and  $(dE/dX)_{\max}$  is the energy deposit at depth  $X = X_{\max}$ . An illustrative result of this fitting procedure, leading to the  $X_{\max}$  determination, is shown in the right panel of Fig. 3.9.

The calorimetric energy, i.e. the energy deposited in atmosphere by the electromagnetic particles of the shower, is obtained by integrating eq. 3.7. Finally the total energy, that corresponds to the primary cosmic ray energy, is calculated by correcting for the

---

<sup>12</sup>A good knowledge of temperature, pressure, humidity and air density, together with the aerosol profiles is therefore essential to reconstruct the shower longitudinal profile. For this reason, in the framework of the Pierre Auger Observatory, a continuous monitoring of the atmosphere is accomplished through dedicated facilities and procedures.

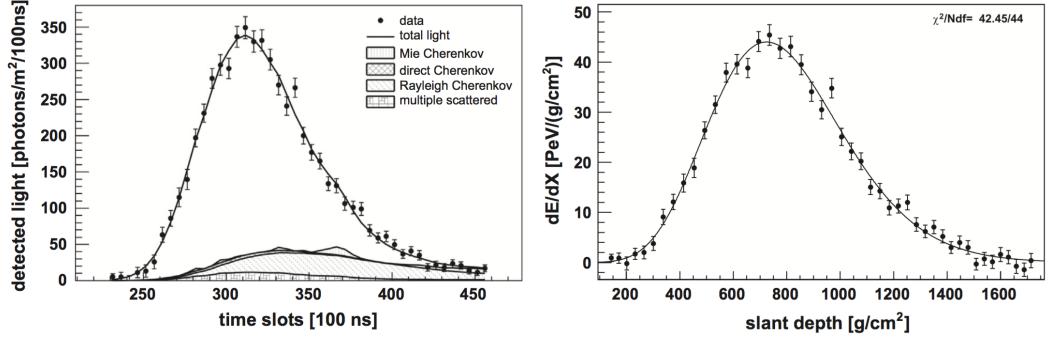


Figure 3.9: Example for the procedure of shower profile reconstruction. On the left, the light collected at the aperture as a function of time, where the different light contributions are shown. On the right, the energy deposit as a function of the atmospheric depth; the line is the resulting Gaisser-Hillas fitting function. Images from [45].

“invisible energy”, carried away by neutrinos and high energy muons, estimated to be 10-15% in the energy range relevant for the FD measurements.

### 3.2.2 SD energy calibration

As explained in sect. 2.2.2, the estimation of the primary energy for events recorded by an array of ground detectors requires the comparison with an independent measurement or with simulations. In fact the energy scale can be determined only through a **calibration** of the energy estimator, characteristic of each experimental setup.

The advantage of the Pierre Auger Observatory is the possibility of *hybrid detection*, i.e. the observation of a shower independently by both the SD and the FD. In the particular case of an event separately reconstructed with both the procedures exposed in sect. 3.1.1 and 3.2.1, the SD energy estimator  $S_{38}$  can be therefore compared with the almost-calorimetric  $E_{FD}$  value from the FD reconstruction.

The SD energy calibration is performed using a sub-set of the hybrid events, selected according with several quality cuts, called *golden* hybrid dataset. As already mentioned, both the SD and FD reconstructions must be successful but also accurate, i.e. must present low uncertainties in the reconstructed parameters. Moreover, only events observed under clear atmospheric conditions<sup>13</sup> are accepted, in order to assure an accurate reconstruction of the shower longitudinal profile. In addition, a fiducial cut on the slant depth range observed by the telescopes is applied<sup>14</sup> to remove any potential bias depending on the mass of the primary particles.

In Fig. 3.10 the correlation between  $S_{38}$  and  $E_{FD}$  for the set of golden hybrid events is shown in grey (the other scatter plots are the results for energy estimators of different datasets). The final step in the calibration is then a fit of this relation, which is apparently

<sup>13</sup>For instance, the maximum values accepted for cloud coverage and the *vertical aerosol optical depth* are respectively 25% and 0.1 g/cm<sup>2</sup>. Events without atmospheric information are immediately rejected.

<sup>14</sup>This selection criteria, described in [53], is the same introduced in the mass composition analysis (sect. 3.5.2) to obtain an unbiased estimation of the  $X_{\max}$  moments as functions of the energy.

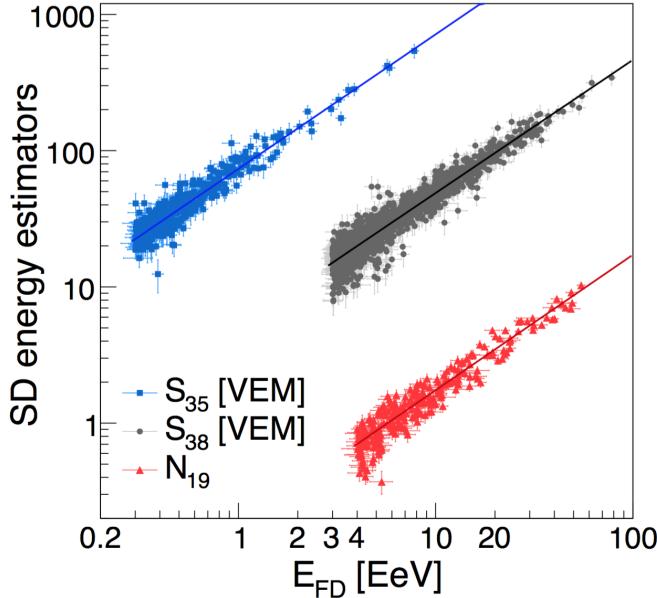


Figure 3.10: Correlation between SD energy estimators and the FD energy values. Each point represents a single event.  $S_{38}$  is the estimator for the SD-1500 vertical dataset, while  $S_{35}$  is a similar estimator employed for the events measured by the *infill* array (750 m spacing). Instead, the  $N_{19}$  estimator is used for the calibration of inclined events ( $\theta \in [60^\circ, 80^\circ]$ ) measured by the SD-1500 array, according with the procedure described in [56]. The lines show the results of the calibration procedures. Images from [57].

well described by a power-law function:

$$E_{FD} = A \left( \frac{S_{38}}{VEM} \right)^B \quad (3.8)$$

More details about the fitting procedure can be found in [54, 55].

The most recent official calibration of the SD-1500 vertical ( $\theta < 60^\circ$ ) dataset has been accomplished for the 35th *International Cosmic Ray Conference* (ICRC-2017). The resulting values [58] are  $A = (1.78 \pm 0.03) \times 10^{17}$  eV and  $B = 1.042 \pm 0.005$ , obtained from a set of 2661 golden hybrid events with  $E_{FD} > 3 \times 10^{18}$  eV recorded in the period from January 2004 to December 2016.

The systematic uncertainty on the SD energy due to the calibration is a few % over the whole range. Instead SD energy resolution is  $\sim 15\%$  and slowly decreasing with energy. It is calculated from the distribution of the ratio  $E_{SD}/E_{FD} = A(S_{38}/VEM)^B/E_{FD}$  for the events used in the calibration, assuming a fixed FD energy resolution of 7.6%. Finally, the uncertainty in the energy scale is  $\sim 14\%$  and approximately energy-independent. It is dominated by the systematics in the absolute FD calibration, as explained in [59].

### 3.3 The Observatory enhancements

After the completion and successful operation of the two main detectors, several enhancements were proposed to extend the energy range of measurements in its lower part,

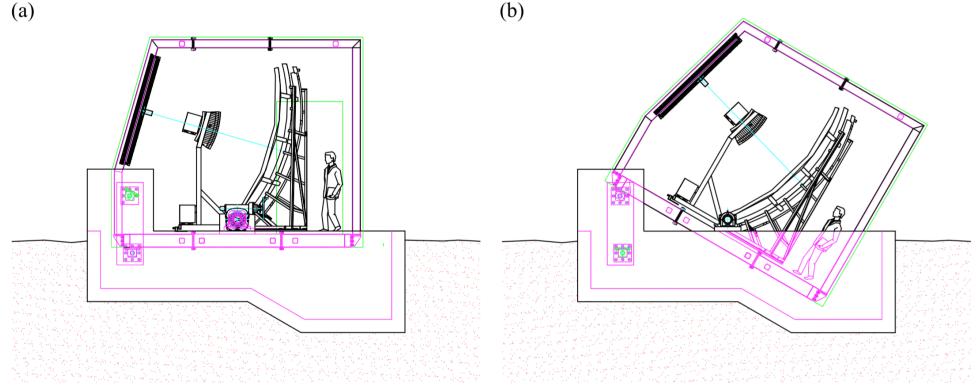


Figure 3.11: Schematic view of one HEAT telescope. (a) Horizontal (downward) mode for service and cross-calibration. (b) Tilted (upward) mode for data taking. Images from [45].

particularly in the region between the knee and the ankle, and to provide different information, that can complement the SD and FD data, about extensive air showers.

### 3.3.1 HEAT

Three additional fluorescence telescopes [60, 45] have been realized near the Coihueco site with the possibility of a *tilted* mode, i.e. they can be inclined by  $29^\circ$  upward to cover an elevation range from  $30^\circ$  to  $58^\circ$ . Considering this characteristic, they were collectively named *High Elevation Auger Telescopes* (HEAT).

The main target of this enhancement was to lower the energy threshold for the shower detection. In fact, in the upward tilted mode, HEAT can observe events down to  $10^{17}$  eV, that develop higher in the atmosphere. An accurate hybrid reconstruction can be accomplished in combination with the *infill*, a denser array with 750 m spacing nested within the standard SD one in an area inside the field of view of the HEAT telescopes, centered 6 km away from the Coihueco site<sup>15</sup>.

In the downward mode, the HEAT field of view overlaps with the Coihueco's one, allowing for useful comparison and cross-calibration.

The HEAT telescopes present almost the same design of the FD ones, but they are placed in three separate blocks which can be moved in the two different operative positions (Fig. 3.11) by an electrically driven hydraulic system.

### 3.3.2 AMIGA

The *Auger Muon and Infilled Ground Array* (AMIGA) [61, 62] is an upgrade of the Observatory designed to directly measure the muon content of air showers in the energy region between  $10^{17}$  and  $10^{19}$  eV. It consists of an array (currently under construction) of 61 detector pairs, each one composed by a WCD of the *infill* array (750 m spacing) and

---

<sup>15</sup>The two apparatuses (HEAT and the *infill* array) have to be placed at a short distance to perform hybrid measurement in the required lower energy range. In fact, a shower will emit fewer fluorescence photons for lower primary energy and thus can be detected only by a near fluorescence telescope.

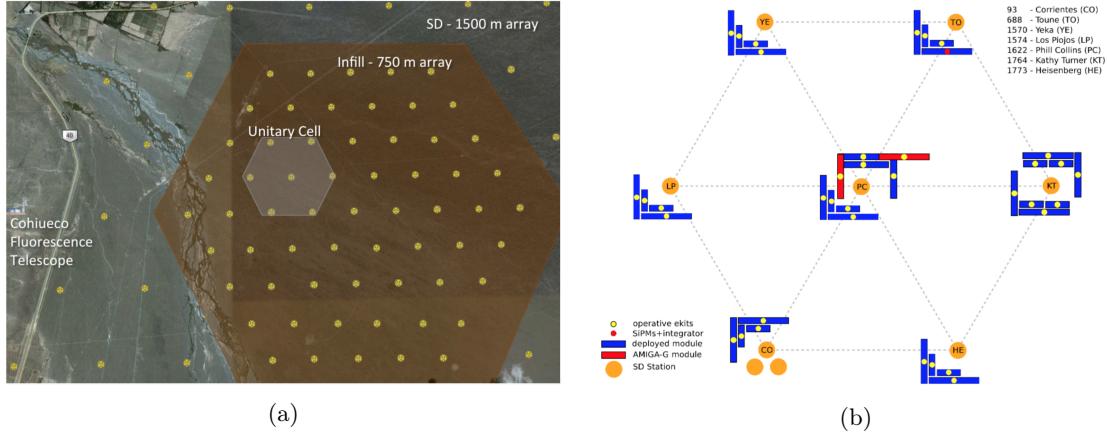


Figure 3.12: (a) Map of the *infill* array, where the area for the complete AMIGA detector is shown with brown background. The small grey shaded area indicates the engineering array hexagon (Unitary Cell). Image from [61]. (b) Layout of the AMIGA Unitary Cell. The scintillators in blue are buried under 2.3 m of soil, while the red ones under 1.3 m. Image from the *AugerPrime* Status Report (01/04/2016).

by a scintillator buried underground. To be effectively shielded by the electromagnetic particles, the scintillators are placed under 2.3 m of soil, approximately correspondent to a vertical depth of  $540 \text{ g/cm}^2$  ( $\sim 20$  radiation lengths).

The AMIGA engineering array, referred to as the *Unitary Cell* (Fig. 3.12b), has been fully operational since March 2015, serving as test bench to identify and solve engineering or acquisition issues as well as to study the performances of the whole detection system. It is composed by seven detector pairs placed at every vertex of a hexagon of the *infill* array, plus its centre. Each WCD is paired with a  $30 \text{ m}^2$  scintillator (also indicated as *muon counter* or MC) segmented in two modules of  $10 \text{ m}^2$  and two of  $5 \text{ m}^2$ . In addition, two positions are equipped with twin MC and one position presents two extra  $10 \text{ m}^2$  scintillators buried at a shallower depth (1.3 m) to study the soil shielding features. The MCs acquisition is triggered by the overlying WCD.

An underground muon detector will provide valuable information to explain the excess of muon observed in data with respect to simulations [63]. At the same time, the determination at event level of a new quantity related to the shower muon content [64] can be used to infer the mass of the primary cosmic rays, especially in combination with the  $X_{\max}$  measurements. For this reason the completion of AMIGA, improved and upgraded [65] with respect to the Unitary Cell, is part of *AugerPrime*, the upgrade of the Pierre Auger Observatory (sect. 3.4).

### 3.3.3 AERA

The *Auger Engineering Radio Array* (AERA) [66, 67] comprises 153 stations spread over an area of  $\sim 17 \text{ km}^2$  located inside the *infill* array. As engineering array, it combines different hardware (24 logarithmic periodic dipole antennas, the rest butterfly antennas),

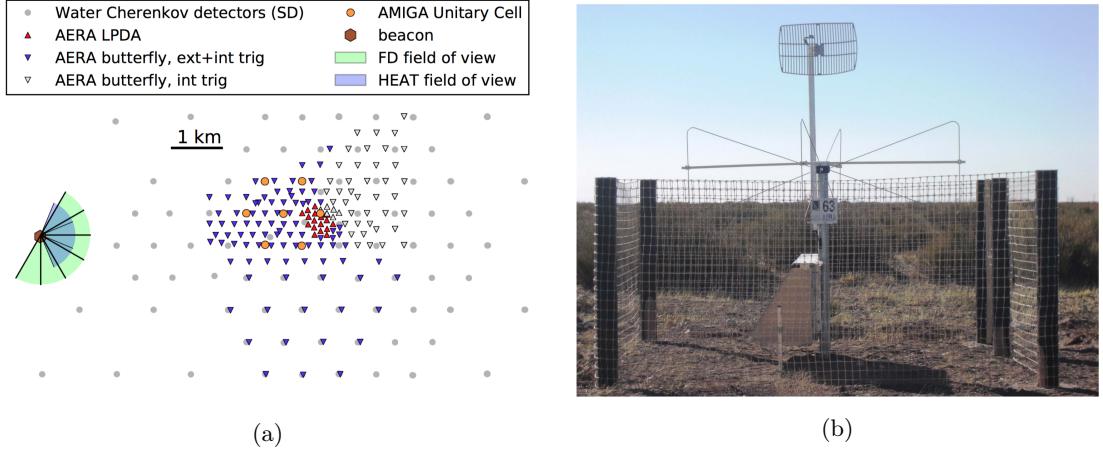


Figure 3.13: (a) Map of the AERA array together with other co-located detectors. Image from [68]. (b) Radio station of AERA: a communication antenna for wireless data transfer, two butterfly antennas for the radio measurements (one aligned east-west and one north-south), a metal box for electronics, a solar panel and a battery for power supply, and a fence protecting against cattle. Image from [67].

electronics and communication systems, spacing between the antennas (dense center of 144 m spacing, then 250 m, 375 m, 750 m grids) and triggers. Each radio station is composed by a two-arms antenna aligned in east-west and north-south directions, sensitive to radiation in the frequency range of 30-80 MHz.

AERA measures the radio emission produced by the electromagnetic part of the shower, with the advantages of a nearly 100% duty cycle and of a negligible attenuation, since the atmosphere is transparent for radio waves at the mentioned frequencies. The background (excluding anthropogenic radio signals) is due to high atmospheric electric fields, occurring for instance in massive rain clouds and during thunderstorms.

The radio emission is mainly produced at a depth around  $X_{\text{max}}$  and emitted in a forward-directed cone around the direction of shower propagation. Since the shape of the radio signals at ground depends on the distance from the emission region, the position of the (electromagnetic) shower maximum can be inferred [68]. Instead, the arrival direction and energy of the primary cosmic ray are reconstructed using the timing information and the radiation density at the individual radio stations. The AERA enhancement already accomplished interesting results, as the development of a new method to measure the total amount of energy transferred from the primary cosmic ray into radio emission [69, 70] and the observation of inclined EeV air showers [71].

### 3.4 The upgrade of the Observatory: *AugerPrime*

To answer many questions still open in the UHECR field, the Observatory has started a major upgrade, called *AugerPrime* [72, 73]. The main goal is to improve the mass composition sensitivity of the surface detector to explore energies above  $10^{19}$  eV, where the fluorescence detector is not adequate due to its limited duty cycle ( $\sim 15\%$ ).

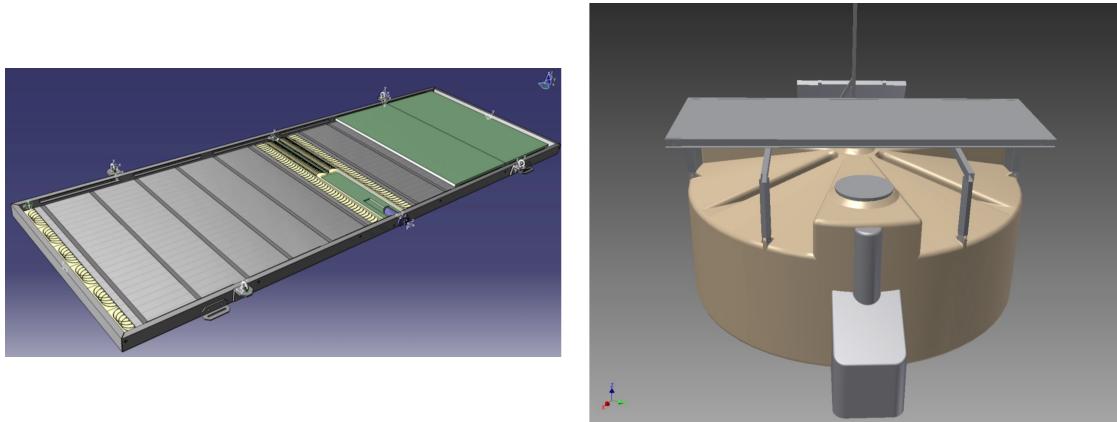


Figure 3.14: (Left) Rendering of an SSD module. Image from [73]. (Right) 3D view of a WCD with a SSD unit on top. Image from [72].

The key element of the upgrade is a new plastic scintillator, to be installed on top of each WCD, called Surface Scintillator Detector (SSD) [74]. It consists of two modules, housed in a box of area  $3.8 \text{ m} \times 1.3 \text{ m}$ , each covering an area of  $1.9 \text{ m}^2$  and triggered by the larger WCD below it (Fig. 3.14).

A thin scintillation detector and a water-Cherenkov tank present very different responses to the shower particles. In particular, the comparison between simultaneous measurements could allow to disentangle the muonic and electromagnetic components of the signal, delivering a mass-correlated information on an event-by-event basis. Therefore the possibility of new studies of the mass composition at the highest energies are foreseen. At the same time, the selection of a sub-sample of events enriched in light elements will become more reliable, enhancing anisotropy studies and so the capability to identify potential sources of UHECRs.

Furthermore, the upgrade of the SD stations is integrated with the installation of a faster SD electronics [75] and of an additional *small photomultiplier* (SPMT) in each WCD. The SPMT is designed to extend the dynamic range [76] from about 600 VEM to more than 30,000 VEM, fixing one drawback of the current SD acquisition: the saturation of signals (see sect. 3.1.1) near the shower core. In effect, more than 40% of the events at energies above  $3 \times 10^{19} \text{ eV}$  present at least one saturated station. The new configuration (WCD with SPMT + upgraded electronics)<sup>16</sup> is expected to deliver less than 2% of saturated events even at the highest energies.

An engineering array [77] of twelve upgraded stations has been operational since 2016, monitoring the performances of new instruments and electronics. The production and deployment of the *AugerPrime* detectors is under-way (thirty more upgraded stations have been deployed as of October 2018) and will be completed within 2019, so that full data taking is planned between 2020 and 2025.

---

<sup>16</sup>Also, for consistency with the associated WCDs, the SSDs are designed to have a dynamic range spanning from single particle signals, needed for calibration, to large signals up to  $\sim 2 \times 10^4$  *minimum ionizing particles* (MIP), the unit for scintillator detectors.

The Observatory upgrade also includes, among its targets, the finalization of the AMIGA array (sect. 3.3.2). In fact, an underground Muon Detector (MD) is necessary to provide direct measurements of the shower muon content and to cross-check and fine-tune the methods that will be implemented to derive the muon information from the upgraded SD detectors (WCD+SSD stations).

Finally, an extension of the FD operations is planned within the AugerPrime program, to be accomplished collecting data when a large fraction of the moon ( $> 70\%$ , see sect. 3.2) is illuminated. The larger night sky background requires a reduction of the PMT gain to avoid excessively high anode currents and the consequent, irreversible deterioration of the PMTs. However, an upgrade of the FD telescopes is not required since the current setup already allows to lower the high voltage supplied to the camera. Preliminary tests showed that the required performances can be satisfied and suggest the feasibility of the described procedure.

## 3.5 Selected results

A review of the latest relevant results from the Pierre Auger Observatory is presented in the following, with a focus on the energy spectrum measurements, the mass composition studies and the arrival directions analyses.

### 3.5.1 Energy Spectrum

The measurement of the UHECRs flux has been one of the first outcomes of the experiment. With more than ten years of data, two spectral features are confirmed beyond doubt [78, 79, 56]: the hardening of the spectrum at about  $5 \times 10^{18}$  eV (the ankle) and a strong suppression at the highest energies, starting around  $4 \times 10^{19}$  eV. The all-particle flux has been updated for the ICRC in 2017 [58], presenting now an exposure exceeding 67,000 km<sup>2</sup> sr yr accumulated since January 2004 until December 2016.

Four independent energy spectra (see Fig. 3.15, left) are obtained from the different dataset collected by the Pierre Auger Observatory instruments. Two separated analysis are performed for vertical (zenith angle  $\theta < 60^\circ$ ) and horizontal ( $\theta \in [60^\circ, 80^\circ]$ ) SD-1500 events respectively. Data from the *infill* (SD-750) allows for the determination of the energy spectrum down to  $10^{17}$  eV, while hybrid events bridge the results from the two arrays, between  $10^{18}$  and  $10^{19.6}$  eV.

The four spectra, in agreement within uncertainties, are combined into a global one (Fig. 3.15, right) and the features of this final energy spectrum are determined fitting it with a model that presents two power-laws around the ankle and a smooth suppression at the highest energies. The ankle is found to be located at  $E_{\text{ankle}} = (5.08 \pm 0.06 \pm 0.8) \times 10^{18}$  eV. The spectral index below the ankle is  $\gamma_1 = 3.293 \pm 0.002 \pm 0.05$ , while above the ankle is  $\gamma_2 = 2.53 \pm 0.02 \pm 0.1$ . The cut-off is characterized using the value of  $E_{1/2}$ , that corresponds to energy at which the integral spectrum drops by a factor of

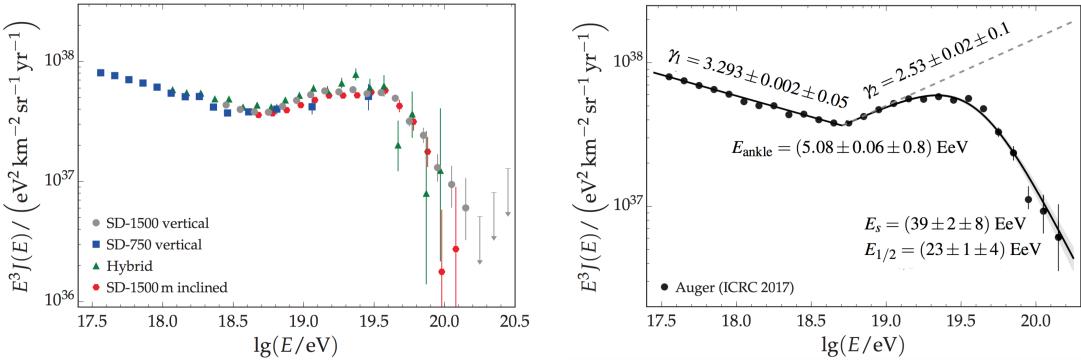


Figure 3.15: Left: The four energy spectra derived from SD and hybrid data. The systematic uncertainty on the energy scale is 14%. Right: The combined energy spectrum and the related fitting parameters (see text). In both panels only statistical uncertainties are shown. Images from [58].

two below what would be the expectation without steepening. This parameter is found to be  $E_{1/2} = (22.6 \pm 0.8 \pm 4.0) \times 10^{18} \text{ eV}$ .

### 3.5.2 Mass composition studies

The mass composition studies are mainly addressed through the  $X_{\max}$  values, as the depth of the shower maximum remains the most robust mass-sensitive EAS observable. In particular, the observed  $X_{\max}$  distributions in small energy intervals are compared with expectations from high energy hadronic interaction models.

The dataset is composed by a selected set of hybrid events above  $10^{17.8} \text{ eV}$ , collected by the standard FD telescopes during stable runs and in good atmospheric conditions. In addition, several quality selection criteria and fiducial FoV cuts [53] are applied to minimize detector-dependent effects in the sampled  $X_{\max}$  distributions. Recently [80] the measurement has been extended down to  $10^{17.2} \text{ eV}$  using data collected by HEAT in coincidence with the Coihueco telescopes: for this reason, the additional dataset at lower energies is called *HeCo*. In the following, the updated analysis [81] presented at the 35th International Cosmic Ray Conference (ICRC-2017) is reported.

The total number of events that passed the selection is 16778 and 25688 for HeCo and Standard-FD respectively. The consistency between the two data sets allows to extend the inferences about the evolution of the average mass composition down to  $\sim 10^{17} \text{ eV}$ , much lower than previous analysis [53]. The  $\langle X_{\max} \rangle$  and  $\sigma(X_{\max})$  obtained from data are shown in Fig. 3.16 as a function of energy. The results can be compared with the behaviours from simulations of proton and iron-initiated showers, performed using three hadronic interaction models tuned according to recent LHC data.

Between  $10^{17.2}$  and  $10^{18.3} \text{ eV}$ ,  $\langle X_{\max} \rangle$  increases by around  $(79 \pm 1) \text{ g/cm}^2$  per decade of energy, while above  $10^{18.3} \text{ eV}$  the elongation rate is significantly smaller ( $(26 \pm 2) \text{ g/cm}^2/\text{decade}$ ). As the first value is larger than what is expected for a constant mass composition ( $\sim 60 \text{ g cm}^{-2}/\text{decade}$ ), it indicates that the mean primary mass is

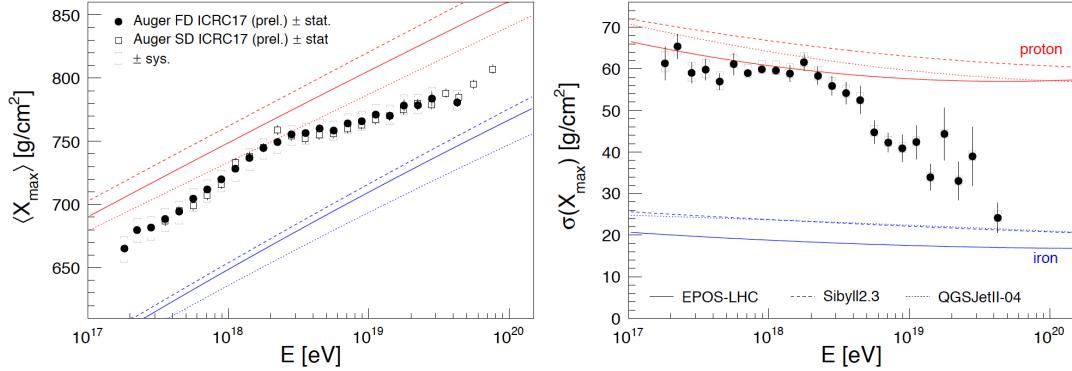


Figure 3.16: The mean (left) and the standard deviation (right) of the measured  $X_{\max}$  distributions as a function of energy compared to air-shower simulations for proton and iron primaries. The result tagged as “FD” (full dots) is obtained using the hybrid dataset [81], while the result tagged as “SD” (empty squares) is obtained through the *risetime* method [47].

getting lighter all the way from  $10^{17}$  to  $10^{18.3}$  eV. Then the trend inverts and the composition becomes heavier. The fluctuations of  $X_{\max}$  start to decrease at around the same energy ( $\sim 10^{18.3}$  eV), being rather constant below.

A new analysis of the evolution of  $\langle X_{\max} \rangle$  with energy has been recently performed [47] using a novel method which exploits the time profiles of the signals from the water-Cherenkov detectors. In [47], a single parameter called *risetime*, namely the time for the signal to increase from 10% to 50% of the final integrated magnitude, is used to characterize the measurement at each station. The depth of shower maximum is then inferred by calibrating this new parameter with measurements from the fluorescence telescopes. For the SD-1500 array, there are 54022 events with energies above  $\sim 0.3$  EeV for which estimates of  $X_{\max}$  have been possible. The values of  $\langle X_{\max} \rangle$  found from this analysis are shown as a function of energy in Fig. 3.16 (left panel, empty squares), where the agreement with hybrid measurements in the overlapping region is more than satisfactory. The result from the surface detector alone actually extends the knowledge of  $\langle X_{\max} \rangle$  to higher energies, but a reduction of statistical and systematic uncertainties is required before further conclusions could be drawn.

A more accurate interpretation of the measurements in terms of mass composition requires a conversion of  $\langle X_{\max} \rangle$  and mostly of  $\sigma(X_{\max})$  into the moments of the distribution of  $\ln A$  (the logarithm of the mass for primary CRs at the Earth’s atmosphere) as explained in sect. 2.3.1. In particular, the second  $X_{\max}$  moment includes the contributions from the intrinsic shower-to-shower fluctuations, smaller for heavier nuclei, and from the mixing of different primary species that reach the Earth.

The mean value  $\langle \ln A \rangle$  and the variance  $\sigma(\ln A)$  are determined from eqs. 2.22 and 2.23 using the EPOS-LHC, QGS-Jet II.04 and Sibyll 2.3 hadronic interaction models. The results are show in Fig. 3.17, where similar trends with energy are observed in the three cases. As already inferred from the mean  $X_{\max}$  values, the average mass of the

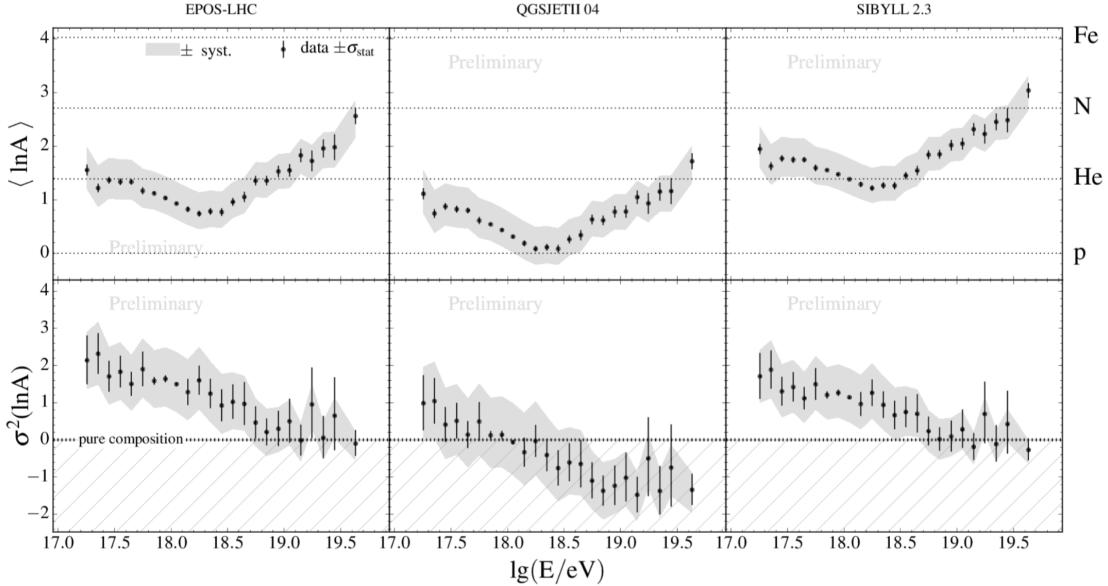


Figure 3.17: The mean (top) and the variance (bottom) of the  $\ln A$  distribution as function of the energy, estimated from data with EPOS-LHC (left), QGS-Jet II.04 (middle) and Sibyll 2.3 (right) hadronic interaction models. Image from [81].

primaries decreases with energy until a minimum value, around  $10^{18.3}$  eV, is reached. Then it becomes increasingly heavier for higher energies. At the same time, the spread in  $\ln A$  is approximately constant below  $10^{18.3}$  eV and then it starts to decrease, indicating the evolution toward a purer mass composition. The combination of the two resulting behaviours strongly suggests that the fraction of lighter primaries becomes smaller for energies above  $10^{18.3}$  eV.

However, one should notice that the negative values of  $\sigma^2(\ln A)$  are unphysical, since are obtained when the corresponding hadronic model predicts  $\sigma(X_{\max})$  values for pure compositions larger than the observed ones (see eq. 2.23). As a consequence, the QGS-Jet II-04 model fails to provide consistent interpretation of data, while EPOS-LHC and Sibyll 2.3 are consistent at least up to the highest energy bins.

### 3.5.3 Anisotropy searches

As explained in sect. 1.3, the interaction of cosmic rays with magnetic fields in the galaxy and in the intergalactic medium makes difficult the identification of candidate sources, even more complicated in the presence of a mixed mass composition [82] (see previous paragraph). However, an analysis of the arrival direction distribution can provide crucial information about the origin and nature of the ultra-high energy cosmic rays. In fact, although particles from individual sources are strongly deflected, it remains possible to detect anisotropies in the distribution of CR arrival directions on large angular scales<sup>17</sup>,

<sup>17</sup>The standard approach is to perform a harmonic analysis over the distributions in right ascension ( $\alpha$ ) and azimuthal angle ( $\phi$ ). In fact, in the presence of a three-dimensional dipole, the analysis in right ascension is sensitive only to the component orthogonal to the Earth's rotation axis. Instead, the dipole

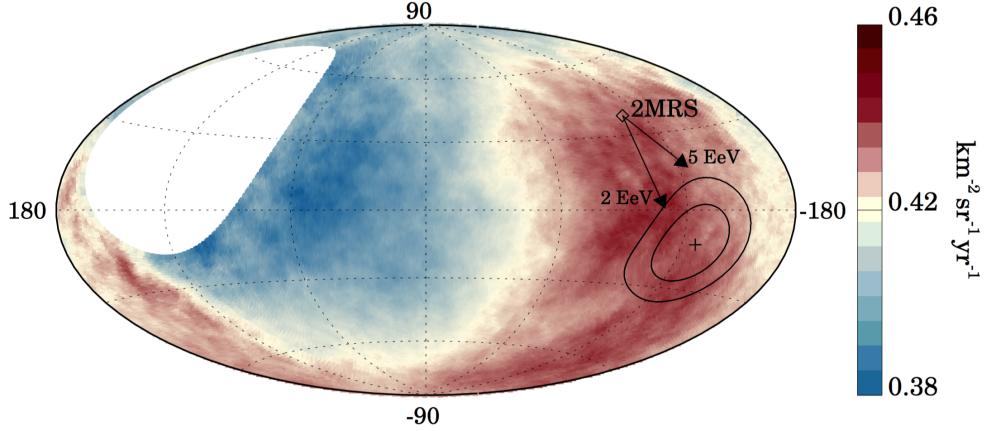


Figure 3.18: Sky map in Galactic coordinates showing the cosmic ray flux for  $E \geq 8$  EeV smoothed with a  $45^\circ$  top-hat function. The Galactic center is at the origin. The cross indicates the measured dipole direction; the contours denote the 68% and 95% confidence-level regions. The dipole in the 2MRS galaxy distribution is indicated. Arrows show the deflections expected for a particular model of the Galactic magnetic field on particles with  $E/Z = 5$  EeV or 2 EeV. Image from [86].

as attempted in several analysis [83, 84, 85] by the Pierre Auger Collaboration.

Updated studies [86, 87] found a statistically significant dipolar large-scale anisotropy for energies above 8 EeV, pointing toward the Galactic coordinates  $(233^\circ, -13^\circ)$ . Such dipolar pattern can be clearly seen in the flux map of Fig. 3.18. The magnitude and direction ( $\sim 125^\circ$  away from the Galactic center) of this anisotropy support the hypothesis of an extragalactic origin for the highest-energy cosmic rays, rather than sources within the Galaxy. In general, models proposing a Galactic origin up to the highest observed energies are in increasing tension with observations. This tension could be alleviated if cosmic rays at a few EeV were dominated by heavy nuclei such as iron, but this would be in disagreement with the lighter composition inferred at these energies.

Furthermore, the highest energy dataset from the Pierre Auger Observatory can provide evidence for anisotropy in the arrival directions of UHECR also on an intermediate<sup>18</sup> angular scale, indicative of an excess from strong, nearby sources. A recent analysis [88] focused on two prominent classes of extragalactic sources: active galactic nuclei (AGNs) and starburst galaxies (SBGs). Interestingly, the contribution from SBGs to the indication of anisotropy is the largest among the tested catalogs. In particular, isotropy is disfavored at  $4.0\sigma$  confidence level with respect to SBGs models, the highest value of the test statistic being for energies above 39 EeV. Nonetheless, additional exposure as well as a deeper understanding about the impact of bulk magnetic deflections are required before considering SBGs as UHECR preferred sources.

---

component parallel to the Earth's rotation axis induces a modulation in the azimuthal distribution of the arrival directions at the array.

<sup>18</sup>"Intermediate" denotes angular scales larger than the experimental resolution,  $\sim 1^\circ$ , and smaller than large-scale patterns,  $>45^\circ$ .

At the same time, the Observatory upgrade will substantially enhance anisotropy studies. In fact the availability of new mass-sensitive observables, provided for each shower, will enable a charge-discriminated analysis, i.e. the exclusion of events likely initiated by heavier nuclei, which constitute a quasi-isotropic background that may mask the signature of individual sources.



# Chapter 4

## Air shower Universality

The concept of Universality in the development of extensive air showers dates back to more than half a century, since from the first studies about purely electromagnetic cascades generated by photons or electrons. The general idea is that the spectra of low-energy particles, mostly electrons and photons, are *similar* in air showers at the same stage of their development (to be better defined in the following). Furthermore, also the angular and the lateral distributions around the shower axis have been demonstrated to show the same behaviour.

A deeper understanding and characterization of such properties (and their limits) can provide an invaluable instrument in the study of EASs and allow a more reliable reconstruction of the physical parameters of cosmic rays, for instance reducing the influence of hadronic interaction models.

In this chapter, the main results about the study of air shower Universality are reported, starting from the application to pure electromagnetic showers and the extension to hadronic ones. Then, the model of Universality elaborated in the framework of the Pierre Auger Collaboration is presented. Finally, the parameterizations implemented for the reconstruction of events measured by the surface detector of the Pierre Auger Observatory are described.

### 4.1 Universality of the shower components

In the last decades the progresses in simulation (due to modern software packages such as AIRES and CORSIKA, updated high-energy interaction models and fast computing) have allowed to accomplish much more reliable predictions of the particle distributions in showers generated by very high energy cosmic rays. By means of these Monte Carlo methods, it was observed [89, 90] that the energy and angle distributions of  $e^+/e^-$  in showers initiated by high energy protons and nuclei present a shape that, to a remarkable degree of precision, is only determined by an age parameter  $s$  defined as:

$$s = \frac{3X}{X + 2X_{\max}} \quad (4.1)$$

where  $X$  is the slant depth traversed by the shower in the atmosphere and  $X_{\max}$  the depth of the maximum in the development of the e.m. part. A critical review and above-all a theoretical interpretation of these (and previous) results was given in [91], where a better motivated and more accurate definition of “age” of a shower is discussed.

A complete investigation about the Universality behaviour of the electron-positron component in extensive air showers is accomplished in [92]. In particular, detailed Monte Carlo simulations are exploited to obtain the electron–positron distributions as functions of energy, mass, zenith angle of the primary particle and on the evolution stage of the shower. The aim is to perform reliable parameterizations of such distributions, that could be used for the analysis of Cherenkov, fluorescence and radio emission observations, and in the interpretation of ground particles measurements.

In [92], the stage of shower development is expressed through the *relative evolution stage*  $t$ , a quantity defined as:

$$t = \frac{X - X_{\max}}{X_0} \quad (4.2)$$

with  $X_0 = 36.7 \text{ g/cm}^2$  the radiation length of electrons in air. Such parameter gives a description at least as good [92] as the shower age parameter  $s$  in eq. 4.1.

The first result concerns the  $e^+/e^-$  energy spectra, which are shown in Fig. 4.1a for the range  $1 \text{ MeV} < \epsilon < 1 \text{ GeV}$  in the secondary particles energy  $\epsilon$ . In these plots the background red curves represent the distributions from showers initiated by different species ( $\gamma$ , p, Fe) at different energies; the three panels allow a comparison for tree different evolution stages. The presence of a universal behaviour is evident; at the same time, the parameterizations (dashed lines) show a very high accuracy, with deviations reported to be generally lower than 10%.

An evidence of Universality in the electron-positrons component emerges also from the analysis of the angular spectrum of these particles, an important factor in the measurements with Cherenkov and radio telescopes. In the top panel of Fig. 4.1b the angular distributions from 20 proton-initiated shower at  $10^{18} \text{ eV}$  are shown. As can be seen, the dispersions near the axis (lower  $\theta$  values) are small, with an increase only for the lower energy particles (small  $\epsilon$  values).

In [92] it is reported that the angular spectra are independent from the primary zenith angle or energy. In addition, the effects of primary mass and shower stage (encoded by the  $t$  parameter) are smaller than the differences between individual showers: then the authors conclude that the parameterizations of the  $e^+/e^-$  angular spectra can be considered only functions of the particles angle and energy.

Another important result is obtained from the study of the electron-positrons lateral spread. The related parameterizations are relevant for ground-array experiments, which are based on the measurements of particle densities at different lateral distances from the shower axis. A Universality description can be found after expressing the lateral spread

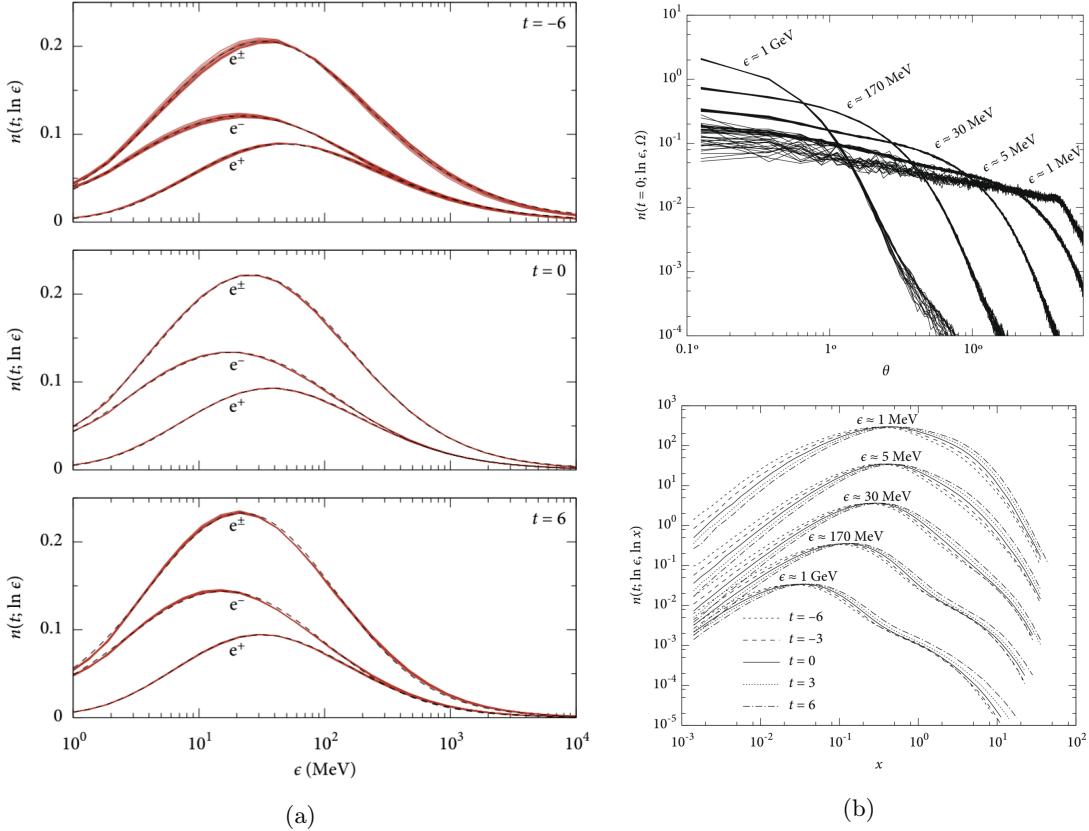


Figure 4.1: (a) Average energy distribution for electrons ( $e^-$ ), positrons ( $e^+$ ) and their sum ( $e^\pm$ ) at different relative evolution stages ( $t = -6, 0, 6$ ). The background red curves represent the distributions from simulated showers initiated by different primaries (p, Fe,  $\gamma$ ) with different energies ( $10^{17}, 10^{18}, 10^{19}$  eV), while the dashed black lines are the corresponding parameterizations (given in [92]). (b, top panel) Angular distributions of  $e^+/e^-$  at the shower maximum ( $t = 0$ ) for different energies  $\epsilon$ , from 20 showers initiated by  $10^{18}$  eV protons. The angle  $\theta$  represents the deviation from the shower axis:  $0^\circ$  is along the primary's trajectory,  $90^\circ$  is perpendicular to the shower axis. (b, lower panel) Average lateral spread distributions for different shower stages  $t$ ; each curve is the average over 20 proton-initiated showers at  $10^{18}$  eV. The curves at 1 GeV are at the actual level; consecutive sets are shifted up by a factor 10. Images from [92].

in terms of the Moli  re radius  $r_M$ , defining a quantity

$$x = \frac{r}{r_M} \approx \frac{r \rho_A(h)}{9.6 \text{ g/cm}^2} \quad (4.3)$$

where  $\rho_A(h)$  is the atmospheric density as a function of height  $h$ .

In the lower right panel of Fig. 4.1b the normalized lateral particle distributions for different shower stages are shown: each line represents the average over 20 proton-initiated showers at  $10^{18}$  eV. Differently from the angular spectra, a dependence on the shower evolution is evident: thus, the related parameterizations are functions also of the parameter  $t$ . Instead, no statistically significant dependences of the lateral spread are found in [92] on the primary zenith angle or energy.

Finally, in Fig. 4.2a the differences arising from showers initiated by different primary masses are highlighted. Such dependence is mostly shown by the second bulge of the

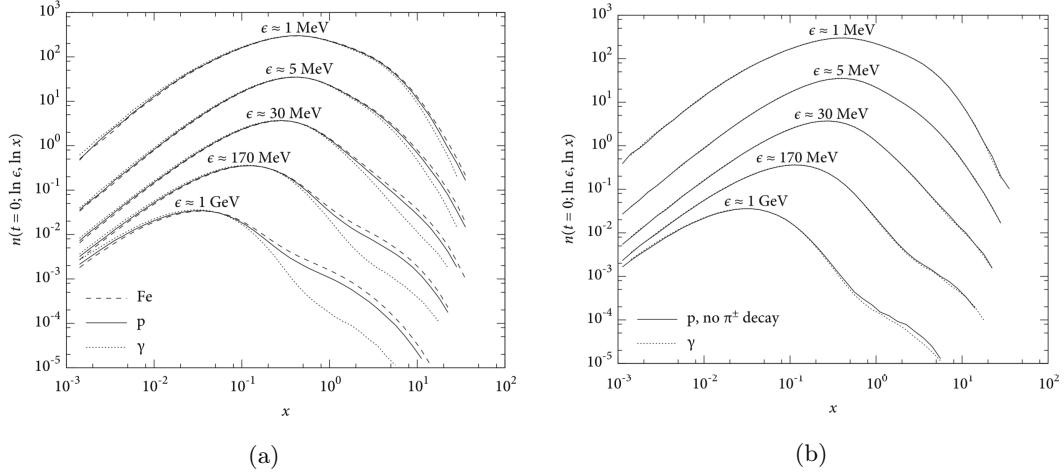


Figure 4.2: (a) Average lateral spread distributions at the maximum ( $t = 0$ ) for different primaries, averaged over 20 showers at  $10^{18}$  eV. (b) Comparison of the lateral spread distributions at the maximum ( $t = 0$ ) averaged over 20 photon showers at  $10^{17}$  eV and 20 proton showers at the same energy but with the  $\pi^\pm$  disabled. The shifts for different energies  $\epsilon$  are as in the lower panel of Fig. 4.1b. Images from [92].

distributions, i.e. the part with higher  $x$  values. The explanation of this effect is related with the separate origin of the two main contributions to the lateral spread distribution. In fact, the left part (small  $x$  values) is made up of  $e^\pm$  produced through bremsstrahlung and pair creation processes (then spread by multiple scattering), which are the main channels for the development of the electromagnetic component of the shower. Instead, the right part (i.e. the second bulge) is due to the pion production channel. This is also demonstrated through Fig. 4.2b, where the  $\pi^\pm$  creation channel was disabled for showers initiated by protons.

The source of this Universality violation is related with the strength of the hadronic part of the cascade for different primary species. The hadronic contribution is relevant mainly in the first stages of the shower development but it leaves a trace in the muonic component, which is also one of the main discriminating factors not only to distinguish photons from hadron primaries, but also among nuclei.

Therefore the importance of deeper studies about the shower muons clearly raised, with the idea to actually extend the Universality concept also to this component. From several studies (see for example [93, 94] and references therein) it emerges that the distributions of muons present an approximately universal behaviour, choosing as parameter to describe the development the quantity

$$X' = X - X_{\max}^\mu \quad (4.4)$$

and expressing the number of muons as:

$$N'_\mu = N_\mu / N_\mu^{\max} \quad (4.5)$$

where  $X$  is the slant depth traversed in atmosphere,  $X_{\max}^\mu$  is the depth of the maximum of the muonic component and  $N_\mu^{\max}$  is the correspondent maximum number of muons.

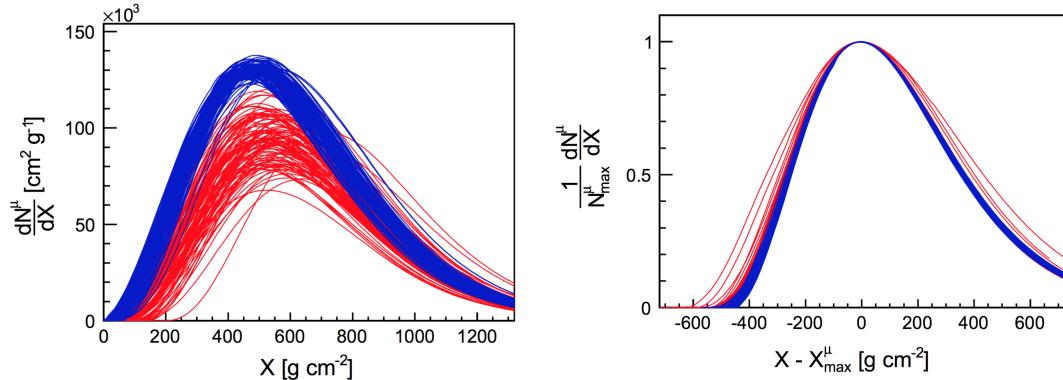


Figure 4.3: (a) Muon production profiles as a function of slant depth ( $X$ ) for 100 proton-initiated (red) and iron-initiated (blue) showers at  $10^{19}$  eV generated with QGSJet-II.03 with zenith  $\theta = 40^\circ$ . (b) The same profiles of the left panel in the  $(X', N'_\mu)$  coordinates of eq. 4.4 and 4.5. Images from [93].

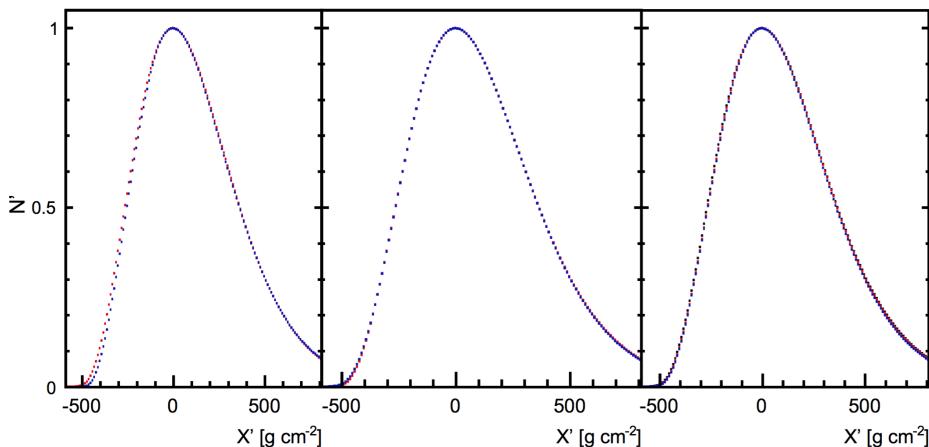


Figure 4.4: Averaged muon production longitudinal profiles in  $(X', N'_\mu)$  coordinates. (Left) Comparison for different primary masses: proton (red) and iron (blue) showers generated with QGSJET-II at  $\theta = 40^\circ$ . (Middle) Comparison for different hadronic interaction models: QGSJet-II (red) and EPOS 1.99 (blue) for proton showers at  $\theta = 40^\circ$ . (Right) Comparison for different zenith angles:  $0^\circ < \theta < 10^\circ$  (black),  $30^\circ < \theta < 40^\circ$  (red),  $45^\circ < \theta < 55^\circ$  (blue) from proton showers generated with QGSJet-II. Image from [93].

The result of this re-normalization is illustrated in Fig. 4.3: on the left panel, the muon production profiles for several proton (red) and iron (blue) initiated shower are depicted, showing their natural differences; on the right panel, the same showers are re-drawn using the eq. 4.4 and 4.5, and a rather universal shape is obtained.

Then the average shape of the longitudinal muon production profile of primary particles with different masses and zenith angles, and also for different hadronic interaction models, are compared. The Universality of the shape is confirmed<sup>1</sup>, as can be observed from Fig. 4.4.

<sup>1</sup>There are, however, some differences between the average shapes in the early phase of the shower, but such minor violation of Universality where expected given the importance, for the early stages of the muonic component, of the very first hadronic interactions of the primary particle.

In conclusion, the fundamental distributions and profiles both of electron-positron and muonic components of an EAS are demonstrated to show a Universal behaviour, i.e. mostly independent from the characteristics of the primary particle and the hadronic interaction models used for simulation. The next step is the development of a Universal model for the reconstruction of showers in UHECR experiments.

## 4.2 Universality of the ground signals

The results presented in the previous section allow to built a new description of EASs based on the Universality paradigm. The basic concept, supported by the universal behaviour observed for the distributions of the shower secondary particles, is that extensive air showers can be characterized (with a remarkable level of accuracy) through a reduced set of macroscopic parameters. In particular, the primary energy ( $E$ ), the depth of shower maximum in the electromagnetic component ( $X_{\max}$ ) and an overall normalization of the muon content ( $R_\mu$ ) are considered sufficient to describe in good approximation the shower properties at ground. It should be noticed that the parameter  $R_\mu$  also accounts for primary composition and hadronic interaction model effects.

As discussed in sect. 2.2.2, an array of surface detectors performs a sampling of the shower at ground level, which means at only one stage of its development. The actual measurement is represented by the signals recorded by every detector. So the idea is to analyse the signals from simulated events looking for an extension of air shower Universality that could apply also to these quantities.

In this model, the stage of shower development is defined by a parameter different from the previous age  $s$  or relative evolution  $t$  values. The best Universal behaviour was found [95, 96] using the variable  $DX$ , defined as the distance from the detector to the shower maximum measured (in g/cm<sup>2</sup>) along the shower axis, that is:

$$DX = X_{\text{gr}} \sec \theta - X_{\max} \quad (4.6)$$

where  $X_{\text{gr}}$  is the vertical depth of the detector and  $\theta$  the zenith angle of the shower<sup>2</sup>. A sketch to illustrate the definition of  $DX$  is given in Fig. 4.5.

Initially [96, 98] the models were developed only for the signal at the optimal distance  $S(r_{opt})$ , a common reconstructed parameter employed as indicator of the shower size (see sect. 2.2.2). Then  $S(r_{opt})$  was parameterized as the sum of two contributions: (i) an electromagnetic part, which corresponds to the e.m. particles of the shower and thus is better-understood and known to be approximately dependent only on  $E$  and  $X_{\max}$ ; (ii) a

---

<sup>2</sup>For a single ground detector at a distance  $r$  from the shower axis,  $DX$  is:

$$DX = X_{\text{gr}} \sec \theta - X_{\max} - r \cos \zeta \tan \theta \rho_{\text{air}}$$

where  $\zeta$  is the azimuthal angle of the detector in the shower-plane (such that  $\zeta = 0^\circ$  corresponds to a tank below the shower axis) and  $\rho_{\text{air}} \approx 10^{-3}$  g/cm<sup>3</sup> is the density of air at ground level. However, the signals are usually averaged over the azimuth, allowing to use the simplified expression in eq. 4.6.

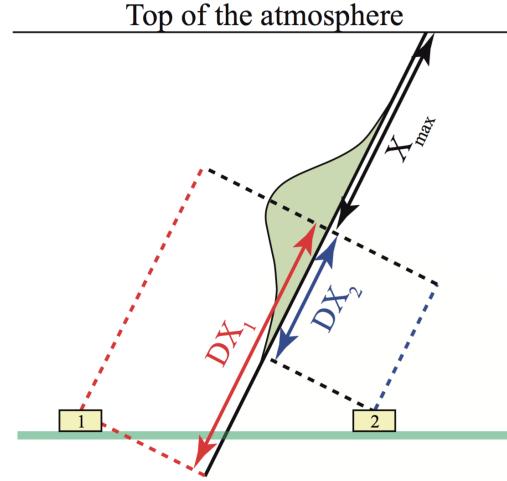


Figure 4.5: Sketch to illustrate the definition of  $DX$ , the distance from a detector to the shower maximum along the shower axis. Image from [97].

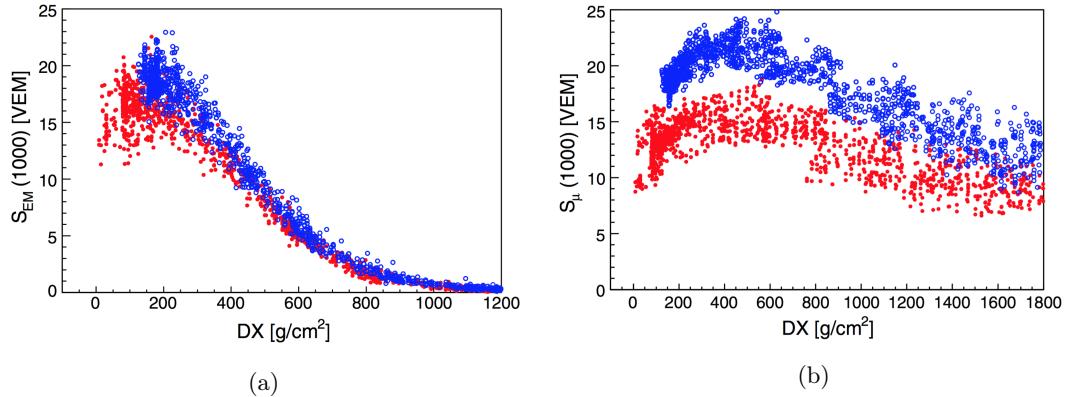


Figure 4.6: (a) Electromagnetic shower-plane signals at  $r = 1000$  m represented as a function of  $DX$  for proton (red dots) and iron (blue circles) showers at  $10^{19}$  eV simulated with QGSJetII/Fluka and with discrete zenith angles in the range  $[0^\circ, 60^\circ]$ . (b) The same as left panel but for the muonic part of the shower-plane signals. One should notice that  $S_\mu$  includes the contribution from muon decay products. Images from [96].

muonic part, related to the shower muon component, that depends (in addition to  $E$  and  $X_{\max}$ ) on the muon normalization  $R_\mu$ . Moreover, the signal contributions due to particles from muon decay are included in the muonic part.

Some explicative results are reported in Fig. 4.6 for the analysis of *shower plane signals*. Such signals are obtained from simulated showers using fiducial flat detectors (with the same average particle response as ground detectors) placed in the plane perpendicular to the shower axis (the so-called shower-plane)<sup>3</sup>. In effect, signals calculated in the shower-plane are not affected by detector geometrical effects, and therefore independent of the zenith angle and more suitable for Universality studies.

In Fig. 4.6a the electromagnetic component  $S_{EM}$  of the shower-plane signal at 1000 m

<sup>3</sup>Of course for the application to experimental data it will be necessary to account for the differences with signals actually measured on the ground-plane, as discussed in the following.

from the shower core is plotted as a function of  $DX$  for proton and iron-initiated showers at  $10^{19}$  eV; the same in Fig. 4.6b for the muonic part  $S_\mu$ . Both contributions present an evolution with  $DX$ , showing a maximum and a following attenuation. The peak for  $S_{EM}$  falls around  $200 \text{ g/cm}^2$  and then the signal rapidly decreases; instead  $S_\mu$  peaks around  $400 \text{ g/cm}^2$  and it gets attenuated more slowly than the electromagnetic component. Furthermore, the differences between models are found [96] to be around 5-10%.

In both cases only differences in normalization and not in shape are apparent, confirming the existence of a universal behaviour also for the measured signals. However, while for the muonic part  $S_\mu$  the signal normalization was predicted to depend on the primary particle and hadronic model (as evident in Fig. 4.6b), in the e.m. signals a shift (also larger than 15%) between proton and iron showers was not expected.

Such shift cannot be related to the muonic component of the shower (i.e. to the differences in muon content between protons and iron showers) because muon decay products are included in  $S_\mu$ . In [96] the authors explained this result with a violation, in hadronic shower simulations, of a strict linear scaling of the electromagnetic signal with energy. In this scenario, considering valid the superposition model (sect. 2.1.3),  $S_{EM}$  would scale as  $(E/A)^\alpha$  (with  $\alpha < 1$ ). Then the electromagnetic signal of an iron shower will be a factor of  $(1/56)^\alpha$  times larger than that of a proton shower at the same energy, and the observed difference of about 15% can be explained by a value of  $\alpha \sim 0.97$ . This effect is supposedly due to the hadronic part of the shower which feeds the electromagnetic component. For instance, an evolution of the  $\pi^0$  spectrum with primary energy, as well as of the injection rate of energy by pion decay in the e.m. part of the shower, could partially explain the observations.

In the following years, a different Universality model was elaborated [97, 99] in order to explicitly take into account the relation between neutral pions and the electromagnetic component of the shower, and between charged pions and the muonic part of the shower. With the aim of building a robust Universality parameterization of the ground signals, it was introduced a shower description that presents four components:

- the purely electromagnetic component (electrons, positrons and photons);
- the muonic component;
- the electromagnetic component from muon interactions and muon decay;
- the electromagnetic component from low-energy hadrons (*jet component*)<sup>4</sup>.

The methodologies implemented to discriminate between the different components in simulation are reported in [99] and will not be discussed in this work.

Adopting these new definitions, the disagreement in the electromagnetic signal  $S_{EM}$  previously observed for different primaries is approximately ruled out. As mentioned,

---

<sup>4</sup>Shower simulations show that a not-negligible fraction of the secondary particles reaching the ground are originated from hadronic interactions at low-energies ( $\sim \text{GeV}$ ). Such particles fall at large distances from the shower core and present large angles with respect to the shower axis, since their lateral distribution is mostly determined by the momentum of the mother hadronic particle.

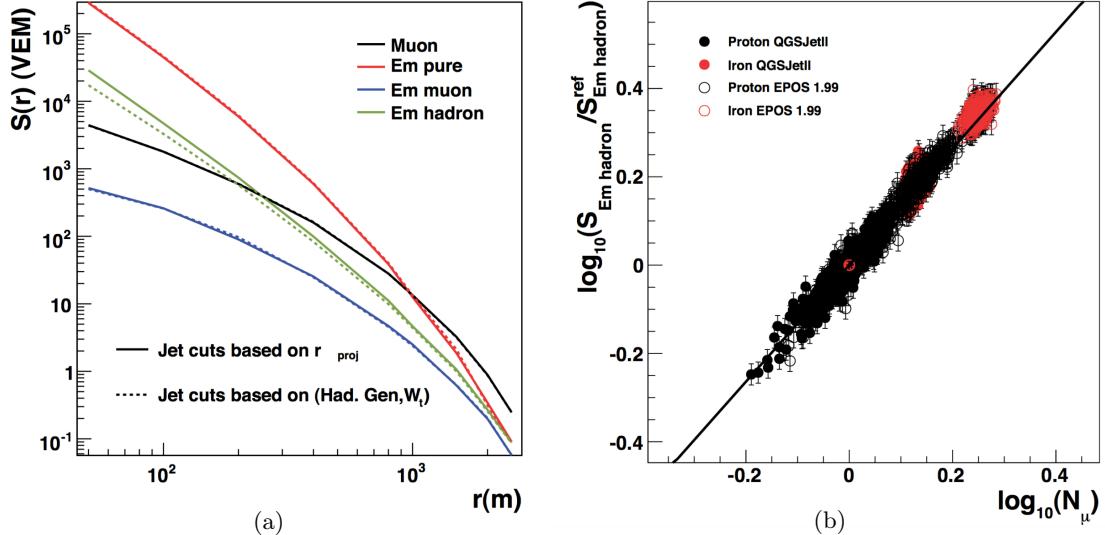


Figure 4.7: (a) The lateral distribution in the 4 Universality components for a simulated shower initiated by a proton with 10 EeV, zenith angle of  $45^\circ$  and  $X_{\max} = 1010 \text{ g/cm}^2$ . The two different lines (solid and dashed) are obtained with two different methods for the selection of the jet-component, as described in [97, 99]. (b) The correlation, at 1000 m from the shower core, between  $N_\mu$  and the e.m. from hadrons signals  $S_{e.m.(had)}$  with respect to a reference signal  $S_{e.m.(had)}^{\text{ref}}$ . Shower at 10 EeV energy and with zenith angle in the range 0 to  $50^\circ$  are employed. Images from [97].

this problematic was mainly ascribed to the energy transfer from the hadronic to the electromagnetic shower parts, which clearly depends on the employed primary and interaction model. Since the (purely) electromagnetic part is now obtained subtracting the jet component from the total electromagnetic signal, the result is very close [99] to the expectation for a photon-induced shower with hadronic interaction channels disabled, meaning that a much more universal behaviour has been recovered.

In Fig. 4.7a the lateral distribution, split in the 4 components, is shown for one simulated shower. The e.m. from hadron component actually exhibits a different radial dependence with respect to the purely e.m. one, further accounting for the Universality violation previously found in [96].

In this new Universality model the signal and its components are still studied as functions of  $DX$  (the distance from the maximum along the shower axis, eq. 4.6). Instead the relative muon content  $R_\mu$  is replaced by a parameter  $N_\mu$  defined as:

$$N_\mu = \frac{S_{0,\mu}(r, DX, E)}{S_{0,\mu}^{\text{ref}}(r, DX, E)} \quad (4.7)$$

a quantity that can be considered as an effective muon content for the shower with respect to a reference primary and model. In effect, in the last expression,  $S_{0,\mu}$  indicates the muonic component signal and “ref” the mean obtained for the reference primary and model (proton QGS-Jet II in [97]). Moreover, the pedix “0” specifies that the signals are obtained simulating the shower in an array of ideal spherical detectors (i.e. detectors

with a projected area of  $10 \text{ m}^2$  regardless the incoming particle direction) with a response function chosen equal to that of a tank from the Pierre Auger Observatory for vertical incident particles. This definition is similar to the shower-plane signals in [96] and has the same motivation, that is to separate the effects due to the detector geometry from the physical behaviour of the signals.

Finally both the e.m. from hadrons and e.m. from muons components are found to be dependent on  $N_\mu$ , condensing on this parameter the primary and interaction model dependences of the electromagnetic part of the shower. The correlation between  $N_\mu$  and  $S_{0,em(hadr)}/S_{0,em(hadr)}^{\text{ref}}$  is shown in Fig. 4.7b at  $r = 1000 \text{ m}$  for 10 EeV showers. In this plot, the fit is accomplished using only data points from the reference model (proton QGS-Jet II), but it is apparent that the other models/primaries not only present the same functional form but actually lie along the same line. Such empirical finding can be interpreted as another proof for the validity of air shower Universality applied to ground signals, since it implies that an iron shower is not distinguishable from a fluctuation of a proton one.

The results obtained in the field of Universality, partially reported in this section, assured that an analytical description of the surface detector signals is possible<sup>5</sup>. The complete parameterization, as implemented in the framework of the Pierre Auger Collaboration, is presented in the following.

### 4.3 Universality of the total integrated signal

The 4-components description of air showers allowed to built a model for the parameterization of integrated signals measured by the WCDs [99] of the Pierre Auger Observatory, later extended [100] for the scintillator detectors of the *AugerPrime* upgrade, and for the underground muon detectors of AMIGA [101]. In the following, the discussion will be focused on the WCD parameterization (in its updated version given in [101]), the only one of interest for this work.

An EAS will be characterized by its primary energy  $E$  and zenith angle  $\theta$ , the slant depth of the maximum in the shower development  $X_{\max}$  and a relative number of muons  $N_\mu$ . Moreover, the stage of shower development is still evaluated through the variable  $DX$ , re-named  $\Delta X$  in [101].

To parameterize the signal, proton and iron-induced showers simulated with CORSIKA are employed. The hadronic interaction models used are QGS-Jet II.03 for the high-energy and FLUKA for the low-energy part. The primaries have zenith angles  $\theta$  ranging from 0 to  $60^\circ$  and energies between  $10^{18.6}$  and  $10^{20}$  eV. For each couple<sup>6</sup> of energy and zenith angle 120 showers were simulated: 10 showers per monthly atmospheric

<sup>5</sup> At least for core distances from 100 to 2000 m and zenith angles between 0 and  $60^\circ$ , that are the ranges explored in the above-mentioned references.

<sup>6</sup>In particular,  $\theta = [0^\circ, 12^\circ, 25^\circ, 36^\circ, 45^\circ, 53^\circ, 60^\circ]$  and  $E = [10^{18.6}, 10^{19}, 10^{19.5}, 10^{20}] \text{ eV}$ .

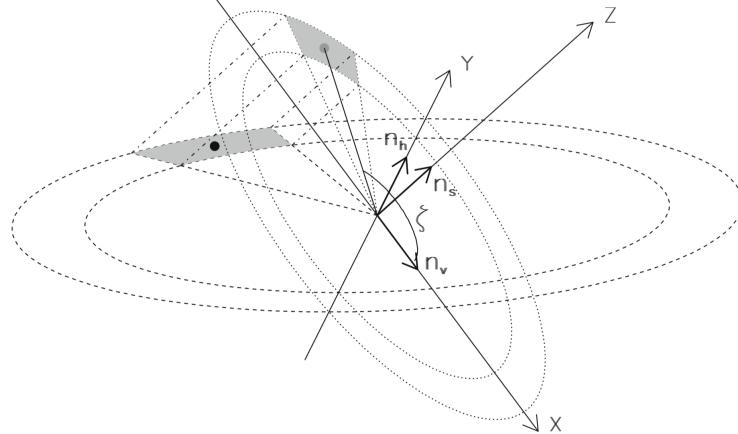


Figure 4.8: Geometrical definition of the sampling region around a detector (black point): filled in grey, it is the projection onto the ground, along Z-axis, of a region in the shower plane (X, Y) delimited by a circular crown and an angular sector. The shower frame is defined by: Z along the shower axis; X in the vertical plane going through the shower axis; Y in the horizontal plane. Image from [102].

model. The signal is studied in a coordinate system where the X and Y-axes lie on the shower plane, and the Z-axis points upwards. The shower plane is defined as the plane perpendicular to the shower axis and containing the surface detector where the signal have to be evaluated. The ground level is set at 1452 m and the magnetic field at the location of the Pierre Auger Observatory is taken into account.

The *resampling* (or un-thinning) procedure described in [102] is performed to recover the particles densities at ground after the thinning of the simulated showers. However, the signals exploited to build the model are calculated averaging<sup>7</sup> over each *sampling area*, obtained projecting at ground a slice of a ring in the shower plane, as shown in Fig. 4.8. The rings are centered at core distances<sup>8</sup>  $r \in [100, 2500]$  m while the slices in  $\psi$  (the azimuthal angle in the shower plane, with  $\psi = 0^\circ/180^\circ$  corresponding to upstream/downstream) are chosen<sup>9</sup> using the expected first order particle symmetry in density  $\rho(r, \psi) = \rho(r, 360^\circ - \psi)$ .

Finally, the average signal  $S$  released in a ground detector at the position  $(r, \psi)$  by a shower with a certain set of  $(E, X_{\max}, N_\mu, \theta)$  values is parameterized as:

$$S(r, \psi | E, X_{\max}, N_\mu, \theta, \rho_{\text{ground}}^{\text{air}}) = \sum_{i=1}^4 S_0^i(r, \Delta X | E) \times f_{\text{mod}}^i(r, \psi | \theta) \times \\ \times f_{\text{atm}}^i(r | \rho_{\text{ground}}^{\text{air}}) \times f_{\text{conv}}^i(r, \Delta X, \psi | \theta) \times f_{\text{fluct}}^i(r | N_\mu) \quad (4.8)$$

---

<sup>7</sup>The resampling procedure requires a Poissonian extraction of the number of particles actually reaching a detector, with mean proportional to the weight of the thinned particles reaching the sampling area. Instead, the Universality parameterization is realized using signals obtained directly with the mean number of particles (no Poisson extraction). The choice of using such average signals was made to reduce the statistical fluctuations. See next section, in particular Fig. 4.15, for a more detailed explanation.

<sup>8</sup>In particular  $r = [100, 200, 400, 800, 1000, 1500, 2000, 2500]$  m.

<sup>9</sup>The bins in  $\psi$  are :  $[-30^\circ, 30^\circ]; [30^\circ, 60^\circ] \cup [300^\circ, 330^\circ]; [60^\circ, 120^\circ] \cup [240^\circ, 330^\circ]; [120^\circ, 150^\circ] \cup [210^\circ, 240^\circ]; [150^\circ, 210^\circ]$ .

where  $i$  refers to the 4 Universality shower components and  $\rho_{\text{ground}}^{\text{air}}$  is the ground air density. In addition, one should notice that  $\Delta X = \Delta X(r, \psi, \theta, X_{\max})$ . Separately for each component, the terms in the last equation indicate:

- $S_0$ , the signal in an ideal detector<sup>10</sup> simulated without ground and for  $\theta = 0^\circ$ ;
- $f_{\text{mod}}$ , a change in the normalization of  $S_0$  due to the differences in atmospheric attenuation arising at every different position of the WCDs with respect to the core and the inclined shower axis;
- $f_{\text{atm}}$ , a factor that takes into account the variations of the atmospheric profile, both on a daily and seasonal basis;
- $f_{\text{conv}}$ , the conversion from ideal to realistic detectors, which differ in the responses and projected areas as functions of the incoming direction of the shower particles;
- $f_{\text{fluct}}$ , a factor that accounts for signal fluctuations using the correlations between each component and  $N_\mu$ .

A deeper discussion about the derivation of these factors and the choices for their functional form can be found in [99, 100, 101] and is omitted here. Only the first ( $S_0$ ) and last ( $f_{\text{fluct}}$ ) terms will be further described.

It is fundamental to stress that all the results for the Universality parameterization have been obtained using only showers generated with the reference primary and model, i.e. proton - QGSJet II.03 in [101]; therefore the accuracy in the predictions for different combinations of primaries and models needs to be tested.

The signal in an ideal detector is parameterized, for each component and each core distance, with a Gaisser-Hillas type function in the form:

$$S_0^i = S_{\text{ref}}^i \left( \frac{E}{10^{19} \text{ eV}} \right)^{\gamma^i} \left( \frac{\Delta X - \Delta X_0^i}{\Delta X_{\text{ref}} - \Delta X_0^i} \right)^{\frac{\Delta X_{\max}^i - \Delta X_0^i}{\lambda^i(E)}} \exp \left( \frac{\Delta X_{\text{ref}} - \Delta X_0^i}{\lambda^i(E)} \right) \quad (4.9)$$

where  $\Delta X_{\text{ref}} = 400 \text{ g/cm}^2$ ,  $S_{\text{ref}}$  is the signal for  $\Delta X = \Delta X_{\text{ref}}$  (but it is also a function of  $r$ ) and  $\lambda^i(E) = \lambda_0^i + f_\lambda^i \cdot (E/10^{19} \text{ eV})$ . The parameters  $S_{\text{ref}}$ ,  $\lambda_0$ ,  $\Delta X_{\max}$  and  $\gamma$  have been fitted as functions of the core distance  $r$  separately for each component. Instead, respectively for the muonic, pure electromagnetic, electromagnetic from muon and electromagnetic from hadrons components, the values of  $f_\lambda$  are fixed to  $[0, 0, 0.04, -0.02]$  and the values of  $\Delta X_0$  to  $[-250, -500, 250, -800] \text{ g/cm}^2$ .

The parameterization of  $S_0$  as function of  $\Delta X$  is shown in Fig. 4.9 for a core distance  $r = 1000 \text{ m}$  and primary energy  $E = 10^{19} \text{ eV}$ . The solid lines are the 4 fit results, while each point corresponds to a sampling area of the simulated showers for the reference primary and model (protons, QGSJet II-03); the colours correspond to different zenith angles. The Universality behaviour is evident, even if each component present a different dependence on  $(r, \Delta X, E)$ .

---

<sup>10</sup> Also in this updated parameterization, such ideal detector is spherical and with a projected area of  $10 \text{ m}^2$  regardless the incoming particles directions. The response function is naturally chosen equal to that of a WCD for vertical incident particles.

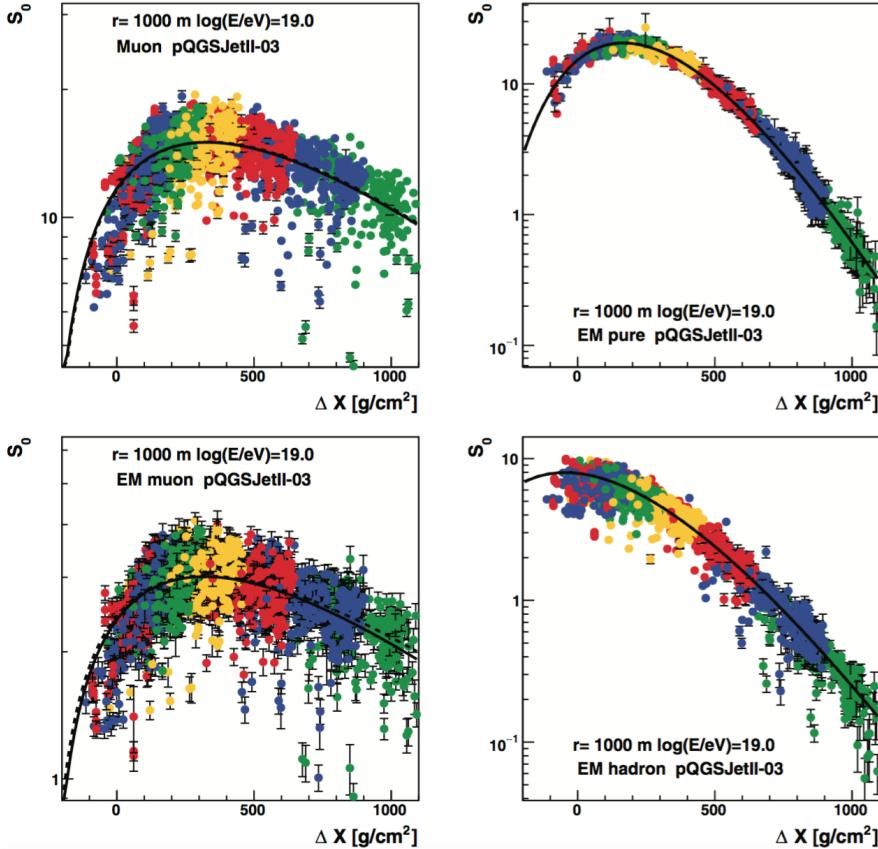


Figure 4.9:  $S_0$  as a function of  $\Delta X$  for the four Universality components. The core distance is  $r = 1000$  m and the primary energy is  $10^{19}$  eV. The different colours indicate different zenith angles. Each point corresponds to a sampling area in a simulated shower; only sampling areas with central  $\psi = 0^\circ, 90^\circ$  and  $180^\circ$  are reported in this plot. The errors are calculated from the Poisson fluctuations of the number of weighted particles generating the signal. Image from [101].

As exposed in the last paragraphs of the previous section, the signal fluctuations of the e.m. from muons and e.m. from low-energy hadrons components, due to the hadronic part of the shower, can be analysed as functions of  $N_\mu$ , defined in [101] as:

$$N_\mu = \frac{S_{0,\mu}(r = 1000 \text{ m}, \psi = 90^\circ | \Delta X, E)}{S_{0,\mu}^{\text{ref}}(r = 1000 \text{ m}, \psi = 90^\circ | \Delta X, E)} \quad (4.10)$$

Eventually, for a complete development of the parameterization, this correlation is investigated also for the pure electromagnetic component.

In Fig. 4.10 the correlations among the three electromagnetic components and the muonic one are shown for the signals  $S_0^i$  with respect to the reference signals. It is necessary to note that the value of  $N_\mu$  is exactly equal to  $S_{0,\mu}/S_{0,\mu}^{\text{ref}}$  only for  $r = 1000$  m and  $\psi = 90^\circ$ . The different colours distinguish among different combinations of primary and interaction models, but the fits (solid lines) are performed using only the reference primary and model (red points, proton - QGSJet II-03) according to the expression:

$$\frac{S_{0,e.m.(k)}}{S_{0,e.m.(k)}^{\text{ref}}} = 1 + \left( \frac{S_{0,\mu}}{S_{0,\mu}^{\text{ref}}} \right) \alpha_k \quad (4.11)$$

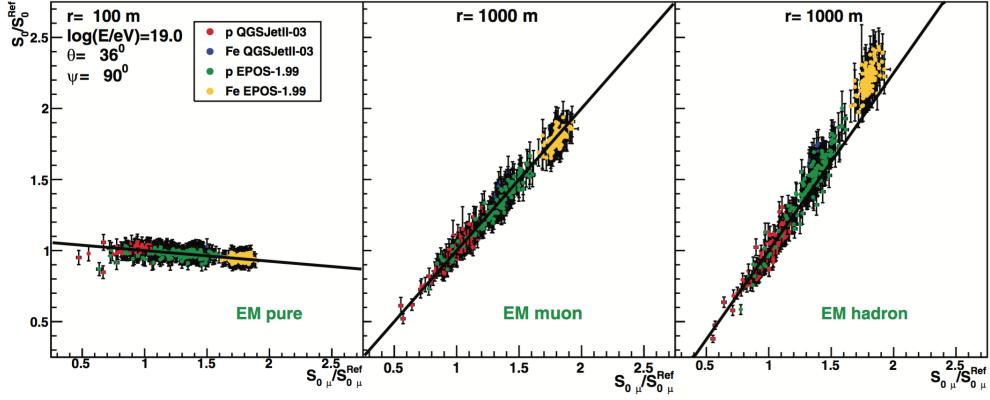


Figure 4.10: Plot of the three electromagnetic components versus the muonic component for the signals  $S_0^i$  predicted in an ideal detector, with respect to the signals for the reference primary and model (proton, QGSJet II.03). The left panel shows the pure e.m. component at  $r = 100$  m, while the middle and right panels show respectively the e.m. from muons and e.m. from low-energy hadrons components at  $r = 1000$  m. To realize these plots, showers with  $\theta = 36^\circ$  and  $E = 10^{19}$  eV have been employed, and only the average signals arising from the sampling areas centered around  $\psi = 90^\circ$  are shown. The different colours indicate different combinations of primary and interaction model. The solid lines indicate the fits accomplished with the expression in eq. 4.11 using only the points (in red) from the reference model and primary. Image from [101].

where  $k = [\text{e.m. pure}, \text{e.m.}(\mu), \text{e.m.(had)}]$  and  $\alpha_k$  depends in principle on the core distance. Also in this case, the fact that the fits provide a good description for different couples of primary/model is a proof of the underling air shower Universality.

Finally, the fluctuation term of eq. 4.8 is parameterized as:

$$f_{\text{fluct}}^i = 1 + (N_\mu - 1) \alpha_i \quad (4.12)$$

where  $\alpha^i$  are the same factors of eq. 4.11. In particular, the values of  $\alpha_{\text{e.m.pure}}$  and  $\alpha_{\text{e.m}(\mu)}$  are fixed respectively to -0.075 and 1 (while  $\alpha_\mu = 1$  by definition). Instead, for the e.m. from hadrons component it is:

$$\alpha_{\text{e.m.}(had)} = 1.25 - 0.13 \exp(-6r/1000 \text{ m})$$

To verify the overall validity of the ground signals parameterization described in this section, the average ratio of the simulated signals to the prediction is analysed, calculating the predicted signal through the Monte Carlo values of energy,  $X_{\max}$  and  $N_\mu$ .

The reference for this test is [103], where the author investigated the accuracy of the signal model using a different library of simulated showers. It consists of simulations with proton, carbon and iron primaries, using QGSJet II.03 and Epos-1.99 as hadronic interaction models. For each mass and interaction model, showers are generated with fixed energy and zenith angle<sup>11</sup>. Moreover, the signals are calculated in an ideal SD array

<sup>11</sup>In particular,  $E = [10^{17}, 10^{17.5}, 10^{18}, 10^{18.5}, 10^{19}, 10^{19.5}, 10^{20}]$  eV and  $\theta = [0, 37, 48, 55, 60]$  degrees (i.e. five zenith angles equidistantly distributed in  $\sec \theta$  for the range  $0^\circ - 60^\circ$ ). To account for the atmospheric variations, 10 showers were simulated for each monthly atmospheric model and for each of the above-mentioned couples of energy and zenith angle.

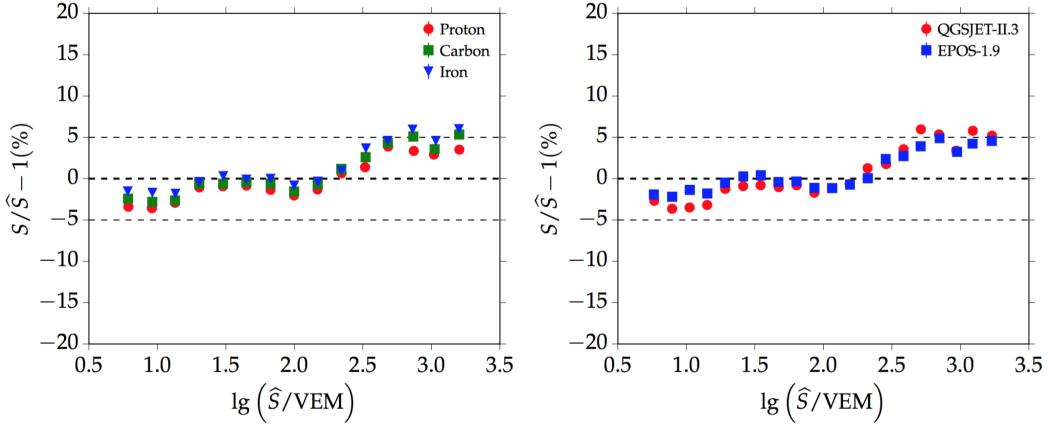


Figure 4.11: Relative signal residuals as a function of the predicted signal  $\hat{S}$  for different shower primaries (left) and hadronic interaction models (right) Images from [103].

with *dense stations* placed in rings at different distances from the core<sup>12</sup>. The dense ring of stations at  $r = 1000$  m is employed to estimate the relative muon content  $N_\mu$  of each simulated shower. Finally, the relative signal residuals  $S/\hat{S} - 1$ , where  $S$  is the simulated signal and  $\hat{S}$  the predicted one, are used for this comparison.

At this point, it is fundamental to address two problems in the description of the WCD integrated signals given by the Universality parameterization. Due to the trigger threshold of each detector of the array, the distribution of residuals is distorted for very low signals (approximately below 10 VEM). In addition the signal in a WCD can be saturated, depending on its time-dependent shape, in the range between 1000 to 2000 VEM. Such signals cannot be directly integrated and thus they do not follow the behaviour encoded in the parameterization. Both types of problematic WCD signals are neglected in the following comparison.

The first check, reported in Fig. 4.11, concerns the residuals as a function of the predicted signal for different primaries and interaction models. As can be seen, there are not significant biases since the differences are always below 5%, giving a good empirical proof of Universality (on average).

However, the analysis of the residuals for different energies and zenith angles (even if again as a function of  $\log(\hat{S})$ ) gives a poor result, as shown in Fig. 4.12. Now a trend emerges more clearly: on average, small signals are apparently overestimated while large signals are underestimated. Moreover, the model is not accurate for energies lower than  $10^{18.5}$  eV: this result was partially expected, considering that only showers with energies above  $10^{18.6}$  eV have been employed in the procedure of parameterization. Also a slightly inaccurate result is obtained for higher zenith angles.

These trends are confirmed analysing the residuals as a function of  $\Delta X$ , as in Fig. 4.13. Again, the parameterization of the signal is not accurate for energies below  $10^{18.5}$  eV and

<sup>12</sup>The rings have radius 200 m, 400 m, 600 m, 800 m, 1000 m, 1222 m, 1494 m, 1826 m, 2232 m and 2728 m centered around the Monte Carlo core. The first 3 rings contain 4 stations each and the rest 8.

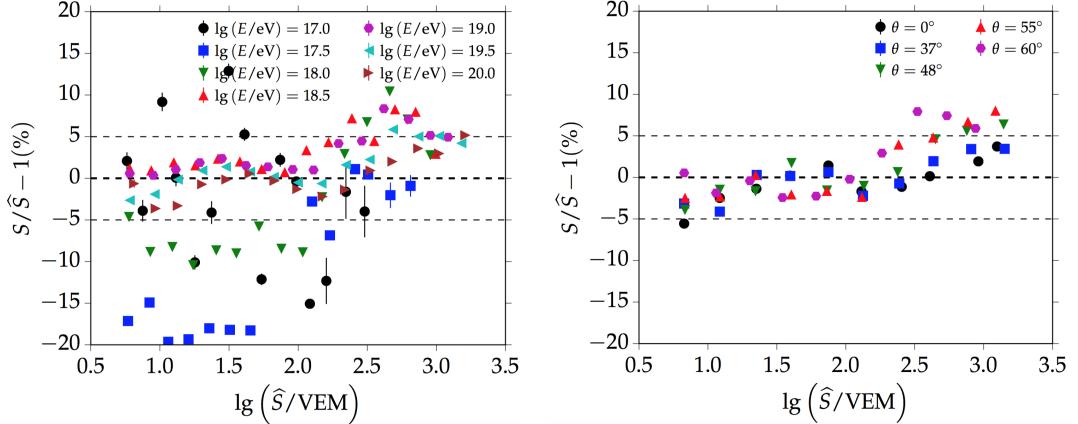


Figure 4.12: Relative signal residuals as a function of the predicted signal  $\hat{S}$  for different primary energies (left) and zenith angles (right). Images from [103].

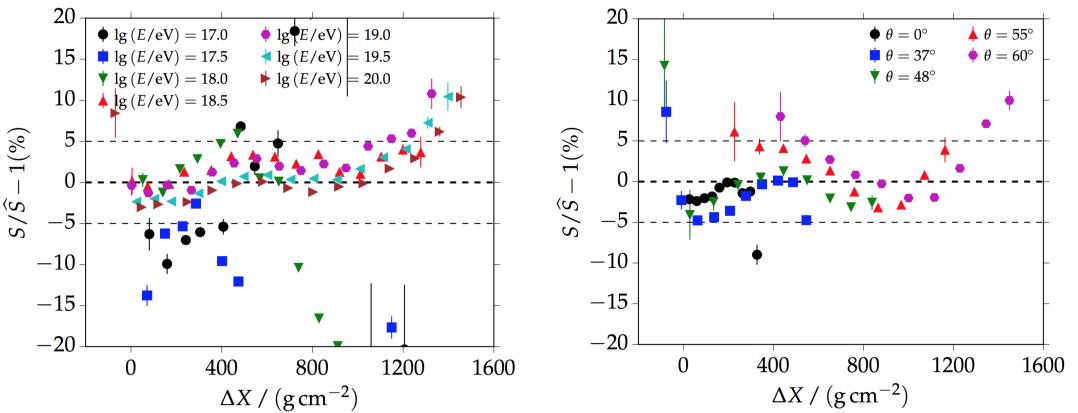


Figure 4.13: Relative signal residuals as a function of  $\Delta X$  for different primary energies (left) and zenith angles (right). Images from [103].

slightly inaccurate for higher zenith angles. Furthermore, the prediction is valid only for values of  $\Delta X > 0$  (i.e. for simulated showers with  $X_{\text{max}}$  above ground) and becomes less and less accurate for values of  $\Delta X$  larger than  $\approx 1000 \text{ g/cm}^2$ .

The last comparison (Fig. 4.14) shows the residuals as a function of the radial distance  $r$ . One can observe that deviations well above 5% are present in signals from stations very close and very far from the core. In effect, a limit of the ground signal parameterization is clearly related with the range core distances used to built it. This is a know inaccuracy of the model, that should be taken into account. In addition, in Fig. 4.14 the inaccuracy of the model for energies below  $10^{18.5} \text{ eV}$  is again evident.

In conclusion, the ground signal parameterization is anyway considered enough robust for primary energies above  $10^{18.5} \text{ eV}$ , showing on average deviations lower than  $\pm 5\%$  (except for extreme values of  $\Delta X$ , signal  $\hat{S}$  and core distance  $r$ )<sup>13</sup>.

---

<sup>13</sup>However, an extension to lower energies is under development, being fundamental for the application of Universality studies to the AMIGA detectors.

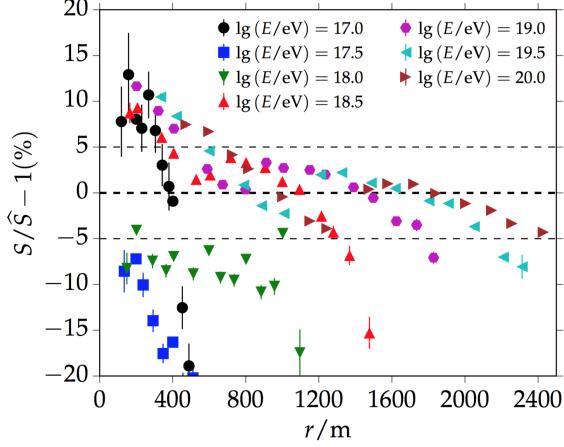


Figure 4.14: Relative signal residuals as a function of core distance  $r$  for different primary energies. Image from [103].

#### 4.4 Universality of the time-dependent signal

A natural extension for the Universality of the integrated signals, as described in the previous section, is a parameterization of the signal shape as a function of time. The initial studies (see for example [104, 105]), accomplished after the definition of the 4-components model, resulted in a first complete description presented in [106], then improved in [107]. Eventually an extended parameterization has been developed in [103]: such version is employed in this work and described in the following.

The final aim of the Universality parameterizations is to build a method for the reconstruction of EASs exploiting the footprints measured by the surface detector of the Pierre Auger Observatory<sup>14</sup> through a fit of the signals and of their time shape.

The model was developed making an ansatz directly for the signal time structure observed in data or obtained from complete simulations that already take into account also the detector response. In particular, separately for each shower component, the integral of the signal is obtained from the previous Universality parameterization while the shape as a function of time will be described with a **Log-Normal distribution**:

$$\frac{dS_i}{dt}(t) = \frac{1}{\sqrt{2\pi}(t - t_{0,i})s_i} \exp \left[ -\frac{(\ln(t - t_{0,i}) - m_i)^2}{2s_i^2} \right] \quad (4.13)$$

where  $i$  denotes the component,  $m$  and  $s$  are the *location* and *shape* parameters (proportional but not equal to the first two central moments of the distribution) and  $t_0$  is a time below which no particles are expected. In the following, the time dependent signal measured or simulated in a WCD will be also called *trace*.

The traces are simulated in bins of 25 ns width (equal to the actual integration time of the WCDs), including saturation effects and a signal background. Instead, the

---

<sup>14</sup>Moreover, an extension of the Universality model to take into account the new detectors to be added within the Observatory upgrade, described in sect. 3.4, has been foreseen during the realization of these parameterizations.

simulated traces for each specific Universality components do not present saturation and background. Finally, to avoid distortions due to trigger thresholds, only traces with an expected total signal above 4 VEM (according with the ground signal Universality model) are kept for the following studies.

The time structure of the signals depends on the primary and shower characteristics and on the detector position; these dependences will be encoded in the values of  $m$ ,  $s$  and  $t_0$ . Suitable parameterizations of the Log-Normal parameters are built employing the extended library of simulated showers described in the previous section, also used in [103] to test the accuracy of the predicted ground signals.

Initially, a model for the origin of times as a function of  $(E, X_{\max}, \theta)$  has been studied. Then, using the predicted  $t_0$ , the values of  $(m, s)$  can be obtained and parameterized as functions of  $(r, \theta, \psi, E, \Delta X)$ . To better understand the required steps, a preliminary digression about the simulation of signal in a detector is necessary.

As previously mentioned, the un-thinning procedure is employed to obtain the traces (called *resampled* traces, Fig. 4.16) that simulate the result of a single measurement. In this procedure only a small number of shower particles is followed until ground level, assigning them a weight larger than one (a more detailed explanation about the thinned shower propagation is given in [102]). In effect the simulated (thinned) particles that reach a certain sampling area<sup>15</sup> are considered representatives for particles of the same type, energy and direction, that would reach the corresponding region of the shower plane in a complete simulation (without thinning) and in a quantity proportional to the final weights, as shown in Fig. 4.15. Then the number of particles that actually enter in a WCD for a single simulated event and for each (thinned) particle is extracted according with a Poisson distribution with mean dependent on the thinned particle weight and on the ratio between the detector surface the sampling area. Instead, in the development of the Universality model, *average* signals have been employed. Such traces, realized with a different procedure [107], can be considered equivalent to the average of many measurements of the same shower in the same position.

The first step in the development of the time model is a parameterization of the origin of times, to predict for each detector a start time of the corresponding trace. With this aim, it was introduced a simplified model to describe the shower development, based on the approximation of a spherical shower front propagating at the speed of light from a certain point (corresponding to the origin of times) high in the atmosphere. In this picture, a convenient quantity to be parameterized is  $D_{TO}$ , defined as the distance from the origin of times to  $X_{\max}$ , calculated along the shower axis in kilometers and different for each shower component. In fact the first interaction point  $X_{\text{first}}$  is not suitable in this description, as a sufficient lateral spread in the shower particles (large enough to have

---

<sup>15</sup> As exposed in the previous section, a sampling area is a slice of a ring in the shower plane surrounding a detector. For instance, a sampling area centered at 800 m and  $\phi = 90^\circ$  includes the surface between  $r \approx 713$  m and  $r \approx 897$  m from the shower core and with  $\psi \in [60^\circ, 120^\circ]$ .

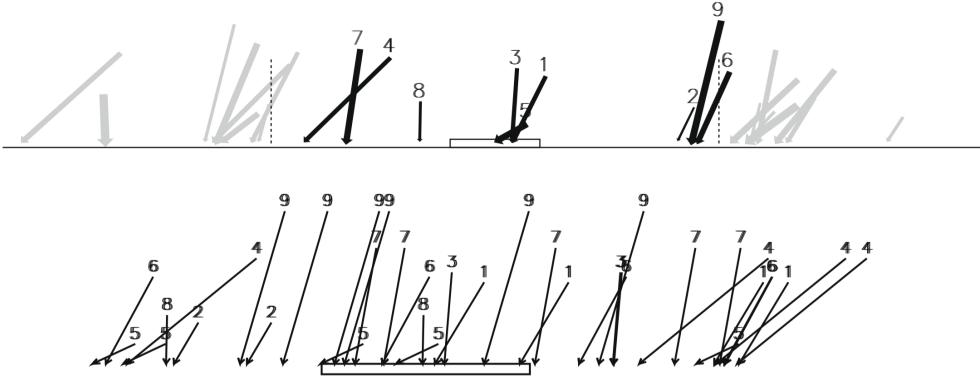


Figure 4.15: Schematic view of the resampling procedure. In the top panel, the thinned particles are represented as arrows, with length proportional to energy and width proportional to weight. Only the thinned particles arriving inside the sampling region (delimited by the dashed lines) around the selected detector (horizontal rectangle) are used. In the bottom panel, the selected thinned particles are converted into a flux of particles with the same direction and energy, with an intensity proportional to the weight, falling on the ground in the sampling region at positions generated uniformly and independently. This resampling gives a fair representation of the particles hitting a detector. It is also equivalent to a Poissonian extraction of a number of particles considered to directly reach the detector, with mean equal to the *resampled weight*, defined as the thinned particle weight times the ratio between the WCD surface and the total area of the sampling region. Image from [102].

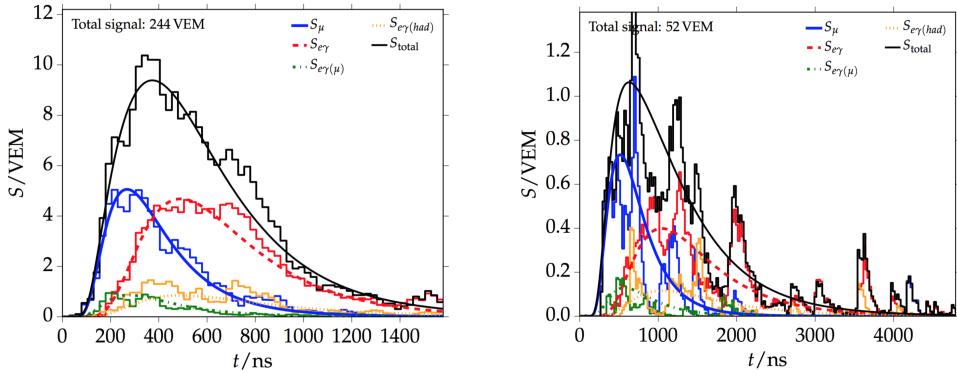


Figure 4.16: Resampled traces for a proton initiated shower ( $E = 10^{19.5}$  eV,  $\theta = 36^\circ$ ), simulated in a detector nearer (left panel) and one further (right panel) from the shower core, using the un-thinning procedure (Fig. 4.15). In black the total signal, sum of the 4 Universality component represented with different colours. The correspondent lines are the best fit results for the reconstruction of the event, obtained with a procedure exposed in sect. 5.3. Image from [103].

a proper shower front) is reached only after the propagation for 100-200 g/cm<sup>2</sup> deeper than  $X_{\text{first}}$ . The optimal point to describe the origin of time in this simplified scenario is somewhere between  $X_{\text{first}}$  and  $X_{\text{max}}$ , and it is expressed by the quantity  $D_{TO}$ .

The parameterization of  $D_{TO}$  is performed (separately for each Universality component) using only the average trace in the sampling area centered at  $r = 800$  m and  $\phi = 90^\circ$  for each simulated shower. Initially the 1%, 10% and 50% time quantiles, respectively indicated as  $t_1$ ,  $t_{10}$  and  $t_{50}$ , are calculated; then the Log-Normal distribution of eq. 4.13

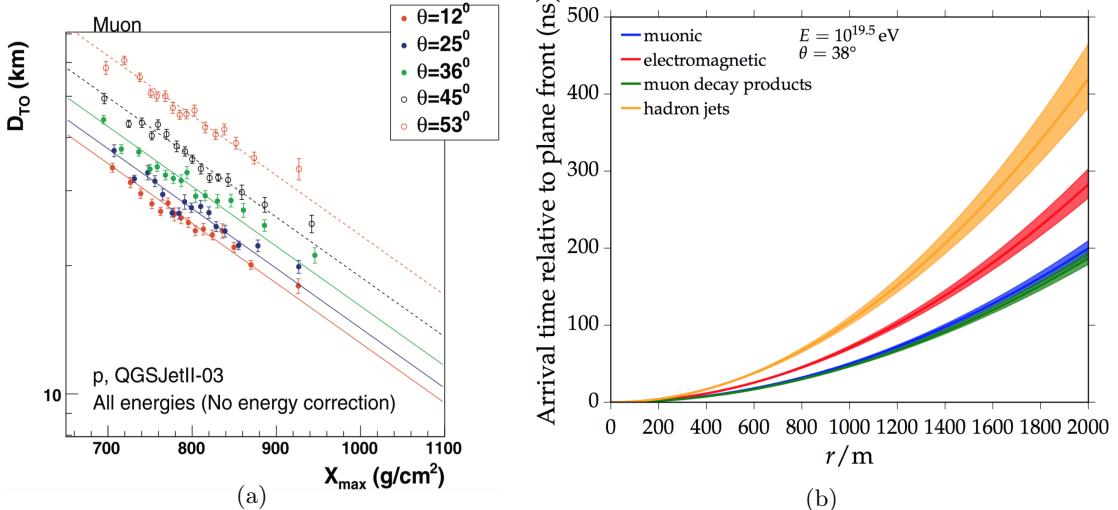


Figure 4.17: (a) Mean  $D_{TO}$  for the muonic component as a function of  $X_{max}$  for different zenith angles. Showers of all energies are included. Image from [107]. (b) The start time  $t_0$  as a function of the core distance, for a fixed energy and zenith angle (as indicated in the plot). The values are relative to the arrival time of the shower front, travelling at the speed of light from the point identified by the parameterized  $D_{TO}$ , at the core position. The results for the 4 shower components are shown with different colours. Image from [103].

that better reproduces this set of values is obtained through a fit<sup>16</sup>. Each value of  $t_0$ , which is associated to one simulated shower and one component, is used to calculate the corresponding  $D_{TO}$ . Finally, these distances are exploited in the parameterization of a mean value of the origin of times as a function of the shower parameters. The ansatz, as reported in [107], presents 3 parameters for each shower component and two different functional forms, in particular:

$$D_{TO}(\theta, E, X_{max}) = \left( D_0(\theta) + D_1 \log_{10} \frac{E}{10^{19} \text{ eV}} \right) \exp \left[ -\frac{X_{max} - 750 \text{ g/cm}^2}{D_2} \right] \quad (4.14)$$

for the muonic and electromagnetic-from-muon components, and

$$D_{TO}(\theta, E, X_{max}) = D_0(\theta) + D_1 \log_{10} \frac{E}{10^{19} \text{ eV}} + D_2 \frac{X_{max} - 750 \text{ g/cm}^2}{200 \text{ g/cm}^2} \quad (4.15)$$

for the pure and electromagnetic-from-hadron components.

The difference among the 4 shower components originates in the different propagation of electromagnetic particles with respect to muonic ones. In fact, after being produced, muons reach the ground almost without interactions, while e.m. particles are scattered several tens of times before entering in a detector.

The result of the parameterization is shown in Fig. 4.17a for the muonic component. It is important to stress that only proton-initiated showers simulated using QGS-Jet II.03 have been used. The start times  $t_0$  at each distance  $r$  from the shower core are obtained

<sup>16</sup>In effect this is a pseudo-fit, since for each average trace there is one set of  $(t_0, m, s)$  that gives exactly the simulated quantiles  $(t_1, t_{10}, t_{50})$ .

from the parameterized value of  $D_{TO}$  using the above-described simplified model for the shower front propagation. An example is given in Fig. 4.17b, where the difference in the start times of the 4 Universality component is evident, and can be clearly related with the shift in the arrival times of the particles that constitute the 4 components, due to the different processes and interaction occurring during the propagation.

In the next step, the parameterization of the Log-Normal  $(m, s)$  values as functions of the shower quantities is accomplished through a proper (chi-squared) fit of the simulated time-dependent signals for each component. To reduce statistical and thinning fluctuations, a weighted average of traces within each  $(r, \psi, \theta, E, \Delta X)$  bin<sup>17</sup> is preliminarily calculated. Since the traces are not in phase, they are firstly resampled to occupy the same set of time bins (each one of 25 ns width) and then normalized, as only the shape is of interest for this parameterization. At this point the weighted average over the normalized traces in each  $(r, \psi, \theta, E, \Delta X)$  bin is calculated, separately for each shower component, in every time bin. The uncertainty is estimated as the standard deviation of the mean of all normalized signals within each time bin.

The traces resulting from such average and normalization are finally fitted with a Log-Normal distribution. In this way, for every  $(r, \psi, \theta, E, \Delta X)$  set a couple  $(m, s)$  is obtained. Since the fluctuations among traces grow at large times, only bins below the 95% quantile of the average normalized trace are used in the minimization. An example of this fitting procedure is shown in Fig. 4.18.

The parameterizations of the Log-Normal parameters are accomplished in two steps. Initially a fit of the  $m$  and  $s$  values is performed separately in each interval of radial distance, with an empirical (but physically motivated) functional form

$$f_{m,s} = f_{\Delta X}(\Delta X) + f_{\text{geo}}(\theta, \psi, \Delta X) + f_{\lg E}(E, \Delta X) \quad (4.16)$$

where:

$$f_{\Delta X}(\Delta X) = a_{\Delta X} + \Delta X_{\text{ref}}[b_{\Delta X} + \Delta X_{\text{ref}}(c_{\Delta X} + d_{\Delta X}\Delta X_{\text{ref}})] \quad (4.17)$$

$$f_{\text{geo}}(\theta, \psi, \Delta X) = \sin \theta (a_{\text{geo}} \cos \psi + a_{\text{geo}} \Delta X_{\text{ref}}) \quad (4.18)$$

$$f_{\lg E}(E, \Delta X) = \lg E_{\text{ref}}(a_{\lg E} + b_{\lg E} \Delta X_{\text{ref}}) \quad (4.19)$$

with  $\Delta X_{\text{ref}} = \Delta X/(750 \text{ g/cm}^2)$  and  $\lg E_{\text{ref}} = \lg(E/10^{19} \text{ eV})$ . The first function ( $f_{\Delta X}$ ), which is also the leading one, highlights how the time structure of the signal in a WCD (mainly in its initial part, as recently studied in [47]) is strictly correlated with the position of the maximum in the shower development. The term  $f_{\text{geo}}$  expresses the effect

---

<sup>17</sup>The binning in core distance  $r$ , azimuth angle  $\psi$ , zenith angle  $\theta$  and primary energy  $E$  corresponds to the fixed values of the Monte Carlo shower library previously described (see also Appendix E.1 of [103]). Instead, the distribution of the distances from shower maximum to detector  $\Delta X$  substantially changes for every  $(r, \psi, \theta, E)$  set, then a compromise between a uniform bin width and a reasonable number of entries has to be found in each case. The method used in [103] was to choose  $\Delta X$  intervals that include at least 30 and less than 60 traces.

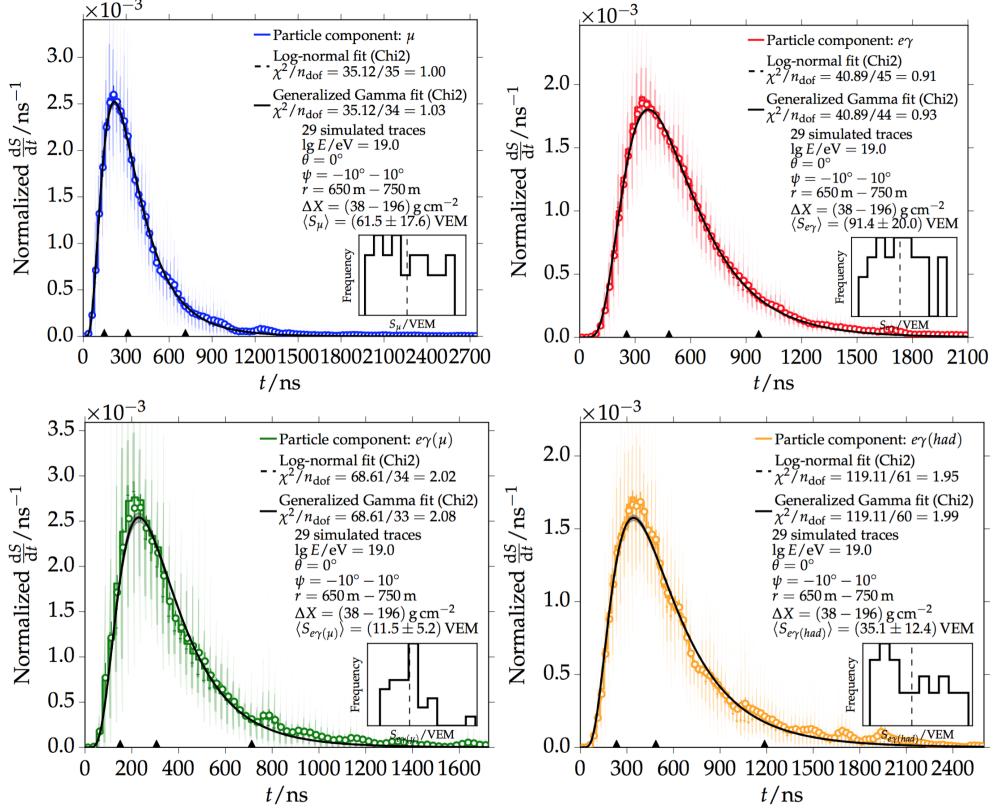


Figure 4.18: Average normalized time dependent signals for showers with an energy of  $10^{19}$  eV and zenith angle of  $0^\circ$  initiated by proton, carbon and iron primaries. The 29 traces employed in these plots are simulated in the sampling areas centered at  $r = 700$  m and  $\psi = 0^\circ$ , and the correspondent detectors present  $\Delta X \in (38, 196)$  g/cm $^2$ . In blue (top left panel), the muonic component; in red (top right) the pure electromagnetic component; in green (bottom left) the e.m. component from muon decay; in orange (bottom right) the e.m. from low-energy-hadron-jets component. The darker error bars are the  $1\sigma$  uncertainty, while the more transparent ones represent the  $3\sigma$  region. The black triangles at the bottom of each plot represent the 10%, 50% and 90% time quantiles. The inner plots illustrate the distributions of the signal intensity separately for each component. The black lines show the Log-Normals obtained from the minimization. Actually also a fit with a Gamma distribution, that present an additional parameter  $l$  (but reduces to a Log-Normal for  $l = 0$ ) is attempted. Since the difference between the two results is very small and the uncertainties on the third parameter  $l$  are large, the Log-Normal description is preferred. Images from [103].

of azimuthal asymmetry [108] (see also Appendix B.1), while the last term ( $f_{\lg E}$ ) encodes potential differences directly related with the primary energy.

In a second moment, the values of each parameter in these empirical functions is fitted with an expression function of  $r$ , in particular a polynomial in the form:

$$f_r(r) = \sum_{i=0}^n a_i \left( \frac{r}{1000 \text{ m}} \right)^i \quad (4.20)$$

where the order  $n$  of the polynomial can change for each different parameters.

However, many of the 64 parameters implied by eq. 4.16 (8 for  $f_m$  and 8 for  $f_s$ , separately for each shower component) can be assumed equal to zero, both because of

physical arguments and/or their negligible effect on the parameterization accuracy. For the location parameter functions  $f_m$ , the parameters actually employed are:

$$\begin{aligned}\mu &: a_{\Delta X}, b_{\Delta X}, c_{\Delta X}, a_{\text{geo}} \\ \text{e.m.(pure)} &: a_{\Delta X}, b_{\Delta X}, c_{\Delta X}, d_{\Delta X}, a_{\text{geo}}, a_{\lg E} \\ \text{e.m.}(\mu) &: a_{\Delta X}, b_{\Delta X}, c_{\Delta X}, a_{\text{geo}} \\ \text{e.m.(hadr)} &: a_{\Delta X}, b_{\Delta X}, a_{\text{geo}}, a_{\lg E}\end{aligned}$$

while for the shape parameter functions  $f_s$ :

$$\begin{aligned}\mu &: a_{\Delta X}, b_{\Delta X}, c_{\Delta X}, a_{\text{geo}}, a_{\lg E} \\ \text{e.m.(pure)} &: a_{\Delta X}, b_{\Delta X}, c_{\Delta X}, d_{\Delta X}, a_{\text{geo}}, b_{\text{geo}}, a_{\lg E} \\ \text{e.m.}(\mu) &: a_{\Delta X}, b_{\Delta X}, a_{\text{geo}}, b_{\text{geo}}, a_{\lg E} \\ \text{e.m.(hadr)} &: a_{\Delta X}, b_{\Delta X}, a_{\text{geo}}, b_{\text{geo}}, a_{\lg E}\end{aligned}$$

Only these 42 parameters have been actually fitted as functions of the core distance  $r$ .

The final results is a complete description of the location and shape parameters  $m = m(r, \psi, \theta, E, \Delta X)$  and  $s = s(r, \psi, \theta, E, \Delta X)$  that, together with the parameterization of the origin of times  $D_{TO}$ , closes the model for the time structure of the ground signals. To test the accuracy in the predicted values, the relative residuals  $(m - \hat{m})/\hat{m}$  and  $(s - \hat{s})/\hat{s}$  have been studied as functions of the shower and detector parameters, where  $m$  and  $s$  are the results of the average traces fit in each  $(r, \psi, \theta, E, \Delta X)$  bin (as in Fig. 4.18) while  $\hat{m}$  and  $\hat{s}$  are the model predictions.

In Fig. 4.19 the residuals for the location parameter  $m$  are shown as functions of  $\Delta X$  for the 4 shower components<sup>18</sup>: as can be noticed, the biases are lower than 1% in the vast majority of cases. Non trivial trends are present in the muonic components ( $\mu$  and e.m.( $\mu$ )) for small zenith angles and  $\Delta X$  values, that correspond to a shower maximum approaching ground level. The cause of these minor shifts, that anyway are inside the  $\pm 1\%$  interval, is unclear<sup>19</sup>. At the same time, the model accuracy in the pure electromagnetic component predictions slightly decreases for high values of  $\Delta X$ ,  $r$  and  $\theta$ . Such biases are likely related with the growing attenuation of the e.m. signal due to the particles absorption in the atmosphere. Since smaller signals are subjected to larger fluctuations, the parameterizations become inevitably less accurate. The same reasoning can be applied to the e.m. from hadronic jets component, that shows the larger biases and uncertainties. In effect, this is also the component with the lower signal magnitude, and therefore with the most difficult parameterization.

The residuals for the shape parameter  $s$  (not shown here, see [103]) present the same behaviours of Fig. 4.19 respectively for each shower component. However their magnitude

<sup>18</sup> Several other plots of the residuals, as functions of different quantities, can be found in [103].

<sup>19</sup> A hypothesis is related with the choice of parameterizing also the muonic components as functions of the distance (in g/cm<sup>2</sup>) from the e.m. maximum of the shower ( $X_{\max}$ ) instead of the muonic maximum ( $X_{\max}^\mu$ ) that was not available in the employed library of simulated showers. Another possible explanation is that, similarly to what was observed in the total ground signals parameterization, these shifts are due to the exclusion of events with  $X_{\max}$  below ground.

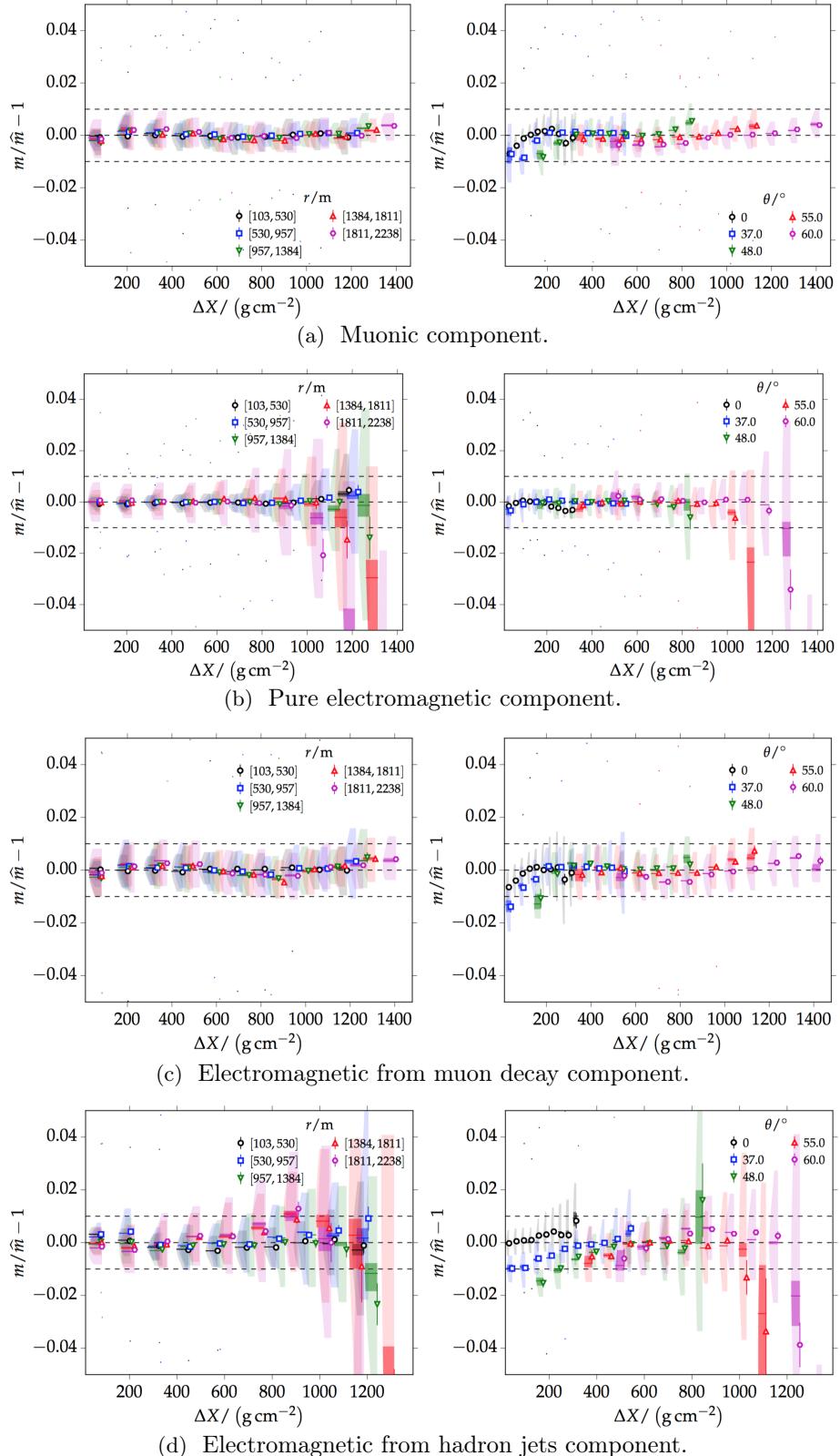


Figure 4.19: Relative residuals for the location parameter  $m$  as functions of the depth between the shower maximum and the detector ( $\Delta X$ ) for different core distances  $r$  (left panels) and zenith angles (right panels). Showers of all energies and primaries are included. The dashed lines correspond to  $\pm 1\%$ , while the background *violin* plots give additional information about the distributions within each bin. Moreover, the small dots above and below each violin distribution represent values beyond  $3\sigma$  from the average in the correspondent bin. Images from [103].

is generally larger, with biases (below 5% on average) that can reach 15% near the minimum and maximum values of the shower and detector quantities employed in the parameterizations. Moreover, the e.m. from muons and e.m. from hadrons components, which present the smaller signal magnitudes, still have the less accurate predictions.

Another necessary test concerns the comparison among the residuals for different primary masses and hadronic interaction models. In [103], the presence of systematic shifts related with these characteristics of the simulations has been investigated studying the difference between the measured (in simulation) and predicted 50% time quantiles ( $t_{50}$ ) as a function of the core distance, separately for each primary (H, C, Fe) and hadronic model (QGSJet-II.03, Epos-1.99). Such analysis also allows to better quantify the overall accuracy of the time model.

In Fig. 4.20a the muonic component residuals ( $t_{50} - \hat{t}_{50}$ ) are reported. The deviations between model and simulation are generally small (between 5 and 10 ns) as well as the shifts between the two interaction models, while among primaries slightly more evident biases are present. This outcome is probably caused by differences in the signal time structure for distinct primary species, which give rise to unaccounted effects also in the Log-Normal parameters ( $m, s$ ) and thus in the predicted quantiles.

In Fig. 4.20b the same analysis is presented for the pure electromagnetic component. In this case, the time residuals clearly increase above  $r \sim 1000$  m, independently from the

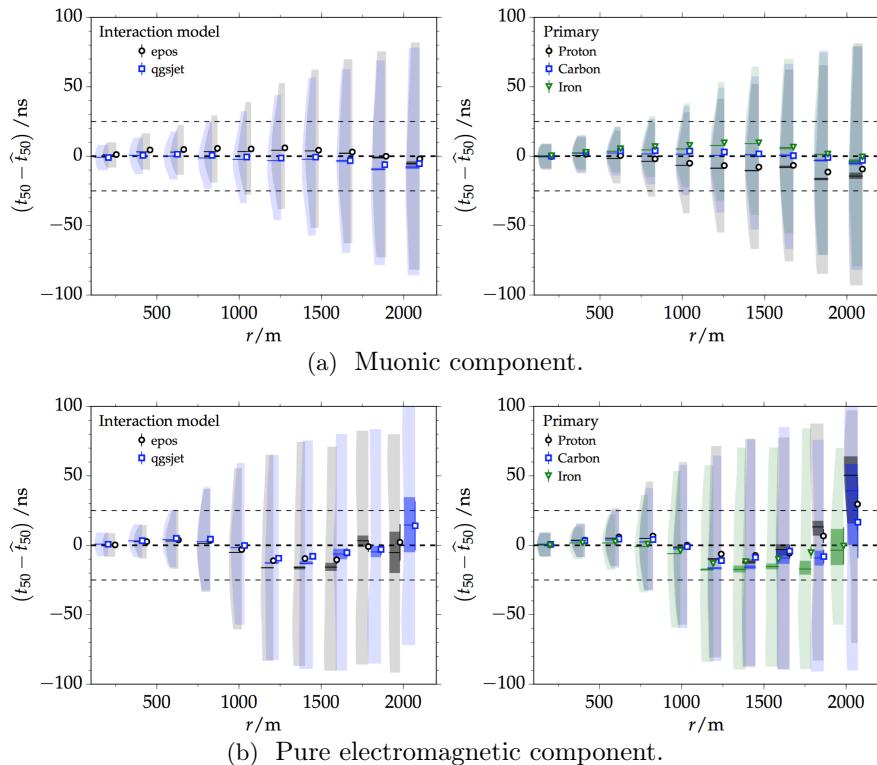


Figure 4.20: Residuals for the 50% time quantile as functions of the core distance  $r$  for different hadronic interaction models (left panels) and primary species (right panels). The dashed lines correspond to  $\pm 25$  ns. Images from [103].

primary or hadronic interaction model. These shifts are mainly the result of inaccuracies in the parameterization for low signals, since the e.m. part of the shower rapidly attenuates at large radial distances. However, the impact on the total trace is substantially reduced because the electromagnetic component constitutes a large fraction of the total signal (comparable with the muonic one) only near the shower core.

The differences in  $t_{10}$  (the 10% quantile) present a similar behaviour but smaller magnitudes, confirming that the early part of the traces (also called *rising* part) is well predicted. Instead, the time model is not able to reproduce accurately the late part of the traces, since the 90% quantile presents much larger residuals.

Nevertheless, the positive validations reported in the last part of this section assure a more than sufficient accuracy for the parameterization of the time-dependent signals in the WCDs. Together with the model for ground signals, exposed in sect. 4.3, the time model allows the development of a powerful method of EAS reconstruction, which can exploit not only the integrated signals but also the measured traces to obtain the most possible information from the surface detector measurements of the Pierre Auger Observatory. Such new reconstruction procedure (and its improvement) is the subject of this thesis and it is introduced in the next chapter.

## Chapter 5

# The Universality reconstruction

In the framework of the Pierre Auger Collaboration, the composition studies are accomplished through the analysis of the  $X_{\max}$  distribution<sup>1</sup> as a function of the energy, a physical quantity measured with good quality in hybrid events, i.e. showers detected by both fluorescence (FD) and surface (SD) detectors. However, the duty cycle for the fluorescence telescopes is about 15% and reduces to only few % because of the necessary quality cuts. So the available number of events is insufficient to extend our knowledge about the mass of primaries in the highest energy range (namely above  $10^{19}$  eV).

Many efforts have been put in the development of a composition analysis which makes use exclusively of data recorded with the surface detector of the Pierre Auger Observatory, operational nearly 100% of the time. Nonetheless this detector only samples the shower at ground level so the determination of  $X_{\max}$ , which is characteristic of the shower development in atmosphere, is noticeably difficult to be accomplished. The strategy is to search for correlations among primary mass and observables actually measurable with the surface detector, even if an extensive use of simulations is required. As a consequence composition studies based on SD-only data are heavily affected by systematics, arising from both measurements and simulations (e.g. due to uncertainties in the hadronic interaction models), and so far incapable of conclusive physical predictions.

To overcome these limits, a different approach is desirable. For this reason, a new reconstruction procedure has been developed inside the Pierre Auger Collaboration, based on the paradigm of Universality in the shower development. The aim is to exploit surface detector data not only to extract observables that are considered correlated with the primary mass, but to accomplish a complete characterization of each EAS. The main innovation of the so-called Universality reconstruction is to make use of the measured traces, i.e. the time-dependent signals acquired by water-Cherenkov detectors, since they hold information about the longitudinal development of the cascade and so about  $X_{\max}$  and the primary mass in general. In the first section, the characteristics of the Universality reconstruction are reported.

---

<sup>1</sup>In fact, the first two moments can offer an overall behaviour of  $X_{\max}$  with energy, but the actual composition studies require the exploitation of complete distributions.

This fitting procedure is still under development, therefore it should be proven over a set of events which are already completely known by means of independent and well-established methods. In this work, such high quality dataset was identified with the golden hybrids, described in the second section. The original version of the Universality procedure was then performed over this selected set of golden hybrids, reconstructing the events from their SD-only measurements. The third section deals with the comparison between the results of the standard reconstruction and of the original (and presently available) Universality fit. However, this new reconstruction shows low accuracy and needs further development to produce physics outcomes beyond the present knowledge.

Since the Universality parameterizations (both of the total signal and of the traces) will not be modified, the Universality procedure has been adjusted with the aim of obtaining outcomes more similar to the standard results for golden hybrids and at the same time provide a goodness-of-fit indicator for the eventual selection of events reconstructed with good quality. These improvements are analysed starting from the fourth section until the end of this chapter.

The final goal is to extend the reconstruction to the entire set of events detected by the surface detector. The accuracy of the final result will depend on the performances of the Universality fit, evaluated in this chapter through the reconstruction of the selected set of golden hybrids. Also for this reason, in the following the analysis will be almost always focused on the higher energetic range, i.e. above  $10^{19}$  eV, where the FD cannot reach sufficient statistics due to its limited duty cycle and the Universality reconstruction would possibly represent a true advancement.

## 5.1 The Universality fitting procedure

The Universality reconstruction employed in this work consists in a search for the set of macroscopic shower parameters ( $E, X_{\max}, N_{\mu}, \theta, \phi, \vec{x}_{\text{core}}, t_{\text{core}}$ ) which better reproduce the data gathered by the surface detector for an event. Such procedure is accomplished through a fit of the integrated signal and of the trace measured by each WCD participating into the event, using the Universality parameterizations introduced in the previous chapter. They offer a robust backbone for the shower reconstruction, as demonstrated through tests on MC events [101, 107] simulated from CORSIKA showers on the Pierre Auger Observatory surface detector array with spacing 1500 m (SD-1500).

Two reconstruction approaches have been implemented in the `Offline` framework of the Pierre Auger Collaboration. In the first one [109, 110] (*identified as Bariloche reconstruction*) the fit is performed using selected quantiles of the time-dependent signals; the second approach [106, 103, 111] (*Karlsruhe reconstruction*) is instead characterized by a direct, bin-to-bin fit of the traces.

In this work, exclusively the *Karlsruhe* method has been employed. For this recon-

struction<sup>2</sup>, two different procedures have been implemented to carry on the fit:

- an **iterative** reconstruction, consisting of 8 steps; in each step some physical quantities are fixed or constrained (to the standard reconstruction values or to the results of a previous step) and the remaining ones are fitted;
- a **global** reconstruction, where the energy is fixed to the standard SD reconstruction value while all the other quantities are fitted at the same time.

The fit is performed after a complete and successful standard SD reconstruction (see sect. 3.1.1) of each event, which gives the initial values of the parameters. In both cases (iterative and global), the two components of the Universality minimization are:

- the **LDF fit**, where the station integrated signals are fitted using the sum of 4 LDF, one for each shower component;
- the **shape fit**, where the traces are directly fitted with the sum of 4 parameterized log-normal distributions, one for each shower component.

The actual fitting procedure requires the definition of suitable likelihoods for this two terms, which exploit the information hold by SD measurement through likelihood minimization. Likelihood terms depending on physical parameters are not considered in the original version of the Universality fit, instead prescribing intervals of values they can assume during the fitting procedure:

- zenith  $\theta \in [\theta_{\text{SD}} - 10^\circ, \theta_{\text{SD}} + 10^\circ]$ ; if  $\theta < 10^\circ$  the lower limit becomes  $0^\circ$ ;
- azimuth  $\phi \in [\phi_{\text{SD}} - 10^\circ, \phi_{\text{SD}} + 10^\circ]$ ;
- muonic content  $N_\mu \in [0, 3.5]$ ;
- $X_{\text{max}} \in [200, 2000] \text{ g/cm}^2$ ;
- core time  $t_{\text{core}} \in [t_{\text{core}}^{\text{SD}} - 10^6, t_{\text{core}}^{\text{SD}} + 10^6] \text{ ns}$ ;
- no hard limits for the core position.

A first step toward the study of performances of this procedure concerns the choice of the events to be used. Since the Universality reconstruction is accomplished fitting the traces and the integrated signals from the surface detector, it is necessary in any case to select events with a good quality in the SD-only measurements.

Furthermore, the accuracy in the determination of  $X_{\text{max}}$  (together with the other shower parameters) has to be studied with respect to a reference reconstruction. In this work, instead of a comparison between the measured  $X_{\text{max}}$  distributions and the ones obtained through the Universality fit, it was decided to perform a comparison on an event-by-event basis. Such strategy can be applied only on simulations or hybrid events: the former ones provide the true values, while the latter ones allow a complete reconstruction of the shower longitudinal development through the hybrid reconstruction (described in sect. 3.2.1), that requires the measurement of a shower by the fluorescence detector and at least one SD station.

---

<sup>2</sup>In particular, this reconstruction procedure is tagged as “*UnivFitterKG*” (short for *Karlsruhe Group*) inside the “*UniversalityFitter*” module of the *Offline* framework. In this work it was used the latest stable version (v3r3) of the *Offline* software and the related revision (30790) of the “*UnivFitterKG*” code.

To fulfill both the above-mentioned conditions, the analysis will be focused on the Universality fit of **golden hybrid events**, which present both standard SD and hybrid successful reconstructions (see also sect. 3.2.2).

In addition to the standard quality selection cuts, only **not saturated** events with  $E_{\text{SD}} > 10^{18.5}$  eV will be reconstructed. A saturated trace provides much less information to the minimization, since the time bins with saturated signals cannot be used in a likelihood calculation for the shape component and the total signal cannot be directly used in the LDF fit. Moreover, in a saturated event (i.e. an event with at least one saturated trace) the “hottest” WCD is very near to the core while the closer tanks are more or less at the same core distance and thus present approximately the same signal. This “degeneracy” weakens any reconstruction method based on the sampling technique, which is the paradigm of the SD data taking. As a consequence, the Universality fit in these cases is generally less accurate.

The method used in this work is to reconstruct unsaturated golden hybrids treating them as SD only events: in this way it is possible to perform an “*a posteriori*” test studying the differences between the reference hybrid reconstruction and the Universality one. In particular, such “*a posteriori*” analysis will focus on the events with an initial energy from the SD reconstruction  $E_{\text{SD}} > 10^{19}$  eV, the range of true interest for the advancement of mass composition studies.

## 5.2 The selected set of golden hybrid events

As mentioned in the previous section, a high quality set of golden hybrid events will be used to evaluate the Universality fit performances. Such dataset is selected with the same cuts used for the energy calibration of the vertical (zenith angle  $\theta < 60^\circ$ ) events used for the ICRC-2015, thus observed in the period 2004-2013. Instead, the events employed for mass composition studies by the Pierre Auger Collaboration are selected from the hybrid dataset. The two datasets are naturally different, as not every hybrid event is a golden hybrid<sup>3</sup>; a comparison is therefore necessary and will be addressed in this section.

The first step is to construct the  $X_{\text{max}}$  distributions for each interval wide 0.1 in the logarithm of the reconstructed energy<sup>4</sup>, separately for the golden events and the hybrids used in the  $X_{\text{max}}$  analysis. For the latter selection, the reference is the composition paper [53] published in 2014: in the following, the events presented as “PRD hybrids” correspond to the dataset used in this paper.

In Fig. 5.1 the  $X_{\text{max}}$  distributions of golden hybrids and PRD-hybrids in the higher

---

<sup>3</sup>As the energy increases, one would expect that the two dataset become more and more similar, as a larger number of stations is hit by each event. However, also in the high energy range the golden hybrid events are less than 2/3 of the simple hybrid ones because of the more severe quality cuts applied to obtain the golden dataset.

<sup>4</sup>In fact the overall composition of cosmic rays is not supposed to change for small energy differences, and so the distribution of  $X_{\text{max}}$  can be related with the composition in each energy bin. See sect. 2.3.1 for more details.

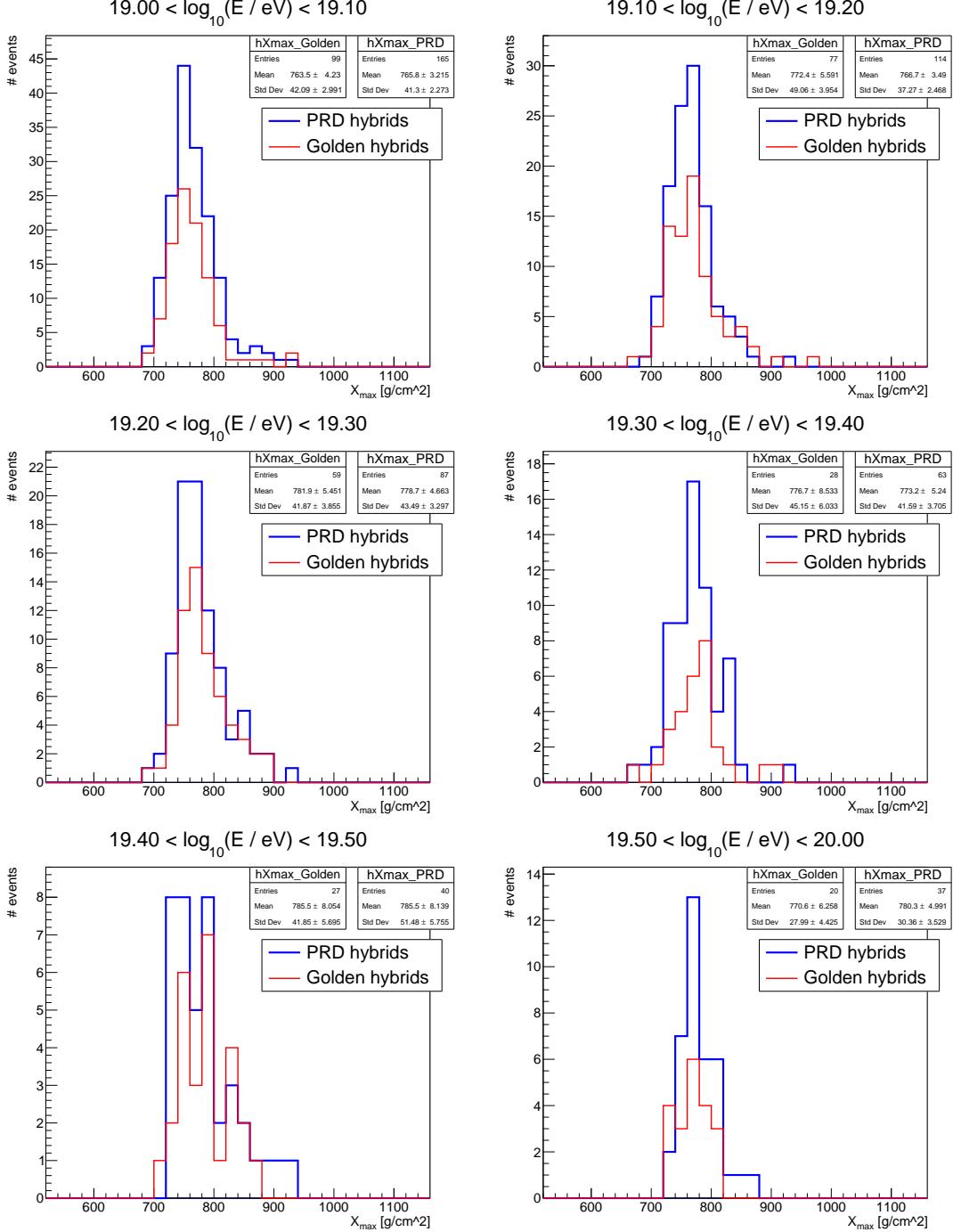


Figure 5.1: The  $X_{\max}$  distributions for the two set of selected hybrid events: in blue, the dataset used in the composition analysis [53]; in red, the golden hybrids selected for the energy calibration of the surface detector performed for ICRC2015.

energy range (above  $10^{19}$  eV) are shown. At first glance the agreement is enough good since the shape of the distributions, compared separately in each energy bin, are compatible despite the low number of events; such similarity naturally improves for lower energies, that is for larger statistics. Moreover, each couple of histograms presents values of mean and standard deviation which are compatible within uncertainties.

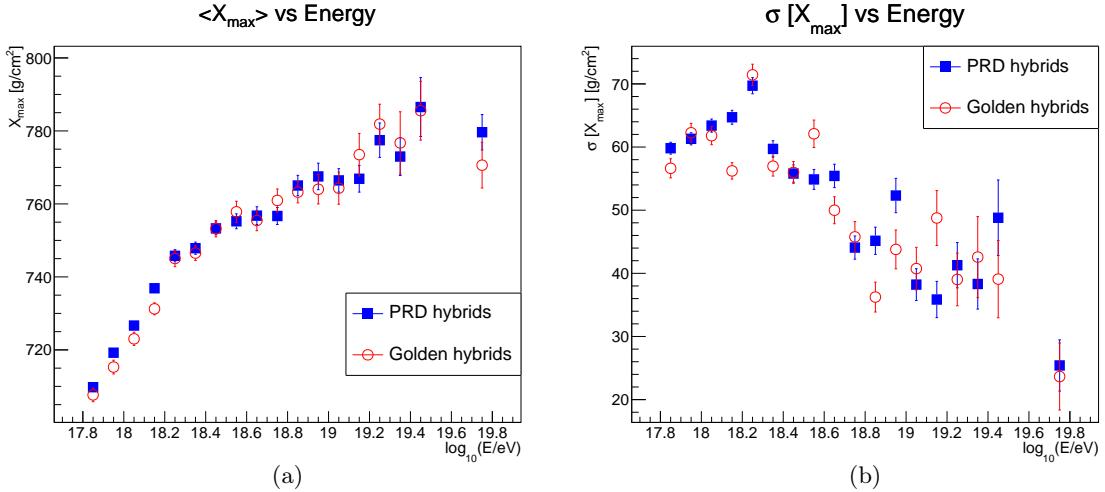


Figure 5.2: Comparison between (a) mean  $\langle X_{\max} \rangle$  and (b) dispersion  $\sigma(X_{\max})$  of the  $X_{\max}$  distributions for the two datasets, as indicated in the legend. The values in each energy bin are obtained from the histograms (as those shown in Fig. 5.1) where the acceptance correction is already applied. See the text for a complete description of the calculation.

A deeper comparison has been performed studying the first two moments of the  $X_{\max}$  distributions, i.e. the mean  $\langle X_{\max} \rangle$  and dispersion  $\sigma(X_{\max})$  as functions of the energy, separately for the two different datasets.

To get the  $X_{\max}$  moments, a correction must be preliminarily applied to the  $X_{\max}$  distributions due to the not-flat acceptance of the fluorescence detector. In fact, the probability to detect an event is not uniform for arbitrary values of the shower maximum depth. So residual distortions in the  $X_{\max}$  distributions are corrected assigning to each event a weight equal to the inverse of its acceptance<sup>5</sup>.

Another ingredient, necessary in the evaluation of  $\sigma(X_{\max})$ , is the detector resolution, which describes the broadening of the true distribution due to the statistical fluctuations of the  $X_{\max}$  measured through the hybrid reconstruction<sup>6</sup>.

The unbiased moments shown in Fig. 5.2a and 5.2b are obtained from the  $X_{\max}$  distributions (like the histograms shown in Fig. 5.1) using the equations for the weighted moments reported in the appendix of [53]. In particular,  $\sigma(X_{\max})$  is calculated as:

$$\sigma(X_{\max}) = \sqrt{\sigma^2(X_{\max})_{\text{meas}} - \sigma_{\text{det}}^2} \quad (5.1)$$

where  $\sigma(X_{\max})_{\text{meas}}$  is the dispersion of the distribution corrected for the acceptance and  $\sigma_{\text{det}}$  is the parameterized detector resolution.

The mean  $X_{\max}$  for the two datasets is considered compatible above  $\sim 10^{18.5}$  eV, as the values per energy bin are in agreement within uncertainties. The comparison between  $\sigma(X_{\max})$  for the two classes of hybrids is instead less satisfying, as not negligible shifts are present also in the higher energy bins (of interest for the following Universality

<sup>5</sup>The acceptance was evaluated through detailed simulations of the atmosphere and of the surface and fluorescence detectors and parameterized as a function of total energy and  $X_{\max}$ , as exposed in [53].

<sup>6</sup>Also the detector resolution has been evaluated [53] by means of dedicated simulation and parameterized as a function of the energy.

analysis). However, the number of events is naturally small in the energy range above  $\sim 10^{19}$  eV and barely sufficient to perform a study of the dispersion; thus, large bin-to-bin fluctuations are expected (and observed also in the PRD-hybrids results for  $\sigma(X_{\max})$ ).

Eventually the agreement between the two set of events is considered enough solid, taking into account the compatibility both in the shape of the distributions per energy bin and in the mean  $X_{\max}$  values while overlooking the  $\sigma(X_{\max})$  discrepancies for the above-mentioned reasons.

### 5.3 The original Universality fit

As exposed in sect. 5.1, the strategy to test the performances of the Universality fit (and of different versions of the procedure that will be developed) consists in a comparison of the results with the standard reconstruction of the selected golden hybrid events. In particular, it was decided to study on an event-by-event basis the differences between the physical parameters obtained through the Universality fit of the SD-only data and the same quantities from the hybrid reconstruction. The distributions of these differences can be considered an indicator for the global quality of the procedure, but anyway this method does not allow a quality selection over the successfully fitted events, which instead requires the definition of a goodness-of-fit estimator. The determination of a proper estimator will be addressed in the following sections.

Another important characteristic, that will be highlighted in the different versions of the Universality procedure, is the *efficiency*, intended as the number of event successfully fitted (i.e. where the minimization reached convergence) over the total number of events of the chosen set. The efficiency does not give information about the quality of the reconstruction, but is however a relevant indicator to evaluate the power of a certain fitting procedure.

The starting point is the study of performances for the latest version of the reconstruction available in the Offline framework of the Pierre Auger Collaboration. In particular, the efforts have been concentrated on the global Universality fit, where the energy of the event is fixed to the SD estimation while all the other physical parameters of the shower ( $X_{\max}, N_{\mu}, \theta, \phi, \vec{x}_{\text{core}}, t_{\text{core}}$ ) are fitted at the same time. This version of the Universality reconstruction will be named “*UnivKG*”.

The following analysis is focused on the subset of golden hybrids with  $E_{\text{SD}} > 10^{19}$  eV (where  $E_{\text{SD}}$  is the energy reconstructed through the standard SD procedure<sup>7</sup>), which counts 228 unsaturated events. The *UnivKG* reconstruction, without any modification, presents a very high efficiency over the chosen dataset: 225 over 228 not saturated golden hybrids are reconstructed from their SD-only measurements.

---

<sup>7</sup>Golden hybrids present also a fluorescence measurement of the energy, that is calorimetric and so more accurate than the SD estimation, which is instead obtained from a calibration procedure (see sect. 3.2.2). However, the final aim of the Universality fit is to correctly reconstruct events detected only by the ground array; so the energy cut has to be necessarily made on the SD reconstructed value.

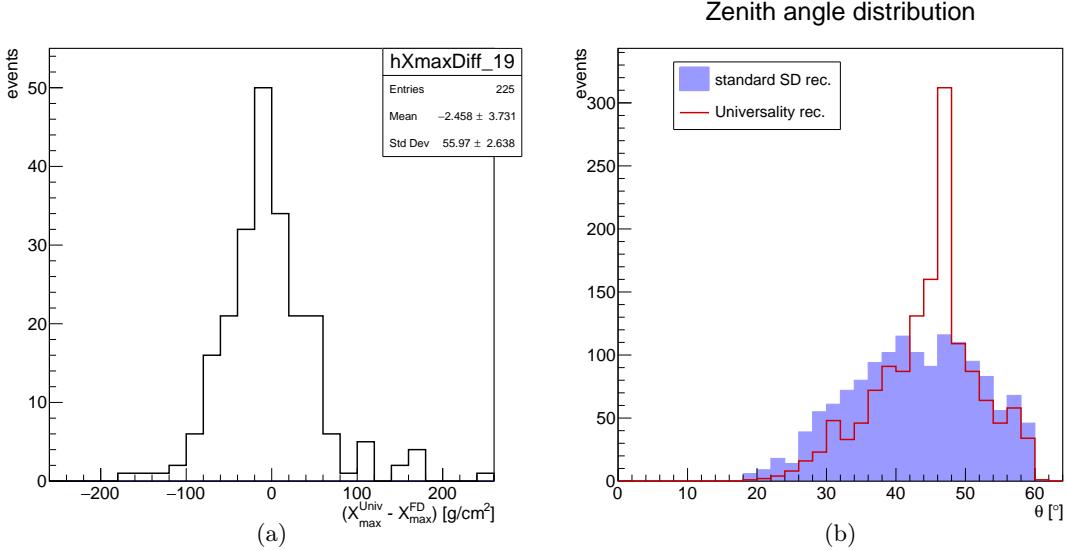


Figure 5.3: (a) Distribution of differences between the depth of shower maximum reconstructed with the Universality global fitting procedure “*UnivKG*” ( $X_{\text{max}}^{\text{Univ}}$ ) and from the fluorescence data after the complete hybrid reconstitution ( $X_{\text{max}}^{\text{FD}}$ ) for the described set of not-saturated vertical ( $\theta < 60^\circ$ ) golden hybrids with  $E_{\text{SD}} > 10^{19}$  eV. Only the values from successfully fitted events are included. (b) Reconstructed zenith angle for the unsaturated golden hybrids with  $E_{\text{SD}} > 10^{18.5}$  eV that are successfully reconstructed with the original Universality procedure. The filled histogram is the distribution obtained from the standard SD reconstruction, while the red histogram shows the results from the Universality fit.

In Fig. 5.3a it is shown the distribution of differences between the values of the depth of shower maximum reconstructed with the global Universality fit ( $X_{\text{max}}^{\text{Univ}}$ ) and from the standard hybrid reconstruction of the fluorescence profile ( $X_{\text{max}}^{\text{FD}}$ ). The bias in the reconstructed  $X_{\text{max}}^{\text{Univ}}$  is negligible, while the standard deviation of the distribution is  $\text{RMS}(X_{\text{max}}^{\text{Univ}} - X_{\text{max}}^{\text{FD}}) = 56.0 \text{ g/cm}^2$ . This outcome highlights the low resolution of this Universality procedure, meaning that the  $X_{\text{max}}$  determination is not very accurate.

The values of the other fitted parameters can be compared with the result both from the SD standard reconstruction and from the hybrid one, except the relative muonic number  $N_\mu$  which is not currently obtained by means of either method.

Then a systematic shift towards zenith angles around  $45^\circ$  is observed for the axis direction reconstructed with the *UnivKG* fit, as shown in Fig. 5.3b for the set of golden hybrids with  $E_{\text{SD}} > 10^{18.5}$  eV (such behaviour is present also in the highest energy range, but it is much less evident because of the lower statistics). As a consequence, a sizeable fraction ( $\sim 10\%$ ) of the successfully fitted events with  $E_{\text{SD}} > 10^{19}$  eV presents an angle between the two axis directions larger than  $4^\circ$  and a distance between the two core positions larger than 200 m. The fitting procedure is indeed allowed to change the shower geometry, as the SD determinations of the parameters are used exclusively as initial values. However, the presence of many events reconstructed with a geometry substantially different (i.e. beyond the known uncertainties) from the SD determination can indicate potential problems of the method.

## 5.4 A first change in the Universality fit: deviance minimisation

The first adjustment is the implementation of a new likelihood function (i.e. the function minimized during the fitting procedure) which can also be used as a goodness-of-fit estimator. This is a necessary step in order to perform an analysis of the reconstruction results properly based on the quality of the minimization procedure and not on the outcome values of the physical quantities (“*a posteriori*”). For this reason the likelihood function to be minimized was changed into a **deviance**, defined as the logarithm of the likelihood ratio, i.e. the ratio between the considered model and the saturated one (see [112] and references therein):

$$D = -2 \ln \frac{\mathcal{L}}{\mathcal{L}_{\text{sat}}} \quad (5.2)$$

This estimator is equivalent to a generalized  $\chi^2$  method, as described in [113].

The statistical description of the two likelihood terms (introduced in sect. 5.1) of the Universality reconstruction is inherited from the *UnivKG* version of the fit. In particular:

- in the LDF fit, each measured total signal is compared to the predicted integrated signals using a poissonian likelihood;
- in the shape fit, the measured content of each time bin from each trace is compared with the predicted value using a gaussian likelihood (meaning that an error over the signal in each bin has to be defined).

Then the total deviance to be minimized in the Universality procedure becomes:

$$D = \sum_k^{N_{\text{stations}}} \left( D_{\text{LDF}}^k + \sum_i^{n_{\text{bins}}^k} D_{\text{shape}}^{k,i} \right) \quad (5.3)$$

It is important to stress that the traces are not independently fitted, as the minimum eventually found does not give the best fit function for each single tank but the set of physical macroscopic parameters of the cascade that better reproduces all the traces and the integrated signals at the same time.

In the case of Poisson distributed binned data the deviance is equal to:

$$D_P = 2 \sum_{i=1}^N \left( x_i \ln \frac{x_i}{\nu_i} + \nu_i - x_i \right) \quad (5.4)$$

where  $x_i$  is the measured content of each bin while  $\nu_i$  is the correspondent expected value.

For the specific case of the LDF component, the quantity actually showing a poissonian behaviour is  $S/f_S^2$ , where  $S$  is the total signal and  $f_S$  is the experimental total signal uncertainty factor (a more detailed explanation is given in the following). Then the expression 5.4 becomes:

$$D_{\text{LDF}}^k = 2 \left( \frac{1}{f_S^2} \cdot S_{\text{tot}}^k \ln \frac{S_{\text{tot}}^k}{\hat{S}_{\text{tot}}^k} + \frac{1}{f_S^2} \cdot \hat{S}_{\text{tot}}^k - \frac{1}{f_S^2} \cdot S_{\text{tot}}^k \right) \quad (5.5)$$

where  $S_{\text{tot}}^k$  the integrated signal of the  $k$ -th station, and  $\hat{S}_{\text{tot}}^k$  the expected signal for the  $k$ -th station, rising from the Universality parametrisation.

Instead, for gaussian distributed binned data the deviance is equal to:

$$D_G = \sum_{i=1}^N \left( \frac{x_i - \nu_i}{\sigma_i} \right)^2 = \chi^2 \quad (5.6)$$

where  $x_i$  and  $\nu_i$  have the same meaning of the previous case and  $\sigma_i$  are the errors assigned to the measured bin values. It is interesting to notice that this expression is exactly equal to the  $\chi^2$  definition, a usual goodness-of-fit estimator. In the shape fit the traces are described as gaussian distributed binned data<sup>8</sup>, so a trace of  $n_{\text{bins}}^k$  bins from a station  $k$  contributes to the likelihood with a gaussian deviation given by:

$$D_{\text{shape}}^k = \sum_{i=1}^{n_{\text{bins}}^k} \left( \frac{s_i^k - \hat{s}_i^k}{\epsilon(s_i^k)} \right)^2 \quad (5.7)$$

where  $s_i^k$  is the signal for the  $i$ -th bin of the trace,  $\epsilon(s_i^k)$  the assigned error (defined in the following) and  $\hat{s}_i^k$  the expected value in that bin from the Universality time-model.

Since the Universality parameterization presents some limits of validity, a selection over the signals and traces, mainly based on the core distance of the WCD which measured them, is necessary. Moreover, requirements over the minimum bin signal and over the minimum number of bins in a trace are also introduced (already in the original version of the Universality fit), intuitively needed to properly fit the time-dependent signals with a complicate function. In particular:

- a trace is used for the shape fit if at least 5 bins have a signal above 0.7 VEM and the station it belongs to has a core distance  $r \in [100, 2200]$  m;
- a station is used for the LDF fit if  $r \in [100, 2500]$  m and if the associated trace has not been already included in the shape fit.

The choice of a poissonian description for the LDF term is related with the experimental observation (see for example [114, 115, 116]) of approximately poissonian fluctuations in the integrated signal measured by the water-Cherekov detectors. In effect, the variance  $\sigma[S_{\text{tot}}]$  of the integrated signal measured by a WCD has been demonstrated to scale as  $\sqrt{S_{\text{tot}}}$  comparing measurements from multiplets of detectors (which detect the shower particles from roughly the same location) placed at various positions in the SD array. In particular, it is found:

$$\sigma[S_{\text{tot}}/\text{VEM}] = f_S \sqrt{S_{\text{tot}}/\text{VEM}} \quad (5.8)$$

where the integrated signal is expressed in VEM, and the uncertainty factor  $f_S$  is parameterized in [114] as a function of the zenith angle of the event (see Fig. 5.4):

$$f_S = 0.865 \cdot [1 + 0.593 \cdot (\sec \theta - \sec 35^\circ)] \quad (5.9)$$

---

<sup>8</sup> Actually the traces could be converted into poissonian binned distributions of photoelectrons, but the cut on the minimum signal at 0.7 VEM for the shape fit, makes the two cases (gaussian and poissonian) indistinguishable since  $1 \text{ VEM} \approx 94$  photoelectrons.

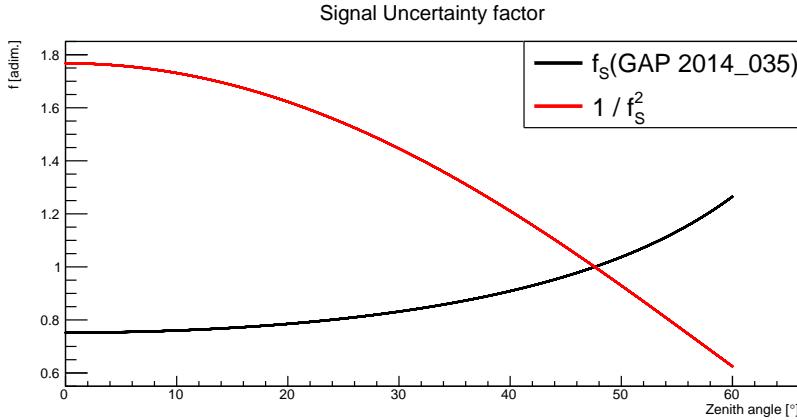


Figure 5.4: In black, the uncertainty factor  $f_S$  for the total measured signal (eq. 5.9). In red, the factor that approximately transforms the integrated signal (measured in VEM) into an effective number of particles  $n_{\text{eff}}$  (eq. 5.10).

Such a(n almost) poissonian behaviour of the total signal is clearly due to the poissonian fluctuations in the number of particles which hit a detector. With this concept in mind, one can search for a factor to approximately transform the integrated signal (measured in VEM) into an effective number of particles<sup>9</sup>, i.e. a quantity  $n_{\text{eff}}$  which obeys to poissonian statistics. From eq. 5.8, it is easily obtained that:

$$n_{\text{eff}} = \frac{1}{f_S^2} \times S \quad (5.10)$$

In fact, the uncertainty of  $n_{\text{eff}}$  can be written as:

$$\sigma[n_{\text{eff}}] = \frac{1}{f_S^2} \sigma[S] = \sqrt{\frac{1}{f_S^2} \times S} = \sqrt{n_{\text{eff}}} \quad (5.11)$$

showing the expected poissonian behaviour. The “conversion” factor  $1/f_S^2$  is shown in Fig. 5.4. It is evident at this point that the quantity actually exploited in the LDF deviance (eq. 5.5) is the number of effective particles  $n_{\text{eff}}$ .

The bin signals  $s_i$  (measured by each WCD in intervals of 25 ns) are assumed to have the same fluctuations as for the total signal. Then the uncertainty  $\epsilon(s_i)$  for each bin used in the shape likelihood calculation is evaluated as:

$$\epsilon(s_i) = f_S \times \sqrt{s_i} \quad (5.12)$$

and can be considered “quasi poissonian”. Also in the original version of the Universality fit a “quasi poissonian” bin signal uncertainty was assumed, but using a factor different from  $f_S$  for the actual calculation of  $\epsilon(s_i)$  (see [111]).

The assumption of “quasi poissonian” errors for the bin content has a physical meaning, even if not explicit: it implies that the bin-to-bin correlations are not considered. In

---

<sup>9</sup>The actual number and energy spectrum of secondary particles inducing a certain signal is not known and quite difficult to estimate, due to the different sensitivities of a WCD to various particle types.

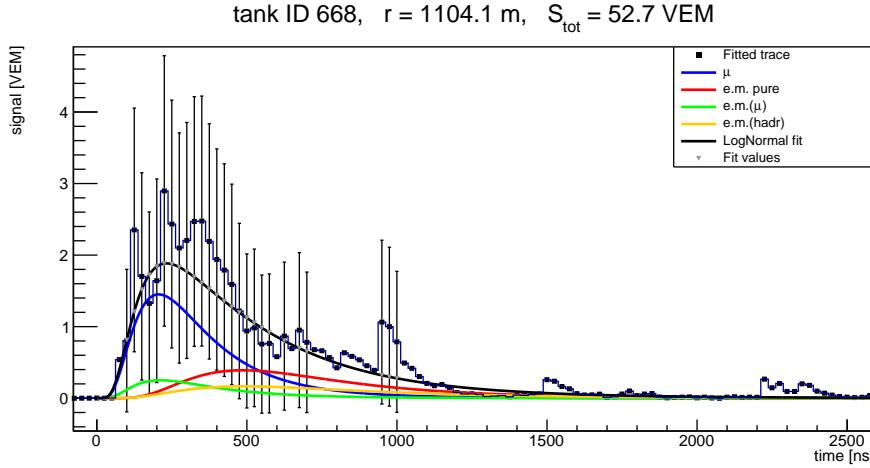


Figure 5.5: Trace fit obtained with the Universality global reconstruction of an event. The bin contents  $s_i$  are reported as points, where only the ones with error bars (“quasi poissonian” errors) were actually used in the fit. The coloured lines are the four log-normal components obtained from the Universality minimization; the sum (in black) is the fitting function for this trace.

fact, given the experimental (approximately) poissonian behaviour of the total integrated signal, a poissonian behaviour also for the signal in each bin can be obtained only in the hypothesis of totally not correlated bin signal values. A “visual” proof of the inadequacy of this choice can be observed in Fig. 5.5, that shows the result of the global fitting procedure for a trace with medium signal of a selected golden hybrid event. The bin-to-bin fluctuations become almost negligible with respect to so large uncertainties; the consequence is a loss of information for the fit itself, since these fluctuations are physical. Moreover, the effect of the minimum bin signal cut (0.7 VEM) is clearly visible: the first bins of the trace are not used, while it is possible to have large gaps of neglected bins between peaks at larger times.

In the following the “quasi poissonian” assumption for the bin content will be argued to be generally inaccurate to describe signals released by the shower particles in small temporal intervals. Since a correct evaluation of the bin signal uncertainties is critical for the fitting procedure, it represents a fundamental issue to face in the improvement of the Universality reconstruction.

Other minor modifications to the original fitting procedure have been applied together with the introduction of the deviance minimization:

- the upper limit on the  $N_\mu$  value is removed;
- the stations with a total measured signal below 12 VEM are not used in the fit; this selection is introduced to assure full trigger efficiency of the WCD and to employ the Universality parameterization of integrated signals only in a range where potential biases are below few percent (see Fig. 4.12 and [117]);
- an interval of bins to be actually fitted in each trace is introduced, in order to neglect isolated small signals in the traces tail that are not correlated to the shower

- particles; the start bin is put in correspondence with the trigger time of each detector, while the stop is placed 60 bins (1500 ns) after the start;
- the minimization is accomplished in two steps, exploiting the MINUIT tools; in a first step, the minimum is found with a low level strategy, fast but inaccurate; re-starting the fit from this minimum with an higher level strategy (slower but more accurate) allows to reach convergence in almost all cases.

#### 5.4.1 The results for the modified Universality reconstruction

In order to check the effect of switching from Log-Likelihood into deviance minimization, the Universality reconstruction has been run again on the same set of golden hybrid events previously described. The results are expected to be different as the relative weights between the likelihood terms are altered, even if the fundamental aspect of the global Universality procedure is maintained:  $(X_{\max}, N_\mu, \theta, \phi, \vec{x}_{\text{core}}, t_{\text{core}})$  are fitted at the same time while the energy is fixed to the SD reconstruction value.

With all the described improvements, a high efficiency above  $10^{19}$  eV is maintained: 223 over the 228 unsaturated events are successfully fitted.

In Fig. 5.6a it is shown the distribution of differences  $X_{\max}^{\text{Univ}} - X_{\max}^{\text{FD}}$  obtained from this modified version of Universality procedure. The mean value presents now a negative bias  $\langle X_{\max}^{\text{Univ}} - X_{\max}^{\text{FD}} \rangle = -15.5 \text{ g/cm}^2$ ; on the other hand the RMS( $X_{\max}^{\text{Univ}} - X_{\max}^{\text{FD}}$ ) is reduced to  $41.8 \text{ g/cm}^2$ . Therefore a systematic bias is introduced, even if the choice of deviance results into an improved accuracy for the reconstruction of the other shower parameters according to the “*a posteriori*” analysis.

In fact, the geometry obtained through this Universality fit is now more similar to the standard reconstructions with respect to the original version. The biased result for the zenith angle determination, shown in Fig. 5.3b, is not present in the procedure with deviance minimization, as one can see in Fig. 5.6b for the complete ( $E_{\text{SD}} > 10^{18.5}$  eV) set of golden hybrids. This improvement is mainly due to the new definition of the bin signal uncertainty (eqs. 5.9 and 5.12) employed in this work. As a consequence, the angular distance between hybrid and Universality reconstructed axis directions is on average reduced, with angular shifts never larger than  $5^\circ$  for the highest energies ( $E_{\text{SD}} > 10^{19}$  eV). Also the distances between core position determinations are reduced on average, with few reconstructed events showing differences larger than 200 m.

At this point, it is possible to study the quality of this modified Universality procedure not only using an “*a posteriori*” analysis, i.e. the distributions of differences between Universality and hybrid reconstructed quantities, but also exploiting the deviance value at the minimum found for each successfully fitted event. As already highlighted, this quantity and its components follow a  $\chi^2$  statistics and can be used as goodness-of-fit estimator. In particular, the ratio of deviance over a suitable number of degrees of freedom (defined in the following) is expected to be  $\approx 1$  in the case of a good fit, as for

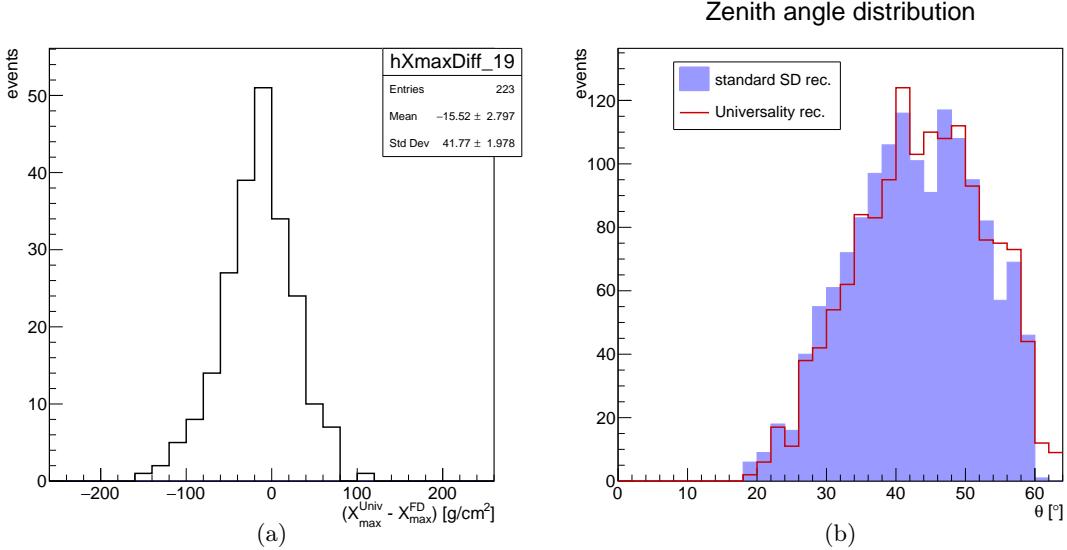


Figure 5.6: (a) Distribution of differences between  $X_{\max}^{\text{Univ}}$  from the Universality global fit with deviance minimization and  $X_{\max}^{\text{FD}}$  from standard hybrid reconstruction. See Fig. 5.3a for further description. (b) Reconstructed zenith angle for events successfully fitted with the Universality procedure through deviance minimization. See Fig. 5.3b for further information.

the  $\chi^2/\text{NDoF}$ . Values of this ratio much larger than unity usually indicate that the used model is not able to explain data; instead, values much smaller than unity often entail an overestimation of the uncertainties, and thus a likely inaccurate result.

The distribution of the total LDF deviance  $D_{\text{LDF}}$  over the number of stations used for the LDF fit is shown in Fig. 5.7a. It is quite compatible with the behaviour expected for a  $\chi^2$  distribution with few degrees of freedom: in effect, the vast majority of the reconstructed events presents less than 3 stations in the LDF component<sup>10</sup>.

Instead, the distribution of the total shape deviance  $D_{\text{shape}}$  of each event with respect to the overall number of fitted bins from every trace is shown in Fig. 5.7b. The values are noticeably small, with an average value around  $\approx 0.17$ . This result will be demonstrated to be caused by the assignment of excessive uncertainties to the bin signals in the trace fit, an inference which can be also deduced from e.g. Fig. 5.5.

In conclusion, the introduced changes generally improve the accuracy of the Universality procedure, evaluated according to the “a posteriori” distributions of the reconstructed quantities with respect to the corresponding hybrid values. At the same time, a bias is found in the Universality determination of the shower maximum depth; therefore the reconstruction still needs adjustments to obtain a more satisfactory result.

---

<sup>10</sup>In many cases all the traces can be exploited in the shape part or the remaining stations have an integrated signal below 12 VEM (and then are not used anyway): these events are not included in the distribution of Fig. 5.7a.

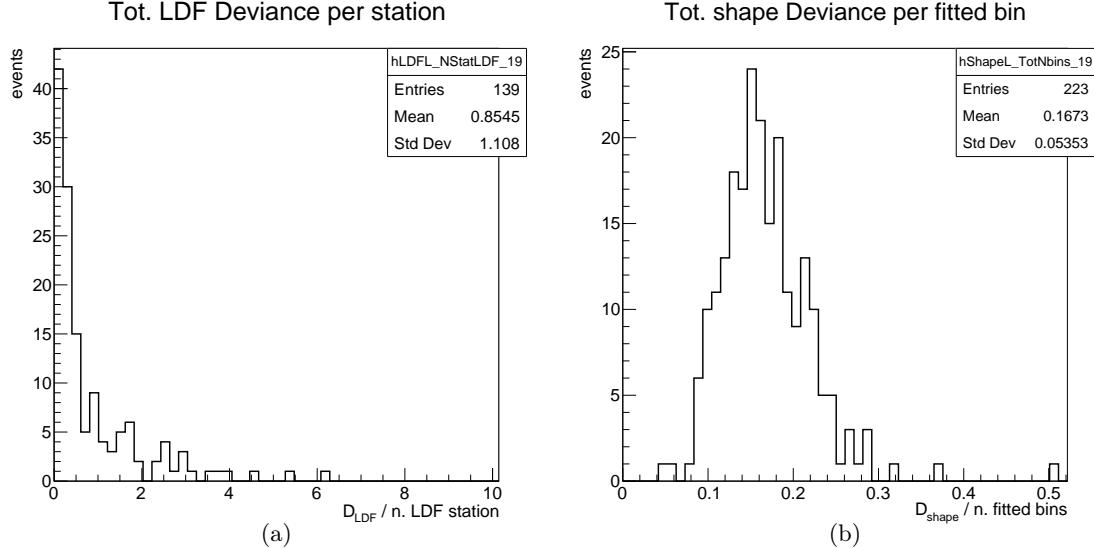


Figure 5.7: (a) Distribution of the ratio  $D_{\text{LDF}}$  over the number of stations used for the LDF fit. (b) Distribution of the ratio  $D_{\text{shape}}$  (sum of the deviance for each station used in the shape fit) over the total number of bins (sum of the bins used from each traces). Each entry represents the value for a successfully reconstructed event from the selected set of unsaturated golden hybrids with  $E_{\text{SD}} > 10^{19}$ . In particular, only events with at least one station used in the LDF component are included in the LDF deviance distribution.

## 5.5 Study of the bin signal fluctuations

The next step in the improvement of the Universality reconstruction is to better estimate the errors assigned to each bin of the traces. The needed reduction can be quantified requiring the compatibility between uncertainties and bin signal fluctuations.

In principle, the correct method to evaluate the magnitude of such fluctuations consists in the study, for several typologies of signals and events, of the behaviour of a sample of traces all generated by one simulated shower in the same detector at fixed conditions. In this way the differences among traces would be only ascribed to the random fluctuations in the shower particles.

This hard task can be accomplished using a full Monte Carlo approach, but it is largely time consuming. A convenient strategy, employed in this work, is instead to elaborate a model and consequently an algorithm for the production of random mock traces that resemble the measured ones, a method much faster than the simulation of cosmic ray cascades.

### 5.5.1 “Quasi poissonian” case

A first, simple model can be built exploiting the “quasi poissonian” assumption for the bin signal uncertainties which, as anticipated in sect. 5.3, also implies that the bin signals are considered uncorrelated.

Starting from a single trace of an event successfully fitted in the Universality reconstruction, a mock trace can be created through the following steps:

1. calculate for each bin  $i$  the value of the actual poissonian variable  $\mu_i$ , related with the predicted bin signal  $\hat{s}_i$  from the Universality time model through eq. 5.10, as

$$\mu_i = \hat{s}_i / f_S^2 \quad (5.13)$$

2. extract for each bin  $i$  a value from a Poisson distribution with mean  $\mu_i$ ; a random quantity in the interval  $[-0.5, 0.5]$  (negative results put equal to zero) is added in order to prevent unrealistic traces with only discrete values of the signal;
3. “reconvert” the extracted value in a proper bin signal multiplying by  $f_S^2$ .

In the following this procedure will be referred as “quasi poissonian” model.

Two traces have been chosen as representatives for higher and lower integrated signals, and will be used in the following to generate several mock realizations in accordance with the “quasi poissonian” model. The results are shown in Fig. 5.8 and Fig. 5.9, where one panel presents the original trace and the other panels eight different mock ones; the continuous line represents the fit function for the true trace (output of the Universality reconstruction) from which the “quasi poissonian” realizations are obtained.

It is clear that the mock traces are not representative of the true one (even more evident in the case of lower integrated signal) since at first sight one can recognise the original signal. Therefore, this model is apparently not able to generate reasonable traces as random realizations from the Universality parameterized Log-Normal functions.

The true mistake in this approach is supposed to lie in the assignment of “quasi poissonian” uncertainty for the bin signals, even if directly inherited from the original Universality procedure. In fact, only the integrated signal can be considered to behave

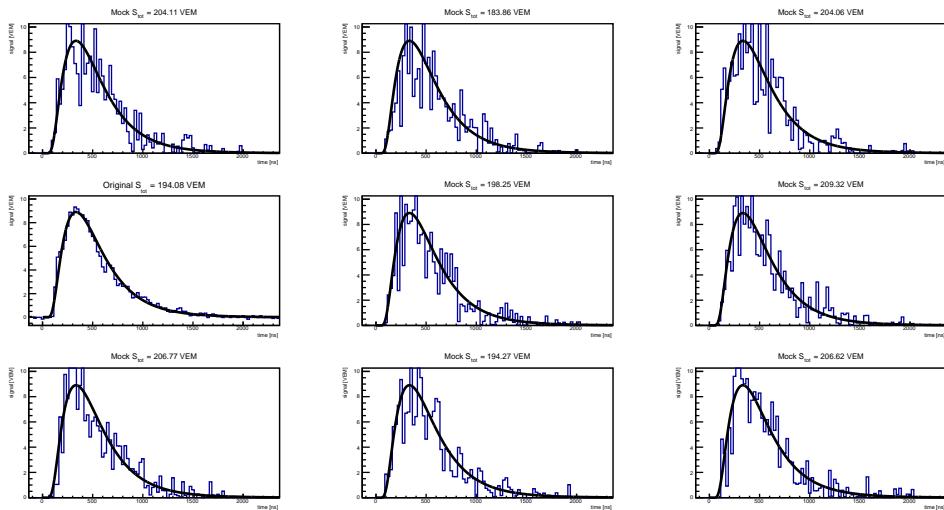


Figure 5.8: Time-dependent signal from station 543 of a sample event (ID *auger\_93505411300\_sd\_8864733*). The original trace has a measured signal  $S_{\text{tot}} = 194$  VEM and is represented in the 4th panel. The other eight panels show different random realizations of the same trace using the “quasi poissonian” model. The continuous line is the sum of the 4 Log-Normals obtained from the Universality shape fit of the true trace.

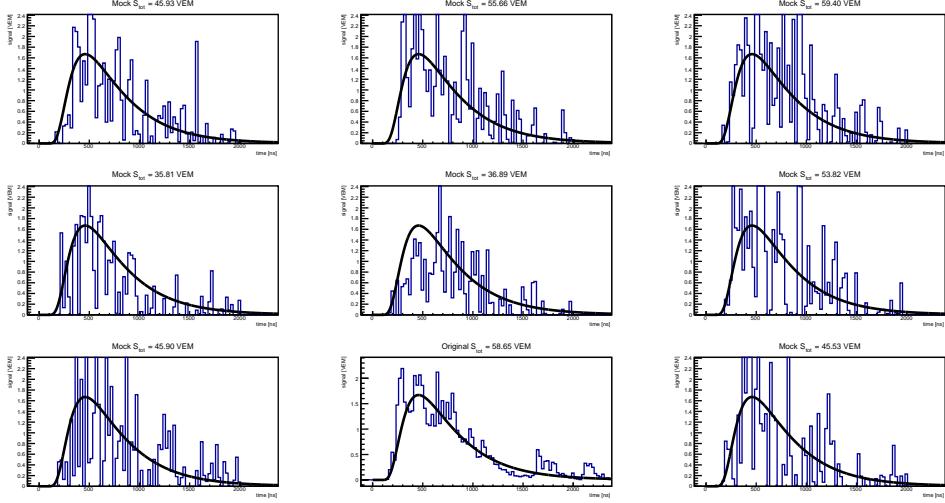


Figure 5.9: Same as Fig. 5.8 from station 549 of the same event. The original trace has a measured signal  $S_{\text{tot}} = 58$  VEM and is plotted in the 8th panel.

approximately as a Poisson variable (as experimentally observed) but the bin signals can not be considered independent.

### 5.5.2 Toy “cluster model”

As shown in the previous paragraph, the “quasi poissonian” assumption for the bin signals fluctuations brings to evident inconsistencies. Its main weakness is to ignore the correlations of neighbour bins due to signal decay time and physical correlations that exist among the shower particles entering in a Cherenkov detector.

In effect, it is undeniable that the measured time dependent signal is composed by a great number of substructures, ideally one for each particle traversing a tank and releasing a signal. To intuitively understand this statement, in Fig. 5.10 (taken from [103]) it is shown a simulated trace for a proton shower of  $10^{19.5}$  eV, separated in the 4 Universality components. In this trace several isolated peaks are clearly visible, suggesting a very simple model to generate credible mock traces and re-define (as a byproduct) the magnitude of the errors assigned to each bin of the trace signal.

The basic idea is that a trace can be obtained as the sum of many **clusters** of signals, i.e. substructures which correlate the signals of neighbour bins, separately for each Universality component. At this preliminary stage the clusters’ origin is left unspecified: in general they should not be considered as correspondent to single particle signals (except for the muonic component, as exposed in the following). Although the existence of such structures is partially justified considering the delay processes happening inside a tank<sup>11</sup>, a proper physical explanation has to be further pursued.

<sup>11</sup>The signal released by a single particle already correlates the contents of more temporal bins (each one 25 ns wide). In fact, the diffusion time of the emitted Cherenkov photons is about 60-70 ns (thus on average 3 bins) before being absorbed. Also photomultipliers and electronics play a role into smoothing and stretching the signal, extending its duration above 3 bins.

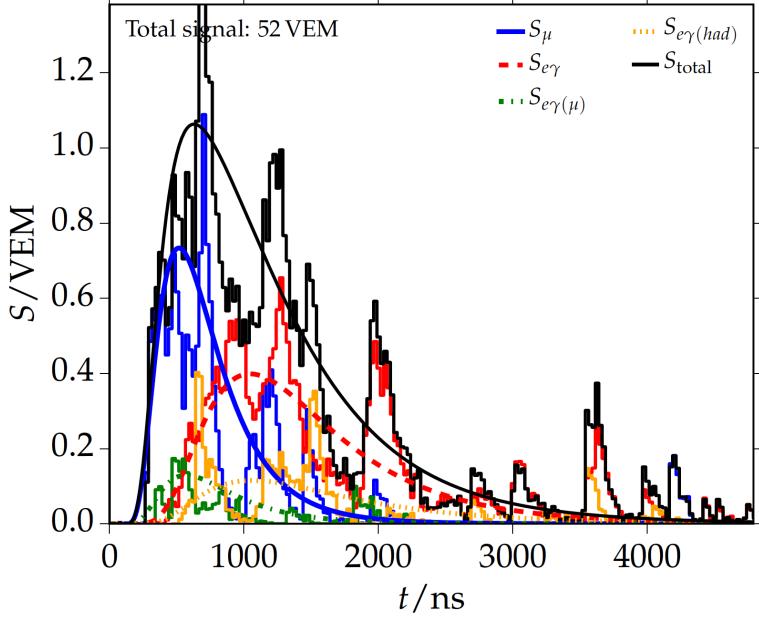


Figure 5.10: Universality reconstruction of a simulated proton shower with energy  $10^{19.5}$  eV. The shape fit result, shown in black, is obtained as sum of the 4 Universality components as reported in the legend. Image from [103].

With the aim to create a simple toy model (hereafter referred to as “cluster model”) to generate credible realizations of a trace, based on the Universality parameterizations, the following assumptions are made:

- the basic block of each trace is a cluster with mean integrated signal  $\langle S_{\text{clu}} \rangle = (1 \times f_S^2)$  VEM, ideally correspondent to the signal produced by one effective particle according to eq. 5.10;  $f_S = f_S(\theta)$  is the uncertainty factor described in eq. 5.9 (with  $\theta$  the zenith angle of the event).
- the cluster’s time distribution is assumed to be a log-normal with scale parameter  $m = 100$  ns and shape parameter  $s = 0.4$  (the location parameter is put equal to 0), approximately deduced from Fig. 5.10 focusing on the peaks above 2500 ns.
- each cluster has a mean content of  $\langle n_{p.e.} \rangle = 94 \times f_S^2$  photoelectrons, since the average number of p.e. per VEM [45] is  $\sim 94$ . The actual number of photoelectrons in a cluster is extracted from a Poisson distribution with mean equal to  $\langle n_{p.e.} \rangle$ .
- the mean number of clusters for each Universality component  $i$  is  $\langle N_i^{\text{clu}} \rangle = S_i / S_{\text{clu}}$ , where  $S_i$  is the predicted signals from the fit<sup>12</sup>. The number of clusters  $N_i^{\text{clu}}$  for the  $i$ -th component of an actual mock trace is extracted, separately for each component, from a Poisson distribution with mean equal to  $\langle N_i^{\text{clu}} \rangle$ .
- the distribution of the clusters temporal positions in the trace corresponds, separately for each shower component, with the related log-normal function obtained

---

<sup>12</sup>The component signals  $S_i$  have to respect the condition  $\sum_i S_i = S_{\text{tot}}$  where  $S_{\text{tot}}$  is the measured integrated signal.

from the shape fit. Therefore, the central time position of each cluster is extracted on these fitted functions.

The first assumption is undoubtedly crude, since only muons present an average single particle signal around 1 VEM; therefore only the muonic part of the trace can be considered to behave approximately as described in this toy model, with clusters roughly representing single muon signals<sup>13</sup>. However, choosing  $S_{\text{clu}} = (1 \times f_S^2)$  VEM, the mean number of clusters for each Universality component results equal to:

$$\langle N_i^{\text{clu}} \rangle = \frac{S_i}{S_{\text{clu}}} = \frac{S_i}{f_S^2} = n_{\text{eff}}^i \quad (5.14)$$

where  $n_{\text{eff}}^i$  corresponds to an effective number of particles for the  $i$ -th component, as described in sect. 5.4 for the total signal (see eq. 5.10). In effect, a parallel between  $n_{\text{eff}}$  and  $\langle N^{\text{clu}} \rangle$  seems reasonable: like the effective number of particles, the clusters represent substructures of the signal without a direct interpretation as physical entities, introduced mainly to create a simple model for the observed traces.

At this point, for a trace fitted with the Universality reconstruction one can generate several mock traces using the toy cluster model, as done in the previous section in the case of “quasi poissonian” errors. In Fig. 5.11, eight different realizations are shown for the trace with higher signal; the continuous line again represents the shape fit result. The same in Fig. 5.12 for the trace with lower signal. In these “mosaics”, it is evident that the cluster model gives a better representation of the time dependent signal with respect to the “quasi poissonian” hypothesis, since the true trace is not easily distinguishable from the mock ones for both low and high signals. In particular the bin-to-bin fluctuations are similar to the ones in real signals; the same for the peak structures, more visible at large times where clusters are less numerous and less overlapped.

Furthermore it can be demonstrated that the integrated signal of a set of mock traces generated from the same original one through the cluster model results distributed as in the experimental observations, in particular with a dispersion equal to  $f_S \sqrt{S_{\text{tot}}}$ . The mathematical proofs are reported in Appendix A.2, while an empirical confirmation is given in the following<sup>14</sup>.

In Fig. 5.13 the distributions of the integrated signal ( $S_{\text{mock}}$ ) for a data sets of 10,000 mock realizations are shown: on the left, for the example trace with higher signal (same

<sup>13</sup>The number of emitted Cherenkov photons is approximately proportional to the length traversed by the particle in water, until its speed  $\beta (= v/c)$  does not substantially changes because of the energy losses. Such approximation is fully valid for  $\mu^\pm$  with  $\sim$ GeV energies: in particular, the signal released by a vertical muon (by definition 1 VEM) is proportional to the WCD height (1.20 m), since the energy losses are around 240 MeV. Instead,  $e^\pm$  arrive with much lower energies and are absorbed. However, the largest part of energy losses only occurs in the last few centimeters of water, meaning that the emission of Cherenkov radiation can be considered (roughly) constant during the  $e^\pm$  propagation and thus proportional to the range of such particles in water. For instance, a 300 MeV electron has a range of  $\approx 61 \text{ g/cm}^2$ , meaning that the released signal is around half of a VEM. At 100 MeV, the range drops to  $\approx 32.6 \text{ g/cm}^2$ , so the signal is already around 0.25 VEM. At 10 MeV, the range is only a few  $\text{g/cm}^2$  and the signal is well below 0.05 VEM.

<sup>14</sup>Nevertheless, this is strictly dependent on the assumption  $S_{\text{clu}} = (1 \times f_S^2)$  VEM for all the clusters, as demonstrated in the appendix A.2.3.

as Fig. 5.11); on the right, for the one with lower signal (same as Fig. 5.12). Each distribution presents a mean equal to the total signal and a variance approximately equal to the square root of the predicted signal multiplied by the uncertainty factor. In particular,  $RMS(S_{low}) = 6.2 \approx (f_S \times \sqrt{S_{low}}) = 6.1$  and  $RMS(S_{high}) = 11.4 \approx (f_S \times \sqrt{S_{high}}) = 11.2$ . The results from the cluster model are thus consistent with the experimental quasi-poissonian behaviour for the fluctuations of the total signal.

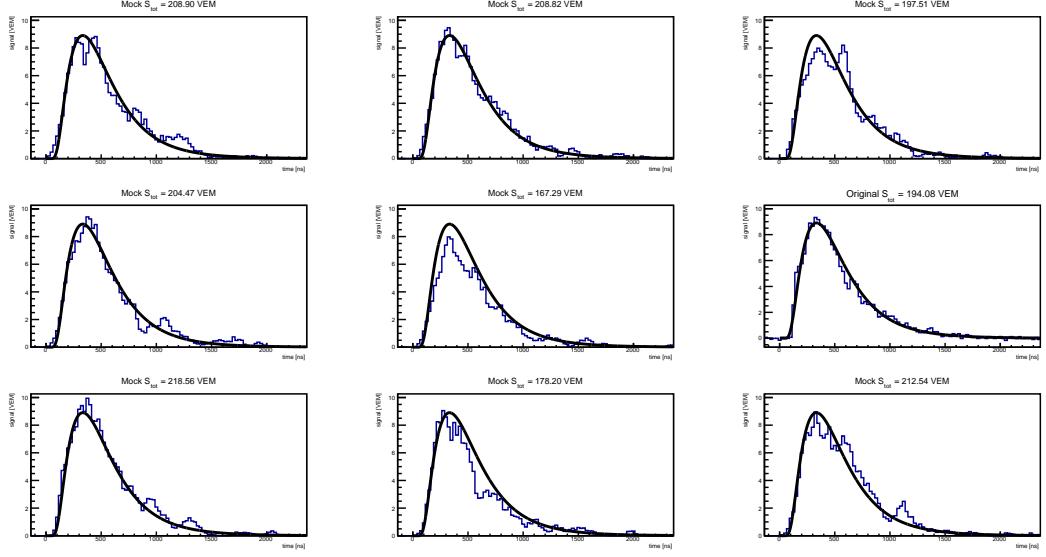


Figure 5.11: Time-dependent signal from station 543 of a sample event (ID *auger\_93505411300\_sd\_8864733*, same as Fig. 5.8) with zenith angle  $\theta = 24.6^\circ$  ( $f_S(\theta) = 0.80$ ). The trace in the 6th panel is the actual one ( $S_{tot} = 194$  VEM), while the other panels show different random realizations of the same trace using the toy cluster model. The continuous line is the shape fit results from the Universality reconstruction of the original event.

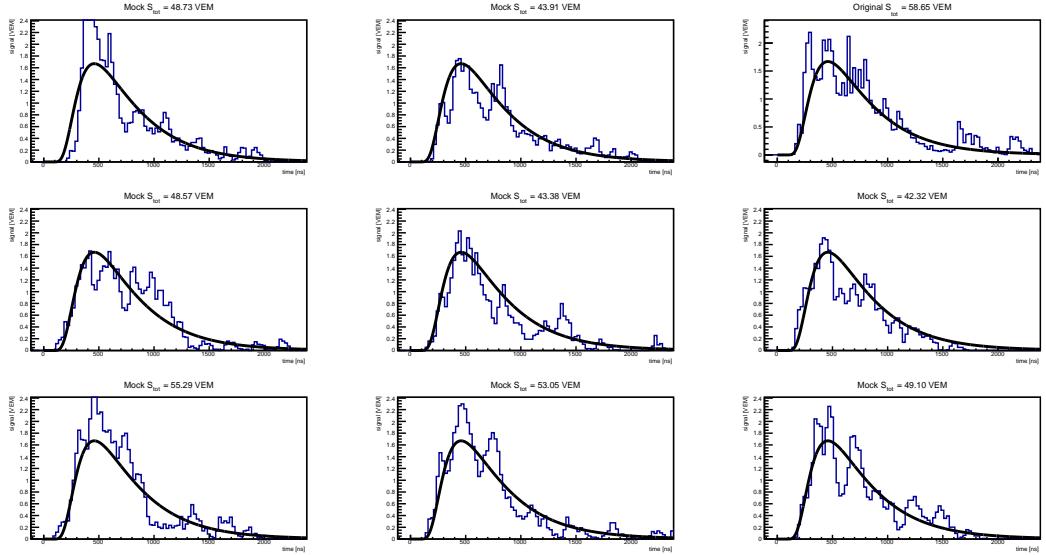


Figure 5.12: Same as Fig. 5.11 from station 549 of the same event ( $S_{tot} = 58$  VEM). The original trace is plotted in the 3rd panel.

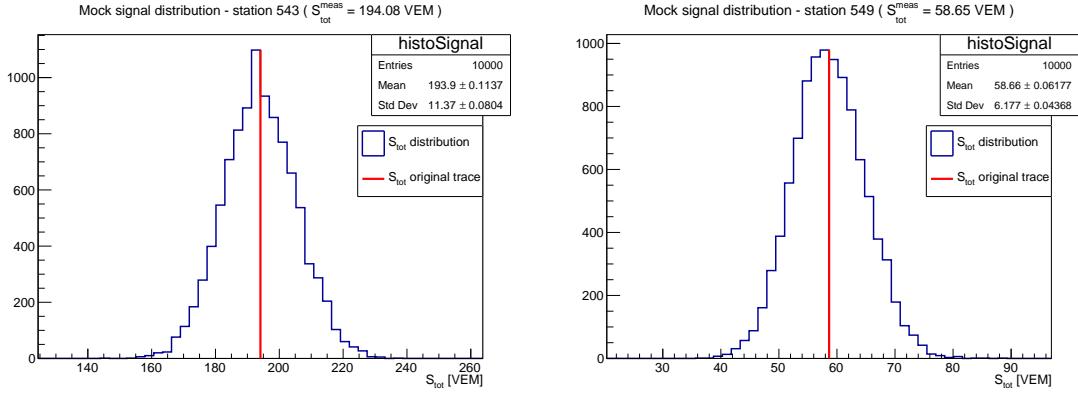


Figure 5.13: Distributions of the total signal according to the toy cluster model, for a data sets of 10,000 mock realizations of the 2 traces from the sample event: on the left the higher signal trace, on the right the lower signal one. The vertical line shows the measured signal.

### 5.5.3 The bin signal uncertainties from the toy cluster model

Using the toy cluster model, one can evaluate the bin content fluctuations studying the dispersion of the signal in each bin, calculated over a large set of mock traces.

The main obstacle in this approach is that the cluster model requires the 4 Universality log-normal functions and related component signals, which are obtained after the fit procedure; instead, one wants to know what uncertainty assign at each bin before the minimization. To overcome this problem, the magnitude of the bin content fluctuations is initially studied using several sets of mock traces, each set generated from a single Log-Normal function. In this way it is possible to evaluate potential changes in the fluctuations, resulting from the different parameters of the originating Log-Normals. Finally the uncertainties for a proper trace (sum of 4 Log-Normals), which is considered to approximately maintain a Log-Normal shape, will be obtained.

The first step consists in understanding which combinations of the Log-Normal parameters are effectively present in time dependent signals successfully fitted by means of the Universality reconstruction. Therefore, taken the results from the last reconstruction (deviance minimization), it was realised a scatter plot of the location parameter  $m$  vs the shape parameter  $s$  of every trace used in the shape fit from the successfully reconstructed events. The plot is shown in Fig. 5.14 with different colours for each Universality component<sup>15</sup>. Looking at this plot, it is decided to study Log-Normals with parameters approximately in the range  $\log_{10}(m) \in [2, 3.1]$  and  $s \in [0.35, 0.65]$ . The other ingredients required to use the cluster model are: the Log-Normal integral  $S_{LN}$ , from which the mean number of clusters is calculated; the zenith angle, necessary to establish the value  $f_S(\theta)$  of the signal uncertainty factor. Integrated signals in the range between 10 and 1000 VEM are employed, with zenith angles in the interval  $[0, 60]$  degrees.

<sup>15</sup>One can notice that the muon component has  $m$  lower and  $s$  higher than the e.m. components, meaning that the signal arrives earlier in the detector and it is narrower in time (as expected from

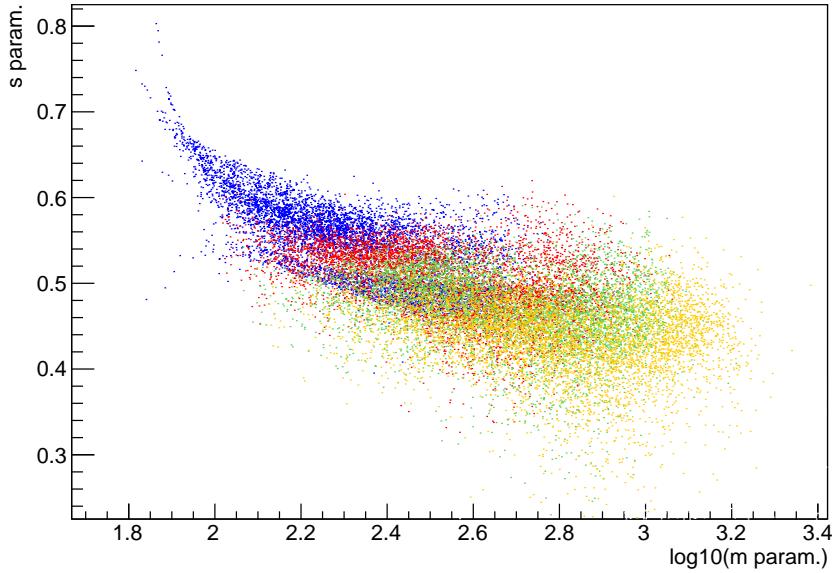


Figure 5.14: Scatter plot for the log-normal  $m$  and  $s$  values obtained from the Universality parameterization, for the traces used in the shape fit of unsaturated events successfully reconstructed in the minimization procedure using Deviance, with energies above  $10^{18.5}$  eV. In different colours the 4 Universality components: (blue) muonic, (red) pure electromagnetic, (green) e.m. from muons, (orange) e.m. from hadrons.

At this point for each Log-Normal, characterized by a set of parameters  $(m, s, S_{LN}, \theta)$ , 1000 mock traces are realized through the cluster model. The idea is to compare the bin content dispersion with the bin signal errors assigned in the Universality fit (i.e.  $\varepsilon(s_i) = f_S \times \sqrt{s_i}$ ) to evaluate a factor  $\delta$  which quantifies the needed uncertainty correction. In order to do so, for each bin the mean content  $\bar{s}_i$  and the dispersion  $\sigma[s_i]$  are calculated over the set of mock realizations, and then the ratio  $r_i = \sigma[s_i]/(f_S(\theta) \times \sqrt{\bar{s}_i})$ . This procedure is reported in Fig. 5.15 for the case with  $m = 500$ ,  $s = 0.5$ ,  $S_{LN} = 100$  VEM and  $\theta = 45^\circ$ . One can notice that the ratio is quite flat in the whole time range.

Finally the values  $r_i$  are gathered into an histogram (lower right panel of Fig. 5.15). The mean of this histogram is considered as the (reducing) factor  $\delta$  for a chosen Log-Normal  $(m, s, S_{LN}, \theta)$ , which expresses how much on average the bin signal fluctuations are different from the bin content uncertainties used in the shape fit.

From this procedure it arises that the factor  $\delta$  does not substantially change inside the chosen parameter space, meaning that for any combination of  $(m, s, S_{LN}, \theta)$  it assumes values in the (quite) small range [0.41, 0.44]. This is a very significant results, since in principle the value of  $\delta$  could be a function of the parameters<sup>16</sup>  $m$ ,  $s$ ,  $S_{LN}$  and  $\theta$ .

Therefore the factor  $\delta$  is put equal to 0.42 and considered constant, meaning that the

---

shower physics).

<sup>16</sup>It could also change for different model assumptions, but in this work only the above-described version of the cluster model has been used.

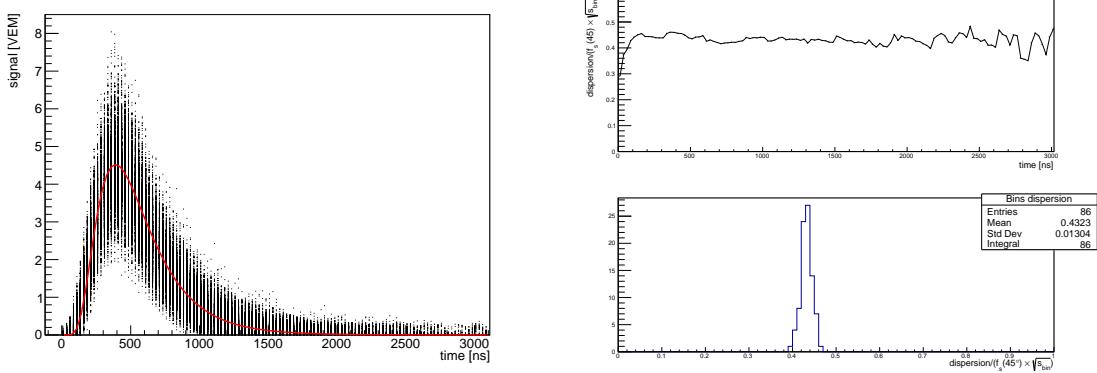


Figure 5.15: (Left) Superposition of 1000 mock traces obtained by means of the cluster model from a Log-Normal function with  $m = 500$ ,  $s = 0.5$ ,  $S_{LN} = 100$  VEM and  $\theta = 45^\circ$ , drawn in red. (Right) In the upper panel, the ratio  $r_i$  (bin signal dispersion over the “quasi poissonian” bin signal error) as a function of time, obtained from the set of 1000 mocks shown on the left. In the lower panel, the histogram with the values of  $r_i$  for the bins with a predicted signal (from the Log-Normal) bigger than 0.1 VEM. The mean of this histogram gives the value of the factor  $\delta$  for the chosen set of Log-Normal parameters.

bin signal errors for a single Universality component  $k$  can be written as:

$$\varepsilon(s_i^k) \approx \delta \times f_S \times \sqrt{s_i^k} = 0.42 f_S \times \sqrt{s_i^k} \quad (5.15)$$

In the hypothesis of absence of correlation among the 4 components, the dispersion of the signal in a bin of the total trace is simply given by the quadratic sum of the dispersion for each single component. As the value of  $\delta$  is constant and independent from the various parameters<sup>17</sup>, using eq. 5.15 the total uncertainty results simply:

$$\varepsilon(s_i) = \sqrt{\sum_k \left( \delta f_S(\theta) \sqrt{s_i^k} \right)^2} = \delta f_S(\theta) \sqrt{s_i} \quad (5.16)$$

where  $s_i^k$  is the bin content in VEM for each component. Thus a factor known *a priori*, independent from the actual characteristics of the measured trace, is obtained. Such result is perfect to be used in the Universality fit, as exposed in the next section.

## 5.6 A second change in the Universality fit: reduced bin signal uncertainties

In the last section it was found that the bin content uncertainty should be reduced, in first approximation, of a factor  $\delta \approx 0.42$  to match the signal fluctuations observed in the mock traces generated through the cluster model. The effect of this additional change on the Universality procedure is verified running a new reconstruction of the same set

<sup>17</sup>In fact, since the reducing factor is (approximately) independent: (a) from the shape of the Log-Normal, it can be used for any signal component; (b) from the signal intensity, it can be used for any station; (c) from the zenith angle, it can be used for any event.

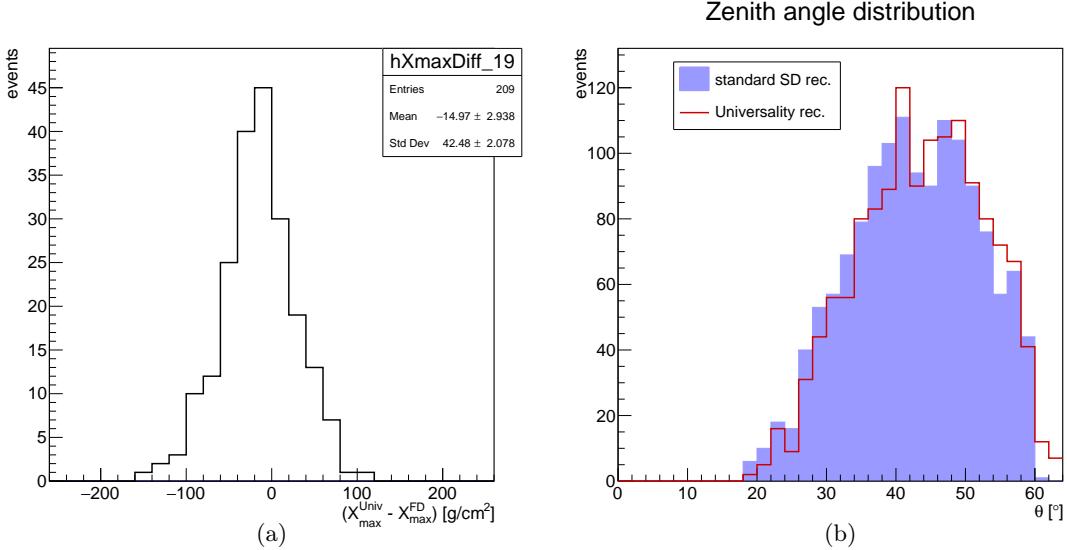


Figure 5.16: (a) Distribution of differences between  $X_{\max}^{\text{FD}}$  from standard hybrid reconstruction and  $X_{\max}^{\text{Univ}}$  from the Universality global fit with reduced bin signal uncertainties. All the other improvements in the minimization procedure are maintained. See Fig. 5.3a for further description. (b) Reconstructed zenith angle for events successfully fitted with the updated Universality procedure. See Fig. 5.3b for further information.

of golden hybrid events previously described. The resulting shower parameters are not expected to substantially change, even if the shape fit will be more constrained to the actual traces. Instead, the output deviances will increase since smaller error in eq. 5.7 with unchanged measurements and (approximately equal) predictions for the bin signals will naturally give larger values of  $D_{\text{shape}}$ .

The efficiency above  $10^{19}$  eV is reduced to  $\sim 92\%$  (209/228 unsaturated events are successfully reconstructed). This reduction was actually expected, since smaller uncertainties makes more difficult for the minimizer to find the Log-Normal functions (i.e. the set of parameters) that can fit in a satisfactory way all the traces at the same time.

In Fig. 5.16a the distribution of  $X_{\max}^{\text{Univ}} - X_{\max}^{\text{FD}}$  obtained from this new version of Universality fit is shown. The mean value ( $\langle X_{\max}^{\text{Univ}} - X_{\max}^{\text{FD}} \rangle = -15.0 \text{ g/cm}^2$ ) and the dispersion (RMS( $X_{\max}^{\text{Univ}} - X_{\max}^{\text{FD}}$ ) =  $42.5 \text{ g/cm}^2$ ) are practically unmodified with respect to the deviance minimization with “quasi poissonian” bin signal errors. At the same time, also the reconstruction accuracy of the other shower parameters is unchanged according to the “*a posteriori*” analysis. For instance the zenith angle determination, shown in Fig. 5.16b, results as good as in the previous version of the Universality procedure. Therefore the differences between hybrid and Universality results for the core position and axis direction are also stable on average.

The new likelihood distributions are reported in Fig. 5.17 for the unsaturated events of higher ( $> 10^{19}$  eV) energies successfully fitted with deviance minimization and reduced bin errors. The distribution of  $D_{\text{LDF}}$  per station remains as good as in the previous version of the reconstruction, as required since this contribution is untouched. Instead,

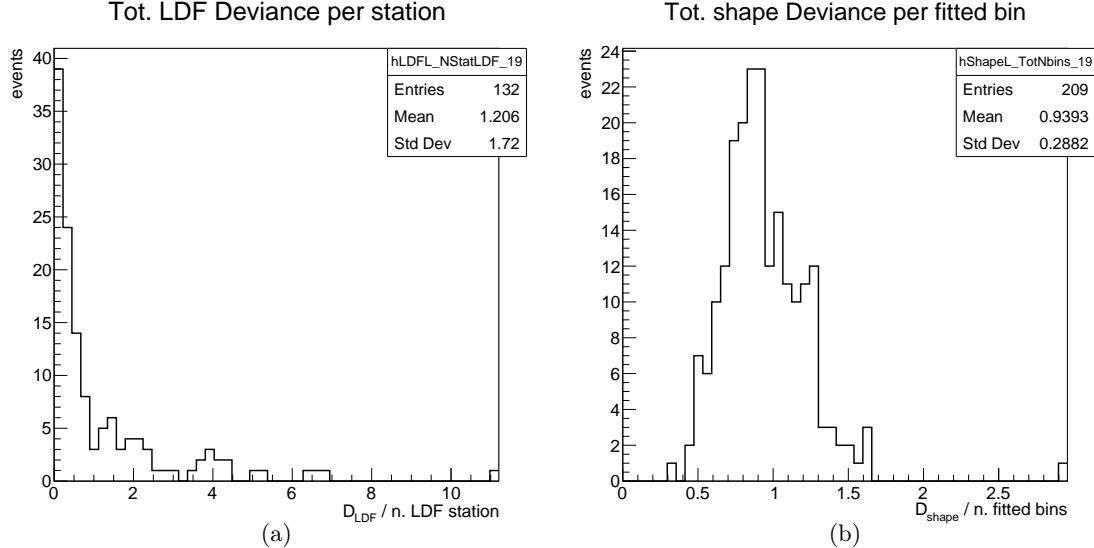


Figure 5.17: (a) Distribution of the ratio  $D_{\text{LDF}}$  over the number of stations used for the LDF fit. (b) Distribution of the ratio  $D_{\text{shape}}$  (sum of the deviance for each station used in the shape fit) over the total number of bins (summed from every used trace). See Fig. 5.7 for more details.

the values of  $D_{\text{shape}}$  per fitted bin are noticeably increased, showing now a distribution centered approximately around 1. Therefore the new uncertainties assigned to the bin signals are of the right order of magnitude, since the effect of errors overestimation is remarkably reduced.

In conclusion, this second change of the Universality fit is supported by the deviance study, in addition to the “a posteriori” distributions. However, the  $X_{\max}$  obtained through this procedure remains biased with respect to the fluorescence reconstructed values; as a consequence, further modifications of the procedure will be attempted before performing any physical analysis, in particular mass composition studies.

## 5.7 A different paradigm for the Universality reconstruction

At this point, additional adjustments in the fit function or in the likelihood terms are apparently insufficient to further improve the Universality procedure. Then the development of a different method to perform the minimization seems necessary.

One characteristic assumption of the Universality reconstruction (as performed until this moment) is the choice of fixing the energy to the SD value ( $E_{\text{SD}}$ ). This choice appears arbitrary, as the other physical quantities are instead fitted inside large intervals (as reported in sect. 5.3). Moreover, the uncertainties on the geometrical parameters obtained through the standard SD procedure are never taken into account.

Then the idea is to introduce a new fitting procedure where all the parameters already determined in the standard SD reconstruction (i.e. axis, core and energy) are fitted, introducing at the same time physical constraints around the SD values. In this way the SD reconstruction will be actually employed as a pre-fit stage for the Universality fitting

procedure, and not only to assign the initial values for the minimization.

The task of constraining the parameters of the Universality fit can be accomplished through additional likelihood terms, where the shift between each quantity and its SD counterpart is compared with a reference resolution obtained from the standard SD reconstruction. Therefore this updated Universality fit, in the following tagged as **constrained reconstruction**, preliminarily requires the determination of such resolutions for the mentioned physical quantities, that are: energy  $E$ ; axis direction  $\vec{a}$ ; core position, in two planar components  $x$  and  $y$ ; core time  $t$ .

### Energy resolution

The SD energy resolution is essentially related with the calibration (see sect. 3.2.2) and on average it is equal, as reported in Tab.1 of [58], to 15% of  $E_{\text{SD}}$  for the set of vertical events with energy higher than  $3 \times 10^{18}$  eV measured by the SD-1500 array. For the purpose of the energy constraint, such value can be assumed constant<sup>18</sup> over the entire energetic range of interest.

### Angular resolution

About the angular resolution of the surface detector, several studies found a dependence on the zenith angle and on the number of tanks interested by the event. Then four angular resolution functions (Fig. 5.18b) are obtained as approximations of the behaviours published in [51] and shown in Fig. 5.18a.

### Core resolution

To evaluate the resolutions of the core position and of the core time, the strategy employed

---

<sup>18</sup>The energy resolution possibly improves for increasing energy but, following a conservative approach, in this work it was preferred to fix its value to the published one (15% of  $E_{\text{SD}}$ ).

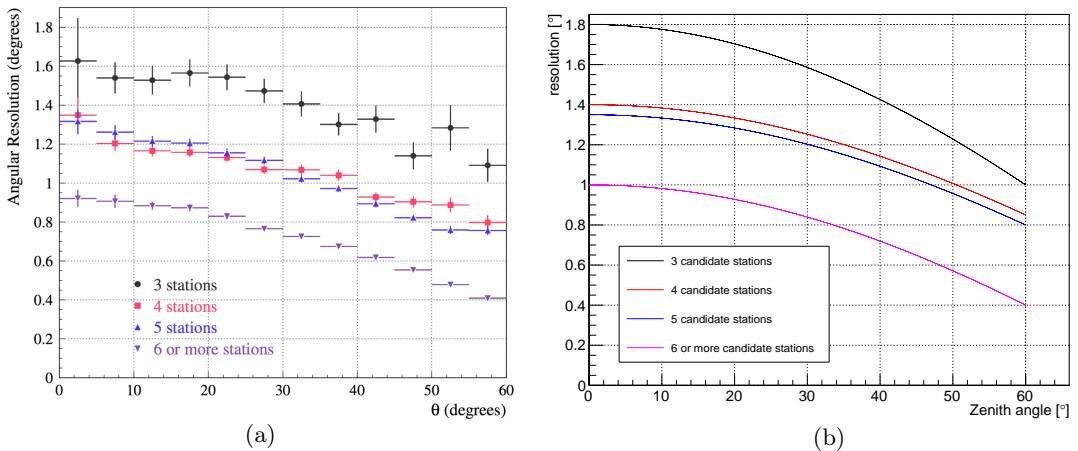


Figure 5.18: (Left) Angular resolution as a function of the zenith angle ( $\theta$ ) for events with an energy above 3 EeV, and for various station multiplicities (taken from [51]). (Right) Angular resolution functions obtained as approximations (linear in  $\cos \theta$ ) of the experimental behaviours.

in this work consists in a comparison between the SD and hybrid reconstructed values. The core time resolution can be directly obtained from the distribution of the core time shift ( $t_{\text{SD}} - t_{\text{hyb}}$ ), while the core position resolution has to be studied separately in two planar components, considering the altitude approximately constant.

However, a well known systematic arises from the comparison of SD and hybrid reconstructed core positions, caused by the azimuthal asymmetry of the signals measured by the ground array [108, 118]. The shower asymmetry does not affect the hybrid reconstruction, but induces a systematic shift<sup>19</sup> of the SD core position toward the “early” region, i.e. the region of the array where detectors record the passage of the shower front earlier (identified in Fig. 5.19 by polar angles  $|\zeta| < \pi/2$  in the shower plane). Moreover, this shift enlarges for increasing zenith angle, until it reaches a maximum (for very large values of  $\theta$ ); then it decreases for more inclined showers. A more detailed explanation about the azimuthal asymmetry effect is given in Appendix B.1.

The shift between SD and hybrid reconstructed core positions can still be easily studied in the shower coordinate system (shown in red in Fig. 5.19). In fact, in this coordinate system the azimuthal asymmetry affects only the  $x$  component of the SD core position, since the  $x$  axis points toward the early region and the  $y$  axis is perpendicular to it. Therefore, the analysis of the differences  $(x_{\text{SD}} - x_{\text{hyb}})$  and  $(y_{\text{SD}} - y_{\text{hyb}})$  between the two (SD and hybrid) core positions projected on the shower plane allows to quantify the effect of azimuthal asymmetry. In particular, a not negligible bias is expected in the  $(x_{\text{SD}} - x_{\text{hyb}})$  distribution, while a null bias should be observed in the  $(y_{\text{SD}} - y_{\text{hyb}})$  one.

This analysis, reported in Appendix B.4, has been accomplished using the selected

---

<sup>19</sup>The shower asymmetry is still not taken into account in the standard SD reconstruction [119, 120].

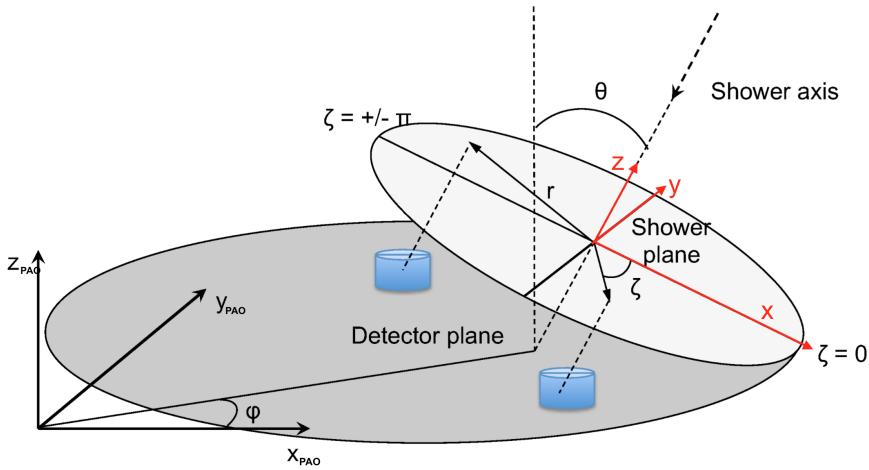


Figure 5.19: Schematic view of the shower geometry, where two different coordinate systems are indicated. On the left, the fixed coordinate system of the Pierre Auger Observatory, where the azimuth ( $\phi$ ) and zenith ( $\theta$ ) of an event are defined. In red, the coordinate system relative to an event (*shower coordinate system*), where the shower plane is perpendicular to the shower axis. Image taken from [118] and modified.

set of golden hybrids (sect. 5.2) with  $E_{\text{SD}} > 10^{19}$  eV. As expected, a bias of  $\sim 23$  m toward the early region is found in the  $x$  component (Fig. B.6a)<sup>20</sup> while the mean of the  $(y_{\text{SD}} - y_{\text{hyb}})$  distribution is compatible with zero (Fig. B.6b). Then the constraint will be centered in a position shifted, with respect to  $x_{\text{SD}}$ , of  $x_0 = -23$  m in the  $x$  axis direction of the shower coordinate system, in order to correct on average the effect of azimuthal asymmetry on the SD reconstructed core. At the same time, the values of the core position resolutions are obtained from the standard deviations of the two distributions of differences, resulting equal to  $\sigma_x = 58.5$  m and  $\sigma_y = 66.5$  m.

A similar analysis has been accomplished for the shift between the SD and hybrid core times, as exposed in Appendix B.5. The result is a bias of about -89 ns in the  $(t_{\text{SD}} - t_{\text{hyb}})$  distribution, again due to the shower asymmetry. Therefore also the core time constraint will be centered around a shifted time with respect to the SD core time of the event (in particular of  $t_0 = 89$  ns) in order to correct the systematic difference on average. The resolution, evaluated from the standard deviation of the  $(t_{\text{SD}} - t_{\text{hyb}})$  distribution, is  $\sigma_t = 215$  ns.

In conclusion, the correct probability distribution has to be chosen for each constraint, based on the statistical behaviour of the related physical quantity:

- for the energy, the constraint is a gaussian distribution with mean equal to  $E_{\text{SD}}$  and dispersion equal to the resolution (15% of  $E_{\text{SD}}$ );
- for the axis, the actually constrained value is the angle between the Universality reconstructed axis and the SD one ( $\widehat{\vec{a}_{\text{Univ}}}, \widehat{\vec{a}_{\text{SD}}}$ ); then the constraint for this angle is a Rayleigh distribution with variance calculated in accordance with the functions plotted in Fig. 5.18b;
- for the core time, the constraint is a gaussian distribution with mean  $(t_{\text{SD}} + t_0)$  (with  $t_0 = 89$  ns) and dispersion  $\sigma_t = 215$  ns;
- for the core position, two separate gaussian constraints on the  $x$  and  $y$  components (calculated on the shower coordinate system) are necessary; the mean values are  $(x_{\text{SD}} + x_0)$  (where  $x_0 = -23$  m) and  $y_{\text{SD}}$ , while the dispersions are  $\sigma_x = 58.5$  m and  $\sigma_y = 66.5$  m.

The corresponding deviance terms, that will be added to the total likelihood in each step of the minimization, are:

$$D_{\vec{a}} = \left( \frac{\widehat{\vec{a}_{\text{Univ}}}, \widehat{\vec{a}_{\text{SD}}}}{\sigma_{\vec{a}}} \right)^2 - 2 \ln \left( \frac{\widehat{\vec{a}_{\text{Univ}}}, \widehat{\vec{a}_{\text{SD}}}}{\sigma_{\vec{a}}} \right) - 1 \quad \text{axis constraint} \quad (5.17)$$

$$D_E = \left( \frac{E_{\text{Univ}} - E_{\text{SD}}}{0.15 E_{\text{SD}}} \right)^2 \quad \text{energy constraint} \quad (5.18)$$

$$D_t = \left( \frac{t_{\text{Univ}} - (t_{\text{SD}} + t_0)}{\sigma_t} \right)^2 \quad \text{core time constraint} \quad (5.19)$$

---

<sup>20</sup>This value is actually the magnitude of the azimuthal asymmetry effect averaged over the zenith angle for the chosen set of golden hybrid events.

$$D_{\vec{c}} = \left( \frac{x_{\text{Univ}} - (x_{\text{SD}} + x_0)}{\sigma_x} \right)^2 + \left( \frac{y_{\text{Univ}} - y_{\text{SD}}}{\sigma_y} \right)^2 \quad \text{core position constraint} \quad (5.20)$$

where  $\sigma_{\vec{c}}$  is the angular resolution as a function of the zenith angle  $\theta_{\text{SD}}$  and of the number of stations of the event (see Fig. 5.18b). The deviance for the axis constraint is obtained from the Rayleigh distribution, following the procedure exposed in Appendix A.1.

It is important to underline that this new approach to the Universality fit, based on the introduction of constraints on all the physical parameters already reconstructed with the standard SD procedure, could lead to different reconstruction outcomes in comparison with the previous methods. In fact, the minimization path covered for each event can be considerably altered by the introduction of the energy as fitted parameter.

## 5.8 Constrained Universality reconstruction

After the changes discussed in the previous section, the Universality reconstruction is performed again on the selected set of golden hybrid events, maintaining deviance minimization and bin signal errors reduction. The efficiency remains high, as 211 over 228 unsaturated events above  $10^{19}$  eV are successfully fitted. However, only 200 events are reconstructed in both the previous method and in the updated approach, meaning that 15 minimizations which previously failed to converge are now successful while 13 events cannot be fitted in the new conditions.

The first important result is a recovery of the bias in the mean value of the  $X_{\text{max}}^{\text{Univ}} - X_{\text{max}}^{\text{FD}}$  distribution: for this new version of Universality fit, it is compatible with zero since  $\langle X_{\text{max}}^{\text{Univ}} - X_{\text{max}}^{\text{FD}} \rangle = (-0.2 \pm 3.5) \text{ g/cm}^2$  (see Fig. 5.20a). On the contrary, the dispersion is negatively affected by the described changes since  $\text{RMS}(X_{\text{max}}^{\text{Univ}} - X_{\text{max}}^{\text{FD}}) = 51.1 \text{ g/cm}^2$  (a value anyway lower than in the original Universality procedure).

About the geometrical parameters, one can notice that the zenith angle distribution (Fig. 5.16b) for the reconstructed events remains noticeably similar to the corresponding SD distribution, as in the previous versions of the Universality fit. At the same time, the other geometrical quantities show small improvements in the reconstruction accuracy according to the “a posteriori” analysis, as shown in the following.

A first quantity of interest is the angle between the hybrid and Universality axis directions. In Fig. 5.21, the distributions of this angular shift are again shown for the previous (deviance minimization with reduced bin uncertainty) and the updated (constrained reconstruction) versions of the Universality fit. The separation between the two shower axes results on average lower in the newer fitting procedure, even if this improvement is more impressive for events below  $10^{19}$  eV.

The shift between Universality and hybrid core positions is instead analysed separately in the two components along the  $x$  and  $y$  axis of the shower coordinate system (Fig. 5.19 in red). More precisely, two “a posteriori” distributions for the core position are built using the differences  $(x_{\text{Univ}} - x_{\text{hyb}})$  and  $(y_{\text{Univ}} - y_{\text{hyb}})$ . In Fig. 5.22 the distributions

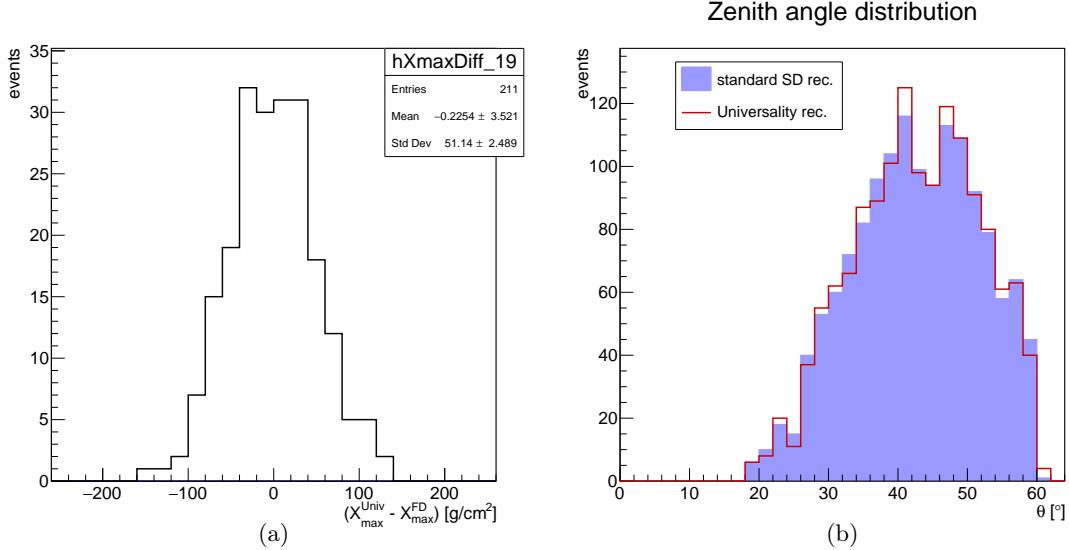


Figure 5.20: (a) Distribution of differences between  $X_{\text{max}}^{\text{FD}}$  from standard hybrid reconstruction and  $X_{\text{max}}^{\text{Univ}}$  for the constrained reconstruction. See Fig. 5.3a for further description. (b) Reconstructed zenith angle for events successfully fitted with the new Universality procedure. See Fig. 5.3b for further information.

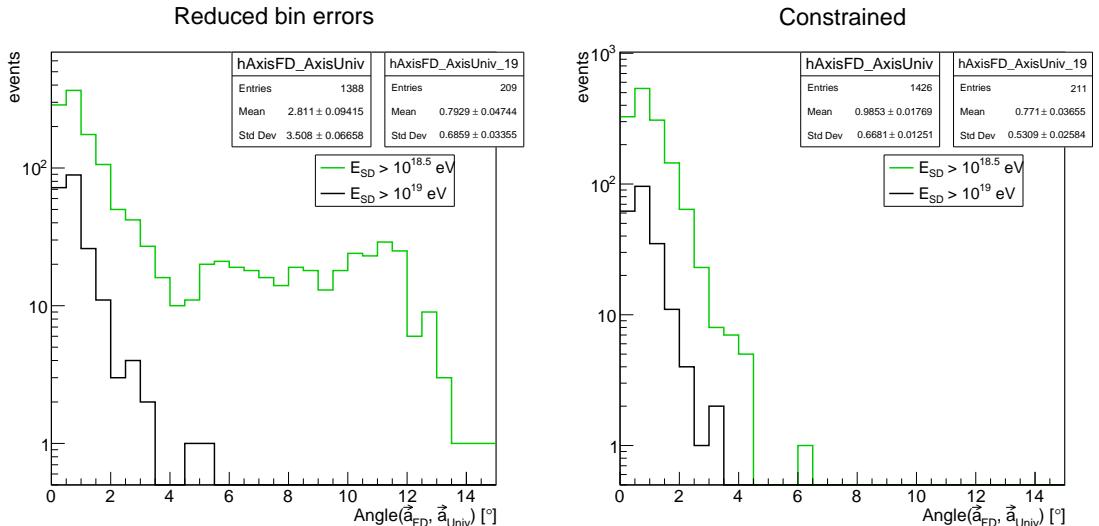


Figure 5.21: Distribution of the angles between the Universality reconstructed axis direction and the corresponding hybrid one (labelled as FD) for the selected set of golden hybrids. On the left, the result for the previous version of the Universality procedure (deviance minimization with reduced bin signal uncertainties); on the right, the result for the constrained reconstruction.

for the previous version of the fit (deviance minimization with reduced bin uncertainty) and for the constrained reconstruction are compared. The mean of the distribution in the  $y$  direction (lower panels) is fully compatible with zero in the updated version of the Universality fit, while it was slightly biased in the previous one. Instead a bias in the  $x$  coordinate is present, as can be seen in the upper panels, and only slightly reduced (from about -11.3 m to -9.7 m). Such negative bias reveals that the Universality fit over-corrects the systematic difference between SD and hybrid core positions due to azimuthal

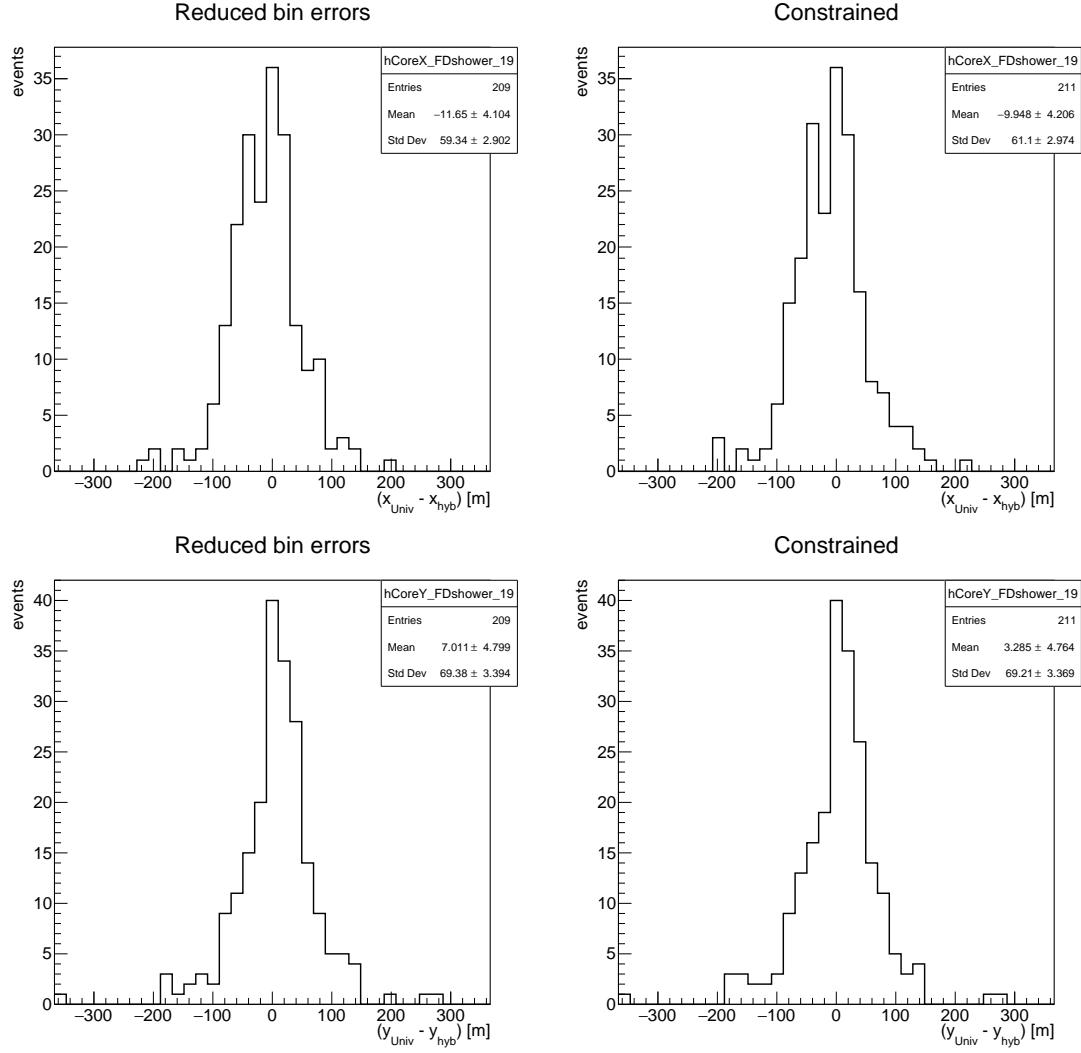


Figure 5.22: Distributions of the differences between Universality and hybrid core positions in the shower coordinate system (see Fig. 5.19). On the upper panels, the differences in the  $x$  component; on the lower panels, the differences in the  $y$  component. On the left, the results for the previous version of the Universality procedure (deviance minimization with reduced bin signal uncertainties); on the right, the results for the constrained reconstruction.

asymmetry. In fact, the shift of  $\sim 23$  m toward the “early” region is replaced by a shift in the opposite direction, even if of half the size. This behaviour is still unexplained.

Another element to consider in the comparison of the constrained Universality reconstruction with the previous procedures is the deviance analysis. The distributions of the shape and LDF deviance terms maintain approximately the same mean values and, more importantly, the same behaviours, as can be seen comparing Fig. 5.23 with Fig. 5.17. Then the improvements achieved with the previous adjustments are preserved.

The next step is the analysis of the newly introduced component of the deviance, due to the constraints on the physical parameters. In particular, the quantity of interest is  $D_{\text{pars}}$ , that is the sum of the four terms ( $D_{\vec{a}}$ ,  $D_E$ ,  $D_t$ ,  $D_{\vec{c}}$ ) defined in the previous section. The weight of this likelihood contribution with respect to the total deviance

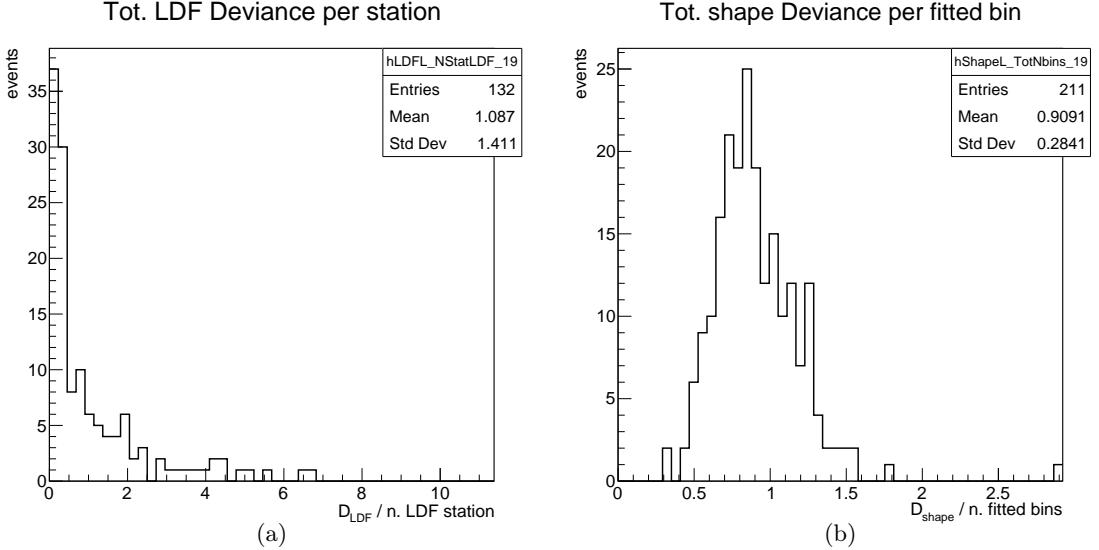


Figure 5.23: (a) Distribution of the ratio  $D_{\text{LDF}}$  over the number of stations used for the LDF fit. (b) Distribution of the ratio  $D_{\text{shape}}$  over the total number of fitted bins. Each entry represents the value for an event successfully reconstructed with the constrained Universality fit. See Fig. 5.7 for more details.

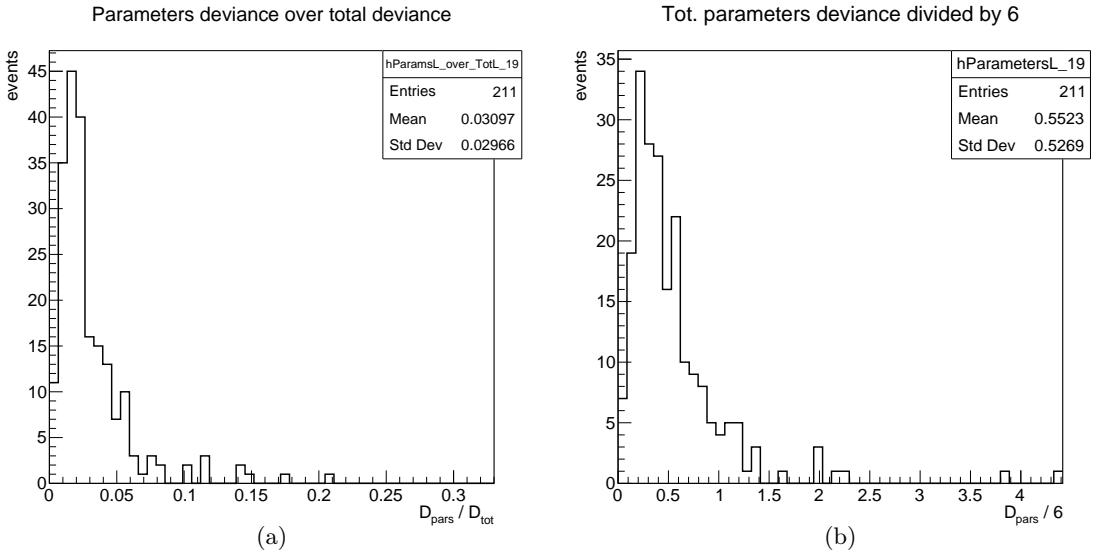


Figure 5.24: (a) Distribution of the ratio  $D_{\text{pars}}/D_{\text{tot}}$ , where  $D_{\text{pars}}$  is the sum of the contributions from the 4 constrains (axis, energy, core time and core position). (b) Distribution of  $D_{\text{pars}}/6$ , where 6 is the number of parameters constrained in the updated Universality fit. Each entry represents the value for an event successfully reconstructed with the constrained fit.

is almost always lower than 5% (see Fig. 5.24a), meaning that the leading term in the minimization procedure reasonably remains the shape deviance.

In Fig. 5.24b the distribution of  $D_{\text{pars}}/6$  is shown; the values are divided by 6, i.e. the number of physical parameters involved in the constraints, to recover a  $\chi^2$  interpretation also in this case. Therefore the value of such ratio quantifies how much the Universality reconstructed parameters differ from the SD counterparts in each event: a value around 1 indicates an average shift of the same magnitude of the resolutions, while lower and

	fitted events	bias( $X_{\max}$ ) [g/cm <sup>2</sup> ]	RMS( $X_{\max}$ ) [g/cm <sup>2</sup> ]	$\frac{D_{\text{LDF}}}{N_{\text{LDF}}}$	$\frac{D_{\text{shape}}}{N_{\text{shape}}}$
<i>UnivKG</i> (sect. 5.3)	225	-2.4 ± 3.7	56.0 ± 2.6	-	-
Deviance (sect. 5.4)	223	-15.5 ± 2.8	41.8 ± 2.0	0.85	0.17
Reduced $\varepsilon(s_{\text{bin}})$ (sect. 5.6)	209	-15.0 ± 2.9	42.5 ± 2.1	1.21	0.94
Constrained (sect. 5.8)	211	-0.2 ± 3.5	51.1 ± 2.5	1.09	0.91

Table 5.1: Recap of the results for the selected set of unsaturated vertical ( $\theta < 60^\circ$ ) golden hybrid events with energy above  $10^{19}$  eV (228 showers in total). The “fitted events” column reports the number of events successfully reconstructed in each Universality method (minimization converged). The bias( $X_{\max}$ ) column refers to  $\langle X_{\max}^{\text{Univ}} - X_{\max}^{\text{FD}} \rangle$ , the mean value of the distribution of differences, while the RMS( $X_{\max}$ ) column refers to the values of  $\text{RMS}(X_{\max}^{\text{Univ}} - X_{\max}^{\text{FD}})$ . The other columns show the average deviance values with respect to a suitable *number of degrees of freedom* in the analogy between  $\chi^2$  and deviance: mean ratio  $D_{\text{LDF}}$  over the number of stations used in the LDF fit ( $N_{\text{LDF}}$ ) for each event with  $N_{\text{LDF}} > 0$ ; mean ratio  $D_{\text{shape}}$  over the total number of bins used in all the trace fits ( $N_{\text{shape}}$ ) for each event.

higher values indicate respectively smaller and larger shifts. The  $D_{\text{pars}}/6$  distribution has a mean value smaller than unity, while its shape corresponds with the expected one for a  $\chi^2$  distribution with a low number of degrees of freedom. This result confirms that the constrained Universality reconstruction (almost) never finds a minimum which is far from the SD one, at least relatively to the employed resolutions.

In tab. 5.1 the quantities of major interest for the comparison of the different versions of the Universality reconstruction exposed in this work are collected.

### 5.8.1 The Universality fitted energy

One fundamental difference among the constrained reconstruction and the previous versions of the Universality procedure is the fit of the energy. Then the “a posteriori” analysis of the reconstructed energy  $E_{\text{Univ}}$  with respect to the hybrid value  $E_{\text{FD}}$ , considered as the “true” energy of each event, has to be accomplished.

A negative bias (-4.6%) is apparent in the distribution of the ratio  $(E_{\text{Univ}} - E_{\text{FD}})/E_{\text{FD}}$  for the selected set of golden hybrids (see Fig. 5.25a) with  $E_{\text{SD}} > 10^{19}$  eV. Moreover the standard deviation of the histogram ( $\sim 22.6\%$ ) is substantially larger than the energy resolution used in the fit constraint<sup>21</sup> (15% of  $E_{\text{SD}}$ ). Since this value of the energy resolution is obtained from the ratio between SD and hybrid energies (see sect. 3.2.2), the values of  $E_{\text{Univ}}$  actually result more dispersed than the SD counterparts with respect to the hybrid reconstructed energy  $E_{\text{FD}}$ .

To further evaluate the reliability of the energy reconstruction, also the differences between Universality and SD energies are studied, using the distribution of the ratio

<sup>21</sup>The intrinsic energy resolution of the hybrid reconstruction (7.6%, see for example [57]) has not been taken into account, but the dispersion of the “a posteriori” distribution would remain in any case larger than the expectation.

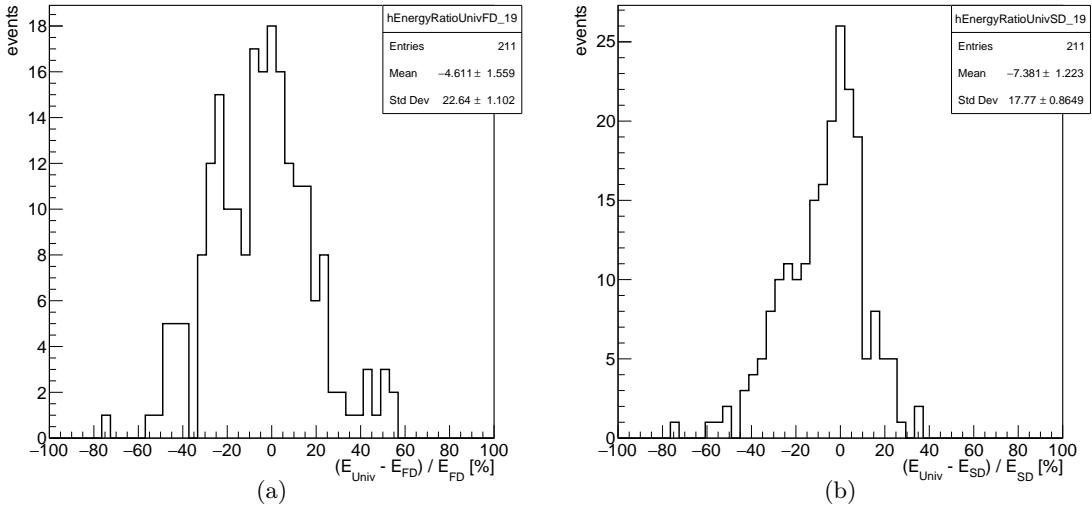


Figure 5.25: (a) Distribution of the ratio  $(E_{\text{Univ}} - E_{\text{FD}})/E_{\text{FD}}$ , that quantifies the energy shift between Universality and hybrid reconstructions. (b) Distribution of the ratio  $(E_{\text{Univ}} - E_{\text{SD}})/E_{\text{SD}}$ , that quantifies the energy shift between Universality and standard SD reconstructions. In both plots, each entry represents the value for a golden hybrid event successfully fitted with the constrained Universality procedure.

$(E_{\text{Univ}} - E_{\text{SD}})/E_{\text{SD}}$ , shown in Fig. 5.25b. A standard deviation slightly larger ( $\sim 17.8\%$ ) than the resolution given as constraint is obtained. However, the relevant result is again the presence of a negative bias (-7.4%), which confirms that the Universality reconstructed energy  $E_{\text{Univ}}$  is on average smaller than the SD one, although the constraint has been centered around  $E_{\text{SD}}$ .

To better understand the effect of the energy constraint and its role in the final outcome of the Universality fit, several constrained reconstructions with different energy resolutions are performed using the same set of golden hybrid events. The results, reported in tab. 5.2, show that the reliability of the energy reconstruction rapidly deteriorates for increasingly wider constraints<sup>22</sup> since both the biases and the dispersions of the “a posteriori” distributions (realized as those in Fig. 5.25) become larger and larger.

Such outcome suggests that the constrained Universality fit is not able to accurately reconstruct the energy. The consequence is evident at the limit of a very large energy resolution, that is when the constraint is removed. As can be observed in Fig. 5.26, the absence of an energy constraint causes the distribution of the Universality-SD energy ratio to get completely scattered, with a majority of the events reconstructed with an energy much smaller than the SD value (-22% on average). As a consequence, the  $(X_{\max}^{\text{Univ}} - X_{\max}^{\text{FD}})$  distribution results also noticeably worse (compare Fig. 5.26b with Fig. 5.20a).

A possible explanation for this behaviour is a degeneracy of the Universality fit due to the (anti-)correlation between energy and  $N_{\mu}$ , which is encoded [107] in the Universality

<sup>22</sup>It was verified that the constraints on the geometrical parameters do not show the same behaviour, meaning that an enlargement of the related resolutions does not substantially worsen the fit results.

resolution	bias hyb.	RMS hyb.	bias SD	RMS SD
[% $E_{SD}$ ]	[%]	[%]	[%]	[%]
15%	-4.6	22.6	-7.4	17.8
20%	-7.6	27.7	-9.6	25.1
30%	-11.0	36.2	-12.1	35.2
50%	-11.8	45.9	-13.1	44.8
$\infty$	-21.0	47.7	-22.0	46.7

Table 5.2: Results for the distribution of the relative energy shifts using the constrained Universality reconstruction with increasingly larger energy resolutions. The selected set of unsaturated vertical ( $\theta < 60^\circ$ ) golden hybrid events with energy above  $10^{19}$  eV (228 showers in total) has been used in all these cases. “bias hyb.” and “RMS hyb.” refers to the bias and standard deviation of the  $(E_{\text{Univ}} - E_{\text{FD}})/E_{\text{FD}}$  distribution, while “bias SD” and “RMS SD” refers to the bias and standard deviation of the  $(E_{\text{Univ}} - E_{\text{SD}})/E_{\text{SD}}$  distribution.

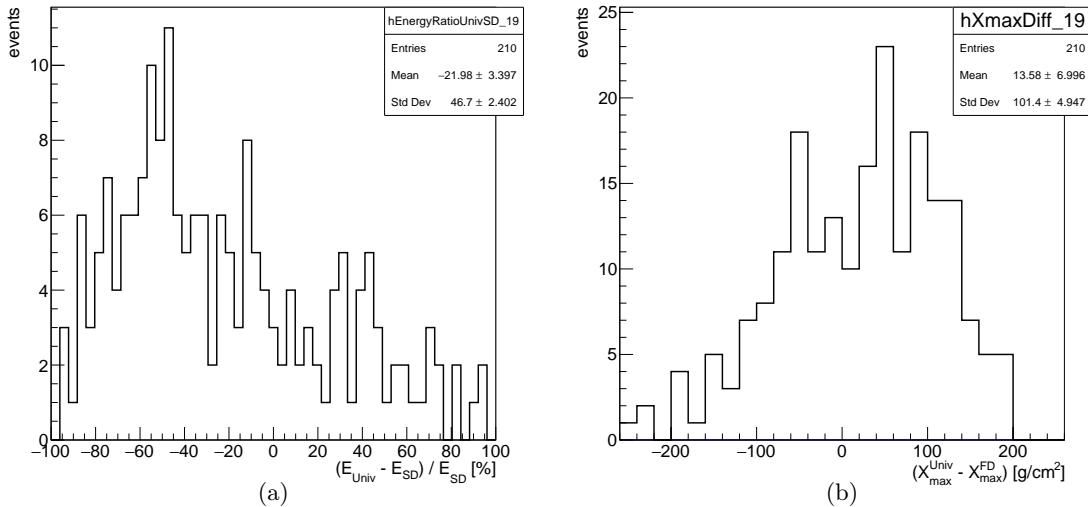


Figure 5.26: (a) Distribution of the ratio  $(E_{\text{Univ}} - E_{\text{SD}})/E_{\text{SD}}$ . (b) Distribution of differences between  $X_{\text{max}}^{\text{Univ}}$  and  $X_{\text{max}}^{\text{FD}}$  from the standard hybrid reconstruction. For both plots, the Universality reconstruction has been performed without energy constraint, while all the other improvements are maintained. See Fig. 5.25b and Fig. 5.20a for more details.

parameterizations<sup>23</sup> (as detailed in the next section) and proven to be effective also in updated studies (see for example [121]).

The distribution of  $r_{E-N_\mu}$ , the correlation coefficient<sup>24</sup> between  $E_{\text{Univ}}$  and  $N_\mu$ , is shown in Fig. 5.27a for the golden hybrids successfully reconstructed with the constrained fit. Since the majority of events presents values of  $r_{E-N_\mu} \approx -1$ , the mentioned anti-correlation between energy and relative muon content is confirmed. The overall effect of this correlation on the Universality minimization is evident in Fig. 5.27b, where each

<sup>23</sup>It should be underlined that this correlation represent an intrinsic limit of the method due to shower physics. In fact, practically the same effect at ground can be produced by a shower of low energy generated by an heavy primary particle (which naturally gives a bigger muonic component) or by a lighter primary of higher energy.

<sup>24</sup> $r_{E-N_\mu}$  is defined as the ratio between the corresponding covariance matrix element and the product of the energy and  $N_\mu$  uncertainties at the minimum.

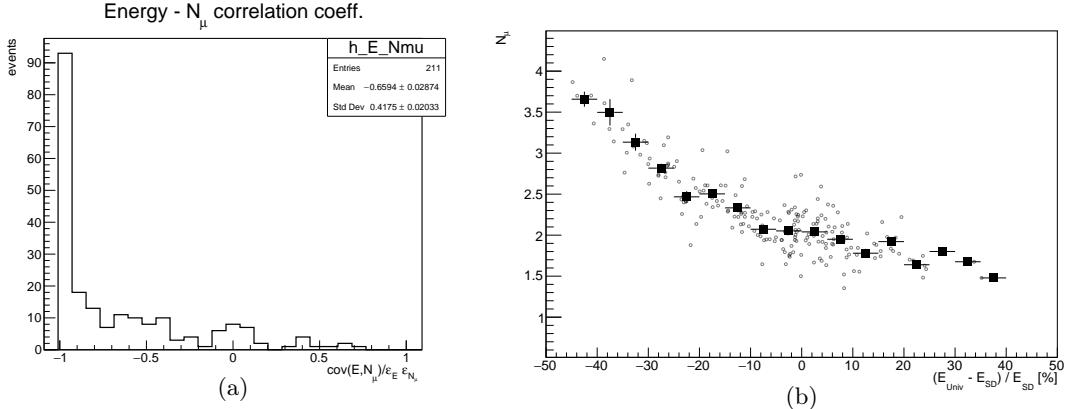


Figure 5.27: (a) Distribution of  $r_{E-N_\mu}$ , correlation coefficient between  $E_{\text{Univ}}$  and  $N_\mu$  separately for each minimization. The values are obtained by the ratio of the covariance matrix element  $\text{cov}(E, N_\mu)$  divided by the product of uncertainties in energy ( $\varepsilon_E$ ) and  $N_\mu$  ( $\varepsilon_{N_\mu}$ ) at the minimum. (b) Scatter plot of the energy shift between Universality and SD reconstructions versus the fitted  $N_\mu$  values. Each small empty dot represents a unsaturated golden hybrid event successfully fitted, while the squares are the mean  $N_\mu$  value for each bin in the energy shift, wide 5%.

event successfully fitted with the constrained reconstruction is represented as a small empty dot placed in correspondence of its reconstructed  $N_\mu$  and relative energy shift  $(E_{\text{Univ}} - E_{\text{SD}})/E_{\text{SD}}$  values. The average behaviour (represented by the black profile) clearly shows that events with larger fitted values of  $N_\mu$  have also a final energy  $E_{\text{Univ}}$  smaller than the SD value  $E_{\text{SD}}$ , and viceversa.

The presence of a strong correlation between energy and  $N_\mu$  is visible also in single events, where different  $(E_{\text{Univ}}, N_\mu)$  couples can produce nearly the same prediction for the signals at ground level. Such behaviour is evident in Fig. 5.28, where the total deviance around the minimum<sup>25</sup> for a typical event is shown as a function of  $E_{\text{Univ}}$  and  $N_\mu$ . To simplify the interpretation of this plot, the two parameters (energy and relative muon content) are expressed in terms of the relative shift from the value at the minimum (respectively  $E^{\min}$  and  $N_\mu^{\min}$ ) found through the constrained reconstruction, i.e. using the quantities  $(E - E^{\min})/E^{\min}$  and  $(N_\mu - N_\mu^{\min})/N_\mu^{\min}$ .

From this figure, observing the colour scale, one can notice that in a large region around the minimum (that is the point (0,0) in this representation) the differences in the total deviance are smaller than unit<sup>26</sup>. Therefore the minimum is actually degenerate, meaning that the reconstructed values of the parameters (in particular of energy and relative muon content) found by the constrained reconstruction are not well defined. In fact, a 10% shift on the energy value  $E_{\text{Univ}}$  could be accomplished with negligible effect on the deviance (and therefore on the other reconstructed parameters) if the  $N_\mu$  value is

<sup>25</sup>The value of  $D_{\text{tot}}$  in each point of Fig. 5.28 is obtained keeping the values of the geometrical parameters and of  $X_{\text{max}}^{\text{Univ}}$  fixed to the values reconstructed in the constrained fit, and changing with 1% steps the values of  $E_{\text{Univ}}$  and  $N_\mu$  inside a  $\pm 25\%$  range. The values of  $D_{\text{tot}}$  are calculated, meaning that the fit is not performed again in each point.

<sup>26</sup>Since the typical values of  $D_{\text{tot}}$  are in the order of few hundreds, the differences among the values in the darkest blue region of Fig. 5.28 are in the order of few %.

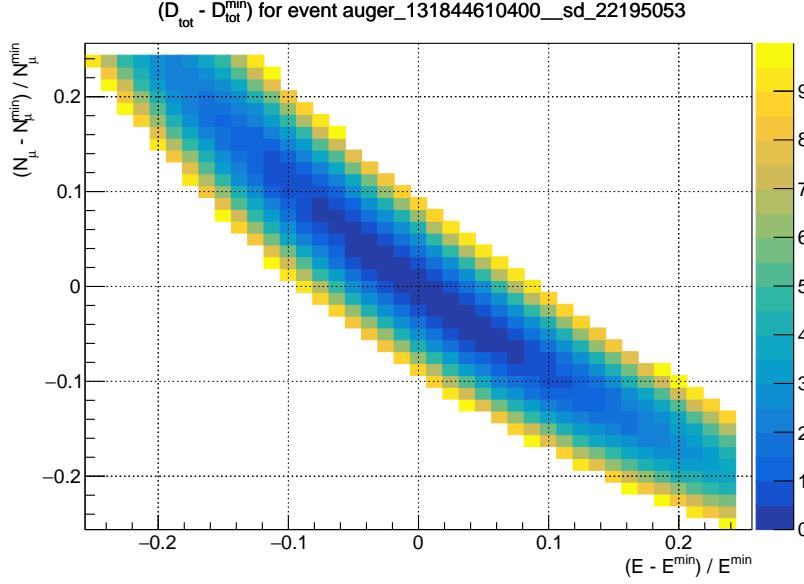


Figure 5.28: Representation of the shift in total deviance from the minimum ( $D_{\text{tot}} - D_{\text{tot}}^{\text{min}}$ ) as a function of energy and  $N_\mu$  (all the other parameters are fixed to the values found at the minimum) for an unsaturated golden hybrid event with  $E_{\text{SD}} > 10^{19}$  eV successfully fitted with the constrained reconstruction.

shifted correspondingly.

Then it is clear why the energy determination is heavily led by the constraint: if the deviance around the minimum is almost flat, even the deviance contribution  $D_{\text{pars}}$  due to the constraints (sect. 5.7) become relevant in the minimization procedure, despite its relatively small magnitude (see Fig. 5.24a). In particular, a small difference in the energy constraint (and thus in the values of  $D_E$ ) induced by a change of the energy resolution can lead to substantially different reconstruction outcomes, as deduced from the results in tab. 5.2, even if the preference toward values  $E_{\text{Univ}} < E_{\text{SD}}$  remains unexplained.

One can conclude that the reconstructed energy  $E_{\text{Univ}}$  excessively depends on the assumption for the energy resolution, while the lack of an energy constraint leads to results in evident tension with the standard reconstruction.

### Possible interpretation

The origin of the (anti)correlation between energy and  $N_\mu$  can be found analysing the parameterization of the integrated signal, exposed in sect. 4.3. In fact, the shapes of the Log-Normals (i.e. the  $m$  and  $s$  parameters) do not depend on  $N_\mu$  and only weakly on  $E$ . Therefore energy and relative muon content affect the fit only through the integrated signals, that are used as normalization factors of each Log-Normal.

Such signals can be demonstrated to be at least partially function of the product  $N_\mu \times E_{\text{Univ}}$  (instead of being function of the two quantities separately), clearly leading to a strong anti-correlation during the minimization.

Using eqs. 4.8, the predicted signal in an ideal detector can be re-written as:

$$S(r, \Delta X, E) = S_{e\gamma} + S_\mu + S_{e\gamma(\mu)} + S_{e\gamma(hadr)} \quad (5.21)$$

where each component signal, following eqs. 4.9 and 4.12, is modeled as:

$$S_i(\Delta X, r, E) \propto K_i(\Delta X, r) \left( \frac{E}{10^{19} \text{ eV}} \right)^{\gamma_i(r)} [1 - (1 - N_\mu) \alpha_i(r)] \quad (5.22)$$

and the terms that do not depend on energy and/or  $N_\mu$  are contained in the factor  $K_i$ <sup>27</sup>.

The values of the parameters  $\alpha_i$  and  $\gamma_i$  are respectively:

- $\alpha_{em} = -0.075$ , thus the pure electromagnetic component is almost independent from the muon content.
- $\alpha_\mu = \alpha_{em(\mu)} = 1$ , meaning that the muon and e.m. from muon components strongly depend on the  $N_\mu$  value.
- $\alpha_{em(had)}$  is actually a function of the station distance from the core ( $r$ ), but in the range of interest for the Universality fit assumes only values approximately between 1.20 and 1.25 (in the following  $\alpha_{em(had)} = 1.2$  is assumed).
- $\gamma_{em} \approx 1$ , neglecting a small dependence on  $r$ ;
- $\gamma_\mu$  and  $\gamma_{em(\mu)}$  are equal and assume values between [0.92, 0.94] as functions of  $r$ ; in first approximation, a value of 0.93 is assumed.
- $\gamma_{em(had)}$  ranges in the same interval of the muonic exponents, but with different values as function of  $r$ ; also in this case a value of 0.93 is assumed.

Finally, in a first approximation, one can write:

$$S_{em} \propto K_{em}(\Delta X, r) (0.925 - 0.075N_\mu) \left( \frac{E}{10^{19} \text{ eV}} \right) \quad (5.23)$$

$$S_\mu \propto K_\mu(\Delta X, r) N_\mu \left( \frac{E}{10^{19} \text{ eV}} \right)^{0.93} \quad (5.24)$$

$$S_{em(\mu)} \propto K_{em(\mu)}(\Delta X, r) N_\mu \left( \frac{E}{10^{19} \text{ eV}} \right)^{0.93} \quad (5.25)$$

$$S_{em(had)} \propto K_{em(had)}(\Delta X, r) (1.2N_\mu - 0.2) \left( \frac{E}{10^{19} \text{ eV}} \right)^{0.93} \quad (5.26)$$

From these expression it is evident that three over four components (i.e. the muonic, e.m. from muon and e.m. from hadrons) are roughly functions of the product of energy and  $N_\mu$ , apart from the dependencies on  $\Delta X$  and  $r$ . As a consequence, a large fraction of the predicted signals<sup>28</sup> directly depends on the product  $N_\mu(E/10^{19} \text{ eV})$ , and not separately on the two quantities. Then it is clear that similar integrated signals are predicted

<sup>27</sup> Actually the pure electromagnetic and e.m. from hadron components present also other terms with a small energy dependence, that anyway can be neglected in this first-order analysis.

<sup>28</sup> A station very near to the shower core is actually dominated by the electromagnetic signal. However, the muonic and the pure e.m. components have the same magnitude already at core distances around 1000 m, and the former grows with increasing distance. Therefore in high energy events, which usually hit several stations, the muonic component plays a main role in most of the recorded signals.

for couples of  $E$  and  $N_\mu$  which present nearly the same product.

In conclusion, the results discussed in this section support the idea that the constrained Universality reconstruction is still unable to provide a reliable determination of the shower energy (and in general of the non-geometrical quantities).

The preference toward energies  $E_{\text{Univ}}$  lower than the SD value (see Fig. 5.25 and 5.26) is still unexplained, although it is a probable indication of internal tensions in the current Universality parameterizations. Indeed, it is known that simulations do not correctly reproduce air showers measurements at ground level. In particular, in the framework of the Pierre Auger Observatory, the average hadronic shower signal is found [63] to be  $1.33 \pm 0.16$  ( $1.61 \pm 0.21$ ) times larger than what predicted using the LHC-tuned model EPOS-LHC (QGSJetII-04). As a consequence, the energy estimator  $S(1000)$  obtained from simulated events is systematically shifted toward lower values. Since the Universality parameterizations were built using simulated showers, the observed disagreement might be due to a discrepancy between measured and simulated signals, not effectively factorised out in the current Universality method by means of the  $N_\mu$  parameter.

To go beyond these limits, an optimization of the Universality model is required<sup>29</sup>. In particular, improved parameterizations should allow to remove the observed biases in the reconstructed quantities, even if the degeneracy of the minimum (mostly due to the energy- $N_\mu$  correlation) would probably still be present. For this purpose, data from scintillator detectors (SSDs of the *AugerPrime* upgrade, sect. 3.4) are instead necessary to disentangle the muonic content of the shower and then possibly remove the mentioned degeneracy.

---

<sup>29</sup>A remake of the parameterizations, to extend the energy range of validity and to add the description of the AMIGA muon detectors (sect. 3.3.2) is currently under-way, thanks to the Karlsruhe group of the Pierre Auger Collaboration.



## Chapter 6

# The Universality reconstruction of the SD-1500 dataset

The previous chapter dealt with the description and improvement of the Universality fit, a new method developed in the framework of the Pierre Auger Collaboration to reconstruct the macroscopic parameters of an EAS using only the measurements gathered by the surface detector. The goal is the extraction of mass-sensitive quantities from a dataset much larger than the hybrid one, employed in the current mass composition studies. In this chapter, the reconstruction of the entire set of events measured by the surface detector (SD-1500) up until the end of 2016 will be accomplished.

As a first step, it is necessary to establish the final version of the Universality procedure to be employed in the reconstruction of the SD-1500 dataset. Considering the reconstruction issues highlighted in the previous chapter, a simplified but faster fitting procedure will be defined in the first section, where the results for the updated golden hybrid dataset (i.e. adding the events between 2013 and 2016) are also shown.

In the second section, the SD-1500 dataset that will be used in the following analysis is defined, pointing out the preliminary event selection and the quality cuts to be applied after the Universality reconstruction.

In the last section of this chapter, the study of the  $X_{\max}$  distributions as functions of the energy is accomplished. Initially the mean  $\langle X_{\max} \rangle$  and the dispersion  $\sigma(X_{\max})$  as functions of the energy are presented. Then a more precise information, even though model dependent, about the primary composition is obtained from the study of the average logarithmic mass  $\ln A$  and its dispersion. Finally, a fit of the shape of the  $X_{\max}^{\text{Univ}}$  distributions using a 4-masses scenario is performed, in order to estimate the fractions of light, intermediate and heavy primary nuclei reaching the Earth.

### 6.1 A simplified Universality fit

At the end of the previous chapter (sect. 5.8.1), it was shown that the current version of the Universality fit is still unable to accomplish a second-level event reconstruction. In

particular, the energy cannot be fitted together with the other shower parameters since the accuracy (and reliability) of the reconstruction would result heavily affected. Then a step back is required, meaning that the energy  $E_{\text{Univ}}$  will be again fixed to the value  $E_{\text{SD}}$  obtained from the standard SD reconstruction. The other changes introduced to improve the Universality fit are maintained, in particular:

- the deviance minimization, introduced in sect. 5.4;
- the reduction of the bin signal uncertainties, as defined in sect. 5.5.3;
- the core constraints in  $x$ ,  $y$  and  $t$ , described in sect. 5.7.

Furthermore, the reconstruction will be simplified reducing the number of fitted parameters, in particular fixing  $\theta$  and  $\phi$  to the values obtained from the SD fit. This choice is considered reasonable since the angle between the Universality reconstructed shower axis and the hybrid one is almost always smaller than the known angular resolution (defined in sect. 5.7) as shown in Fig. 5.21, thus the fit of  $\theta$  and  $\phi$  does not add relevant information to the reconstruction. Instead, the core position and core time are left as fitted parameters since the Universality procedure corrects to some extent the systematic shift of the SD core position due to azimuthal asymmetry.

In this Universality reconstruction (tagged as “simplified” in the following) the minimization is accomplished in a shorter amount of time but the results are equivalent to the version described in sect. 5.6, i.e. before the constrained Universality fit. In particular, the core parameters are well reconstructed and the bias in the  $X_{\text{max}}^{\text{Univ}}$  distribution is still present, as can be seen in Fig. 6.1a (compare with Fig. 5.16a) for the golden hybrid dataset used in the previous chapter (described in sect. 5.2).

Since the analysis of the mass composition will be accomplished on the SD-1500 dataset collected before January 2017 (i.e. the set of events employed for the study of the energy spectrum in [58]), also the set of golden hybrids to use in the evaluation of the Universality reconstruction performances has to be updated. In fact in the previous chapter the analysis was performed using the set of golden hybrids collected before 2014, available at the beginning of my PhD work. However, the energy calibration of the surface detector has been revised between the 33rd and the 35th ICRC<sup>1</sup> (i.e. between 2015 and 2017), mainly adding the golden hybrids collected between 1st January 2014 and 31st December 2016.

The unsaturated golden hybrids with energy above  $10^{19}$  eV and zenith  $\theta < 60^\circ$  used for the updated (ICRC-2017) energy calibration are 334. Applying the simplified Universality reconstruction over this set of events, a good efficiency is maintained (294/334 events are successfully fitted). However, a difference arises in the study of  $X_{\text{max}}^{\text{Univ}}$  with respect to the previous golden hybrid dataset. In fact, as can be seen in Fig. 6.1b, the bias has increased, reaching  $-25 \text{ g/cm}^2$ .

This noticeable shift is probably due to the changes in the energy scale and in the

---

<sup>1</sup>International Cosmic Ray Conference.

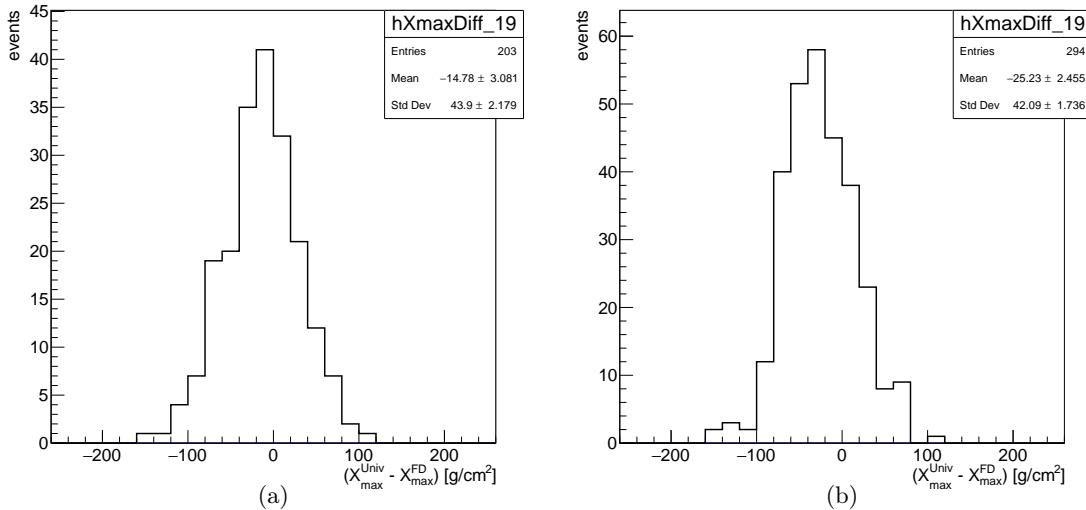


Figure 6.1: Distribution of differences between  $X_{\max}^{\text{FD}}$  from standard hybrid reconstruction and  $X_{\max}^{\text{Univ}}$  for the simplified version of the Universality fitting procedure. On the left, the result for the unsaturated events with  $E_{\text{SD}} > 10^{19}$  eV from the old golden hybrid dataset (used for ICRC-2015 calibration and described in sect. 5.2). On the right, the result for the updated dataset (used for the ICRC-2017 calibration).

energy calibration, performed between 2015 and 2017. In fact, several improvements in the hybrid reconstruction [58] caused an increase in the FD energy of few %. At the same time, the quality cuts employed to define the golden hybrid dataset for the calibration have been slightly modified, causing an additional change in the calibration parameters<sup>2</sup>.

Since the Universality energy is fixed to the SD value, a different energy calibration is expected to produce evident changes in the reconstruction of the non-geometrical<sup>3</sup> parameters (i.e. shower maximum depth and  $N_{\mu}$ ) because of the unresolved issues of the Universality model, conjectured in sect. 5.8.1.

## 6.2 The SD-1500 dataset

The dataset employed in this work is a subset of the complete SD-1500 vertical ( $\theta < 60^\circ$ ) dataset employed in the realization of the energy spectrum presented at the 35th ICRC [58]. In particular, only events with energies above  $10^{19}$  eV measured between 1st January 2004 and 31st December 2016 are included, for a total of 13966 shower.

### 6.2.1 Event selection

To assure a good quality in the Universality reconstruction, additional selection cuts have to be applied to the above-mentioned dataset.

---

<sup>2</sup>Both this effects are anyway well below the systematic uncertainty in the energy scale, estimated around 14% in [59] and confirmed in [58].

<sup>3</sup>The geometrical quantities, in particular the core position and core time, are instead reconstructed with the same accuracy as for the ICRC-2015 golden dataset.

## Saturation

As explained in sect. 5.1, the presence of a saturated station greatly reduces the power of the Universality fitting procedure: for instance, a saturated trace cannot be used in the shape fit, which is fundamental for the reconstruction of the non-geometrical parameters ( $X_{\max}$  and  $N_{\mu}$ ). Then events with a saturated station (representing around 30% of the dataset) are excluded from the following analysis.

However, the fraction of events excluded grows with energy, as the probability of saturation also increases with energy: for instance, around 50% of the events above  $10^{19.5}$  eV present a saturated signal. Therefore this selection cut clearly represents a limit of the current Universality reconstruction, since it greatly reduces the available statistics in the higher energy range, which is of major interest for the extension of mass-composition studies.

Such issue will be solved within the upgrade of the Pierre Auger Observatory (*Auger-Prime*), since each WCD will be equipped with a small-PMT (see sect. 3.4) extending the dynamic range of signal acquisition to more than 30,000 VEM and thus allowing the measurement of traces without saturation in almost all the events. However, the implementation of a method to efficiently include the saturated signals in the Universality fit is a fundamental point to be addressed for any future analysis that would use the measurements recorded before the upgrade.

## Field of view

A known issue of the present Universality model (see sect. 4.3) is that only simulations where the maximum is reached above the ground level of the Pierre Auger Observatory ( $\approx 875 \text{ g/cm}^2$ ) have been used to build the parameterization. Therefore the class of events with deeper  $X_{\max}$  values is under-represented in this set of simulation, resulting in a loss of predictability power for the Universality reconstruction. In fact, the reconstructed  $X_{\max}^{\text{Univ}}$  value of events with a maximum near or below ground is expected to be underestimated, since (using a metaphor) the Universality method only contains the features of showers with  $X_{\max}$  above ground.

Since lighter nuclei develop deeper in the atmosphere with respect to heavier ones, a larger fraction of light-initiated showers are excluded from the parameterizations with respect to heavier ones. As a consequence, the Universality reconstruction is expected to be less reliable (i.e. to show larger biases) for showers initiated by lighter nuclei (as observed in [111] using simulated events). The net result is a negative bias<sup>4</sup> in the  $X_{\max}^{\text{Univ}}$  distribution, which critically depends on the primaries composition.

This issue is present at any energy, but affects less the events with large zenith angles. In fact, given a certain primary mass and energy, the probability of a shower to develop

---

<sup>4</sup>Such bias is completely different from the bias observed in the  $(X_{\max}^{\text{Univ}} - X_{\max}^{\text{FD}})$  distribution for the golden hybrid dataset, as explained in the following.

with a maximum above ground (and then to have been included in the parameterizations) grows with increasing zenith angle. Then it should be possible to define a range in the zenith angle where the bias is minimal.

In this work, the reconstruction is considered reliable at a certain energy and zenith angle if a proton-initiated shower has at least 90% probability to develop with a maximum above ground<sup>5</sup>. To evaluate the zenith angle (function of the energy) at which proton-initiated showers have exactly 90% probability to develop with  $X_{\max}$  above ground, a parametric model for the shower maximum distribution in the form of a generalized Gumbel distribution  $g(X_{\max}|E, A)$  [122] is employed. The Gumbel parameterizations have been determined with CONEX shower simulations and provide a reasonable description of the  $X_{\max}$  distribution in a wide energy range. In particular, the updated parameterization described in [123] will be used.

The result (red line in Fig. 6.2) shows that the mentioned issue (and the related bias) is effective for angles approximately<sup>6</sup> below  $20\text{--}25^\circ$ . Then the magnitude of such bias cannot be directly evaluated on data using the golden hybrid dataset. In fact, because of the FD field of view cut [53], showers with zenith below  $\approx 20^\circ$  are not selected for the hybrid dataset. Instead, the set of events measured by the surface detector is heavily affected, presenting a flat (in  $\cos^2 \theta$ ) distribution.

The solution to un-bias the  $X_{\max}^{\text{Univ}}$  result for the following analysis of the SD-1500 dataset is therefore to exclude the events with zenith angle below  $\theta_{\text{cut}}$  defined as in Fig. 6.2 (i.e. the angle where at least 90% of the proton shower develop with  $X_{\max}$  above ground) which is parameterized as a function of the energy<sup>7</sup>. This selection causes another relevant reduction of the statistics, since about 20% of the unsaturated events are excluded. Moreover, this cut is again more severe for increasing energy (since the minimal angle grows with energy), representing an additional limit of the current Universality reconstruction for the study of  $X_{\max}$  at the highest energies.

Finally, only events with  $\theta < 55^\circ$ , independently from their energy, will be included for the following analysis. This cut allows to remove many poorly reconstructed events<sup>8</sup> and to reduce biases<sup>9</sup>.

<sup>5</sup>In the same conditions, showers initiated by heavier primaries have a larger probability to present a maximum above ground, since proton-initiated showers develop deeper in the atmosphere.

<sup>6</sup>The behaviour in Fig. 6.2 is obtained from an average between the results for 3 different hadronic interaction models used to generate the Gumbel distributions.

<sup>7</sup>In particular, the parameterization for  $\theta_{\text{cut}}(E)$  is:

$$\theta_{\text{cut}} = a + b \lg E + c (\lg E)^2 + d (\lg E)^3 \quad (6.1)$$

with  $\lg E = (\log_{10}(E/eV) - 19)$ ,  $a = 21.26^\circ$ ,  $b = 0.77^\circ$ ,  $c = 8.80^\circ$  and  $d = -4.50^\circ$

<sup>8</sup>It was noticed that the uncertainties in the  $X_{\max}^{\text{Univ}}$  determination suddenly increase for zenith angles close to  $60^\circ$  (in some cases the relative uncertainty  $\epsilon(X_{\max}^{\text{Univ}})/X_{\max}^{\text{Univ}}$  is more than 100%). Instead, events with  $\theta < 55^\circ$  present a relative uncertainty always well below 20%.

<sup>9</sup>In sect. 4.3 it was mentioned that the ground signals are reproduced less accurately by the Universality parameterization at large zenith angles. In effect, a larger  $X_{\max}^{\text{Univ}}$  bias is found for events with  $\theta < 55^\circ$  with respect to the golden hybrid dataset as a whole. This issue is likely due to the stronger attenuation of signals in more horizontal events: the shower maximum is further from ground and the

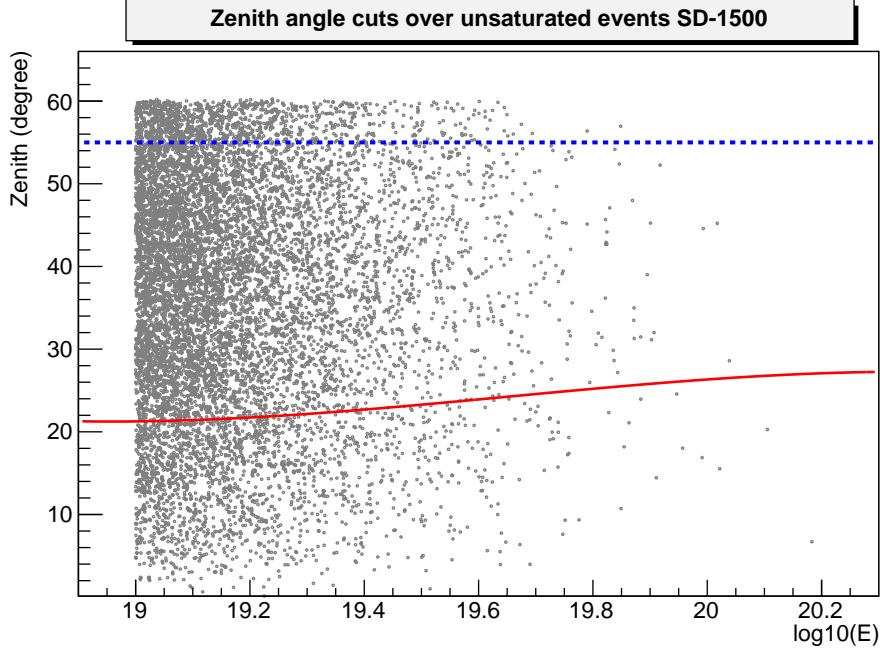


Figure 6.2: Representation of the zenith angle cuts applied on the SD-1500 dataset to obtain an un-biased evaluation of the  $X_{\max}^{\text{Univ}}$  distributions. The continuous red line shows the low zenith cut according with eq. 6.1, while the dotted blue line shows the cut at  $\theta = 55^\circ$  for inclined events. The points represent the unsaturated events measured by the surface detector with an energy  $E > 10^{19}$  eV and used for the ICRC-2017 energy spectrum.

### 6.2.2 Quality cuts

The simplified Universality reconstruction is then applied to the selected SD-1500 dataset, i.e. to 6854 unsaturated events with  $E > 10^{19}$  eV and zenith angle between  $\theta_{\text{cut}}(E)$  and  $55^\circ$ . The minimization converges around 90% of the times, as expected from the efficiency observed for the unsaturated golden hybrids (sect. 6.1).

At this point, a quality selection of the reconstructed events has to be performed. In particular, the goal is to exclude outliers, i.e. events which are successfully fitted but present unacceptable values of the reconstructed physical quantities (usually near the boundaries given in the minimization).

#### Minimum predicted pure electromagnetic signal in the hottest tank

It was noticed that events reconstructed with very small  $X_{\max}^{\text{Univ}}$  values (i.e. near the low boundary of  $200 \text{ g/cm}^2$ ) also present negligible predicted signals for the pure electromagnetic component, even in the station with larger signal (usually called *hottest* tank). In each event the hottest tank is the detector closest to the core and it is inevitably hit by the e.m. part of the shower, meaning that a null e.m. pure signal is quite unrealistic.

Therefore a large fraction of outliers is removed requiring the predicted pure e.m. sig-

---

electromagnetic part is largely absorbed before reaching the detectors.

cut	SD-1500	golden hybrids
	$E_{\text{SD}} > 10^{19}$ eV	$E_{\text{SD}} > 10^{18.5}$ eV
ICRC-2017	13966 (100%)	2573 (100%)
unsaturated	9767 (69.9%)	2105 (81.8%)
$\theta < 55^\circ$	8691 (89.0%)	1920 (91.2%)
$\theta > \theta_{\text{cut}}(E)$	6854 (78.8%)	1885 (98.2%)
fit converged	6115 (89.2%)	1767 (93.7%)
$\hat{S}_{\text{e.m.pure}}^{\text{hottest}} > 1$ VEM	6064 (99.2%)	1766 (99.9%)
$D_{\text{tot}}/N_{\text{tot}} < 4$	6050 (99.8%)	1766 (100%)
surviving %	6050 (43.3%)	1766 (68.6%)

Table 6.1: Table of event selections and quality cuts for the datasets used in this work. The simplified Universality reconstruction (sect. 6.1) is employed. The percentages refer to the previous line, except the last ones which refer to the initial number of events.

nal in the hottest tank ( $\hat{S}_{\text{e.m.pure}}^{\text{hottest}}$ ) to be larger than 1 VEM<sup>10</sup>. In this way, also several events reconstructed with unrealistic couples of  $X_{\text{max}}^{\text{Univ}}$  and  $N_\mu$  (small values of the former compensated by large values of the latter) are rejected.

### Total deviance cut

A second quality selection is accomplished using the deviance values. In particular, the ratio between the total deviance of the event  $D_{\text{tot}}$  (sum of the shape and LDF components, plus the contribution due to the core constraints) and the total number of “degrees of freedom”  $N_{\text{tot}}$  of the fit (given by the sum of the bins employed in the shape fit of each trace plus the number of stations employed in the LDF fit) is used.

Large values of this ratio can be naturally related with an inaccurate result for the fit, recalling the interpretation of the deviance as a generalized  $\chi^2$ . Maintaining a conservative approach<sup>11</sup>, the selection is performed rejecting events with  $D_{\text{tot}}/N_{\text{tot}} > 4$ .

The overall result of event selection, reconstruction and quality cuts is shown in Table 6.1. As can be noticed, only 43% of the initial SD-1500 events (above  $10^{19}$  eV) have survived. Such great reduction is almost totally due to the event selection performed before the Universality reconstruction: the requirement of unsaturated events and the field of view cut immediately halve the available dataset. Moreover, these two selection cuts become more severe for increasing energy, making difficult an actual extension of the mass-composition analysis at the highest energies.

To solve these issues, a revision of the Universality parameterization (to include

---

<sup>10</sup>This quantity is smaller than 1% of the average signals released in the hottest tank of a shower at energies above  $10^{19}$  eV, so the probability of rejecting not-outlier events is considered very low.

<sup>11</sup>The idea is to not affect the golden hybrids, which are considered good quality events and should not need additional quality cuts. Since the selected set of unsaturated golden hybrids from ICRC-2017, successfully fitted with the simplified Universality procedure, does not present values of  $D_{\text{tot}}/N_{\text{tot}}$  above  $\sim 4$ , the deviance cut for the SD-1500 dataset is set at this value.

showers with  $X_{\max}$  below ground) together with improvements in the fitting procedure (to exploit saturated stations in the reconstruction) is advisable.

### 6.3 Mass composition

In this section, the analysis of the  $X_{\max}^{\text{Univ}}$  distributions is presented, stressing the implications over mass composition at the highest energies.

A preliminary step is the evaluation of the resolution of the Universality method ( $\sigma_{\text{Univ}}$ ), which quantifies the broadening of the “true”  $X_{\max}$  distribution induced by the reconstruction. Moreover the simplified Universality fit (described in sect. 6.1) produces a bias in the reconstruction of  $X_{\max}$ , highlighted in the analysis of the golden hybrids dataset (Fig. 6.1b). The value of such bias have to be studied as a function of the energy and subtracted from the  $X_{\max}^{\text{Univ}}$  obtained for each event.

Both the bias correction and the resolution are evaluated using the reconstruction outcome for the updated (ICRC-2017) set of golden hybrids, described in sect. 6.1. Moreover, the event selection exposed in sect. 6.2.1 has to be applied, meaning that only unsaturated golden hybrids with a zenith angle above  $\theta_{\text{cut}}(E)$  (eq. 6.1) and below  $55^\circ$  (i.e. the range shown in Fig. 6.2) are selected for the following analysis.

Since only 263 events with  $E > 10^{19}$  eV are usable, an evaluation of bias and resolution based solely on the highest energy events is considered unreliable. For this reason, all the selected and successfully fitted golden hybrids with energy above  $10^{18.5}$  eV will be exploited, allowing to derive the parameterizations of the bias and of the resolution as functions of the energy.

To estimate the bias, the distributions of the differences  $(X_{\max}^{\text{Univ}} - X_{\max}^{\text{FD}})$  are made separately for each energy bin wide 0.1 in  $\log_{10}(E/\text{eV})$  (except the last one, containing all the events with  $E > 10^{19.4}$  eV). The value of the bias at each energy (corresponding to the mean energy of the events included in each bin) is taken equal to the mean of the distribution. Then the parameterization is accomplished using a function linear in the logarithm of the energy, i.e. in the form:

$$\text{bias}(E) = a + b[\log_{10}(E/\text{eV}) - 19] \quad (6.2)$$

The resolution of the Universality method ( $\sigma_{\text{Univ}}$ ) is instead obtained using the dispersion of the  $X_{\max}^{\text{Univ}}$  distribution in each energy bin (chosen with the same intervals as for the bias evaluation). In fact, recalling eq. 5.1 and considering  $\sigma_{\text{Univ}}$  in the same way as the detector resolution  $\sigma_{\text{det}}$  for the hybrid reconstruction, it is obtained:

$$\sigma_{\text{Univ}}^2 = \sigma^2(X_{\max}^{\text{Univ}})_{\text{meas}} - \sigma^2(X_{\max}) \quad (6.3)$$

where  $\sigma(X_{\max}^{\text{Univ}})_{\text{meas}}$  is the standard deviation of the  $X_{\max}^{\text{Univ}}$  distribution in a certain energy bin and  $\sigma(X_{\max})$  is the “true” dispersion of  $X_{\max}$  at the same energy, measured through the ICRC-2017 hybrid dataset [81] subtracting the FD detector resolution (as explained

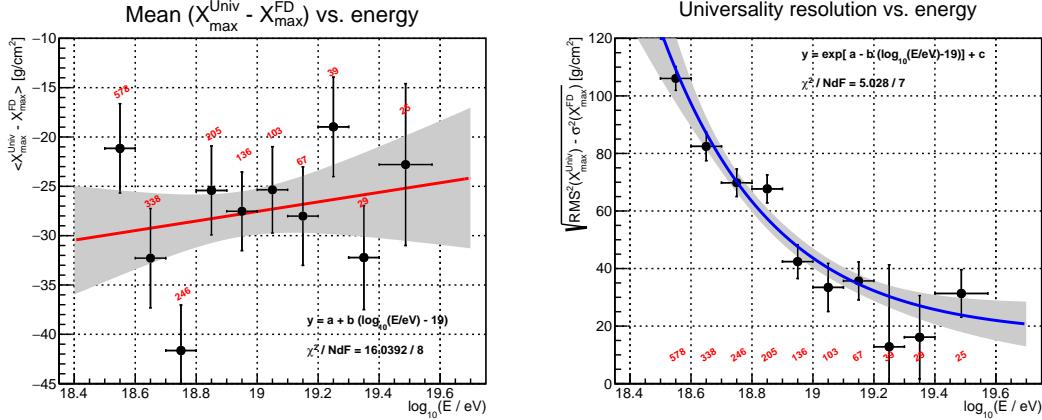


Figure 6.3: Fits of the bias correction (left) and of the Universality resolution (right) as functions of the energy. The golden hybrids for the calibration of the ICRC-2017 dataset, with energy above  $10^{18.5}$  eV, selected according with the cuts exposed in table 6.1, and reconstructed through the simplified version of the Universality fit, are used. The number of events in each energy bin is shown in red. The points in the left panel represent the mean of the distribution of the differences ( $\langle X_{\max}^{\text{Univ}} - X_{\max}^{\text{FD}} \rangle$ ) in each energy bin; instead, the points in the right panel represent the resolution  $\sigma_{\text{Univ}}$  in each energy bin, calculated according with eq. 6.3. The expressions employed to fit the two behaviours, and the resulting goodness-of-fit, are shown in the panels. The grey areas indicate the uncertainty in the fit results at 68% confidence level.

	a	b	c
bias	$-27.55 \pm 1.55$	$4.84 \pm 6.03$	-
resolution	$3.29 \pm 0.54$	$2.75 \pm 1.10$	$16.79 \pm 12.56$

Table 6.2: Table of fit results for the bias correction and the resolution obtained from the simplified Universality reconstruction of the ICRC-2017 golden hybrid events. The function employed to fit the bias is eq. 6.2, while for the resolution is eq. 6.4.

in sect. 5.2). Then the parameterization of  $\sigma_{\text{Univ}}$  with energy is accomplished fitting the values with an exponential plus a constant:

$$\sigma_{\text{Univ}} = \exp \{a - b[\log_{10}(E/\text{eV}) - 19]\} + c \quad (6.4)$$

It should be stressed that the results for  $\sigma_{\text{Univ}}$  (shown in the right panel of Fig. 6.3) clearly indicate that the resolution of the Universality fit rapidly becomes the dominant contribution in the dispersion of the  $X_{\max}^{\text{Univ}}$  distribution for energies below  $10^{19}$  eV. As a consequence, the current Universality reconstruction is not accurate enough to allow the study of the second moment of the  $X_{\max}$  distribution in the lower energy range.

The parameters obtained from the fit of the bias and from the fit of the resolution are reported in table 6.2.

Finally, the  $X_{\max}^{\text{Univ}}$  distributions for the selected SD-1500 dataset are shown in Fig. 6.4, where the bias correction (eq. 6.2) is already applied. The mass composition analyses presented in the following sections are accomplished using these distributions.

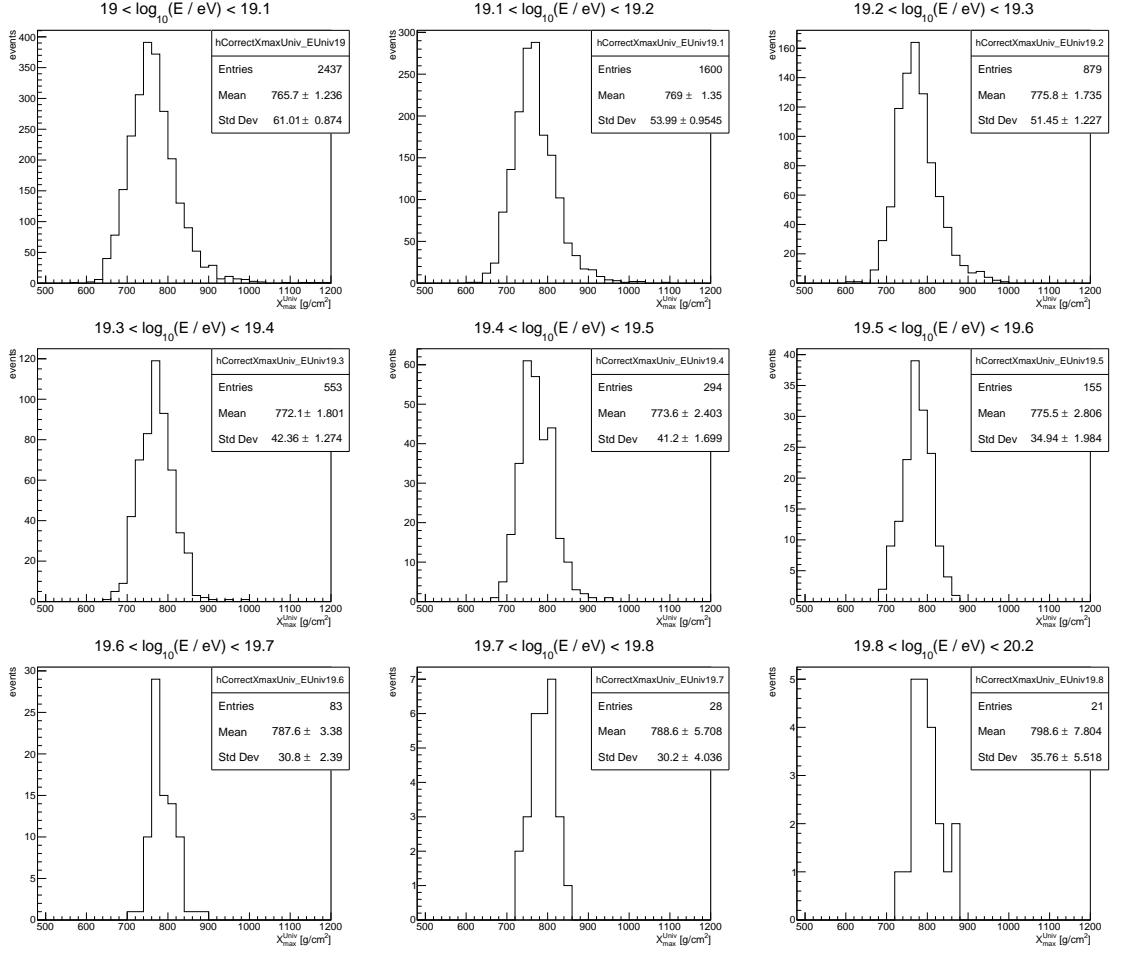


Figure 6.4: The  $X_{\max}^{\text{Univ}}$  distributions for the SD-1500 dataset after event selection and quality cuts (as described in sect. 6.2) and bias correction.

### 6.3.1 The first two moments of the $X_{\max}$ distribution

As exposed in sect. 2.3.1, the first moment of the  $X_{\max}$  distribution allows to infer information about the average composition of cosmic rays as a function of the primary energy. The final result from the Universality reconstruction of the surface detector measurements is shown in Fig. 6.5, where the Universality determination of  $\langle X_{\max} \rangle$  is compared with the values obtained from hybrid events, presented by the Pierre Auger Collaboration at the 35th ICRC [81].

The error bars, which indicate the statistical uncertainty, are mostly hidden by the markers except in the higher energy bins; systematic uncertainties are added as square brackets to the hybrid values. Eventually the two results are compatible inside the uncertainties, confirming but not further proving the drift of the primaries mass composition from lighter elements toward heavier ones.

The number of Universality reconstructed events from the SD-1500 dataset goes from roughly 9 to 3 times the hybrids statistics passing from the lower (i.e.  $10^{19}$  eV) to the higher energy bin. The fact that the Universality dataset does not present a statistics

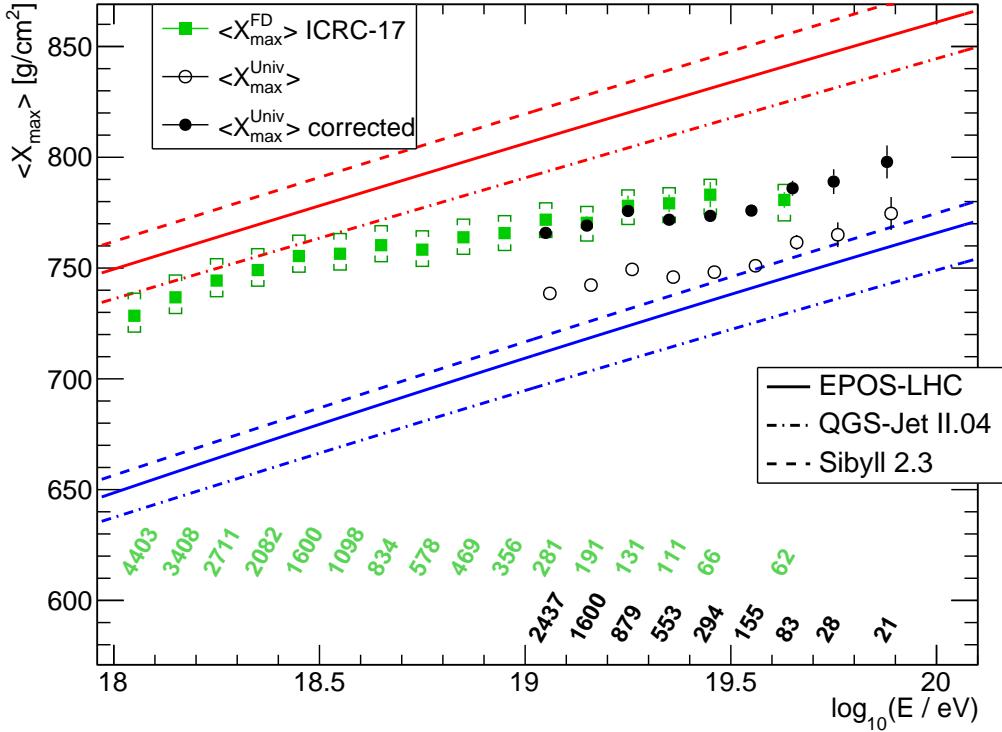


Figure 6.5: First moment of the  $X_{\max}$  distribution ( $\langle X_{\max} \rangle$ ) as a function of the energy. The full black points indicate the result for the selected SD-1500 dataset obtained with the simplified Universality procedure after the bias correction (left panel of Fig. 6.3, eq. 6.2); instead, the empty black points are the un-corrected averages of the  $X_{\max}^{\text{Univ}}$  distributions in each energy bin. The green points show the result [81] for the ICRC-2017 hybrid events. The error bars indicate the statistical uncertainty, while the square brackets are the systematic ones. The different lines show the average predicted for a composition of pure proton/iron primaries with 3 updated hadronic interaction models. In the lower part of the plot is reported the number of events in each energy bin: in black the SD-1500 Universality selection, in green the hybrids.

much larger than the hybrid one is mainly caused by the selection cuts (sect. 6.2.1), that will be possibly removed in future analysis with updated parameterizations.

Nonetheless, the last bin of the hybrid analysis is replaced with 3 bins of normal width (0.1 in  $\log_{10} E$ ) plus an additional bin above  $10^{19.8}$  eV, confirming the power of the Universality method. In particular the last values of  $\langle X_{\max} \rangle$  from the Universality analysis may already indicate that the increase of the primary mass with energy is slowing at the highest energies, but larger statistics and the evaluation of systematic uncertainties are needed before drawing more significant conclusions. When the issues in the parameterization and in the fitting procedure will be solved (see sect. 6.4), it would most likely become possible to reliably evaluate the average composition until the largest energies reached in the SD spectrum (around  $10^{20}$  eV).

In Fig. 6.6 the Universality determination of  $\sigma(X_{\max})$  is shown, in comparison with the results from [81]. The Universality from SD-1500 and the hybrid results are again in agreement, but in this case the extension obtained with Universality presents large uncertainties because of the low accuracy in the determination of  $\sigma(X_{\max})$  in the highest

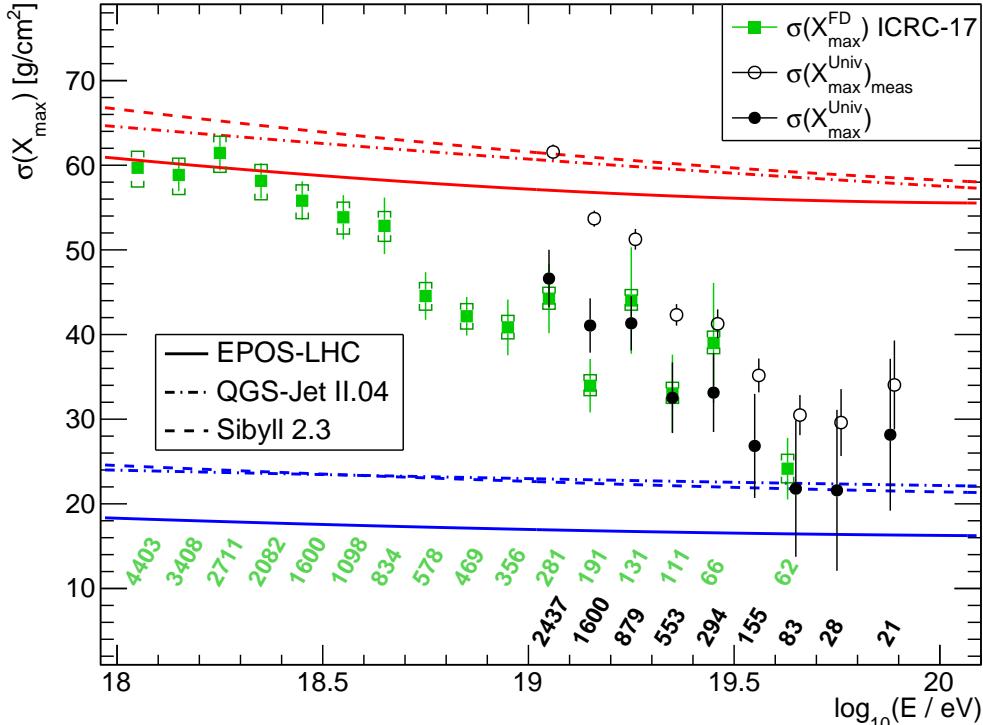


Figure 6.6: Second moment of the  $X_{\max}$  distribution ( $\sigma(X_{\max})$ ) as a function of the energy. The full black points indicate the result for the selected SD-1500 dataset obtained with the simplified Universality procedure after the subtraction of the resolution (right panel of Fig. 6.3, eq. 6.4); instead, the empty black points are the un-corrected standard deviations of the  $X_{\max}^{\text{Univ}}$  distributions in each energy bin, as in Fig. 6.4. The green points show the result [81] for the ICRC-2017 hybrid events. The error bars indicate the statistical uncertainties, while the square brackets are the systematic ones. The lines show the predictions for pure proton/iron composition with 3 updated hadronic interaction models. In the lower part of the plot is reported the number of events in each energy bin: in black the SD-1500 Universality selection, in green the hybrids.

energy bins. Apart for the small number of events which survived the cuts at the highest energies, such large uncertainties are mostly due to the errors assigned at the Universality resolution (Fig. 6.3 on the right) for energies above  $10^{19.5}$  eV, where the value of  $\sigma_{\text{Univ}}$  was obviously extrapolated.

The capability of the Universality reconstruction to reproduce the second moment of the  $X_{\max}$  distribution is very important, because such quantity is difficultly obtained with other methods exploiting the SD-1500 dataset for mass composition studies (see for example [47]). Furthermore, the suggested improvements in the Universality method are also supposed to reduce the resolution  $\sigma_{\text{Univ}}$ , allowing a better estimate (together with a substantially increased statistics) of the second moment of the  $X_{\max}$  distributions.

### 6.3.2 Estimation of the first two moments of the $\ln A$ distribution

The  $X_{\max}$  moments obtained through the Universality reconstruction can be used to evaluate the mean and dispersion of the  $\ln A$  distribution, that is the distribution of the logarithmic masses of primaries at the Earth's atmosphere. In particular, the information

about the masses dispersion complements the average logarithmic mass, easily estimated from  $\langle X_{\max} \rangle$ , allowing to built a more complete (although model dependent) picture of UHECR composition.

To calculate the average logarithmic mass, eq. 2.22 (reported here) is used:

$$\langle \ln A \rangle = \frac{\langle X_{\max} \rangle - \langle X_{\max} \rangle_p}{f_E}$$

where  $\langle X_{\max} \rangle_p$  denotes the mean depth at maximum of proton showers and  $f_E$  is an energy dependent parameter.

Instead the  $\ln A$  dispersion is obtained using eq. 2.23, that is:

$$\sigma_{\ln A}^2 = \frac{\sigma_{sh}^2(\langle \ln A \rangle) - b \sigma_p^2}{b \sigma_p^2 + f_E^2}$$

where  $\sigma_{sh}^2(\langle \ln A \rangle)$  is interpreted as the magnitude of the shower-to-shower fluctuations at the average mass (previously obtained),  $b$  is a parameter depending on hadronic interactions, and  $\sigma_p^2$  is the  $X_{\max}$  variance for proton showers.

The functional forms of the parameterizations for  $\langle X_{\max} \rangle_p$ ,  $f_E$ ,  $\sigma_{sh}^2(\langle \ln A \rangle)$  and  $\sigma_p^2$  are given in [48], while the updated values of the parameters given in [123] will be used in the following<sup>12</sup>.

The two moments of the  $\ln A$  distribution, evaluated from the Universality  $\langle X_{\max} \rangle$  and  $\sigma(X_{\max})$  values exposed in the previous section, are shown for three different hadronic interaction models in Fig. 6.7, in comparison with the corresponding values for the hybrid dataset [81]. The agreement between the two results (Universality and hybrid) is good, as they present the same trends in the common energy range: an increase of the average mass (i.e. a heavier composition) and a decrease of the dispersion (i.e. a purer composition) at least until  $10^{19.5}$  eV.

However, the points added through the Universality method seemingly show a slowdown of these behaviours. On the one hand, the value of  $\langle \ln A \rangle$  from Universality remains approximately constant<sup>13</sup> above  $10^{19.5}$  eV (considering the uncertainties). On the other hand, the last three bins of  $\sigma^2(\ln A)$  from Universality do not show an evident continuation of the previous decrease, even though the uncertainties are too large to draw more precise conclusions.

One should notice that the negative values in the  $\ln A$  dispersion plots (i.e. in the exclusion zones) are obtained when the  $\sigma(X_{\max})$  from data is smaller than the average shower-to-shower fluctuations predicted by a certain hadronic interaction model. As a consequence, QGS-Jet II-04 fails to provide consistent interpretation of data also in the Universality reconstructed dataset, while EPOS-LHC and Sibyll 2.3c are still marginally consistent at the higher energies.

---

<sup>12</sup>Notice that, differently from the ICRC-2017 results given in [81], the updated version (Sibyll 2.3c) of the Sibyll model is employed in [123].

<sup>13</sup>The exact value is different for each hadronic model, but the presence of a constant  $\langle \ln A \rangle$  is evident in all the three models. Moreover, this approximately constant value is compatible with the  $\langle \ln A \rangle$  estimation in the last bin of the hybrid analysis separately for each model.

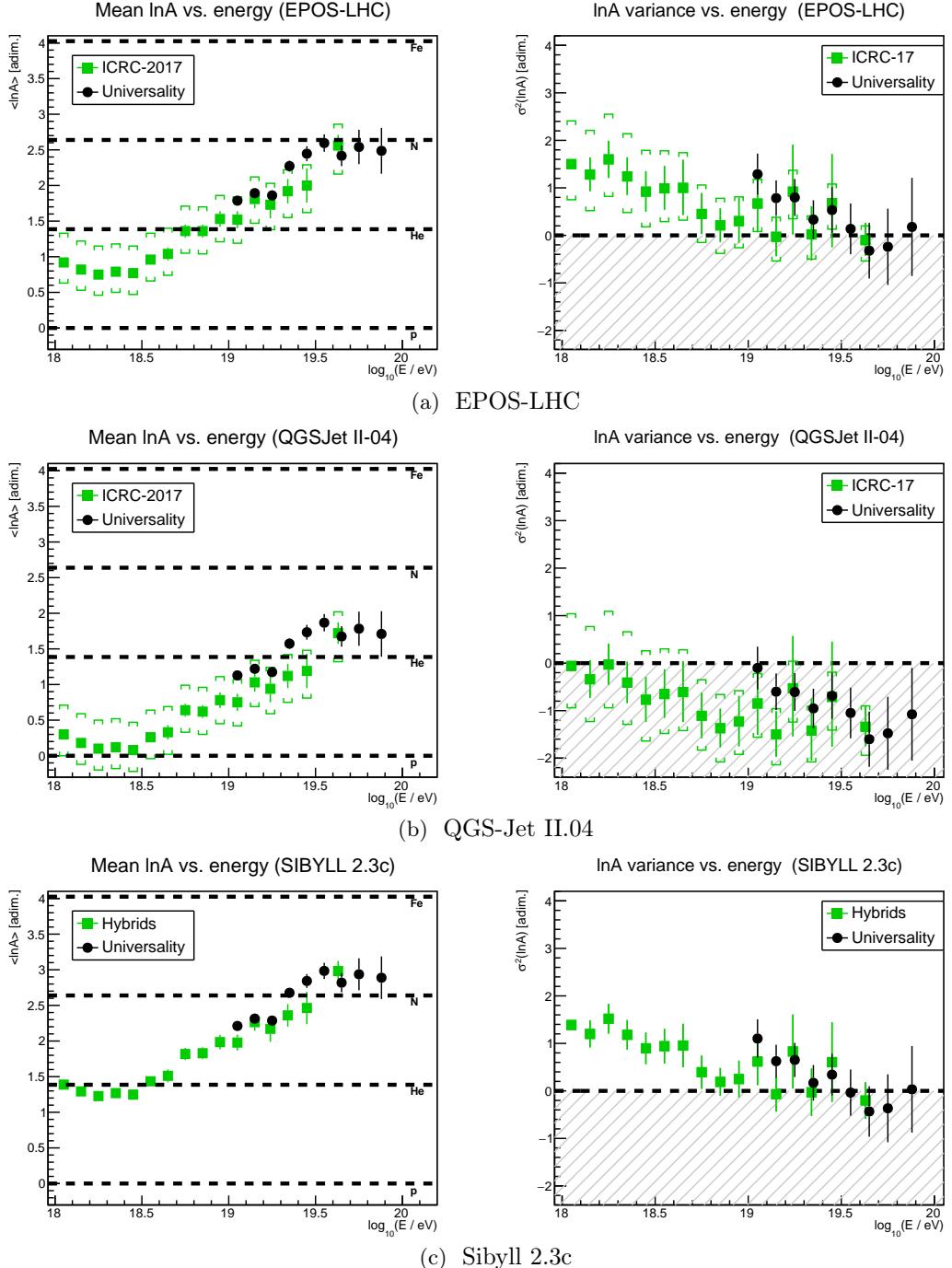


Figure 6.7: The mean (left) and the variance (right) of the  $\ln A$  distribution estimated from the moments of the  $X_{\max}$  distribution. The hadronic interaction models employed are EPOS-LHC (top), QGSJet II.04 (middle) and Sibyll 2.3c (bottom). In black, the values obtained from the Universality reconstruction; in green, the ICRC-2017 values [81] for EPOS-LHC and QGS-Jet, while for Sibyll the calculation has been accomplished using the updated parameters in [123] and the ICRC-2017  $X_{\max}^{\text{FD}}$  moments [81]. The errors bars indicate the statistical uncertainty, that for the Universality values (and in the case of Sibyll also for the hybrids) are calculated propagating the  $X_{\max}$  moments statistical errors and the uncertainties over the parameters of the  $\ln A$  model. The square brackets indicate the systematic uncertainties (not available for Sibyll). The lines in the left panels correspond the  $\ln A$  of the indicated elements, while in the left panels the line at  $\sigma^2(\ln A) = 0$  corresponds to a pure composition.

### 6.3.3 Fit of the mass fractions

To maximize the information that can be extracted from data, the shape of the  $X_{\max}$  distribution can be directly used to infer the fractions of different primary masses, usually choosing representatives of light, intermediate and heavy nuclear species reaching the Earth. Such approach allows to reduce the degeneracies intrinsically related in the description of a distribution only by means of its moments. This is clearly shown in Fig. 6.8, where the two simulated  $X_{\max}$  distributions are produced by different mass compositions but present identical mean and dispersion.

The fitting procedure, introduced in [124], prescribes a comparison of the measured  $X_{\max}$  distribution with the predictions made using a certain set of nuclear masses and varying their fractions. Therefore this method requires the definition of the expected  $X_{\max}$  distributions for each primary mass  $A$ , as functions of the energy  $E$  and of the hadronic interaction model. In [124] such distributions are obtained from CONEX shower simulations. Instead, in this work the expected  $X_{\max}$  distributions are described in terms of generalized Gumbel functions  $g(X_{\max}|E, A)$ , using the parametric model introduced in [122] and recently updated in [123].

To perform the fit, each expected  $X_{\max}$  distribution has to be corrected for detection effects, i.e. the detector resolution. Therefore the Gumbel distributions are convoluted with a gaussian distribution with variance equal to the resolution of the Universality reconstruction ( $\sigma_{\text{Univ}}$ ), evaluated at each energy through eq. 6.4. In Fig. 6.9a the result of this convolution for proton and iron Gumbel distributions at  $10^{19}$  eV are reported, showing an evident increase of the overlapping region due to the large resolution of the current Universality fit.

The zenith cuts introduced in sect. 6.2.1 could affect the  $X_{\max}$  distribution in the same way of the field of view cut exposed in [53], thus requiring the introduction of an

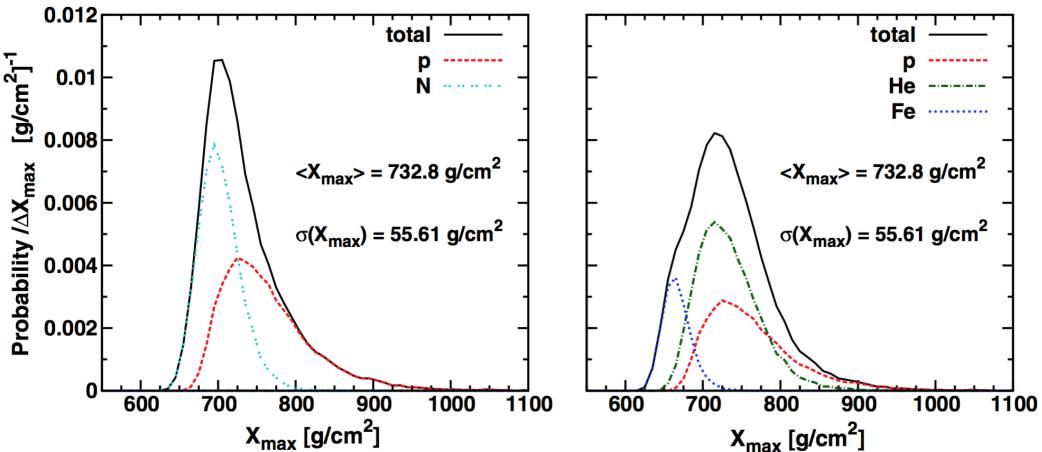


Figure 6.8: Two simulated  $X_{\max}$  distributions in the range  $10^{18.2-18.3}$  eV generated with identical mean and dispersion but from different compositions using the hadronic interaction model EPOS-LHC. Image from [124].

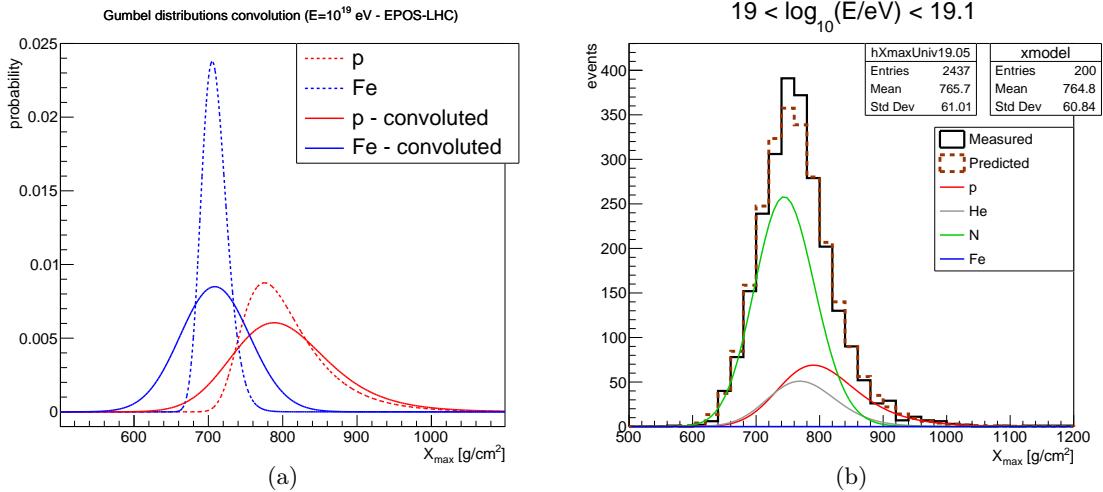


Figure 6.9: (Left) Effect of the Universality resolution over the Gumbel distributions. The Gumbel functions for proton and iron ( $E = 10^{19}$  eV, EPOS-LHC model), parameterized according to [123], are convoluted with a gaussian distribution with variance equal to  $\sigma_{\text{Univ}}(10^{19}$  eV). (Right) Result for the fraction fit in the energy range  $10^{19} - 10^{19.1}$  eV, accomplished using the convoluted Gumbel distributions for (H, He, N, Fe). The black histogram shows the measured  $X_{\max}^{\text{Univ}}$  distribution, while the red-dashed one is the fit result. The 4 coloured lines are the convoluted Gumbel distributions weighted with the corresponding best-fit fractions and normalized to the total number of reconstructed events in the chosen energy bin. The values of  $X_{\max}^{\text{Univ}}$  have been preliminary corrected for the bias (left panel of Fig. 6.3).

acceptance as a function of  $X_{\max}^{\text{Univ}}$  (see [53] and sect. 5.2). However, the issues in the current Universality method affect the reconstruction of simulated events in the same way of measured ones, preventing at the moment a meaningful evaluation of the acceptance.

Finally, a binned maximum-likelihood discriminator is employed to obtain the best-fit fractions<sup>14</sup> with respect to the chosen set of (convoluted) Gumbel distributions: in particular, the same set of four masses (H, He, N, Fe) of previous studies [124, 81] is employed. As in the  $X_{\max}$  moments analysis, the events are binned in intervals of 0.1 in  $\log_{10} E$  and of 20 g/cm<sup>2</sup> (starting from 0) in  $X_{\max}$ . In addition, before the fit the measured  $X_{\max}^{\text{Univ}}$  values are corrected in accordance with the Universality reconstruction bias (left panel of Fig. 6.3, eq. 6.2). An example for the outcome of the fraction fit is given in Fig. 6.9b, showing the result in the range  $10^{19} - 10^{19.1}$  eV.

The fractions of H, He, N and Fe as functions of the energy are shown in Fig. 6.10, where the results obtained from the Universality reconstructed SD-1500 dataset and from hybrid events [123] are compared<sup>15</sup>; in both cases the EPOS-LHC hadronic interaction model is employed<sup>16</sup>.

<sup>14</sup>Independently from the number of masses employed, the fractions must sum to unity.

<sup>15</sup>Such hybrid fractions slightly differ from the values presented at ICRC-2017 [81] because of the model for the  $X_{\max}$  distributions: the convolution of a gaussian with an exponential in [81], the Gumbel distributions in [123].

<sup>16</sup>The fraction fit of the Universality reconstructed SD-1500 dataset has been performed also using other two hadronic interaction models (QGS-Jet II.04 and Sibyll 2.3c). In all the cases the outcome resulted compatible with the fractions obtained from the hybrid dataset. Since this analysis is mostly intended as a feasibility study, only the results for EPOS-LHC will be shown in the following.

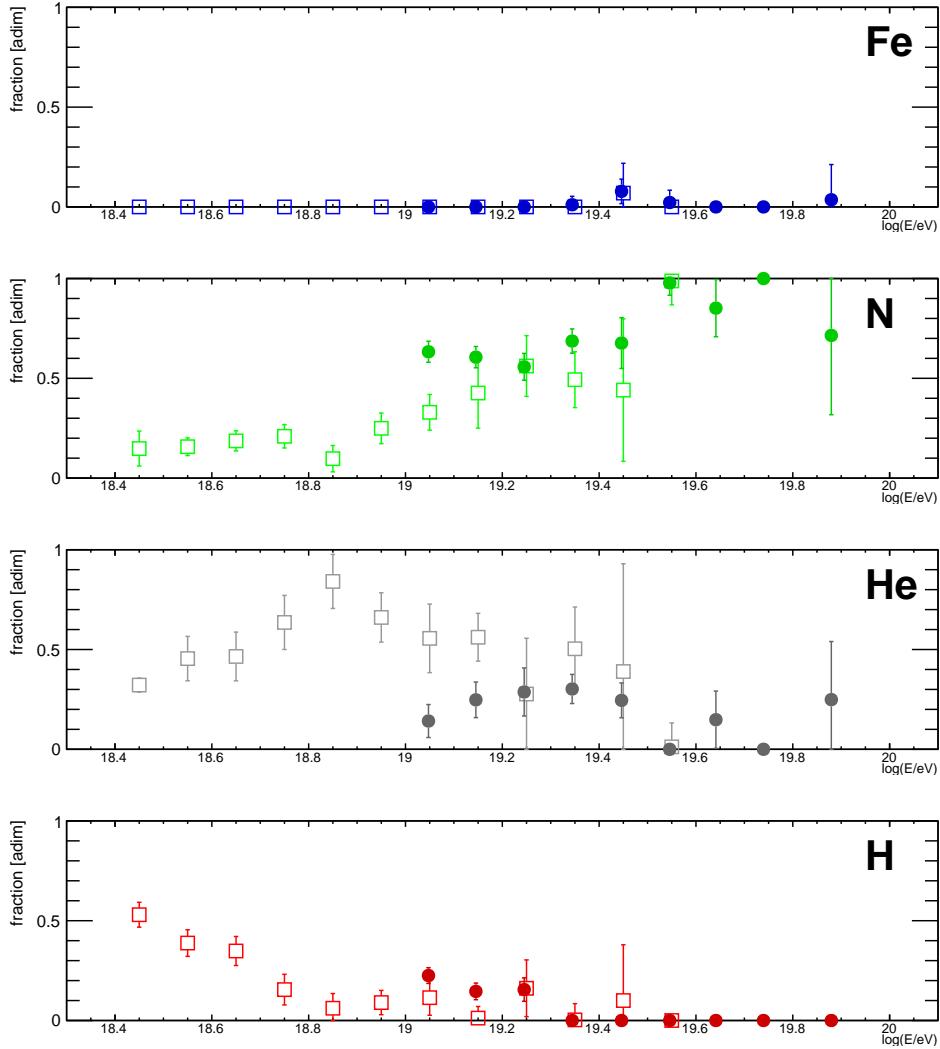


Figure 6.10: Mass fractions obtained from the fit of the  $X_{\max}$  distributions using the Gumbel distribution from the EPOS-LHC hadronic model. The empty squares indicate the result for the hybrid dataset reported in [123], while the full dots are the result for the analysis of the Universality reconstruction of the SD-1500 dataset. The error bars indicate the uncertainties propagated from the errors over the fitted parameters.

The Universality and hybrid results present similar trends in the common energy range, i.e. between  $10^{19}$  and  $10^{19.6}$  eV. In both cases, the fraction of heavy elements (represented by iron) is negligible and a prevalence of nitrogen is observed as the energy increases. However, the helium fractions are smaller in the Universality result than in hybrids (in particular near  $10^{19}$  eV), with correspondingly larger fractions of nitrogen. This behaviour shows the importance of correlations in the fitting procedure, which become even more relevant in the Universality result because of the large resolution  $\sigma_{\text{Univ.}}$ . The disagreement among the exact values of each fraction can be actually explained considering that the uncertainties given in Fig. 6.10 are calculated from the propagation of the errors over the fitted parameters. A more correct procedure would be to generate a

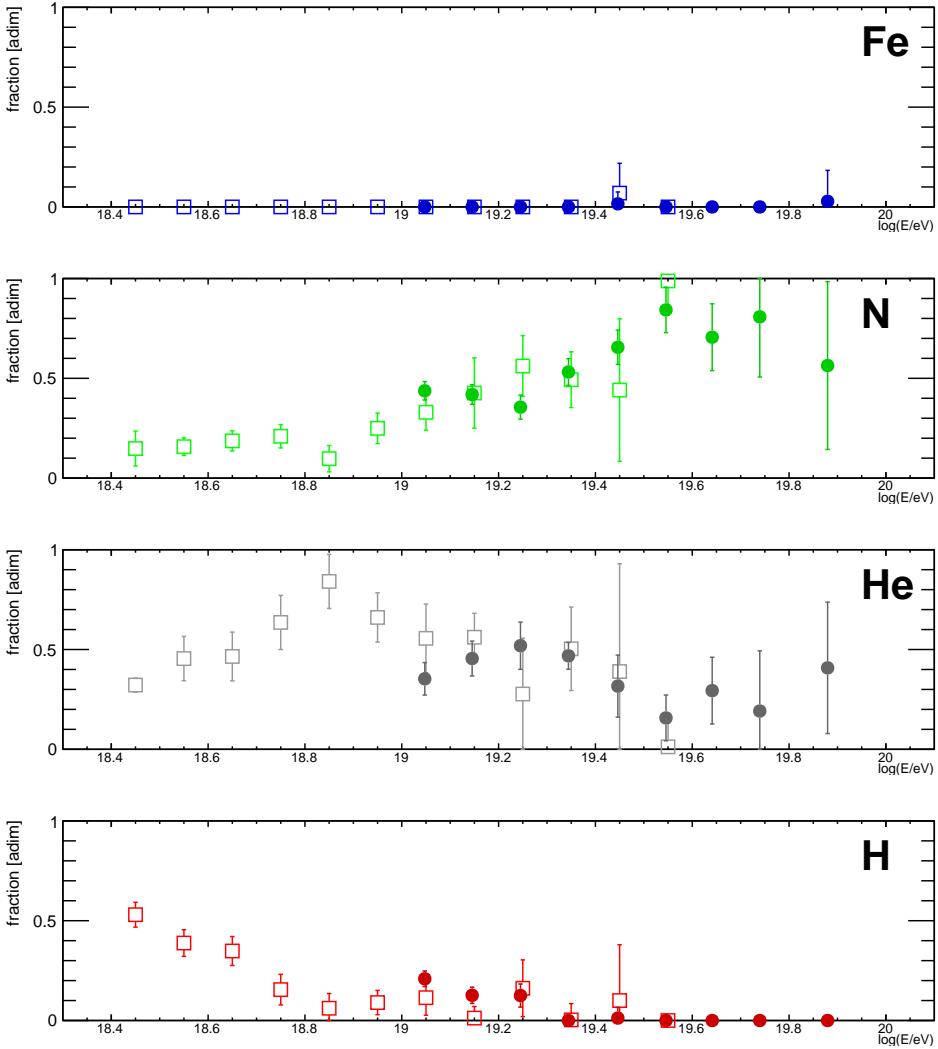


Figure 6.11: Same as Fig. 6.11, but after an additional shift of all the  $X_{\max}^{\text{Univ}}$  values of  $5 \text{ g/cm}^2$  towards larger (i.e. deeper) atmospheric depths.

large number of simulated datasets (mocks) from the best-fit solution and then evaluate the uncertainties with a likelihood method, as described in [124].

Besides, systematic uncertainties have not been evaluated. A hint about their magnitude could be obtained considering for instance the 68% uncertainty region of the bias (see Fig. 6.3), which covers a range roughly between  $2.5$  and  $5 \text{ g/cm}^2$  from the bias values parameterized as a function of the energy. Indeed, an artificial shift of  $5 \text{ g/cm}^2$  applied to all the  $X_{\max}^{\text{Univ}}$  distributions results in a not-negligible shift of the resulting fractions, as can be noticed comparing Fig. 6.11 with Fig. 6.10.

Then it is clear that the main sources of systematic uncertainties are the bias and resolution of the Universality method. On the one hand, the assumption of a gaussian representation for the resolution of the Universality fit could be simplistic and lead to small inconsistencies among the  $X_{\max}^{\text{Univ}}$  distributions and the convoluted Gumbel functions. On the other hand, the choice of performing a rigid shift of the distributions

according to the Universality bias (introduced basically to correct  $\langle X_{\max}^{\text{Univ}} \rangle$ ) could be inappropriate, since more sophisticated methods may be required to recover a mean value compatible with the hybrid measurements while obtaining the correct shape of the  $X_{\max}$  distributions.

The knowledge about the fractions for different group of masses (represented by the four chosen elements) can be anyway employed to separate the energy spectrum in a correspondent number of components, multiplying the flux in each bin for the fractions fitted through the above-described procedure.

Given the current low discrimination power of the fractions fit due to the large value of the Universality resolution  $\sigma_{\text{Univ}}$ , it was decided to consider only two components in the separation of the energy spectrum: a light component, obtained from the sum of proton and helium fractions; a heavy component, from the sum of nitrogen and iron. The result is reported in Fig. 6.12, where the light and heavy spectra obtained from the Universality fractions are compared with the prediction from the hybrid dataset [123] represented by a shaded area.

Observing the two (light and heavy) spectra, the transition from a light-dominated toward and intermediate-heavy-dominated scenario is evident. However, the information added through the fit of the  $X_{\max}^{\text{Univ}}$  distributions seems to disfavour a complete disappearance of the light component above  $\sim 10^{19.5}$  eV.

This is an interesting result, since the search for a flux contribution of protons up to the highest energies is a key science objective of the *AugerPrime* upgrade and a decisive ingredient for prospects of proton astronomy. Furthermore, the knowledge of the proton flux is critical in the estimation of the expected flux of secondary photons and neutrinos.

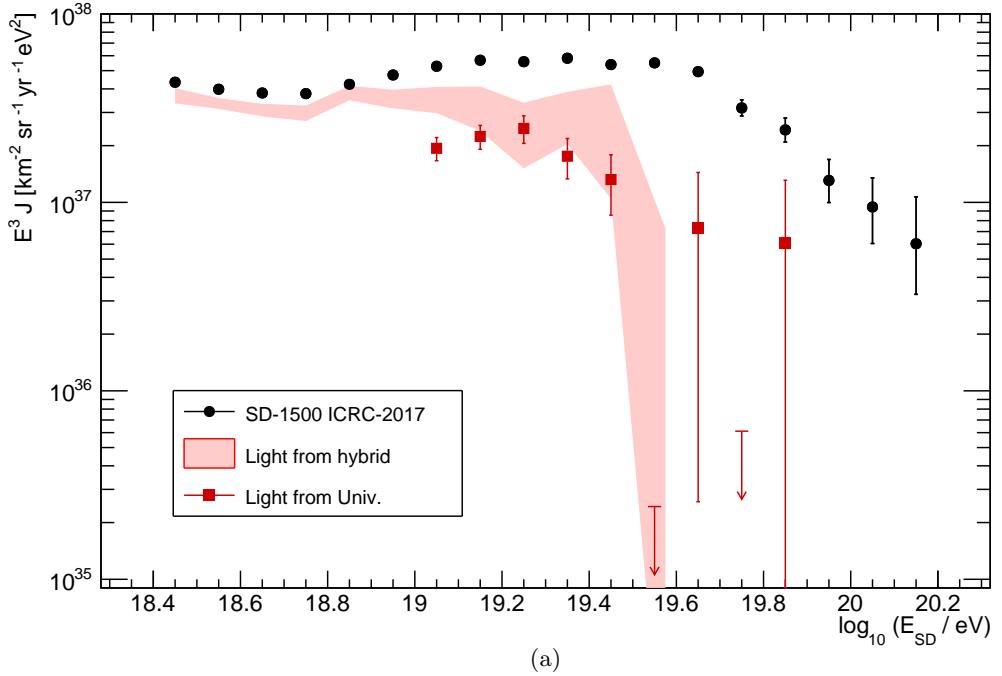
## 6.4 Future prospects

Two main issues in the current Universality reconstruction have been stressed in this chapter. Firstly, the negative bias in the  $X_{\max}^{\text{Univ}}$  reconstruction, pointed out in the analysis of golden hybrids and necessarily corrected before performing the mass composition studies. Moreover, the reduction of the available statistics due to the event selection (sect. 6.2.1), which penalizes especially the highest energy range.

A revision of the Universality model is foreseen to solve these issues. In fact, updated parameterizations to include showers with  $X_{\max}$  below ground would eliminate the need of a field of view cut, while improvements in the minimization procedure could allow the inclusion of saturated signals in the reconstruction.

The *AugerPrime* upgrade (sect. 3.4) will also represent a major advancement in the Universality reconstruction and in its reliability. In fact, the presence of small-PMTs will allow to measure signals without saturation in almost all events, while the data from the array of surface scintillator detectors (SSD) will offer a complementary measurement of the shower at ground with respect to the array of WCDs. Then the reconstruction will

### SD-1500 spectra - EPOS-LHC



### SD-1500 spectra - EPOS-LHC

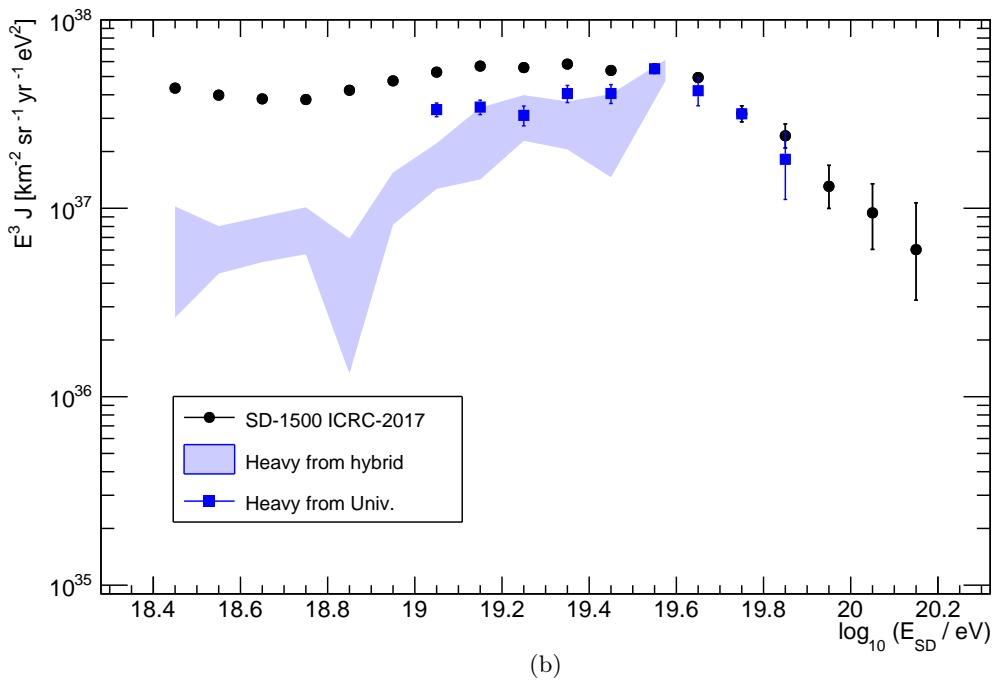


Figure 6.12: Energy spectrum separated in light (top) and heavy (bottom) components. The black dots indicate the energy spectrum presented in [58], while the squares (red for the light component, blue for the heavy one) are obtained using the fractions from the fit of the  $X_{\max}^{\text{Univ}}$  distributions. The shaded areas represent the correspondent result obtained using the fraction fit of the hybrid dataset performed in [123]. As for Fig. 6.10 and 6.11, the systematic uncertainties have not been evaluated.

be possibly enhanced though the development of an improved fitting procedure in which the two measurements are exploited at the same time. Indeed, the comparison between WCD and scintillator signals can allow to disentangle the electromagnetic and muonic contributions, since the two detectors present different responses for the two types of particles (see for example [125]). In this way it will become possible to estimate the relative muon content directly from shower data and constraint the values of  $N_\mu$  in the Universality reconstruction.

The  $N_\mu$  parameter is not exploited in this analysis since a reference for real data is not available and simulations are still untrustworthy: even the updated hadronic interaction models predict a muonic component between 20% and 60% smaller than what is inferred from indirect studies on data (see for instance [63] and references therein). However, the relative muon content is a mass sensitive parameter, therefore the values of  $N_\mu$  could provide another fundamental information for composition studies if properly reconstructed. Moreover,  $N_\mu$  in combination with  $X_{\max}$  can be used for the selection of light-enriched events to strengthen the anisotropy studies.



# Conclusions

In this thesis, a new method for the reconstruction of extensive air showers measured by the surface detector of the Pierre Auger Observatory has been presented, with the aim of obtaining mass-sensitive parameters from a dataset much larger than the hybrid one (tied to the low duty cycle of the fluorescence detectors).

The Universality reconstruction, described in chapter 5, consists in a best-fit of the values of the macroscopic shower parameters: geometry, energy,  $X_{\max}$  and a quantity  $N_\mu$  related to the muon content of the shower. The original version of the fitting procedure, implemented in the `Offline` software of the Collaboration, has been studied and modified using as benchmark the golden hybrid dataset, i.e. an ensemble of high quality hybrid events used for the energy calibration of the surface array.

Several improvements have been accomplished, the more relevant ones being: the introduction of a goodness-of-fit estimator (the deviance) as minimizing function for the reconstruction; the evaluation of a correct magnitude for the signal uncertainties in each time bin (25 ns width) of the measured traces. However, it was noticed that the accuracy in the reconstruction decreases when the energy is fitted together with the other quantities; in this case, a systematic underestimation of  $E_{\text{Univ}}$  with respect to the standard SD result is observed. The reasons behind such issue are still unknown, even if it is argued to be related with internal tensions of the Universality model.

A simplified reconstruction, optimized through the previous analysis, is then established in chapter 6. Such version of the fitting procedure is employed in the reconstruction of the events with energy above  $10^{19}$  eV collected by the surface detector between January 2004 and December 2016. A preliminary step to assure a good quality in the reconstruction is the selection of the dataset. In addition to the exclusion of saturated events, a fiducial field of view cut, function of the zenith angle and of the energy, is required to assure an unbiased measurement of the mass composition. These cuts greatly reduce the available dataset, in particular at the higher energies, where a statistics only 3 times the hybrid one actually survives.

The mass composition studies accomplished in this work are based on the analysis of the  $X_{\max}^{\text{Univ}}$  distributions at different energies (intervals wide 0.1 in the energy logarithm, except the last one which includes the events above  $10^{19.8}$  eV). A bias in the fitted  $X_{\max}^{\text{Univ}}$  values has been parameterized as a function of the energy and subtracted in the final analysis. Moreover, the evaluation of the Universality method resolution ( $\sigma_{\text{Univ}}$ ) is

required to study the true dispersion of the  $X_{\max}$  distributions.

The first two moments of the  $X_{\max}$  distributions, obtained from the Universality reconstructed SD-1500 dataset, show a good agreement with the hybrid result in the common range and extend it with 3 additional energy bins. With the knowledge of  $\langle X_{\max} \rangle$  and  $\sigma(X_{\max})$  from Universality, the study of the mean logarithmic mass ( $\ln A$ ) and its dispersion is accomplished. For energies up to  $10^{19.5}$  eV, the trend toward a heavier and purer composition is evident, in accordance with the hybrid results. A slowdown of this behaviour is observed in the last energy bins, added in this work, but the new information that can be obtained from the current version of the Universality reconstruction is still not supported by a sufficient statical significance.

Finally, the full  $X_{\max}^{\text{Univ}}$  distributions are used to extract the fraction of four representatives of primary masses (H, He, N, Fe) through a best-fit procedure employing the description of the shower maximum distribution with Gumbel functions. Also in this case, the power of the Universality method is evident, showing an agreement with the hybrid results in the common bins and extending the result at higher energies. Using the updated knowledge of mass fractions obtained from Universality, a tentative study of the energy spectrum separated in a lighter and a heavier component is accomplished, which suggests the presence of a light contribution also at the highest energies.

When the issues in the Universality parameterizations will be solved, achieving a substantial reduction of the resolution  $\sigma_{\text{Univ}}$  and possibly removing the need of a bias correction, the results of the reconstruction procedure are expected to lead an extension of the hybrid results and a new horizon for the mass composition studies.

The Universality method will be also substantially enhanced with the *AugerPrime* upgrade, currently under construction. In fact, the new data from SSDs will be exploited in a fully global shower reconstruction and will allow to disentangle the electromagnetic and muonic signals, thanks to the different responses of scintillators and water-Cherenkov detectors. In this way, it will become possible to constrain the muonic component of the shower, allowing a combined mass composition analysis which employs  $X_{\max}$  and  $N_{\mu}$ .

## Appendix A

# Some additional details about the Universality fit

### A.1 The deviance for the axis constraint

The distribution of small angles  $\alpha$  in space, e.g. in scattering processes, can be described by a Rayleigh distribution

$$f(\alpha; \sigma) = \frac{\alpha}{\sigma^2} \exp(-\alpha^2/2\sigma^2) \quad (\text{A.1})$$

where  $\sigma$  is called *scale parameter*.

This distribution can be used also to describe the angle between two different reconstructions of the shower direction of an event (for instance Universality vs SD or SD vs hybrid), in the hypothesis of small angular difference between the two. In this case, the value of the scale parameter can be taken equal to the angular resolution ( $\sigma_{\vec{a}}$ ) experimentally found as uncertainty of the SD axis reconstruction (see sect. 5.7).

The starting point in the calculation of a deviance is eq. 5.2 (see also the paragraph “*Testing goodness-of-fit with maximum likelihood*” in [113]), that defines:

$$D = -2 \ln \frac{\mathcal{L}}{\mathcal{L}^{\text{sat}}} \quad (\text{A.2})$$

The likelihood for a quantity which is supposed to be distributed according with a Rayleigh distribution can be written as:

$$\mathcal{L} = f(\alpha|\sigma_{\vec{a}}) = \frac{\alpha}{\sigma_{\vec{a}}^2} \exp(-\alpha^2/2\sigma_{\vec{a}}^2) \quad (\text{A.3})$$

where  $\alpha$  is the angle between two different determinations of the shower axis.

In order to evaluate the saturated likelihood, it is necessary to distinguish model and data in the actual case of interest, that is the comparison between Universality and SD axis directions. Since the model is undoubtedly the Universality paradigm, the Universality shower axis is the model prediction; instead, the SD axis should be intended as a measurement.

Then the saturated likelihood is obtained for the model that better describes the data, that in the case of a Rayleigh distribution with scale parameter  $\sigma_{\vec{a}}$  is obtained for a Universality axis which is distant exactly  $1\sigma_{\vec{a}}$  from the SD axis, meaning that the saturated model is obtained for  $\alpha = \sigma_{\vec{a}}$ . In effect, this is also the value with maximum probability (i.e. the mode) of a Rayleigh distribution with scale parameter  $\sigma_{\vec{a}}$ .

Then the saturated likelihood is:

$$\mathcal{L}^{\text{sat}} = f(\sigma_{\vec{a}}|\sigma_{\vec{a}}) = \frac{1}{\sigma_{\vec{a}}} \exp(-1/2) \quad (\text{A.4})$$

Finally, the deviance to be used in the axis constraint can be calculated as:

$$D_{\vec{a}} = -2 \ln \frac{\alpha}{\sigma_{\vec{a}}} + \frac{\alpha^2}{\sigma_{\vec{a}}^2} - 1 \quad (\text{A.5})$$

## A.2 Study of mock traces generated through the cluster model

### A.2.1 The expected mean of a set of mock traces

The expectation value for the signal of a set of mock traces, randomly generated from a certain trace using the cluster model (as described in section 5.5.2), can be studied considering fixed the zenith angle  $\theta$  of the event (which fixes also the value of the uncertainty factor  $f_S$ ) and the integrated signals  $S_i$  of each shower component predicted by the Universality fit of the chosen original trace. In this conditions, the mean signal over the generated set of mock traces can be written as:

$$\langle S_{\text{mock}} \rangle = \left\langle \sum_{i=1}^4 \sum_{j=1}^{N_i} S_{\text{clu}}^{i,j} \right\rangle = \left\langle \sum_{i=1}^4 \sum_{j=1}^{N_i} \frac{n_{p.e.}^{i,j}}{\langle p.e./VEM \rangle} \right\rangle \quad (\text{A.6})$$

where  $i$  runs over the 4 Universality components (e.m.,  $\mu$ , e.m.( $\mu$ ), e.m.(h)),  $N_i$  is the number of clusters extracted for each component,  $S_{\text{clu}}^{i,j}$  is the signal of each single cluster, that is obtained from the extracted number of photoelectrons  $n_{p.e.}^{i,j}$  divided by the average number of p.e. per VEM  $\langle p.e./VEM \rangle = 94$ .

The number of clusters  $N_i$  is extracted from a Poisson distribution with mean  $\langle N_i \rangle = S_i / \langle S_{\text{clu}} \rangle = S_i / f_S^2$ , in accordance with the first assumption of the cluster model. Moreover, also the number of photoelectrons  $n_{p.e.}^{i,j}$  is extracted from a Poisson distribution with mean  $\langle n_{p.e.} \rangle = f_S^2 \langle p.e./VEM \rangle$ , in accordance with the second assumption.

Therefore the sum  $\sum_{j=1}^{N_i} n_{p.e.}^{i,j}$  for each component is a sum of *independent identically distributed* (IID) random variables with a random number of terms  $N_i$  that is independently extracted according to a Poisson distribution. Then the result of each sum is distributed according to a *compound Poisson distribution*, and it is in turn a random variable.

With these considerations, and in the hypothesis of independence among the components of the shower, the initial expectation value can be re-written as:

$$\langle S_{\text{mock}} \rangle = \frac{1}{\langle p.e./VEM \rangle} \sum_{i=1}^4 \left\langle \sum_{j=1}^{N_i} n_{p.e.}^{i,j} \right\rangle \quad (\text{A.7})$$

The *Wald's equation* (or identity) [126, 127] states that the expectation value of the sum of a random number  $N$  of IID random variables  $x_i$  (that thus present  $\langle x_i \rangle = \langle x \rangle \forall i$ ) is:

$$\left\langle \sum_{j=1}^N x_i \right\rangle = \langle N \rangle \langle x \rangle \quad (\text{A.8})$$

Then, applying this result to each expectation value of the sum in eq. A.7, one has:

$$\left\langle \sum_{j=1}^{N_i} n_{p.e.}^{i,j} \right\rangle = \langle N_i \rangle \langle n_{p.e.} \rangle \quad (\text{A.9})$$

as  $\langle n_{p.e.}^{i,j} \rangle = \langle n_{p.e.} \rangle = f_S^2 \langle p.e./VEM \rangle \forall i, j$ .

Finally, the expectation value for the mock traces signal is:

$$\langle S_{\text{mock}} \rangle = \frac{\langle n_{p.e.} \rangle}{\langle p.e./VEM \rangle} \sum_{i=1}^4 \langle N_i \rangle = f_S^2 \sum_{i=1}^4 \frac{S_i}{f_S^2} = \sum_{i=1}^4 S_i = S_{\text{fit}} \quad (\text{A.10})$$

where  $S_{\text{fit}}$  is the integrated signal predicted through the Universality reconstruction.

### A.2.2 The expected variance of a set of mock traces

In the same conditions required for the calculation of the expected signal (i.e. fixed zenith and component signals), the variance for a set of mock traces generated from the same original trace through the cluster model can be written as:

$$\sigma^2[S_{\text{mock}}] = \sigma^2 \left[ \sum_{i=1}^4 \sum_{j=1}^{N_i} \frac{n_{p.e.}^{i,j}}{\langle p.e./VEM \rangle} \right] \quad (\text{A.11})$$

Also in this case, considering that the Universality components are independent and that the sums  $\sum_{j=1}^{N_i} n_{p.e.}^{i,j}$  are independent random variables distributed according to 4 compound Poisson distributions, the variance can be re-written as:

$$\sigma^2[S_{\text{mock}}] = \frac{1}{\langle p.e./VEM \rangle^2} \sum_{i=1}^4 \sigma^2 \left[ \sum_{j=1}^{N_i} n_{p.e.}^{i,j} \right] \quad (\text{A.12})$$

So it is required to calculate the variance of a sum of a random number of independent random variables. This task can be accomplished using the *Law of total variance* (or *Law of Iterated Variances*) [126, 127], which states that given 2 random variables  $X$  and  $Y$  (which lies on the same probability space) with  $\sigma^2[Y]$  finite, it is:

$$\sigma^2[Y] = \sigma^2[\langle Y|X \rangle] + \langle \sigma^2[Y|X] \rangle \quad (\text{A.13})$$

where  $\langle Y|X \rangle$  and  $\sigma^2[Y|X]$  are the conditional expected value and conditional variance of  $Y$  given  $X$ , i.e. the expected value and variance of the probability distribution of  $Y$  given a certain value of  $X$ .

In the case of interest for this work, the sum corresponds to the variable  $Y$  and the number of clusters to the variable  $X$ : in fact, the sum outcome is a random variable conditioned by the value of  $N_i$ . Then eq. A.13 becomes:

$$\sigma^2 \left[ \sum_{j=1}^{N_i} n_{p.e.}^{i,j} \right] = \sigma^2 \left[ \left\langle \sum_{j=1}^{N_i} n_{p.e.}^{i,j} \mid N_i \right\rangle \right] + \left\langle \sigma^2 \left[ \sum_{j=1}^{N_i} n_{p.e.}^{i,j} \mid N_i \right] \right\rangle \quad (\text{A.14})$$

However, for a fixed number of terms, the sum of independent poissonian variables follows a poissonian distributions with mean equal to the sum of the means. Therefore:

$$\left\langle \sum_{j=1}^{N_i} n_{p.e.}^{i,j} \mid N_i \right\rangle = N_i \langle n_{p.e.} \rangle \quad (\text{A.15})$$

$$\sigma^2 \left[ \sum_{j=1}^{N_i} n_{p.e.}^{i,j} \mid N_i \right] = N_i \langle n_{p.e.} \rangle \quad (\text{A.16})$$

Then eq. A.14 becomes:

$$\sigma^2 \left[ \sum_{j=1}^{N_i} n_{p.e.}^{i,j} \right] = \sigma^2 [N_i \langle n_{p.e.} \rangle] + \langle N_i \langle n_{p.e.} \rangle \rangle \quad (\text{A.17})$$

But  $\langle n_{p.e.} \rangle$  is a constant value, so:

$$\sigma^2 \left[ \sum_{j=1}^{N_i} n_{p.e.}^{i,j} \right] = \langle n_{p.e.} \rangle^2 \sigma^2 [N_i] + \langle n_{p.e.} \rangle \langle N_i \rangle \quad (\text{A.18})$$

The number of clusters  $N_i$  for each component is distributed according to a Poisson distribution, so  $\sigma^2 [N_i] = \langle N_i \rangle$ , and eq. A.18 can be eventually written as:

$$\sigma^2 \left[ \sum_{j=1}^{N_i} n_{p.e.}^{i,j} \right] = \langle N_i \rangle \langle n_{p.e.} \rangle^2 \left( 1 + \frac{1}{\langle n_{p.e.} \rangle} \right) \quad (\text{A.19})$$

The form chosen for eq. A.19 will be useful in the following. In fact, the complete variance for a mock trace can be now expressed as:

$$\sigma^2[S_{\text{mock}}] = \frac{\langle n_{p.e.} \rangle^2}{\langle p.e./VEM \rangle^2} \left( 1 + \frac{1}{\langle n_{p.e.} \rangle} \right) \sum_{i=1}^4 \langle N_i \rangle \quad (\text{A.20})$$

since  $\langle n_{p.e.} \rangle$  as a constant value.

Finally, remembering that  $\langle n_{p.e.} \rangle = f_S^2 \langle p.e./VEM \rangle$  and that  $\sum_{i=1}^4 \langle N_i \rangle = S_{\text{fit}}/f_S^2$ , the last equation result:

$$\sigma^2[S_{\text{mock}}] = f_S^2 S_{\text{fit}} \left( 1 + \frac{1}{f_S^2 \langle p.e./VEM \rangle} \right) \quad (\text{A.21})$$

While the expected mean value for the signal of a mock trace is exactly equal to the integrated signal obtained from the (Universality fit of) the original trace, the expected variance is slightly different from the experimental value of the measured signal variance, that is equal to  $\sigma^2[S_{\text{meas}}] = f_S^2 S_{\text{meas}}$ . This discrepancy is obviously due to the factor  $(1 + 1/\langle n_{p.e.} \rangle)$ , which is a function of the number of photoelectrons per VEM and of the zenith angle of the event through the factor  $f_S$ .

However, such factor has a small effect of the final evaluation of the signal uncertainty, which is the main application of the cluster model in this work. In fact, comparing the square roots of the two variances (that are considered as the uncertainties of the measured and mock signals), the relative difference is:

$$\frac{\sigma[S_{\text{mock}}] - \sigma[S_{\text{meas}}]}{\sigma[S_{\text{meas}}]} = \sqrt{1 + \frac{1}{f_S(\theta)^2 \langle p.e./VEM \rangle}} - 1 \quad (\text{A.22})$$

Since the value of  $\langle p.e./VEM \rangle$  is fixed and equal to 94 photoelectrons, the magnitude of the difference changes only as a function of the zenith angle. In particular, as the value of  $f_S$  becomes smaller, the value of the difference grows larger. In the worst case scenario, that is when  $\theta = 0^\circ$ , it is  $f_S(0^\circ) \approx 0.75$ . Then the value of the ratio is:

$$\frac{\sigma[S_{\text{mock}}] - \sigma[S_{\text{meas}}]}{\sigma[S_{\text{meas}}]} \approx 0.0094 \quad (\text{A.23})$$

which means that the discrepancy between the experimental signal uncertainty and the dispersion of the mock traces signals is always lower than 1%.

### A.2.3 Effect of a different choice for the single cluster signal

As exposed in sect. 5.5.2, the signal of a single cluster was assumed to have the value  $\langle S_{\text{clu}} \rangle = (1 \times f_S^2) \text{ VEM}$ . The value of this quantity is employed in the calculation of the mean number of clusters  $\langle N_i \rangle = S_i / \langle S_{\text{clu}} \rangle$  for each Universality component, and of the mean number of photoelectrons  $\langle n_{p.e.} \rangle = \langle S_{\text{clu}} \rangle \times \langle p.e./VEM \rangle$  for each cluster.

These quantities ( $\langle N_i \rangle$  and  $\langle n_{p.e.} \rangle$ ) are substituted with their actual values only in the last steps of the two mathematical proofs given in the previous sections. Then these demonstrations are valid independently from the value of  $\langle S_{\text{clu}} \rangle$ , and only the final results (eq. A.10 and eq. A.21) for the expected value and variance of the mock traces signal would change for a different choice of the single cluster signal.

Therefore one can easily study the effect of a different choice for  $\langle S_{\text{clu}} \rangle$ . For instance, let  $\langle S_{\text{clu}} \rangle = (1 \times \alpha) \text{ VEM}$ , with  $\alpha$  constant. Then  $\langle N_i \rangle = S_i / \alpha$  and  $\langle n_{p.e.} \rangle = \alpha \times \langle p.e./VEM \rangle$  (with implied VEM units).

The expected value is actually unchanged, since:

$$\langle S_{\text{mock}} \rangle = \frac{\langle n_{p.e.} \rangle}{\langle p.e./VEM \rangle} \sum_{i=1}^4 \langle N_i \rangle = \alpha \sum_{i=1}^4 \frac{S_i}{\alpha} = \sum_{i=1}^4 S_i = S_{\text{fit}} \quad (\text{A.24})$$

Instead, the variance becomes:

$$\sigma^2[S_{\text{mock}}] = \alpha^2 \left(1 + \frac{1}{\alpha \langle p.e./VEM \rangle}\right) \sum_{i=1}^4 \frac{S_i}{\alpha} = \alpha S_{\text{fit}} \left(1 + \frac{1}{\alpha \langle p.e./VEM \rangle}\right) \quad (\text{A.25})$$

Therefore it is clear that only for  $\alpha \approx f_S^2$  it can be obtained a variance which is compatible with the experimental value.

This limitation is directly related with the hypothesis of independent clusters, i.e. that the time position of each cluster is randomly extracted on the Log-Normal functions obtained from the shape fit. However, it is not possible to elaborate an improved assumption without more precise indications on the time and space correlations among the (electromagnetic) particles entering in a detector, which would require deeper studies about temporal and spatial distributions of particles at ground.

## Appendix B

# Azimuthal asymmetry, core position and core time resolutions

### B.1 The azimuthal asymmetry

The stochastic development of extensive air showers is known to give rise to a symmetry in the shower properties around the shower axis. However, in the case of inclined EAS (i.e. showers with zenith angle  $\theta \neq 0^\circ$ ), particles at different azimuth angles with respect to the shower axis traverse different atmospheric depths, as shown in Fig.B.1a, resulting into an asymmetry in the surface detector measurements [108]. To describe this effect, the concept of “early” and “late” regions, as illustrated in Fig.B.2, is introduced. The detectors labelled as “early” (corresponding to  $|\zeta| < \pi/2$ ) are the first ones to be reached by the shower front, while the detectors in the “late” part ( $|\zeta| \geq \pi/2$ ) reveal particles which traversed more atmosphere.

The asymmetry affects the signals collected by the SD and is due to the combination of the longitudinal evolution of the shower and of geometrical effects, related to the incidence angles of the particles on the WCDs. The first contribution is produced by the

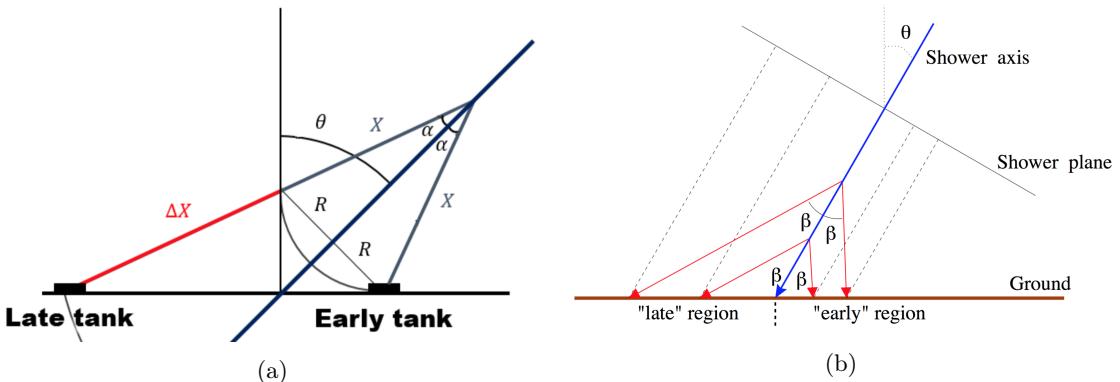


Figure B.1: (a) Difference in the path  $\Delta X$  traversed by particles reaching a late and an early detector, both particles departing with the same angle  $\alpha$  from the shower axis. Image from [118]. (b) Particles going away from the shower axis with symmetric angles  $\beta$ . It can be noticed as they are seen with different zenith angles by the WCD, in particular more vertical in the “early” region. Image from [128].

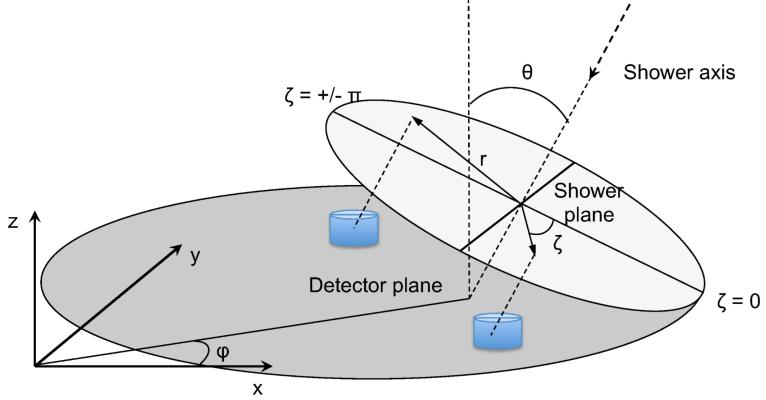


Figure B.2: Schematic view of the shower geometry. The incoming direction of the primary particle defines “early” ( $|\zeta| < \pi/2$ ) and “late” ( $|\zeta| \geq \pi/2$ ) regions. Image from [118].

bigger attenuation of the electromagnetic component ( $e^+, e^-, \gamma$ ) that reaches detectors in the late region as compared to early ones, as illustrated in Fig.B.1a. Thus the measured signals are weaker in the late part and also the ratio of the muon to electromagnetic signals is modified. The geometrical contribution is more easily understood observing Fig.B.1b, where one can see that shower particles reaching the early region are more vertical (with respect to the detector plane) than in the late part. Since muons with larger zenith angle give rise to bigger signal in the WCDs, while the electromagnetic component is anyway absorbed, an asymmetry in the signals between late and early detectors is produced. Geometrical effects predominate at small zenith angles, while for showers with  $\theta > 30^\circ$  attenuation effects are the main contribution.

The effect of the asymmetry in the signal (and in the time spread of the signal) measured by a detector depends by the distance from the core position of the WCD. In fact, the differences between early and late region are small for measurements close to the shower axis. Furthermore, the magnitude of the asymmetry strongly depends on the zenith angle of the shower, as illustrated in Fig.B.3. In fact, while the asymmetry is absent in vertical showers, it grows for increasing zenith. However, it again disappears for very large zenith angles (horizontal showers) since the electromagnetic part is almost completely absorbed before reaching the ground while the muons are attenuated substantially of the same amount in the early and late regions. In summary, the azimuthal asymmetry increases with the zenith angles until a certain inclination, where it reaches a maximum; then it decreases for more inclined showers.

The effect of azimuthal asymmetry is also observed in the signal risetime [108], a parameter that can be used to make inferences about the depth of the shower maximum and thus about the primary mass composition [47].

Despite the exposed results, in the SD standard reconstruction a cylindrical symmetry for the signals at ground is assumed. For this reason the azimuthal asymmetry affects this reconstruction, in particular the determination of the core position. In fact, the

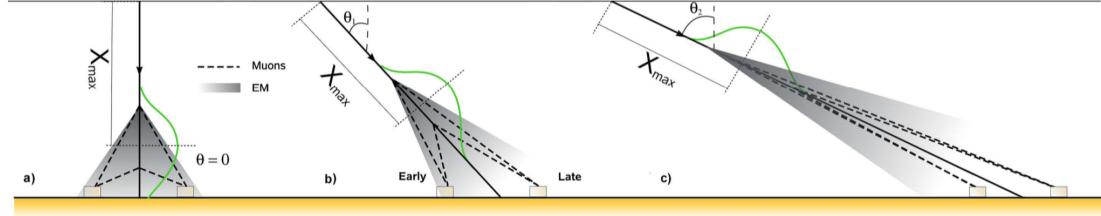


Figure B.3: Schematic view of shower development when arriving at different zenith angles. Image from [129].

assumed symmetry is clearly violated, as can be seen in Fig.B.4 where the isodensity lines are shown for the ground plane and for the shower plane (i.e. a plane perpendicular to the shower axis). Such profiles are eccentric ellipses on the SD array surface, which becomes eccentric circumferences in the shower plane. Therefore the attempt of the SD reconstruction to preserve the supposed cylindrical symmetry generates a shift toward the early region between the real core position and the reconstructed one.

This result can be easily understood considering, in first approximation, the core as the barycenter among the positions of the triggered detectors weighted according to the measured signals. As previously exposed, the signals measured by detectors in the early region are larger than the ones measured in the late region at the same core distance. Therefore it is clear that the core position would be systematically shifted nearer to the detectors in the early region.

The shift in the SD reconstructed core can be studied through simulation or hybrid events. In fact, the hybrid reconstruction is not affected by the azimuthal asymmetry effect, since it does not use the signals measured by the WCDs. In particular golden hybrid events can be reconstructed with both standard SD and hybrid methods, allowing an event-by-event comparison of the core positions. However, in order to correctly evaluate the shift, it is mandatory to define the relationship between the coordinate systems used in the Offline framework.

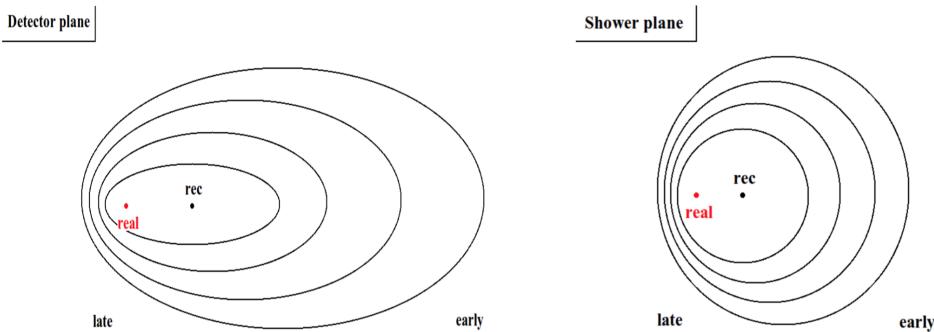


Figure B.4: Isodensity lines in the ground plane and their projection on the shower plane. The effect of the azimuthal asymmetry is evident and it produces a shift between the real core position (red point) and the SD reconstructed position (black point). Image from [118].

## B.2 The shower coordinate system

The reference system for the Pierre Auger Observatory (the so-called *Pampa Amarilla* coordinate system) has the origin in the center of the SD array, at an altitude of 1400 m a.s.l., and the three axes oriented in the easting (X), northing (Y) and vertical (Z) directions. The X-Y plane, approximately corresponding to the ground level, is also called *detector plane*.

However, the reconstruction of an event is more easily performed (and understood) in a coordinate system relative to the particular shower core position and axis direction. So the origin of the *Pampa Amarilla* coordinate system is usually translated into the core position of the event. The angle between the shower axis and the vertical direction defines the zenith ( $\theta$ ) of the event and the angle between the X-axis and the projection of the shower axis on the detector plane is the azimuth ( $\phi$ ).

Then it is defined the *shower coordinate system*, i.e. a reference where the Z-axis coincides with the shower axis; thus the X-Y plane results perpendicular to the shower direction and is usually called *shower plane*. This coordinate system is obtained after the previous translation, applying two rotations:

- a rotation around the vertical axis of an angle equal to the azimuth, to make the new X-axis coincide with the projection of the shower axis on the detector plane;
- a rotation around the updated Y-axis (redefined after the previous rotation) of an angle equal to the zenith, to make the new Z-axis coincide with the shower axis.

These steps can be translated into the following transformation from detector plane ( $dp$ ) to shower plane ( $sp$ ) coordinates:

$$X_{sp} = (\cos \phi \cdot X_{dp} + \sin \phi \cdot Y_{dp}) \cos \theta \quad (B.1)$$

$$Y_{sp} = \cos \phi \cdot Y_{dp} - \sin \phi \cdot X_{dp} \quad (B.2)$$

It is fundamental to notice that the positive direction of the X-axis in the shower plane corresponds, by construction, to the “early” part of the shower, in particular to the line with  $\zeta = 0$  according to Fig.B.2. So the azimuthal asymmetry affects only the X-component of the SD core position in the hybrid shower coordinate system.

## B.3 Comparison of different core positions

In general, the core position of a shower is identified with the intersection between the axis direction and the ground. Therefore the comparison of different core positions, obtained from different reconstructions, is meaningful only if both the points lay on the same plane. When the core positions to be compared are at different altitudes, they have to be projected on a common plane in accordance with the respective shower axis directions (since the axis can result different from two different reconstructions).

In the standard SD reconstruction, the core is obtained from the intersection of the SD axis with ground, so the quote is always around 1400 m a.s.l.. Instead, in the hybrid

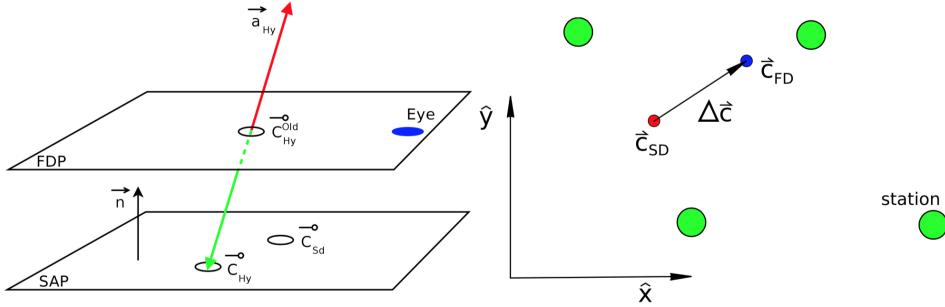


Figure B.5: (Left) Projection of the hybrid core from the eye-centric *fluorescence detector plane* onto the *surface array plane*. (Right) Distance between the core positions in the detector plane. Image from [120].

reconstruction the core is calculated as the intersection of the shower axis with a surface parallel to the ground but with the same altitude of the building that measured the shower (*eye-centric fluorescence detector plane*, FDP), which means few tens of meters higher than the SD array level.

In the case of a golden hybrid event, which is separately reconstructed with both methods, the two core position are therefore systematically shifted in altitude. Thus, before any comparison, each of the two cores have to be re-calculated as the intersection between the shower axis and a suitable surface. In many applications, it is usually chosen the detector plane (also called *surface array plane*, SAP), i.e. the X-Y plane (at 1400 m a.s.l.) of the *Pampa Amarilla* coordinate system, as illustrated in Fig.B.5.

However, in order to properly quantify the effect of azimuthal asymmetry, it is advisable (as clarified in the following section) to accomplish the comparison between the two core positions on the shower plane, that is perpendicular to the shower propagation.

In particular, since the SD and hybrid axes can be different, the study is performed in the hybrid shower coordinate system, which means that the chosen shower plane is perpendicular to the hybrid shower axis and has the origin in the hybrid core position.

The new SD core position is then obtained as the intersection of the SD shower axis with the above-defined hybrid shower plane, similarly to how it was done for the hybrid core in Fig.B.5. In conclusion, only the distance between positions in hybrid shower plane (X-Y plane of the coordinate system referred to the hybrid reconstruction) is of interest in the following analysis.

## B.4 Evaluating the effect of azimuthal asymmetry on the standard SD core position

As mentioned in the previous section, the study of the distance between hybrid and SD core positions is performed in the (hybrid) shower coordinate system. The reason behind this choice is to remove the purely geometrical enhancement of the shift between SD and

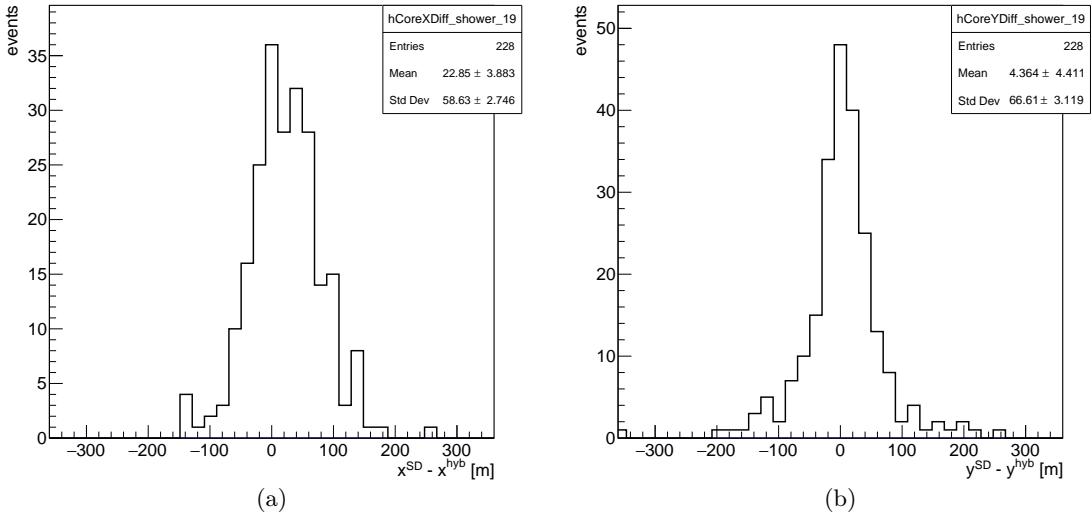


Figure B.6: Distribution of the two components of the distance between SD and hybrid core positions in the hybrid shower coordinate system, calculated for the selected set of unsaturated golden hybrid events with  $E_{SD} > 10^{19}$  eV. (a) X-component, where is evident the bias due to the azimuthal asymmetry. (b) Y-component, which is compatible on average with zero.

real core positions. Indeed, the other characteristics of the azimuthal asymmetry are maintained in the shower plane, even if the distances in the early-late direction (as the shift due to azimuthal asymmetry) are reduced of a factor  $\cos \theta$  according to eq.B.1.

In the shower coordinate system, the X-axis corresponds to the “early-late” direction, so the azimuthal asymmetry affects only this component of the SD core in the hybrid shower coordinate system. Instead, the shifts on the Y-axis are due to the fluctuations between the two reconstruction procedures. As a consequence, the distribution of the X-coordinates values is expected to have a positive mean not compatible with zero, i.e. a bias due to azimuthal asymmetry, while the distribution of the Y-coordinates is expected to be zero on average.

In Fig.B.6 the two distributions are shown and present the expected behaviours: on the X-axis the SD core is biased of approximately 23 m towards the early region (positive X-axis direction), while the distances on the perpendicular direction (Y-axis) are distributed around zero.

The results obtained in Fig.B.6 have been used to establish the core position constraint (sect. 5.7) employed in the constrained Universality reconstruction (sect. 5.8). Considering the differences between the two distributions, it was decided to separate the core constraint into two different gaussian contributions for the two components of the core positions in the shower plane:

- for the X-component, the gaussian constraint is given with a non-zero mean  $x_0 = -23$  m, in order to force a correction of the azimuthal asymmetry effect in the Universality fit with respect to the SD standard reconstruction;
- for the Y-component, the mean of the gaussian constraint is put equal to zero

$(y_0 = 0 \text{ m})$ , since a systematic correction is not needed;

- for both components, the resolutions of the constraints (i.e. the  $\sigma$  parameters of the gaussians) are taken directly equal to the standard deviation of the related distributions:  $\sigma_x = 58.5 \text{ m}$  and  $\sigma_y = 66.5 \text{ m}$ .

It should be noticed that a residual zenith angle dependence of the azimuthal asymmetry is still present in the shower coordinate system (as described in Fig.B.3 and related text); however, it is considered negligible for the aim of a constraint.

## B.5 Standard SD and hybrid core times comparison

Using the same arguments of the previous section, also the values of the core time for the hybrid and standard SD reconstruction are compared in order to evaluate the bias (if needed) and the resolution to be used in the core time constraint (sect. 5.7) of the constrained reconstruction (sect. 5.8).

In Fig. B.7 the result for the distribution of the differences ( $t_{\text{SD}} - t_{\text{hyb}}$ ) is reported. As can be seen, a negative bias is present, meaning that the SD reconstruction of the core time is on average earlier with respect to the hybrid one. This outcome is due to the azimuthal asymmetry: since the core position is systematically shifted toward the “early” region, the core time would be systematically anticipated as well.

Therefore, the constraint to be used in the Universality reconstruction is chosen as a gaussian distribution with mean equal to the opposite of the observed bias (i.e.  $t_0 = 89 \text{ ns}$ ) and variance equal to the standard deviation of the distribution (i.e.  $\sigma_t = 215 \text{ ns}$ ).

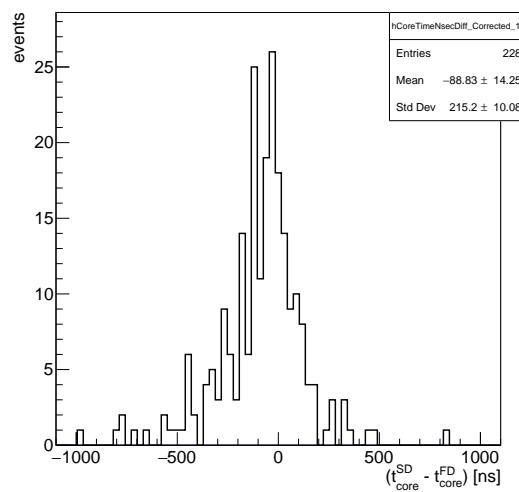


Figure B.7: Distribution of differences between the SD and hybrid core times (tagged as “FD”), calculated for the selected set of unsaturated golden hybrid events with  $E_{\text{SD}} > 10^{19} \text{ eV}$ . The time bias due to the azimuthal asymmetry is evident.



# Bibliography

- [1] Victor F. Hess. “Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten”. In: *Phys. Z.* 13 (1912), pp. 1084–1091.
- [2] Domenico Pacini. “La radiazione penetrante alla superficie ed in seno alle acque”. In: *Il Nuovo Cimento (1911-1923)* 3.1 (1912), pp. 93–100.
- [3] R. A. Millikan. “High frequency rays of cosmic origin”. In: *Proceedings of the National Academy of Sciences* 12.1 (1926), pp. 48–55.
- [4] M. Tanabashi et al. (Particle Data Group). “The Review of Particle Physics”. In: *Phys. Rev. D* 98 (2018), p. 030001.
- [5] Enrico Fermi. “On the origin of the cosmic radiation”. In: *Physical Review* 75.8 (1949), p. 1169.
- [6] Malcolm S. Longair. *High Energy Astrophysics*. Cambridge University Press.
- [7] Thomas K. Gaisser, Ralph Engel, and Elisa Resconi. *Cosmic rays and particle physics*. Cambridge University Press, 2016.
- [8] Anthony M. Hillas. “The origin of ultra-high-energy cosmic rays”. In: *Annual review of astronomy and astrophysics* 22.1 (1984), pp. 425–444.
- [9] Frank M. Rieger, Valenti Bosch-Ramon, and Peter Duffy. “Fermi acceleration in astrophysical jets”. In: *The Multi-Messenger Approach to High-Energy Gamma-Ray Sources*. Springer, 2007, pp. 119–125.
- [10] Glennys R. Farrar and Andrei Gruzinov. “Giant AGN flares and cosmic ray bursts”. In: *The Astrophysical Journal* 693.1 (2009), p. 329.
- [11] Eli Waxman. “Cosmological gamma-ray bursts and the highest energy cosmic rays”. In: *Physical Review Letters* 75.3 (1995), p. 386.
- [12] Jorg P. Rachen and Peter L. Biermann. “Extragalactic ultra-high energy cosmic-rays - part one - contribution from hot spots in Fr-II radio galaxies”. In: *Astronomy and Astrophysics* 272 (1993), p. 161.
- [13] Luis A. Anchordoqui, Gustavo E. Romero, and Jorge A. Combi. “Heavy nuclei at the end of the cosmic-ray spectrum?” In: *Physical Review D* 60.10 (1999), p. 103001.
- [14] Ke Fang et al. “Testing the newborn pulsar origin of ultrahigh energy cosmic rays with EeV neutrinos”. In: *Physical Review D* 90.10 (2014), p. 103005.
- [15] Kumiko Kotera, Elena Amato, and Pasquale Blasi. “The fate of ultrahigh energy nuclei in the immediate environment of young fast-rotating pulsars”. In: *Journal of Cosmology and Astroparticle Physics* 2015.08 (2015), p. 026.
- [16] V. Berezinsky. “Ultra high energy cosmic rays from cosmological relics”. In: *Nuclear Physics B-Proceedings Supplements* 87.1-3 (2000), pp. 387–396.

- [17] J. Abraham et al. “Upper limit on the cosmic-ray photon fraction at EeV energies from the Pierre Auger Observatory”. In: *Astroparticle Physics* 31.6 (2009), pp. 399–406.
- [18] Alexander Aab et al. “Search for photons with energies above  $10^{18}$  eV using the hybrid detector of the Pierre Auger Observatory”. In: *JCAP* 1704.04 (2017), p. 009. DOI: 10.1088/1475-7516/2017/04/009. arXiv: 1612.01517 [astro-ph.HE].
- [19] S. I. Ginzburg and V. L. Syrovatskii. *The Origin of Cosmic Rays*. 1964.
- [20] Ruth Durrer and Andrii Neronov. “Cosmological magnetic fields: their generation, evolution and observation”. In: *The Astronomy and Astrophysics Review* 21.1 (2013), p. 62.
- [21] James W. Cronin. “The highest-energy cosmic rays”. In: *Nuclear Physics B-proceedings supplements* 138 (2005), pp. 465–491.
- [22] Arno A. Penzias and Robert W. Wilson. “A measurement of excess antenna temperature at 4080 Mc/s.” In: *The Astrophysical Journal* 142 (1965), pp. 419–421.
- [23] Kenneth Greisen. “End to the cosmic-ray spectrum?” In: *Physical Review Letters* 16.17 (1966), p. 748.
- [24] Georgi T. Zatsepin and Vadim A. Kuzmin. “Upper limit of the spectrum of cosmic rays”. In: *Journal of Experimental and Theoretical Physics Letters (JETP Letters)* 4.3 (1966), pp. 78–80.
- [25] Veniamin Berezinsky, Askhat Gazizov, and Svetlana Grigorieva. “On astrophysical solution to ultrahigh energy cosmic rays”. In: *Physical Review D* 74.4 (2006), p. 043005.
- [26] R. Aloisio, V. Berezinsky, and A. Gazizov. “Ultra high energy cosmic rays: The disappointing model”. In: *Astroparticle Physics* 34.8 (2011), pp. 620–626.
- [27] Dmitri Ivanov. “Report of the Telescope Array - Pierre Auger Observatory Working Group on Energy Spectrum”. In: *Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*. 2017, pp. 18–24.
- [28] Luisa Arruda, Fernando Barao, and Rui Pereira. “Particle identification with the AMS-02 RICH detector: D/p and anti-D/anti-p separation”. In: (2007). arXiv: 0801.3243 [astro-ph].
- [29] Bruno Rossi. “Über die Eigenschaften der durchdringenden Korpuskularstrahlung im Meeressniveau”. In: *Zeitschrift für Physik* 82.3-4 (1933), pp. 151–178.
- [30] Walther Bothe. “Zur Vereinfachung von Koinzidenzzählungen”. In: *Zeitschrift für Physik A Hadrons and Nuclei* 59.1 (1930), pp. 1–5.
- [31] Bruno Rossi. “Method of registering multiple simultaneous impulses of several Geiger’s counters”. In: *Nature* 125.3156 (1930), p. 636.
- [32] Pierre Auger et al. “Extensive cosmic-ray showers”. In: *Reviews of modern physics* 11.3-4 (1939), p. 288.
- [33] Bruno Rossi. “Misure sulla distribuzione angolare di intensità della radiazione penetrante all’Asmara”. In: *Supplemento a la Ricerca Scientifica* 1 (1934), p. 579.
- [34] Walter Heitler. *The quantum theory of radiation*. Courier Corporation, 1984.
- [35] Antoine Letessier-Selvon and Todor Stanev. “Ultrahigh energy cosmic rays”. In: *Rev. Mod. Phys.* 83 (3 Sept. 2011), pp. 907–942. DOI: 10.1103/RevModPhys.83.907. URL: <http://link.aps.org/doi/10.1103/RevModPhys.83.907>.

- [36] J. Matthews. “A Heitler model of extensive air showers”. In: *Astroparticle Physics* 22.5-6 (2005), pp. 387–397.
- [37] J. Engel et al. “Nucleus-nucleus collisions and interpretation of cosmic-ray cascades”. In: *Physical Review D* 46.11 (1992), p. 5013.
- [38] G. Battistoni et al. “Deviations from the superposition model in a Dual Parton Model with formation zone cascade in both projectile and target nuclei”. In: *Nuclear Physics B-Proceedings Supplements* 52.3 (1997), pp. 123–125.
- [39] M. Ave et al. “Measurement of the pressure dependence of air fluorescence emission induced by electrons”. In: *Astropart. Phys.* 28 (2007), pp. 41–57. DOI: 10.1016/j.astropartphys.2007.04.006. arXiv: astro-ph/0703132 [ASTRO-PH].
- [40] M. Ave et al. “Energy dependence of air fluorescence yield measured by AIRFLY”. In: *Nucl. Instrum. Meth.* A597 (2008), pp. 46–49. DOI: 10.1016/j.nima.2008.08.051.
- [41] A. M. Hillas et al. “Measurement of Primary Energy of Air Showers in the Presence of Fluctuations.” In: *International Cosmic Ray Conference*. Vol. 3. 1971, p. 1001.
- [42] D. Newton, J. Knapp, and A. A. Watson. “The optimum distance at which to determine the size of a giant air shower”. In: *Astroparticle Physics* 26.6 (2007), pp. 414–419.
- [43] J. Hersil et al. “Observations of extensive air showers near the maximum of their longitudinal development”. In: *Physical Review Letters* 6.1 (1961), p. 22.
- [44] Thomas K. Gaisser and Antony M. Hillas. “Reliability of the method of constant intensity cuts for reconstructing the average development of vertical showers”. In: *International Cosmic Ray Conference*. Vol. 8. 1977, pp. 353–357.
- [45] Alexander Aab et al. “The Pierre Auger Cosmic Ray Observatory”. In: *Nucl. Instrum. Meth.* A798 (2015), pp. 172–213. DOI: 10.1016/j.nima.2015.06.058. arXiv: 1502.01323.
- [46] Karl-Heinz Kampert and Michael Unger. “Measurements of the Cosmic Ray Composition with Air Shower Experiments”. In: *Astropart. Phys.* 35 (2012), pp. 660–678. DOI: 10.1016/j.astropartphys.2012.02.004. arXiv: 1201.0018.
- [47] Alexander Aab et al. “Inferences on mass composition and tests of hadronic interactions from 0.3 to 100 EeV using the water-Cherenkov detectors of the Pierre Auger Observatory”. In: *Phys. Rev.* D96.12 (2017), p. 122003. DOI: 10.1103/PhysRevD.96.122003. arXiv: 1710.07249.
- [48] Pedro Abreu et al. “Interpretation of the Depths of Maximum of Extensive Air Showers Measured by the Pierre Auger Observatory”. In: *JCAP* 1302 (2013), p. 026. DOI: 10.1088/1475-7516/2013/02/026. arXiv: 1301.6637 [astro-ph.HE].
- [49] I. Allekotte et al. “The surface detector system of the Pierre Auger Observatory”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 586.3 (2008), pp. 409–420.
- [50] Veberič, Darko. “Estimation of Signal in Saturated Stations of Pierre Auger Surface Detector”. In: *The Pierre Auger Observatory: Contributions to the 33th International Cosmic Ray Conference (ICRC 2013)*. 2013, pp. 23–26.
- [51] C. Bonifazi. “The angular resolution of the Pierre Auger Observatory”. In: *Nucl. Phys. Proc. Suppl.* 190 (2009), pp. 20–25.

- [52] J. Abraham et al. “The Fluorescence Detector of the Pierre Auger Observatory”. In: *Nucl. Instrum. Meth.* A620 (2010), pp. 227–251. DOI: 10.1016/j.nima.2010.04.023. arXiv: 0907.4282.
- [53] Alexander Aab et al. “Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above  $10^{17.8}$  eV”. In: *Physical Review D* 90.12 (2014), p. 122005.
- [54] Pesce, Roberto. “Energy calibration of data recorded with the surface detectors of the Pierre Auger Observatory: an update”. In: *The Pierre Auger Observatory: Contributions to the 32th International Cosmic Ray Conference (ICRC 2011)*. 2011, p. 13.
- [55] Schulz, Alexander. “The measurement of the energy spectrum of cosmic rays above  $3 \times 10^{17}$  eV with the Pierre Auger Observatory”. In: *The Pierre Auger Observatory: Contributions to the 33th International Cosmic Ray Conference (ICRC 2013)*. 2013, p. 27.
- [56] Alexander Aab et al. “Measurement of the cosmic ray spectrum above  $4 \times 10^{18}$  eV using inclined events detected with the Pierre Auger Observatory”. In: *JCAP* 1508 (2015), p. 049. DOI: 10.1088/1475-7516/2015/08/049. arXiv: 1503.07786 [astro-ph.HE].
- [57] Valiño, Inés. “The flux of ultra-high energy cosmic rays after ten years of operation of the Pierre Auger Observatory”. In: *The Pierre Auger Observatory: Contributions to the 34th International Cosmic Ray Conference (ICRC 2015)*. 2015, p. 9.
- [58] Francesco Fenu. “The cosmic ray energy spectrum measured using the Pierre Auger Observatory”. In: *The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*. 2017, pp. 9–16. URL: [http://inspirehep.net/record/1618413/files/1617990\\_9-16.pdf](http://inspirehep.net/record/1618413/files/1617990_9-16.pdf).
- [59] Verzì, Valerio. “The Energy Scale of the Pierre Auger Observatory”. In: *The Pierre Auger Observatory: Contributions to the 33th International Cosmic Ray Conference (ICRC 2013)*. 2013, p. 7.
- [60] Hermann-Josef Mathes, T. “The HEAT Telescopes of the Pierre Auger Observatory: Status and First Data”. In: *The Pierre Auger Observatory: Contributions to the 32th International Cosmic Ray Conference (ICRC 2011)*. 2011.
- [61] Suarez, F. “The AMIGA muon detectors of the Pierre Auger Observatory: overview and status”. In: *The Pierre Auger Observatory: Contributions to the 33th International Cosmic Ray Conference (ICRC 2013)*. 2013.
- [62] Pierre Auger Collaboration et al. “Prototype muon detectors for the AMIGA component of the Pierre Auger Observatory”. In: *Journal of Instrumentation* 11 (2016).
- [63] Alexander Aab et al. “Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory”. In: *Phys. Rev. Lett.* 117.19 (2016), p. 192001. DOI: 10.1103/PhysRevLett.117.192001. arXiv: 1610.08509.
- [64] Figueira, Juan Manuel. “An improved reconstruction method for the AMIGA detectors”. In: *The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*. 2017.

- [65] A. Aab et al. “Muon counting using silicon photomultipliers in the AMIGA detector of the Pierre Auger observatory”. In: *Journal of Instrumentation* 12.03 (2017), P03002.
- [66] Pedro Abreu et al. “Antennas for the detection of radio emission pulses from cosmic-ray induced air showers at the Pierre Auger Observatory”. In: *Journal of Instrumentation* 7.10 (2012), P10011.
- [67] Schröder, Frank G. “Radio detection of air showers with the Auger Engineering Radio Array”. In: *The Pierre Auger Observatory: Contributions to the 33th International Cosmic Ray Conference (ICRC 2013)*. 2013.
- [68] Holt, Ewa M. “Recent Results of the Auger Engineering Radio Array (AERA)”. In: *The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*. 2017.
- [69] Alexander Aab et al. “Measurement of the Radiation Energy in the Radio Signal of Extensive Air Showers as a Universal Estimator of Cosmic-Ray Energy”. In: *Phys. Rev. Lett.* 116.24 (2016), p. 241101. DOI: 10.1103/PhysRevLett.116.241101. arXiv: 1605.02564.
- [70] Alexander Aab et al. “Energy Estimation of Cosmic Rays with the Engineering Radio Array of the Pierre Auger Observatory”. In: *Phys. Rev.* D93.12 (2016), p. 122005. DOI: 10.1103/PhysRevD.93.122005. arXiv: 1508.04267.
- [71] Alexander Aab et al. “Observation of inclined EeV air showers with the radio detector of the Pierre Auger Observatory”. In: *JCAP* 1810.10 (2018), p. 026. DOI: 10.1088/1475-7516/2018/10/026. arXiv: 1806.05386.
- [72] A. Aab et al. “The Pierre Auger Observatory upgrade-preliminary design report”. In: *arXiv preprint arXiv:1604.03637* (2016).
- [73] Daniele Martello. “The Pierre Auger Observatory Upgrade”. In: *The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*. 2017, pp. 147–154.
- [74] Radomir Šmida. “Scintillator detectors of AugerPrime”. In: *The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*. 2017, p. 169.
- [75] Tiina Suomijärvi. “New electronics for the surface detectors of the Pierre Auger Observatory”. In: *The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*. 2017, p. 155.
- [76] Antonella Castellina. “The dynamic range of the AugerPrime Surface Detector: technical solution and physics reach”. In: *The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*. 2017, p. 161.
- [77] Zizhao Zong. “First results from the AugerPrime engineering array”. In: *The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*. 2017, p. 177.
- [78] J. Abraham et al. “Observation of the suppression of the flux of cosmic rays above  $4 \times 10^{19}$ eV”. In: *Phys. Rev. Lett.* 101 (2008), p. 061101. DOI: 10.1103/PhysRevLett.101.061101. arXiv: 0806.4302 [astro-ph].
- [79] J. Abraham et al. “Measurement of the energy spectrum of cosmic rays above  $10^{18}$  eV using the Pierre Auger Observatory”. In: *Phys. Lett.* B685 (2010), pp. 239–246. DOI: 10.1016/j.physletb.2010.02.013. arXiv: 1002.1975 [astro-ph.HE].

- [80] Alessio Porcelli et al. “Measurements of  $X_{\max}$  above  $10^{17}$  eV with the fluorescence detector of the Pierre Auger Observatory”. In: *The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*. 2017.
- [81] Jose Bellido. “Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above  $10^{17.2}$  eV and Composition Implications”. In: *The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*. 2017, pp. 40–47. URL: [http://inspirehep.net/record/1618417/files/1617990\\_40-47.pdf](http://inspirehep.net/record/1618417/files/1617990_40-47.pdf).
- [82] Alexander Aab et al. “Evidence for a mixed mass composition at the ‘ankle’ in the cosmic-ray spectrum”. In: *Phys. Lett. B* 762 (2016), pp. 288–295. DOI: [10.1016/j.physletb.2016.09.039](https://doi.org/10.1016/j.physletb.2016.09.039). arXiv: [1609.08567 \[astro-ph.HE\]](https://arxiv.org/abs/1609.08567).
- [83] Pedro Abreu et al. “Large scale distribution of arrival directions of cosmic rays detected above  $10^{18}$  eV at the Pierre Auger Observatory”. In: *Astrophys. J. Suppl.* 203 (2012), p. 34. DOI: [10.1088/0067-0049/203/2/34](https://doi.org/10.1088/0067-0049/203/2/34). arXiv: [1210.3736 \[astro-ph.HE\]](https://arxiv.org/abs/1210.3736).
- [84] Alexander Aab et al. “Large Scale Distribution of Ultra High Energy Cosmic Rays Detected at the Pierre Auger Observatory With Zenith Angles up to  $80^\circ$ ”. In: *Astrophys. J.* 802.2 (2015), p. 111. DOI: [10.1088/0004-637X/802/2/111](https://doi.org/10.1088/0004-637X/802/2/111). arXiv: [1411.6953 \[astro-ph.HE\]](https://arxiv.org/abs/1411.6953).
- [85] I. Al Samarai. “Indications of anisotropy at large angular scales in the arrival directions of cosmic rays detected at the Pierre Auger Observatory”. In: *The Pierre Auger Observatory: Contributions to the 34th International Cosmic Ray Conference (ICRC 2015)*. 2015.
- [86] Alexander Aab et al. “Observation of a Large-scale Anisotropy in the Arrival Directions of Cosmic Rays above  $8 \times 10^{18}$  eV”. In: *Science* 357.6537 (2017), pp. 1266–1270. DOI: [10.1126/science.aan4338](https://doi.org/10.1126/science.aan4338). arXiv: [1709.07321 \[astro-ph.HE\]](https://arxiv.org/abs/1709.07321).
- [87] A. Aab et al. “Large-scale cosmic-ray anisotropies above 4 EeV measured by the Pierre Auger Observatory”. In: *Astrophys. J.* 868.1 (2018), p. 4. DOI: [10.3847/1538-4357/aae689](https://doi.org/10.3847/1538-4357/aae689). arXiv: [1808.03579 \[astro-ph.HE\]](https://arxiv.org/abs/1808.03579).
- [88] Alexander Aab et al. “An Indication of anisotropy in arrival directions of ultra-high-energy cosmic rays through comparison to the flux pattern of extragalactic gamma-ray sources”. In: *Astrophys. J.* 853.2 (2018), p. L29. DOI: [10.3847/2041-8213/aaa66d](https://doi.org/10.3847/2041-8213/aaa66d). arXiv: [1801.06160 \[astro-ph.HE\]](https://arxiv.org/abs/1801.06160).
- [89] M. Giller et al. “Similarity of extensive air showers with respect to the shower age”. In: *Journal of Physics G: Nuclear and Particle Physics* 31.8 (2005), p. 947.
- [90] F. Nerling et al. “Universality of electron distributions in high-energy air showers—Description of Cherenkov light production”. In: *Astroparticle Physics* 24.6 (2006), pp. 421–437.
- [91] Paolo Lipari. “Concepts of “age” and “universality” in cosmic ray showers”. In: *Physical Review D* 79.6 (2009), p. 063001.
- [92] S. Lafebre et al. “Universality of electron–positron distributions in extensive air showers”. In: *Astroparticle Physics* 31.3 (2009), pp. 243–254.
- [93] S. Andringa et al. “The muonic longitudinal shower profiles at production”. In: *Astroparticle Physics* 35.12 (2012), pp. 821–827.

- [94] L. Cazon et al. “A model for the transport of muons in extensive air showers”. In: *Astroparticle Physics* 36.1 (2012), pp. 211–223.
- [95] Aaron S. Chou et al. “An universal description of the particle flux distributions in extended air showers”. In: *Contributions to the 29th International Cosmic Ray Conference (ICRC 2005)*. 2005.
- [96] Fabian Schmidt et al. “A model-independent method of determining energy scale and muon number in cosmic ray surface detectors”. In: *Astroparticle Physics* 29.6 (2008), pp. 355–365.
- [97] M. Ave et al. “Extensive Air Shower Universality of Ground Particle Distributions”. In: *Contributions to the 32th International Cosmic Ray Conference (ICRC 2011)*. 2011.
- [98] A. Yushkov et al. “Precise determination of muon and electromagnetic shower contents from a shower universality property”. In: *Physical Review D* 81.12 (2010), p. 123004.
- [99] M. Ave et al. *Prediction of the tank response  $S(r, DX, E, \theta, \psi)$  from shower universality*. GAP-2011-087.
- [100] M. Ave. *Ground signal parameterization for water Cherenkov and scintillator detectors and performance of ground arrays that combine both techniques*. GAP-2012-141.
- [101] M. Ave et al. “A generalized description of the signal size in extensive air shower detectors and its applications”. In: *Astroparticle Physics* 87 (2017), pp. 23–39.
- [102] Pierre Billoir. “A sampling procedure to regenerate particles in a ground detector from a “thinned” air shower simulation output”. In: *Astroparticle Physics* 30.5 (2008), pp. 270–285.
- [103] Alexander Schulz. “Measurement of the Energy Spectrum and Mass Composition of Ultra-high Energy Cosmic Rays”. GAP2016-021. PhD thesis. 2016.
- [104] D. Maurel, M. Roth, and J. Gonzalez. “Universality of the time structure of ground particle distributions and its application to the reconstruction of extensive air showers”. In: *Contributions to the 33th International Cosmic Ray Conference (ICRC 2013)*. 2013.
- [105] D. Maurel et al. *Extended shower universality: A model of the time-dependent signal in water Cerenkov tanks*. GAP-2013-022.
- [106] Detlef Maurel. “Mass composition of ultra-high energy cosmic rays based on air shower universality”. GAP2013-105. PhD thesis. 2013.
- [107] M. Ave, M. Roth, and A. Schulz. “A generalized description of the time dependent signals in extensive air shower detectors and its applications”. In: *Astroparticle Physics* 88 (2017), pp. 46–59.
- [108] A. Aab et al. “Azimuthal asymmetry in the risetime of the surface detector signals of the Pierre Auger Observatory”. In: *Physical Review D* 93.7 (2016), p. 072006.
- [109] M. Ave et al. *(Energy,  $N_\mu$ ,  $X_{\max}$ ) from the Bariloche reconstruction*. GAP-2013-109.
- [110] S. Blaess, J. Bellido, and B. Dawson. *Fits to the FD and SD  $X_{\max}$  distributions using the Bariloche Universality reconstruction*. GAP-2015-016.

- [111] Ariel Bridgeman. “Determining the Mass Composition of Ultra-high Energy Cosmic Rays Using Air Shower Universality”. GAP2018-039. PhD thesis. 2018.
- [112] D. Walz and M. Erdmann. *Constraining UHECR source scenarios with spectrum and X<sub>max</sub> distributions*. GAP-2014-034.
- [113] G. Cowan. *Statistical Data Analysis*, Oxford Science Publications.
- [114] L. Molina Bueno, P. Billoir, and I. C. Maris. *Signal variance for the TOTd and MoPS triggers*. GAP-2014-035.
- [115] M. Ave, P. Bauleo, and T. Yamamoto. *Signal Fluctuations in the Auger Surface Detector Array*. GAP-2003-030.
- [116] A. Castellina. *Accuracy of signal measurements in the Auger Surface Detector*. GAP-2003-031.
- [117] Pierpaolo Savina, Lorenzo Perrone, and Carla Bleve. *Validation of Universality model with hybrid events at energies down to 10<sup>18</sup> eV*. GAP-2018-047.
- [118] Ignacio Andres Minaya Flores. “Mass Composition Studies from the Azimuthal Asymmetry in the Risetime of the Surface Detector Signals of the Pierre Auger Observatory”. GAP2016-020. PhD thesis. 2016.
- [119] M. T. Dova, F. G. Monticelli, and H. Wahlberg. *Influence of the reconstructed core position on S(1000) for inclined showers*. GAP-2006-035.
- [120] A. Creusot and D. Veberič. *SD-Hybrid Reconstruction Systematics*. GAP-2008-047.
- [121] F. Gesualdi and X. Bertou. *Determination of an Energy Scale based on the Universality Model using the Scintillator Surface Detectors*. GAP-2017-071.
- [122] Manlio De Domenico et al. “Reinterpreting the development of extensive air showers initiated by nuclei and photons”. In: *Journal of Cosmology and Astroparticle Physics* 2013.07 (2013), p. 050.
- [123] Sergio Petrera and Francesco Salamida. *Update of the X<sub>max</sub> parameterizations for post-LHC hadronic models*. GAP-2018-021.
- [124] A. Aab et al. “Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications”. In: *Phys. Rev.* D90.12 (2014), p. 122006. DOI: 10.1103/PhysRevD.90.122006. arXiv: 1409.5083.
- [125] Alexander Aab et al. “The Pierre Auger Observatory Upgrade - Preliminary Design Report”. In: (2016). arXiv: 1604.03637 [astro-ph.IM].
- [126] William Feller. *An introduction to probability theory and its applications*. John Wiley & Sons, 2008.
- [127] Neil A. Weiss. *A Course in Probability*. Addison-Wesley, 2005.
- [128] X. Bertou and P. Billoir. *On the Origin of the Asymmetry of Ground Densities in Inclined Showers*. GAP-2000-017.
- [129] M.T. Dova et al. “Time asymmetries in extensive air showers: a novel method to identify UHECR species”. In: *Astroparticle Physics* 31.4 (2009), pp. 312–319.

# Acknowledgements

First of all, I want to thank my supervisor prof. Sergio Petrera, who has been a true guide during the development of this work, while teaching me how a physicist should properly carry on research. Without his advices, commitment and patience, this thesis would have been not even half as good as it is.

Then I want to thank Francesco Salamida, for his (scientific and personal) support and his assistance in the issues (usually Offline-related) we faced during this project.

Thanks to prof. Markus Roth for the hospitality at the KIT, and to Ariel Bridgeman for the fruitful exchanges and discussions about the Universality reconstruction.

Thanks also to Massimo Mastrodicasa, who provided me with a useful code to easily employ the Gumbel distributions in the last part of my research.

I'm very grateful to the *Gran Sasso Science Institute* for giving me the opportunity to study the physics of ultra-high energy cosmic rays in L'Aquila.

However, it would have been impossible to reach this result without the support and care of many people also from outside the Auger group.

Therefore many thanks to all my GSSI friends, for all the wonderful time spent together during the three years (and a half) of my PhD. In particular, thanks to the physicists and *urbans* of the XXXI cycle, plus the physicists of the XXXII one: L'Aquila has become a second home for me only because of them.

At the same time, I want to thank my friends in Messina and Catania, who believed in me (and worried for me when the amount of work and stress was growing): despite the distance, they always made me feel their affection.

Last but not least, thanks my family for their infinite love and support.