



HAL
open science

Cosmic-ray astronomy at the highest energies with ten years of data of the Pierre Auger observatory

Lorenzo Caccianiga

► **To cite this version:**

Lorenzo Caccianiga. Cosmic-ray astronomy at the highest energies with ten years of data of the Pierre Auger observatory. High Energy Physics - Experiment [hep-ex]. Université Pierre et Marie Curie - Paris VI, 2015. English. NNT : 2015PA066643 . tel-01346967

HAL Id: tel-01346967

<https://theses.hal.science/tel-01346967>

Submitted on 20 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université Pierre et Marie Curie

École doctorale des Sciences de la Terre et de l'environnement et Physique de l'Univers, Paris

Laboratoire de physique nucléaire et des hautes énergies

**Cosmic-ray astronomy at the highest energies
with ten years of data
of the Pierre Auger Observatory**

Par Lorenzo Caccianiga

Thèse de doctorat de physique

Dirigée par: Piera Luisa Ghia
Co-encadrée par: Julien Aublin

Présentée et soutenue publiquement le 14 septembre 2015

Devant un jury composé de:

Piera Luisa Ghia - Directrice de thèse
Miguel Mostafa - Rapporteur
Elisa Resconi - Rapporteur
Corinne Berat - Examineur
Delphine Hardin - Examineur
Cyril Lachaud - Examineur

Abstract

Identifying the sources of the ultra-high energy cosmic rays (UHECRs, above $\sim 10^{18}$ eV), the most energetic particles known in the universe, would be an important leap forward for both the astrophysics and particle physics knowledge. However, developing a *cosmic-ray astronomy* is arduous because magnetic fields, that permeate our Galaxy and the extra-Galactic space, deflect cosmic rays that may lose the directional information on their sources. This problem can be reduced by studying the highest energy end of the cosmic ray spectrum. Indeed, magnetic field deflections are inversely proportional to the cosmic ray energy. Moreover, above $\sim 4 \cdot 10^{19}$ eV, cosmic rays interact with cosmic photon backgrounds, losing energy. This means that the sources of the highest energy cosmic rays observed on Earth can be located only in the nearby universe (~ 200 Mpc or less). The largest detector ever built for detecting cosmic rays at such high energies is the Pierre Auger Observatory, in Argentina. It combines a 3000 km^2 surface array of water Čerenkov detectors with fluorescence telescopes to measure extensive air showers initiated by the UHECRs. This thesis was developed inside the Auger Collaboration and was devoted to study the highest energy events observed by Auger, starting from the selection and reconstruction up to the analysis of their distribution in the sky. Moreover, since the composition at these energies is unknown, we developed a method to select proton-like events, since high Z cosmic rays are too much deflected by magnetic fields to be used for cosmic-ray astronomy.

In the first chapter, we will give a description of the current general knowledge of cosmic rays: their measurements at Earth, their propagation and consideration on possible acceleration sites. In chapter 2, the principle of detection of UHECRs will be summarized, describing the properties of the extensive air showers (EAS) they produce as they interact with the atmosphere and how they can be detected. Chapter 3 will focus on the Pierre Auger Observatory, describing its technical features and the key results it obtained in 10 years of operation. Chapter 4 will detail the reconstruction of events recorded by the surface detector of the Pierre Auger Observatory, that will be the ones used in this thesis work. Chapter 5 will describe the first part of the work: the building of the dataset of the highest energy events recorded by Auger. A thorough study of all the selected events will be described, together with the improvement to the event selection and reconstruction that were motivated by the findings of this study. In chapter 6 we will present the anisotropy tests to which this dataset has been subjected and the cross-checks made in this thesis work. In chapter 7 we will describe a multi-variate analysis method to separate light from heavy events in the dataset. Finally, in chapter 8, we will perform anisotropy tests on the proton-enriched samples obtained this way.

Contents

1	Cosmic Rays	1
1.1	Cosmic rays at the Earth: energy spectrum, mass composition and arrival directions	2
1.1.1	Spectrum	2
1.1.2	Composition	4
1.1.3	Arrival directions	4
1.2	Cosmic ray propagation	5
1.2.1	Magnetic field deflections	6
1.2.2	Interactions with background radiations	8
1.3	UHECR origin	10
1.3.1	Acceleration mechanism	10
1.3.2	Candidate sources	12
1.3.3	Alternative models	15
2	UHECR Extensive Air Showers: Physics and Detection	17
2.1	Extensive Air Showers	18
2.1.1	Electromagnetic showers	19
2.1.2	Hadronic showers	20
2.1.3	Shower simulations and hadronic interaction models	21
2.2	Principles of Extensive Air Shower Detection	24
2.2.1	Arrays of particle detectors	24
2.2.2	Light detectors	25
2.2.3	Other techniques	26
3	The Pierre Auger Observatory	29
3.1	The site	30
3.2	The Surface Detector	31
3.2.1	Čerenkov stations	31
3.2.2	Data communication and acquisition	32
3.2.3	Maintenance	33
3.2.4	Aperture and exposure	33
3.3	Fluorescence Detector	34
3.3.1	The telescopes	34
3.3.2	FD event reconstruction	34
3.3.3	Atmospheric monitoring	37

3.3.4	Hybrid exposure calculation	37
3.4	Enhancements	38
3.4.1	Low-energy extensions	38
3.4.2	Radio and Microwave detection	39
3.4.3	Planned upgrades	39
3.5	Highlight of the Pierre Auger Observatory results	40
3.5.1	Energy Spectrum	40
3.5.2	Mass composition: depth of shower maximum from FD	42
3.5.3	Mass-sensitive observables from SD and hadronic models	45
3.5.4	Arrival directions	48
3.5.5	Search for neutral particles	50
4	Reconstruction of surface detector events	57
4.1	SD Triggers	58
4.1.1	Station trigger: T1 & T2	58
4.1.2	T3: array trigger	59
4.2	Signal treatment	60
4.2.1	The Vertical Equivalent Muon (VEM)	61
4.2.2	Conversion of the FADC trace into VEM units	61
4.2.3	Saturation recovery	61
4.3	Selection and quality	62
4.3.1	Station selection and T4 physics trigger	63
4.3.2	T5: fiducial trigger	64
4.4	Reconstruction	65
4.4.1	Time fit: first estimate of the arrival direction	65
4.4.2	Lateral distribution function and global fit	65
4.4.3	Uncertainty on the arrival direction	68
4.4.4	Uncertainty on S_{1000}	69
4.5	From S_{1000} to primary energy	70
4.5.1	From S_{1000} to S_{38} : the constant intensity cut	71
4.5.2	From S_{38} to energy: the “golden” hybrid events	72
4.5.3	From S_{38} to energy	73
5	Auger dataset of extremely high energy cosmic rays	75
5.1	Studies on the selection criteria	76
5.1.1	PMT selection	76
5.1.2	Station selection: misidentified non-triggered stations	78
5.1.3	Event selection	81
5.2	Detector aging and its impact on event reconstruction	82
5.2.1	Shape of the signal and effect on the trigger	82
5.2.2	Effects on the event rate	84
5.2.3	Effects on the highest energy dataset	85
5.3	Energy and angular resolutions	86
5.3.1	Angular resolution	86
5.3.2	Energy resolution	87
5.4	The final dataset and consistency checks	90

6	Anisotropy studies	99
6.1	Context of the study	100
6.2	Analysis strategy	102
6.3	Autocorrelation	104
6.4	Cross correlation with astrophysical catalogues	105
6.4.1	Catalogues	105
6.4.2	Cross-correlation with flux-limited samples	106
6.4.3	Cross-Correlation with Bright AGNs	108
6.5	Single-source correlation (CenA)	111
6.6	Search for correlation with neutrino “catalogues”	113
6.6.1	Analysis strategy	113
6.6.2	Magnetic fields deflections	114
6.6.3	Preliminary results	116
6.7	Conclusions	118
7	Multi-variate analysis for mass discrimination	121
7.1	The sample of simulated events	122
7.2	Mass-sensitive observables	123
7.2.1	Longitudinal development	123
7.2.2	Lateral development	126
7.2.3	Muon content	129
7.3	Multi-variate analysis	135
7.3.1	Methods training and fine-tuning	139
7.4	Conclusions	142
8	An application of MVA: correlations test with candidate proton events	145
8.1	The proton-enhanced datasets	145
8.1.1	Cross-check with the Universality method	147
8.2	Correlation tests	149
8.3	Conclusions	150
Appendices		
A	Multi Variate analysis parameters and results plot	157
A.1	Fine-tuned parameters	157
Bibliography		159

Chapter 1

Cosmic Rays

*Omnia qua propter debent per inane quietum
aeque ponderibus non aequis concita ferri*

Titus Lucretius Carus *De Rerum Natura* II, 238-239

Cosmic rays have played an important role in particle physics since their discovery in 1912 by Viktor Hess. Positrons, muons and pions are just few of the particles discovered through cosmic ray studies. In a way, cosmic rays sources were natural particle accelerators before particle accelerators were built on Earth.

After one century of cosmic rays studies there is still one question, perhaps the most important one, that remains unanswered: what is out there, in the universe, that accelerates charged particles up to extreme energies?

The answer to this question is crucial not only to understand the nature of cosmic rays, but also to shed new light on these cosmic accelerators: after years of photon astronomy, we are now attempting to start charged particle astronomy.

In this chapter, a brief resume of our current knowledge on cosmic rays, their energy spectrum, mass composition and possible sources is presented. We will try and follow backwards cosmic rays from their observations at Earth, with their key characteristics (section 1.1), going through their propagation in space (section 1.2) and finally reaching the sources that may have accelerated them (section 1.3).

We will discuss results regarding cosmic rays at all energies, but we will focus in particular on the highest energy cosmic rays¹, since they will be the main focus of the work of this thesis. The results coming from the Pierre Auger Observatory will be described in detail in the next chapter 3.

¹Ultra-high energy cosmic rays, UHECRs, here considered for convention with energy above $\sim 10^{17}$ eV.

1.1 Cosmic rays at the Earth: energy spectrum, mass composition and arrival directions

“Cosmic rays” is a general term which refers to all the high-energy massive particles that are generated outside the solar system². In this section we will describe the current status of cosmic ray measurements at the Earth. In particular, we will go through the energy spectrum of cosmic rays and its characteristics, and then go into composition and arrival direction measurements and how they evolve as a function of energy.

1.1.1 Spectrum

The energy spectrum of cosmic rays spans over more than 10 orders of magnitude in energy and 30 in flux, as can be seen in fig 1.1: for energies above ~ 30 GeV, it is impressively well described by the power law $\frac{dN}{dE} \approx E^{-\gamma}$. Due to this power law, the flux is very high at lower energies (1 particle per m^2 per second at 100 GeV) and decreases quickly as energy rises, up to less than one particle per km^2 per century at 100 EeV³. Due to the huge differences in flux and energy, cosmic rays of different energies cannot be detected in the same way. In particular, up to $\sim 10^{14}$ eV, the flux is high enough to allow direct detection. Because of the atmosphere shielding, direct detection of primary cosmic rays has to be done in the high atmosphere, through balloons, or from the orbit, with satellites. At higher energies, the detection of cosmic rays is made indirectly, measuring the shower of particles that cosmic rays induce when interacting with the high atmosphere. The characteristics of the shower carry information on the primary energy, arrival direction and nature, as we will describe in detail in the next chapter.

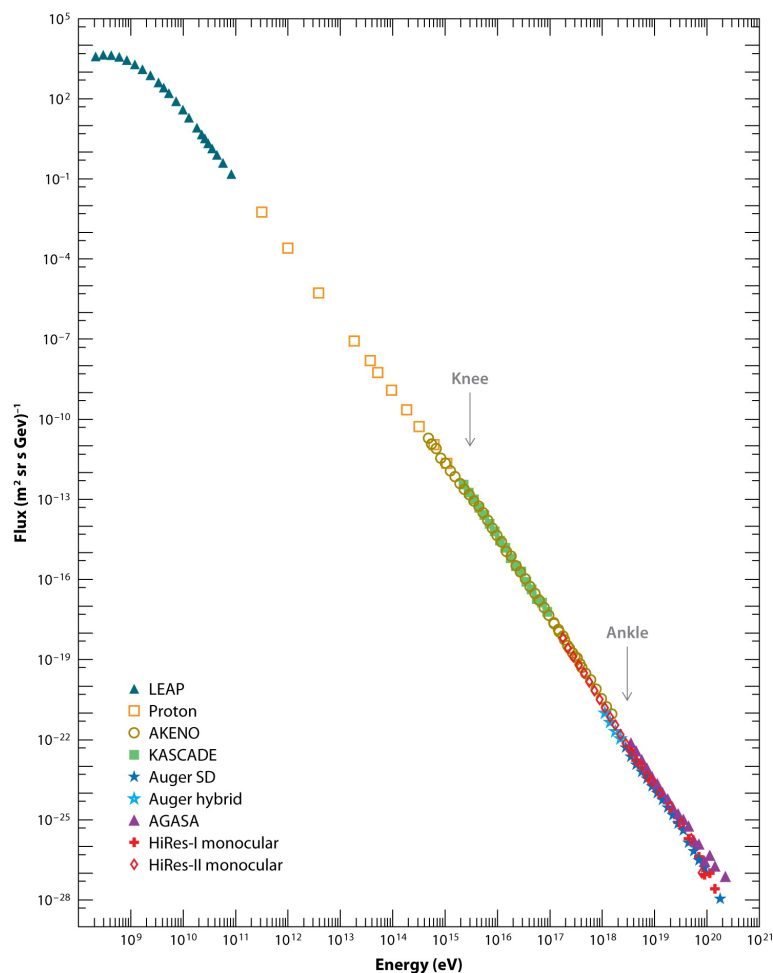
The power law factor γ has a value of $\sim 2.6 - 2.7$ up to $\sim 5 \cdot 10^{15}$ eV, then it rises to about 3.1. At about $4 \cdot 10^{18}$ eV there is another change in the slope, as γ falls back to a value of about 2.7. The points where the slope changes are usually called knee and ankle respectively. The possibility of another slight change in the slope between the knee and the ankle, at around $8 \cdot 10^{16}$ eV, called second knee, is being debated [9].

The knee is thought to be caused by a limit in the energy accessible through the acceleration mechanism in our Galaxy, in particular in the case that the acceleration mechanism is dependent on the charge of the cosmic ray, and the maximum energy protons can reach is about $5 \cdot 10^{15}$ eV. In this scenario, the steepening of the spectrum between the knee and the ankle would be due to the superimposition of the cut-offs of different nuclei, which can be accelerated to higher and higher energies according to their charge.

The ankle could be caused by the transition from Galactic sources to extra-Galactic ones. Galactic sources, would reach their maximum accelerating capabilities, leaving extra-Galactic sources to contribute more and more to the total flux. Another explanation of this feature is the so-called dip model [10]: assuming an extra-Galactic proton component, the ankle is naturally obtained from the interaction of cosmic rays with the cosmic microwave background through $e^+ e^-$ production. The discrimination between the different models relies mainly on mass composition studies.

²Sometimes also particles coming from the Sun, usually known as *solar wind*, are considered as the low-energy cosmic rays.

³1 EeV = 10^{18} eV.




 Beatty JJ, Westerhoff S. 2009.
 Annu. Rev. Nucl. Part. Sci. 59:319–45

Figure 1.1: Cosmic rays flux as a function of energy, as measured by different experiments. Data from LEAP [1], Proton [2], AKENO [3], KASCADE [4], Auger surface detector (SD), Auger hybrid [5], AGASA [6], HiRes-I monocular and HiRes-II monocular [7]. From [8]

At energies higher than $\approx 5 \cdot 10^{19}$ eV a fall in the flux is observed [7, 5]: it is usually explained as the result of the interaction of cosmic rays protons with cosmic background radiation. This hypothesis will be extensively discussed in paragraph 1.2.2. Other possible interpretations of this suppression are a limit of the acceleration capabilities of the sources or photodissociation of heavy primaries. Both these explanations are supported by the evidences of a trend towards a heavier composition at the highest energies ([11, 12] see paragraph 3.5.2). A more detailed review of the most recent measurements of the cosmic ray spectrum at the highest energies will be provided in section 3.5.1.

1.1.2 Composition

The composition of cosmic rays is directly measurable as long as a direct measurement of the cosmic rays is possible, on balloons or satellites up to ~ 100 TeV. In this energy region, the electron fraction is higher 1-2% up to 100 TeV and then decreases rapidly. Protons are the most abundant species (89%), with an important fraction of helium nuclei (10%) and fewer heavier nuclei, up to really high Z (> 92). The relative abundance of nuclei is close to what is observed in the Solar System but it does not exactly match, because of spallation in the interactions with the interstellar medium. The study of the presence of ultra-heavy nuclei and, in particular, of radioactive isotopes in cosmic rays is an active field of study with various experiments going on, both on balloon (ATIC, CREAM, TIGER) and in space (ISS-CREAM, AMS-02) [13, 14, 15, 16, 17].

At higher energies, where detection is made indirectly, the primary mass has to be inferred from the characteristics of the shower. In the energy region between the *knee* and the *ankle*, the two most common techniques to infer the chemical composition are the measurement of the number of muons (used for instance in KASCADE-Grande[18] and IceCube-IceTop[19]) and the detection of the Čerenkov light (used in Tunka-133) [20]. At energies above the ankle the measurement of the shower maximum depth via fluorescence telescopes is the most used technique (see paragraph 3.5.2 for further details). Since these are not direct measurements, to extract an absolute value of the primary mass measurements have to be compared to predictions from simulated particle showers (see paragraph 2.1.3 for further details about these models). In figure 1.2 a summary of the results obtained by these and other experiments is plotted. The figure shows the behavior of the average logarithm of the primary mass A as a function of energy. We can see a trend towards a heavier composition up to $\sim 10^{17}$ eV followed by a rapid decrease towards proton-like primaries for events around $\sim 10^{18}$ eV. Measurements at higher energies will be discussed later in paragraph 3.5.2.

1.1.3 Arrival directions

The third important feature that can be measured in cosmic rays is their arrival directions. Because of magnetic deflections (see next paragraph) we expect a strong anisotropy only for the highest energies. However, if some particular features of the magnetic field at the boundary of the solar system or farther out is present, it might focus cosmic rays along certain lines and the observed arrival direction distribution could consequently be anisotropic also at lower energies.

Indeed, while we can easily assess that, at first order, the arrival directions of cosmic rays are isotropic, large-scale ($90^\circ - 120^\circ$) anisotropies have been detected since the 1930s. Such asymmetries are now observed with amplitudes of the order of $10^{-3} - 10^{-4}$ [21, 22] up to $\sim 10^{14}$ eV. The anisotropy amplitude is energy dependent, but the topological structure remains nearly the same in the energy range from tens of GeV to hundreds of TeV. Generally, an excess region named “tail-in” (distributed around 40° to 90° in Right Ascension (R.A.)) and a deficit region named “loss cone” (around 150° to 240° in R.A.) are observed, in a dipole-like structure. These features were observed, among others, by the Tibet AS γ [23], ARGO-YBL [24] and MILAGRO [25] observatories, although with some disagreement.

These experiments suggested also a smaller scale ($10^\circ - 30^\circ$) anisotropy, with few high

density regions observed. Such anisotropy has been recently confirmed by both IceTop-IceCube [26] and HAWC [27] experiments, in the southern and in the northern hemisphere respectively. So far, no theory of CRs in the Galaxy exists which is able to explain both the observed large- and intermediate-scale anisotropies while at the same time leaving the standard model of CRs and that of the local Galactic magnetic field unchanged [21].

At energies higher than 100 TeV, the anisotropy pattern abruptly changes, losing its dipole-like structure. Recent results from the KASCADE-Grande experiment pose stringent upper limits on the dipole anisotropy at energies higher than $2.6 \cdot 10^{15}$ eV [28]. However, if the dipole amplitude measurements are not significant, it is interesting to see that the dipole phases obtained from this study were in almost all energy bins pointing towards the Galactic center. This is interesting in particular since it agrees inside the statistical uncertainties with the results obtained by the Pierre Auger Observatory at higher energies (above $3 \cdot 10^{17}$ eV), which will be described in more detail in paragraph 3.5.4.

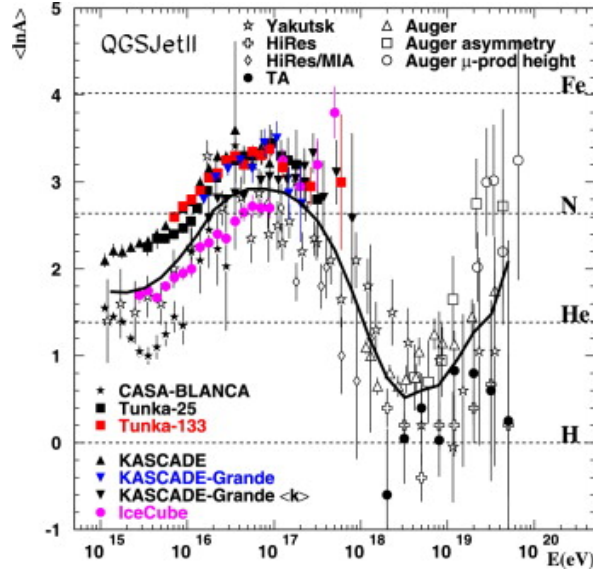


Figure 1.2: Average logarithm of the cosmic ray mass as a function of energy, extracted from the comparison of measurements from different detectors with predictions from the QGSJET-II interaction model. The solid line is drawn as a guidance and is obtained by averaging the values of $\langle \ln A \rangle$ in each bin. From [20].

1.2 Cosmic ray propagation

To correctly interpret the measurements at Earth that were described in the previous section, and link them to the possible sources of cosmic rays, which will be described in the next section, we need to know what happens in between. In particular, since cosmic rays are mainly charged particles, their interactions with the magnetic fields have to be taken into account when studying the arrival direction distribution of cosmic rays. Moreover, at the highest energies, cosmic rays may interact with the cosmic photon background. In this

section we will briefly describe the propagation effects that will be important for deducing information on cosmic ray sources based on the measurements made at the Earth.

1.2.1 Magnetic field deflections

Cosmic rays are mainly charged particles; therefore, as they propagate through magnetic fields, they experience deflections proportional to their charge and inversely proportional to their energy. In the case of a uniform field the deflections can be easily computed as

$$d\theta(E, d) \propto Z \cdot \left(\frac{E}{10^{20} \text{ eV}} \right)^{-1} \cdot \left(\frac{B}{10^{-9} \text{ G}} \right) \cdot \left(\frac{d}{\text{Mpc}} \right)$$

where Z is the charge of the cosmic ray, E is its energy, B is the strength of the magnetic field, d the distance traveled. A proton of about 100 EeV traveling through a nG magnetic field for 1 Mpc is, for example, deflected by less than 1° .

Due to this deflection, identifying the sources of cosmic rays is far more difficult than in photon astronomy. Only the highest energy cosmic rays may travel without suffering too strong a deflection that would prevent the identification of their sources. It is indeed important to notice that an iron nucleus will be deflected 26 times more than a proton of the same energy.

In the following two paragraphs a summary of the present knowledge on Galactic and extra-Galactic magnetic fields is presented.

Galactic magnetic fields

The magnetic field of our Galaxy (GMF) was mostly unknown until few years ago. The present generation of sophisticated and quantitatively-constrained models of the GMF has emerged only recently.

The GMF is mostly studied through measurements of the Faraday rotation of the polarization in radio emission of different sources [29]. The Faraday effect causes a rotation of the plane of polarization which is linearly proportional to the component of the magnetic field in the direction of propagation. Since in this case the effect is caused by free electrons, the formula that describes the rotation along a sight path of length d is written in CGS units as

$$RM = \frac{e^3}{2\pi m_e^2 c^4} \int_0^d n_e(s) B_{||}(s) ds \quad (1.1)$$

where $n_e(s)$ and $B_{||}(s)$ are respectively the density of electrons and the component of the magnetic field parallel to the line of sight, both calculated in each point s along the sight path. If we have an estimate of $n_e(s)$, we can then extract $B_{||}(s)$ from the rotation measurements. If the rotation measurement is made on a pulsating source, this can be done with the measurement itself, since the dispersion caused by these electrons results in a time delay between pulses received at different wavelengths. Measuring at the same time the polarization and the timing of the pulses then allows us to have an estimate of the weighted mean of the magnetic field along the line of sight. This kind of measure may have been challenging until only a few decades ago, but today the number of accessible sources has increased by two orders of magnitude since the 1990s and will likely increase again, thus resulting in much better constraints on GMF [30]. Other constraints on the GMF

may come from maps of polarized and unpolarized Galactic synchrotron emission, such as those produced by WMAP [31] and soon by Planck [32]. Galactic synchrotron emission is complementary to Faraday rotation, since the former depends on the transverse magnetic field, while the latter is sensitive to the parallel one.

The GMF is commonly split into its regular (or coherent) and random components. The regular field is the component with a large coherent length, which can be modeled with simple functions depending only on the position in the Galaxy. The random field, on the other hand, has a small coherence length, and for this reason it averages to zero on scales larger than about a few hundreds of parsecs.

In principle, the strength of the regular component is proportional to the matter density in the Galaxy and decreases with the distance from the Galactic center. On the Galactic plane, the field decrease is well described by the exponential function $e^{-R_{GC}/8.5}$ where the distance from the Galactic center R_{GC} is expressed in kpc. This expression works only for $R_{GC} > 3$ kpc since in the Galactic bulge region the magnetic field is difficult to study. In the vicinity of the Solar system the strength of the regular magnetic field is about $2 \mu\text{G}$. Several models for the GMF regular component exist in literature; all of them predict much larger deflections for cosmic rays coming from the Galactic center region.

An example of deflection map for extremely high energy cosmic rays protons in the Jansson-Farrar 2012 (JF2012) model of the regular GMF [33], which will be described later in section 6.6.2, can be seen in figure 1.3. Random events were generated isotropically all over the sky at energies ranging from 40 EeV to 150 EeV following the cosmic rays spectrum as measured by Auger (see Section 3.5.1). Then the events were backtracked through the magnetic fields and the angular deflection from the original position was plotted. Overall, these studies lead us to conclude that the deflections in the regular Galactic magnetic field may be of the order of tens of degrees even for protons at the highest energies ($E > 40$ EeV), in particular for events coming from the Galactic center region.

Measuring the random component of the GMF is more difficult. In any case, the estimate that near the Solar system the total magnetic field strength is about $6 \mu\text{G}$ suggests, by subtraction, that the random magnetic field strength is about $4 \mu\text{G}$, i.e. twice as big as the regular field.

The Galactic magnetic halo, i.e., the extension of the magnetic field above and below the Galactic plane, is still not well known. Recent measurements suggest that the halo could significantly contribute to cosmic ray scattering [34].

Extra-Galactic magnetic fields

The extra-Galactic magnetic fields (EGMF) are much less known than the Galactic ones. The observations through Faraday rotation measurements lead us to believe that magnetic fields of $\approx 1 - 40 \mu\text{G}$ are present in galaxy clusters, while the upper limit of the average magnetic fields all over space is $1 \mu\text{G}$ if the correlation length of the field is 1 Mpc, i.e., the average distance between galaxies [35].

The structure and strength of the EGMF can strongly affect the propagation of UHE-CRs on large distances (tens-hundreds of Mpc). The uncertainties on EGMF distribution and strength leads to important differences in the simulated deflections induced on UHE-CRs. For a proton with energy $E \gtrsim 100$ EeV deflections from less than one degree up to

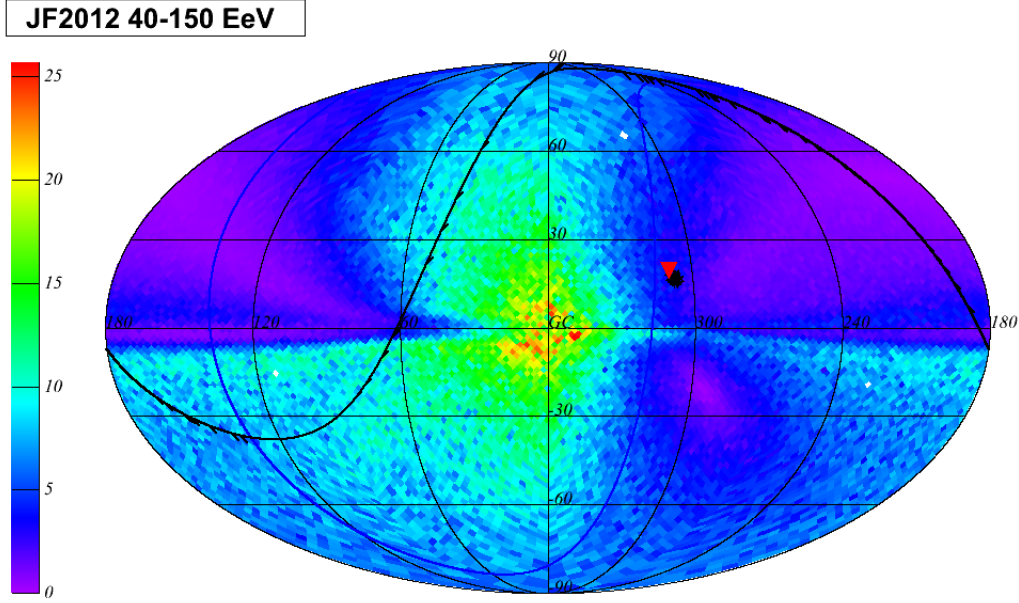


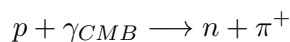
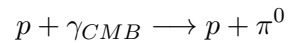
Figure 1.3: Deflections (see scale in degrees on the left) for cosmic rays of energy 40–150 EeV in the GMF as predicted by the Jansson Farrar 2012 model. The map is in galactic coordinates (see section 6.6.2 for more details).

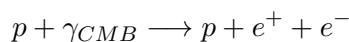
$10^\circ - 20^\circ$ have been suggested [36].

More details about magnetic field deflections and models will be described later in section 6.6.2, as these are relevant for this thesis work. Indeed, we have to keep in mind that, in spite of still incomplete measurements of the magnetic field strength in the Galaxy and scarce information on the magnetic fields in its halo and in the intergalactic medium, cosmic rays may be heavily deflected even at the highest energies, especially in the case of large Z primaries.

1.2.2 Interactions with background radiations

Cosmic rays at the highest energies are expected to suffer a strong flux attenuation due to their interaction with cosmic backgrounds that permeate the whole space. In particular, photons of the Cosmic Microwave Background (CMB) have an average energy of $E_\gamma \approx 10^{-4}$ eV and their energy density is 0.2 eV cm^{-3} [37]. A proton propagating through space may then interact with CMB photons if its energy is beyond a certain threshold. The most probable interactions are pion photoproduction (through Δ^+ resonance) and pair production:





The threshold energy for pion photoproduction is:

$$E_{th} = \frac{m_\pi}{4E_\gamma} (2m_p + m_\pi) \approx 7 \cdot 10^{18} \text{ eV using } E_\gamma = 1 \text{ meV}$$

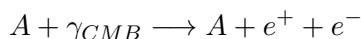
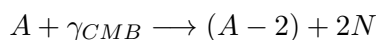
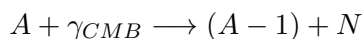
The threshold energy for pair production turns out to be lower:

$$E_{th} = \frac{m_e(m_p + m_e)}{E_\gamma} (2m_p + m_\pi) \approx 5 \cdot 10^{17} \text{ eV using } E_\gamma = 1 \text{ meV}$$

Note that here the *average* value for CMB photon energy was used. Since the CMB spectrum is almost a perfect black body, higher energy photons are available for those interactions, thus reducing threshold energies.

The pion photoproduction effect was first suggested by Greisen [38] and independently by Zatsepin and Kuz'min [39] soon after the discovery of CMB by Penzias and Wilson [40]. It is therefore common to refer to those interactions as the GZK effect.

Nuclei can interact with the CMB, in particular through photodisintegration reactions and pair production:



Threshold energies for those reactions increase with the mass, and therefore lighter components are more quickly disintegrated than heavier ones. Therefore, cosmic rays coming from relatively distant sources (a few tens of Mpc) are expected to be protons or heavy nuclei, in particular iron, with lack of light and middle-weighted components.

The energy losses associated with these interactions imply a suppression of the cosmic ray flux at Earth at the highest energies (the so-called GZK cutoff), as well as a limit to the distance from which a source can contribute significantly to the flux (the so-called GZK horizon). We can indeed compute a mean free path associated with those interactions: $\lambda = (n_\gamma \sigma)^{-1}$ where n_γ is the CMB photon density ($\approx 400 \text{ cm}^{-3}$) and σ the cross section of the interaction ($\approx 10^{-28} \text{ cm}^2$ at threshold energy). For example, for a cosmic ray proton at the threshold energy, λ_p is $\sim 8 \text{ Mpc}$: since a proton loses on average 20% of its energy in each interaction, its energy will decrease by an order of magnitude after traversing 100 Mpc. Taking into account the mean free path (that depends on primary energy and mass), it is possible to evaluate an “attenuation factor”, given by the fraction of events injected by the source with energy above the threshold which still remain above threshold after traversing a distance D . This is illustrated in figure 1.4 for sources that inject protons, silicon and iron. The two graphs show the fraction of protons, silicon and iron cosmic rays with energy $E > 60 \text{ EeV}$ (left panel) and $E > 80 \text{ EeV}$ (right panel) that arrive at Earth from sources that are farther away than a distance D . Around 90% of the flux of protons with $E > 60 \text{ EeV}$ should come from distances smaller than 200 Mpc. This distance shrinks to 90 Mpc if $E > 80 \text{ EeV}$. We can see that the suppression (horizon) is of comparable size for protons and iron nuclei, and is higher (smaller) for intermediate-mass nuclei.

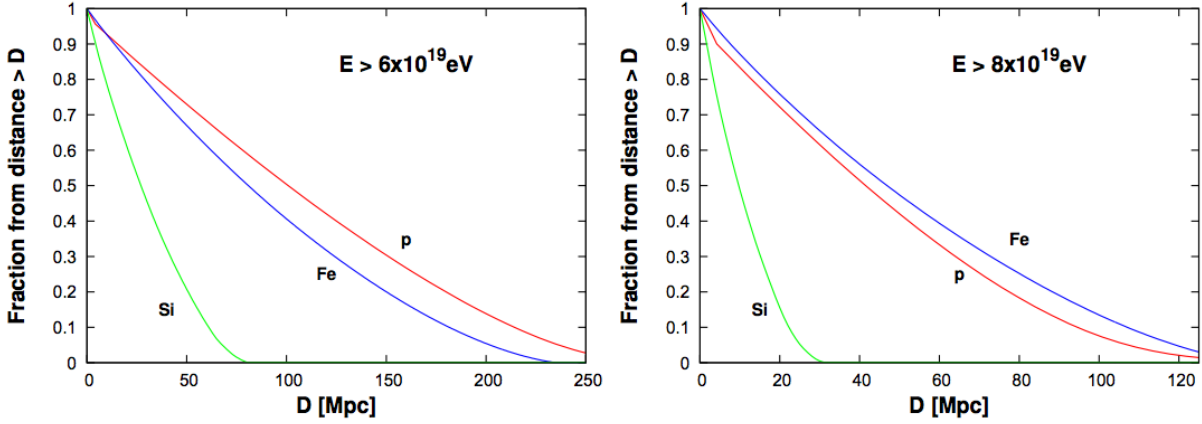


Figure 1.4: GZK horizon: fraction of CRs that arrive at Earth with energy above 60 EeV (left panel) and 80 EeV (right panel), from sources that are farther away than a distance D and that inject protons, silicon and iron, respectively. From [41].

1.3 UHECR origin

After following cosmic rays from Earth back through magnetic fields and cosmic backgrounds, we can ask ourselves the question: which astrophysical objects have the possibility to accelerate cosmic rays up to such high energies? We will present here the most plausible candidates, after briefly describing how acceleration is thought to be achieved.

1.3.1 Acceleration mechanism

In order to understand which astronomical objects are candidate sources of cosmic rays it is important to understand the mechanisms that can accelerate particles to such high energies. All basic acceleration mechanisms require magnetic fields: it is indeed required for the particle to be confined into the source in order to be accelerated. This implies that the magnetic field of the source must be coherent at least in the space necessary to contain the radius of curvature of the particle at the requested energies. The value of this radius directly depends on the strength of the magnetic field, thus leading to an estimation of the highest energy achievable, i.e.

$$\left(\frac{E_{max}}{\text{EeV}}\right) \propto Z \left(\frac{B}{\mu\text{G}}\right) \left(\frac{R}{\text{kpc}}\right)$$

where Z is the particle charge, B the strength of the magnetic field and R is the radius of the acceleration region. This implies that, to accelerate particles to ultra-high energies, either highly condensed objects with huge B or enormously extended objects are needed. Note that energy losses at the source will lower this estimation. This effect varies a lot in the different kind of sources and affects also the mass composition of the accelerated cosmic rays: nuclei for example can photodisintegrate inside the sources.

Fermi acceleration mechanism

The most commonly invoked particle acceleration mechanism in astrophysics, and probably one of the first ones modeled, is the diffusive shock acceleration, also known as Fermi I process.

In 1949 [42], Enrico Fermi proposed a mechanism of acceleration for charged particles through interaction with a magnetized moving gas. Initially, he proposed a mechanism (later named Fermi II) able to explain energy transfer from magnetized clouds to individual particles. The idea was that energetic particles would be scattered by magnetic irregularities; in the interaction with a magnetized cloud the particle changes its direction and gains or loses energy depending on whether the collision is head-on or tail-on. After averaging over a number of particle-cloud interactions, thanks to the fact that the head-on collisions are slightly more probable, the process leads to a net energy gain ($\sim \Delta E/E_0 = \epsilon = 4/3\beta^2$ where $\beta = v/c$ is the velocity of the cloud). The quadratic dependence on the velocity is what gave this method the name Fermi II. Since the velocity of the cloud is not relativistic, $V \ll c \rightarrow \beta \ll 1$, this means that, in this formulation, the acceleration turns out to be very slow. If, however, the interaction happens in a shock area, things change drastically: the velocity of magnetic irregularities, on which particles will interact, is negligible with respect to the fluid velocity on both sides of the shock. This implies that the interaction centers can be considered at rest with the fluid both upstream and downstream. In the shock system, whether a particle moves from upstream to downstream or vice-versa, it always sees the fluid on the other side of the shock as approaching. Therefore, each time an energetic particle crosses the shock, it always suffers head-on collisions with magnetic irregularities on the other side of the shock, thus gaining energy much faster: $\sim \Delta E/E_0 = \epsilon = 4/3\beta_s$ where the velocity in β_s is now the velocity of the shock that is also much higher than the typical velocity of magnetic perturbations in the Galaxy [43]. The linear dependence on the shock speed is the reason for this method to be called Fermi I. The attractive feature of this method is that it reproduces the power-law spectral feature that is observed. Indeed, after n collisions the particle will reach an energy

$$E_n = E_0(1 + \epsilon)^n$$

If after every collision the particle has a probability P to escape from the accelerating region, after n collision the probability that it escapes at the n -th collision will be

$$P_n = P(1 - P)^n$$

Given an initial number of particles N_0 after n collisions, a number

$$N_n = N_0 P_n = N_0 P (1 - P)^n$$

will escape with an energy E_n . Using previous equations, this can be re-written as

$$N_n = N_0 P \left(\frac{E_n}{E_0} \right)^{\frac{\ln(1-P)}{\ln(1+\epsilon)}}$$

This latter equation immediately translates into the differential spectrum of accelerated particles:

$$\frac{dN}{dE} \sim \frac{N_n}{(E_{n+1} - E_n)} = \frac{N_n}{\Delta E} \propto E^{-\gamma}$$

with $\gamma = 1 - \frac{\ln(1-P)}{\ln(1+\epsilon)}$. With reasonable values for P and ϵ this value turns out to be $\gamma \sim 2$, that, taking into account the diffusion of cosmic rays in the Galaxy, is consistent with the observed spectrum at Earth. This schematic and simplified calculation allows us to understand why this simple model is still considered more than 60 years after its proposal. However, it is just a basis on which more complete and complex models for the acceleration mechanism have been developed.

Non linear diffusive shocks

While the sources of the high energy cosmic rays are still unknown, some interesting evidence that cosmic rays up to 10^{17} eV are probably generated in supernovae remnants (SNR) have been recently published [44]. However, using a simple model as the Fermi one, the conditions measured inside a SNR are not enough to explain the acceleration of particles to these energies. Moreover, the previous model was made for a test-particle, meaning that the acceleration of cosmic rays had a negligible effect on the energy balance of the source. This is not compatible with SNRs being the main factories of Galactic cosmic rays, because this would imply that around 10% of the kinetic energy of the blast wave must be converted into accelerated particles. This implies that the test-particle description of the shock acceleration process is not appropriate and the dynamical back-reaction of the particles on the accelerators must be taken into account. So if SNRs are the primary sources of Galactic cosmic rays, a correct description of the acceleration process can only be achieved within the framework of Non-Linear Diffusive Shock Acceleration with Magnetic Field Amplification. This has been done in particular in the last decade and now we have good theoretical models that can correctly describe how acceleration in SNRs take place (see e.g. [43]).

1.3.2 Candidate sources

The elements given in the previous paragraphs are enough to build a scheme to identify and classify possible ultra-high energy cosmic rays sources among astrophysical objects, based on their magnetic fields and their dimensions. The basic condition to impose is that the particles should be contained in the source up to the final energy that they will acquire. This means that the Larmor radius of the particle should not exceed the size of the acceleration region: $r_L = \frac{E_{max}}{ZecB} \leq L$. A comparison of this request with the known measurements of L and B for some candidate sources is shown in the so-called Hillas Plot [45] (figure 1.5), where the typical magnetic fields of various possible sources of 100 EeV protons and irons are plotted as a function of their dimensions.

In the following we provide a list of the astrophysical candidate sources that have been proposed for the acceleration of the highest energy cosmic rays. The first two candidates are both related to Active Galactic Nuclei (AGNs), i.e., galaxies with a nuclear supermassive black hole with accretion processes running. Schematic views of an AGN are shown in figures 1.6 and 1.7 to explain the different parts named in the text below.

1. AGN accretion disk: AGNs have been considered as appealing candidate sources of cosmic rays for the last 50 years [45, 46]. In this case the engine of the acceleration is the accreting disk of matter orbiting around a supermassive black hole ($10^6 - 10^{10} M_\odot$) in the nucleus of the galaxy. Those accreting disks are known to have, in

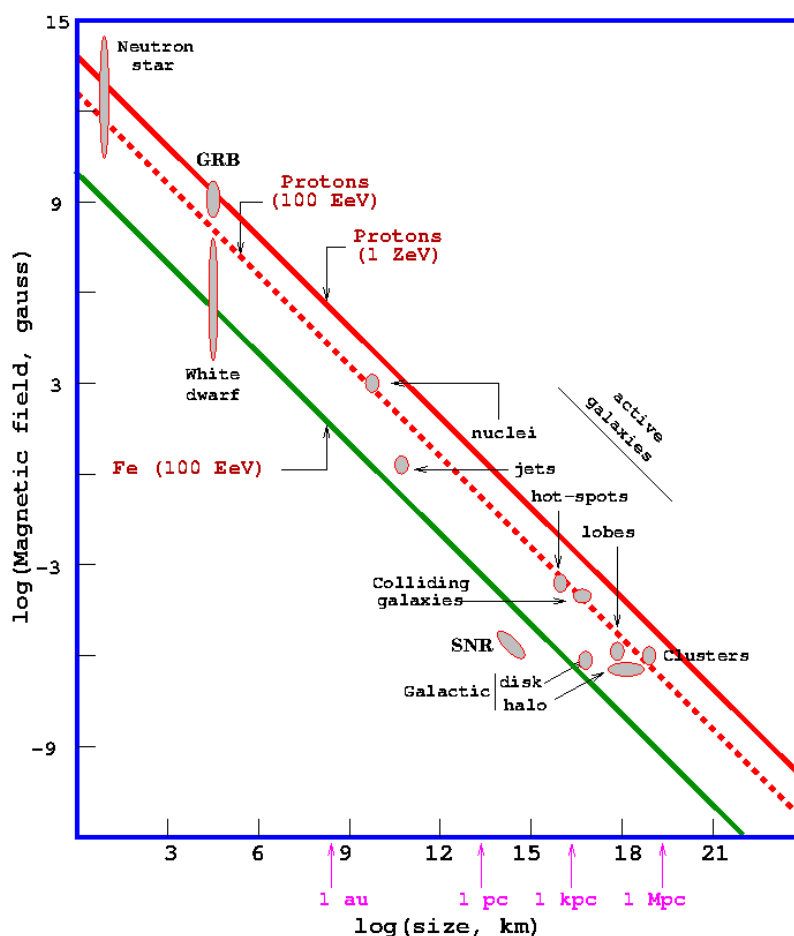


Figure 1.5: The so-called Hillas Plot. In double-logarithmic scale, typical dimensions (x-axis) and magnetic fields (y-axis) of various possible sources are plotted. Lines show the minimum B and R ratios that allow us to accelerate iron nuclei up to 100 EeV (green line), protons up to 100 EeV (dashed red line) and protons up to 1000 EeV (red line) [45].

the jets (with dimensions of the order of $10^{-4} - 10^{-3}$ pc), magnetic fields of the order of few G [47]. These parameters, already marginally consistent with acceleration up to 100 EeV, are unlikely to be achieved under realistic conditions. The high radiation field around the central engine of an AGN is likely to interact with the accelerated protons while energy losses, due to synchrotron radiation, Compton processes, and adiabatic losses, will also take place. The situation is worse for nuclei that will photodisintegrate even faster. For this reason, it is now believed that the inner part of AGN galaxies could only contribute to accelerating particles up to small fractions of EeV [35].

2. FRI and FR II radio-galaxy lobes: in Fanaroff-Riley (FR) radio-loud active galaxies, extended lobes of magnetized plasma originate from the central black hole. The lobes are observed mainly in radio through synchrotron emission, and may slowly vanish with distance from the nucleus (FRI, in this case the lobes may be called

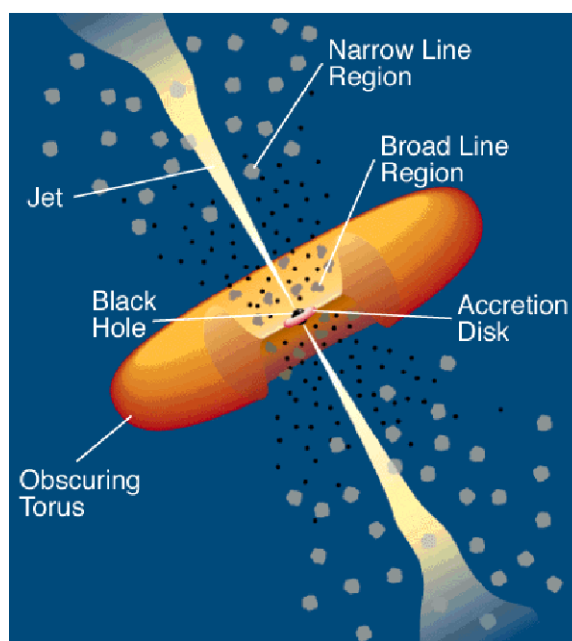


Figure 1.6: Active Galactic Nuclei scheme. It is possible to see the supermassive black hole surrounded by its accreting disk. The jets here represented are typical of Radio Loud AGN and Blazars, and are not present in every active galaxy. Figure from [48].

also plumes) or show high density areas called hot-spots at the lobe edges (FR II). In figure 1.7 maps in radio band of an FRI galaxy (3C31) and an FR II one (3C98) are shown to visualize the differences. The different behavior of the lobes in the two galaxy classes is probably due to the fact that the jets leaving the accretion region are supersonic (FR II, $\beta \sim 0.3$) or subsonic (FRI) [49]. Lobes usually exceed in size the galaxy itself with dimensions up to hundreds of kpc. In the lobes, magnetic fields are usually of less than a μG . On the other hand, hot spots are usually of the order of few kpc and show magnetic fields of the order of hundreds of μG . Both the lobes and the hot spots are candidates to accelerate protons up to hundreds of EeV. If the jets are oriented towards the Earth, these objects are usually called blazars. Blazars are much brighter and can be detected at higher distances than radiogalaxies; but, since the jet is collimated, they are much rarer than radiogalaxies. Nevertheless, blazars and radiogalaxies are the same objects and can be equally considered as candidates for UHCER acceleration.

3. Galaxy clusters: in galaxy clusters, magnetic fields of the order of μG are thought to be present. Those magnetic fields may extend to large dimensions (up to 500 kpc). This should allow acceleration of UHECRs up to 100 EeV. Actually, the fact that long paths in the galaxy cluster are needed increases the probability of interacting with cosmic backgrounds (see section 1.2.2) and lowers the maximum energy expectation to some tens of EeV.
4. Compact objects: cosmic rays could be accelerated directly by compact objects such

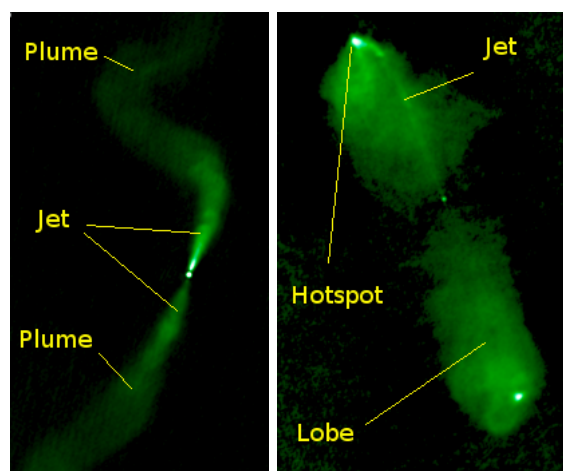


Figure 1.7: A Fanaroff-Riley type I galaxy (left, 3C31) and a type II one (right, 3C98) seen in radio band to show the morphological differences.

as neutron stars, in particular pulsars and magnetars. In this case the maximum achievable energy should be $E_{max} = \frac{\omega}{c} \cdot Z \cdot B_s \cdot R^2$ where ω is the spinning velocity of the object, Z the charge of the accelerated particle, B_s the magnetic field at the surface and R the radius of the object. For example, using values from the Crab Nebula Pulsar: $B_s \sim 10^{12}$ G, $R \sim 10$ km, $\omega \sim 200$ Hz, we obtain a maximum energy of about 10^{18} eV for a proton. Objects with more “extreme” characteristics such as magnetars, where magnetic fields up to 10^{16} G are thought to be present, are then necessary to explain UHECRs [50].

5. Gamma Ray Bursts (GRB): Gamma Ray Bursts are prompt explosions in the γ band. GRBs duration distributions show two different populations which are thought to have different origins: short (< 1 s) GRBs are possibly the result of the merge of two neutron stars of a binary system into a black hole, while long ($> 1 - 2$ s) ones have been associated to hypernovae. Some GRBs may be also single Soft Gamma-ray Repeaters (SGR), a class of γ -emitting neutron stars. In particular, the accreting disk around the neutron star binary system could be a good site for particles acceleration. GRBs are though disfavored as primary UHECR source since the major fraction of them were observed at high redshifts (up to $z = 5$) and the few GRBs observed within the GZK horizon would not explain the observed flux [36].

1.3.3 Alternative models

Models which do not require an astrophysical object that accelerates cosmic rays have been proposed. Those “exotic” models suggest that cosmic rays are produced directly at high energy (so-called Top-down models) as products of supermassive particle decays rather than being accelerated [51, 52].

Those models were invoked to explain the apparent lack of flux suppression observed in the 90s by the AGASA experiment. To explain the observed spectrum, those particles should have high masses ($m > 10^{11}$ GeV), be fairly abundant and have a mean lifetime

long enough to survive on cosmological timescales.

Such particles should decay into quarks and leptons. Quarks would then produce hadrons: mostly pions, and a small fractions of nucleons. Pions would decay into photons, muons, neutrinos (or anti-neutrinos) and electrons (or positrons). For this reason, all these models predict a large production of photons and neutrinos, in particular at high energies: the most recent upper limits on photon and neutrino fluxes are already below the expectation for most of such models. These limits, as measured by the Pierre Auger Observatory, will be shown in detail later in section 3.5.5 and in particular in figure 3.16; here we anticipate the conclusion that at the moment top-down models for the production of UHECRs are strongly disfavored by experimental data.

In this chapter we tried to summarize the current knowledge about the cosmic ray phenomenon. The research in this field is progressing and the recent evidence for the acceleration of cosmic rays up to 10^{17} eV in Supernovae Remnants is surely an outstanding result. However, we still lack the information about what is capable of accelerating particles to energies of the order of tens of Joules or even more.

This thesis work is devoted to trying to add another little piece of information to this knowledge, a little brick that will hopefully help to build a bridge toward the understanding of the origin of cosmic rays.

Chapter 2

UHECR Extensive Air Showers: Physics and Detection

In America lo sai che i cocodrilli vengon fuori dalla doccia?

Samuele Bersani, *Cocodrilli* - from *** (1997)

We have already mentioned in the previous chapter that cosmic rays in different energy ranges require different types of detectors. In particular, the effective detection area is crucial: for cosmic rays of energy $E \lesssim 10^{14}$ eV the flux is high enough for direct detection. This can be performed by satellites or high-altitude balloons in order to avoid atmosphere absorption. At higher energies, given the much lower fluxes, detectors with larger apertures are required. Cosmic ray detection is then performed indirectly: when a high energy cosmic ray interacts with the atmosphere, it creates a shower of secondary particles (EAS: Extensive Air Shower). The EAS can be observed either by detecting the particles that reach the ground or by detecting the electromagnetic emissions of the shower particles in the atmosphere.

In this chapter we will sketch the current knowledge of the extensive air shower physics. In section 2.1 in general how EAS develop in the atmosphere using simple schematic models. We will also describe the basis of EAS simulations, describing the hadronic interaction models as well as their uncertainties. Section 2.2 in turn will be devoted to summarizing the methods that are used to detect the EAS, in particular at the highest energies, and how these methods allow us to extract information about the primary cosmic rays.

2.1.1 Electromagnetic showers

The electromagnetic component of the EAS is the most important in terms both of number of particles and energy. Heitler's model [54] describes electromagnetic cascades as a perfect binary tree: the pair production length for a high energy photon (>1 MeV) and the bremsstrahlung emission length for a high energy electron (>100 MeV) are similar.

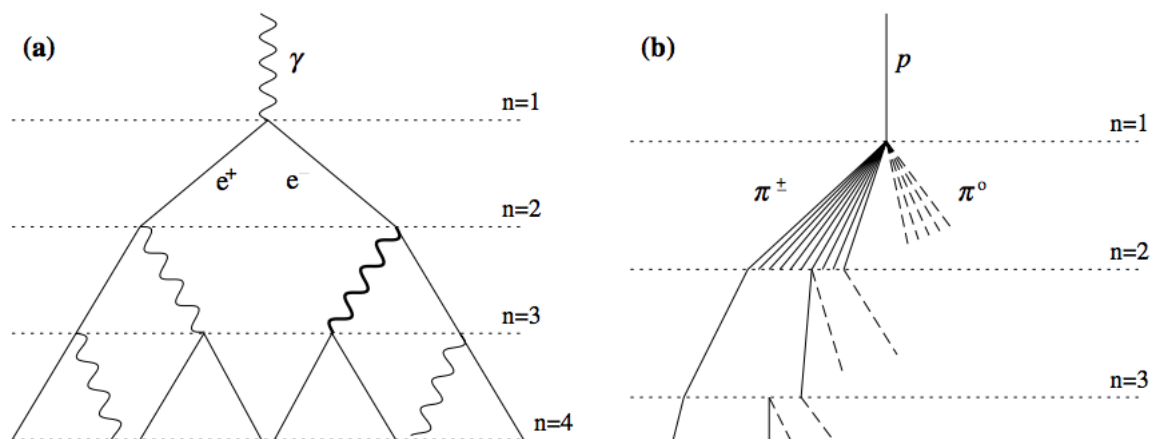


Figure 2.2: Schematic views of an electromagnetic cascade (left) and a hadronic shower (right). In the hadron shower, dashed lines indicate neutral pions which do not re-interact, but quickly decay, yielding electromagnetic subshowers (not shown). Not all pion lines are shown after the $n = 2$ level. Neither diagram is to scale. From [55]

In this simplified model, sketched in figure 2.2, left, we assume that at each step photons produce a e^+e^- pair (with both particles carrying the same energy) while an electron loses half its energy by emitting a γ via bremsstrahlung. If we ignore collision energy losses and neglect cross-section dependencies on energy, we can imagine that after traveling n interaction lengths ($\ln 2 \lambda_\gamma$) the shower is composed of 2^n particles carrying an energy of $E_0/2^n$ each, where E_0 is the primary energy. When the energy of each particle drops under a critical value where the rate of energy loss by ionization processes equals the radiative loss ($E_c \approx 86$ MeV in air) the number of particles stops growing. This point, called X_{\max} , is reached after

$$n = \frac{\ln(E_0/E_c)}{\ln 2} \quad (2.1)$$

generations, at a depth

$$X_{\max} = X_0 + \lambda_\gamma (\ln(E_0/E_c)) \quad (2.2)$$

where X_0 is the depth of the first interaction. The maximum number of particles will be

$$N_{\max} = E_0/E_c \quad (2.3)$$

The rate of change of X_{\max} with respect to the change in E_0 is called *elongation rate*, defined as:

$$D \equiv \frac{d X_{\max}}{d \log_{10}(E_0/\text{eV})} = \ln(10) \frac{d X_{\max}}{d \ln(E_0)} \simeq 2.3\lambda_\gamma \quad (2.4)$$

Given the interaction length value, we may estimate that elongation rate is about 85 g/cm^2 per decade in air for em showers.

The lateral distribution of particles, i.e. the particle density as a function of the distance to the shower axis, is mainly determined by Coulomb scattering of the dominant electromagnetic component and can be approximated by the Nishimura-Kamata-Greisen (NKG) function [56, 57]:

$$\rho(r) = k \left(\frac{r}{r_M} \right)^{-\alpha} \left(1 + \frac{r}{r_M} \right)^{-\eta-\alpha} \quad (2.5)$$

where r_M is the Molière radius ($\sim 80 \text{ m}$ at sea level for low energy particles); η and α are parameters determined experimentally and k is proportional to the shower size.

2.1.2 Hadronic showers

The hadronic component of the shower can be described at a basic level with the modified Heitler model by Matthews [55], as sketched in figure 2.2, right: the key parameter for this model is the hadronic interaction length λ_I . The atmosphere is approximated as layers of fixed thickness $\lambda_I \cdot \ln 2$. In each layer it is assumed that hadronic interactions produce $2N_\pi$ charged pions and N_π neutral ones. The interaction length and pion multiplicity ($3N_\pi$) are assumed to be energy independent. While π^0 decay immediately and feed the electromagnetic component described before, π^+ and π^- interact further until they reach an energy E_c^π ($\approx 20 \text{ GeV}$ in air) where decaying into muons is more likely than undergoing a new interaction. One third of the available energy goes into electromagnetic component while the remaining continues as hadrons.

Pions reach the critical energy after

$$n_c = \frac{\ln(E_0/E_c^\pi)}{\ln(3N_\pi)} \quad (2.6)$$

steps. If we assume that *all* pions decay into muons when they reach E_c^π , we obtain the number of muons in the shower $N_\mu = (2N_\pi)^{n_c}$. Introducing a parameter $\beta = \ln 2N_\pi / \ln 3N_\pi$ (≈ 0.85 for pion multiplicity $N_\pi = 5$) we obtain

$$N_\mu = \left(\frac{E_0}{E_c^\pi} \right)^\beta \quad (2.7)$$

The muon multiplicity is then not linearly dependent on the primary energy but it increases with it at a slower rate.

The determination of the position of the shower maximum is more complex in this case: the dependence of the cross section and multiplicity on energy will modify the rate of change of X_{\max} with the primary energy (the *elongation rate*). A simple approximation

can be derived by assuming that only the first generation of neutral pions contribute to the electromagnetic component:

$$X_{max}^p = X_0 + \lambda_\gamma \ln \left(\frac{E_0}{(3 n_\pi) E_c} \right) \quad (2.8)$$

where n_π is now the number of charged pions produced in the first interaction occurring at depth $X_0 = \lambda_I \ln(2)$. This is only a rough estimate, since we do not consider pions generated in interactions after the first one, and the inelasticity factor, i.e., the fraction of energy carried away in the collision by the leading particle, is also not considered. However with this approximation we can fairly derive the elongation rate for a proton as:

$$D_p = D_{em} + \frac{d}{d \log_{10}(E_0)} (X_0 - \lambda_\gamma \ln(3n_\pi)) \quad (2.9)$$

which can be evaluated as $\approx 58 \text{ g/cm}^2$ per decade.

The hadronic model can be extended to a nuclear primary using the superposition model: the interaction of a nucleus with atomic number A is viewed as the superposition of the interactions of A nucleons of energy E_0/A . Due to this energy fragmentation, showers produced by heavy nuclei will develop higher and faster. Pions will reach their critical energy sooner and this will lead to a higher number of muons with respect to a proton-induced shower. Using this simple assumption, we may then infer the difference on the X_{max} position with respect to a proton:

$$X_{max}^A = X_{max}^p - \lambda_\gamma \ln A \quad (2.10)$$

The number of muons will result to be

$$N_\mu^A = N_\mu^p A^{1-\beta} \quad (2.11)$$

The elongation rate described in equation 2.9 is independent from the primary energy; since here we modeled a high energy nuclei primary as the superposition of nucleons with lower energy, this formula still holds, with an interaction point X_0 that will be higher in the atmosphere. This will show up as parallel lines in an X_{max} vs energy plot, as sketched in figure 2.3. Moreover, for heavy nuclei the fluctuation on the position of X_{max} is smaller than for light ones.

2.1.3 Shower simulations and hadronic interaction models

The simplified models presented above allow us to have an intuition of the development of an EAS. What is used in cosmic ray physics to compare experimental results with theoretical expectations are in fact Monte Carlo simulations of extensive air showers. Those rely on phenomenological models based on experiments at accelerators and on Quantum Chromo Dynamics (QCD) theory.

EAS simulations follow the evolution of the shower in the atmosphere, using interaction models to compute the probability of interaction of each particle and its products, which will be in turn followed until the shower reaches the ground, where a simulation of the detector response is performed. The main problem for building simulations of showers at the highest energies is the lack of direct measurements for high energy interactions. We

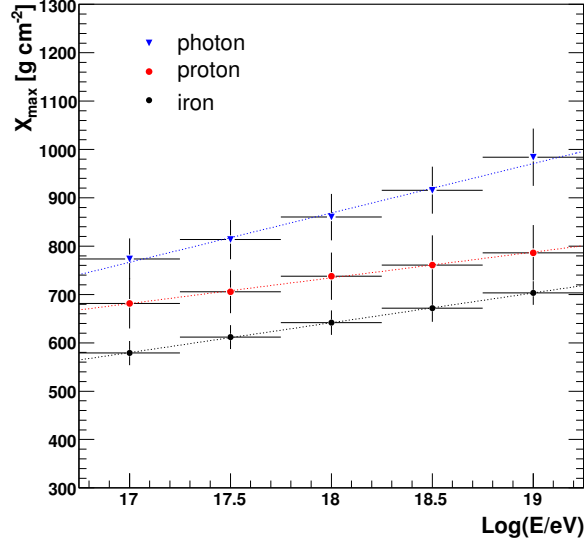


Figure 2.3: Behavior of the average depth of shower maximum as a function of energy for simulated primary photons, protons and iron nuclei (blue, red and black lines respectively).

have to keep in mind that the interaction of a cosmic ray at 10^{20} eV with a nucleus in the atmosphere happens at a center of mass energy of $\sim 3 \cdot 10^{15}$ eV, nearly two orders of magnitude higher than the ones available at run 1 in LHC.

The most common code for EAS simulation is CORSIKA [58] which implements several high and low energy hadronic interaction models. Different models approach the microscopic partons cascades differently. The main high-energy interaction models are SIBYLL[59], QGSJET [60, 61] and EPOS [62]; the commonly used low-energy interaction model is FLUKA [63]. In this work we will use QGSJET and EPOS models because these are the ones that were updated with the measurements from experiments at LHC (in particular TOTEM [64] and LHCf [65]). All high-energy interaction models now in use are based on the Gribov Regge theory of Pomeron exchange in multiple scattering. The Pomeron corresponds to microscopic partons (quarks and gluons) cascades.

The parameters of hadron production that are most important for describing air shower development are the cross section (or mean free path), the multiplicity of secondary particles of high energy, and the production ratio of neutral to charged particles. Until the start of LHC, these parameters were not well constrained by particle production measurements at accelerators. As a consequence, depending on the assumptions on how to extrapolate existing accelerator data, the predictions of hadronic interaction models differed considerably. As said, there are different versions of hadronic interaction models commonly used to simulate air showers. Here we will focus on the two high energy models which were updated to take into account LHC data at 7 TeV: QGSJETII-03 [60, 61] was updated into QGSJETII-04 [66] and EPOS 1.99 [62] replaced by EPOS LHC [67]. The difference between the two models is that QGSJET is based on the Quark-Gluon String model [68] and can treat loop diagrams of Pomeron exchange to infinite order in the Gribov-Regge

picture. The EPOS model on the other hand makes use of energy sharing between the involved partons but treats the higher order loop diagrams with an effective theory. Data from LHC were of key importance in constraining the basic parameters of hadron interaction at the highest energies. An example is shown in figure 2.4 where the proton-proton cross-section is shown as measured in data coming from accelerators and in the extrapolation used in pre-LHC and post-LHC models. We can see that pre-LHC models, in particular QGSJETII-03, used an extrapolation for p-p cross section at 7 TeV center of mass that data from the TOTEM [64] experiment at LHC found to be inexact, although not dramatically. On the other hand, results from LHCf experiment measurements suggest that the shape of the neutron production cross-section was better described by the QGSJETII-03 than in the EPOS 1.99 model [65].

The updated version of both QGSJET and EPOS are now tuned to match the measurements made by LHC experiments. However, the discrepancies between the predictions of pre-LHC models and TOTEM data shown in figure 2.4 allow us to understand how difficult a task it is to extrapolate hadronic interaction models up to the highest energies. Since accelerators capable of reaching such energies are not within the possibilities of current technology, the best constraints for the hadronic models at such energies come from the studies of cosmic ray showers themselves. In section 3.5.3 we will present results obtained by the Pierre Auger Observatory on this aspect.

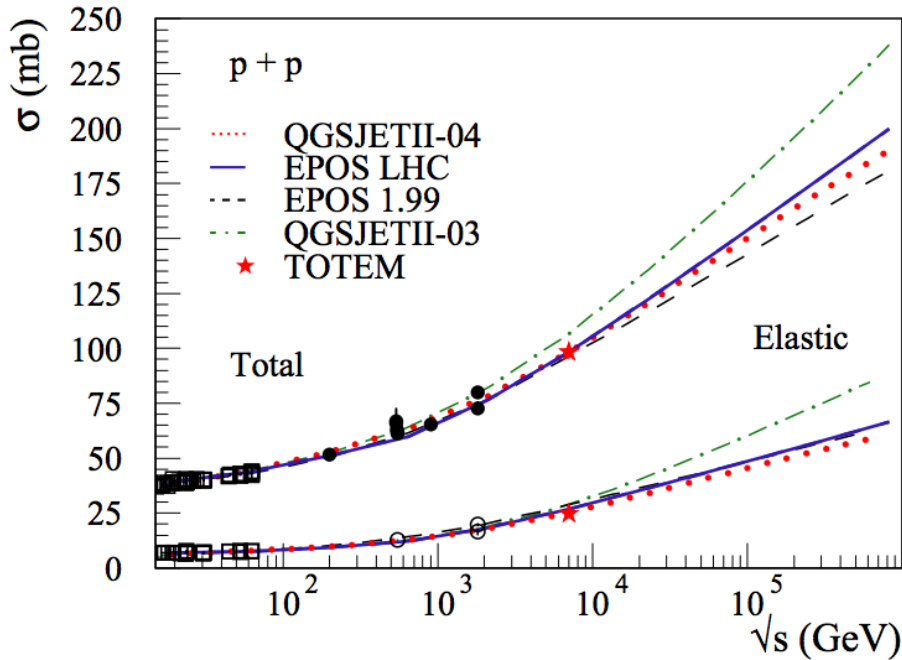


Figure 2.4: Total and elastic p-p cross section calculated with EPOS LHC (full line), QGSJETII-04 (dotted line), EPOS 1.99 (dashed line) and QGSJETII-03 (dashed-dotted line). The red stars are the measurement from the TOTEM experiment at LHC, black dots are earlier measurements done with different accelerators. From [69].

2.2 Principles of Extensive Air Shower Detection

In the following subsections the main methods developed for EAS detection and the most important experiments devoted to detect cosmic rays, with particular focus regarding the highest energies, are presented.

2.2.1 Arrays of particle detectors

As the EAS reaches the ground it can be detected by an array of detectors which sample the shower particles at ground. This idea was born together with the discovery of EAS. Already in the early 30s Bruno Rossi found coincidences in particle detectors surrounded by lead and imagined they could be due to a shower of particles [70]. The discovery of EAS is however usually credited to Pierre Auger who in 1939 found coincidences on a hundred meter scale [71]. That was the birth of the so-called air shower arrays. The first ultra-high energy air shower array was developed by John Linsley at Volcano Ranch (USA), using scintillators deployed on a surface of 8 km^2 . Using that array, the first event in the 10^{20} eV range was detected [72]. In the 90's the Akeno Giant Air-Shower Array (AGASA [73]) was the biggest air shower detector: it consisted of 111 scintillator counters on an area of 100 km^2 . It was also equipped with 27 muon-detector stations for a multi-component detection of the showers. Smaller arrays, exploring a lower energy region, are, among others, the EAS-TOP detector [74], active from 1989 to 2000 and the more recent KASCADE [75] and KASCADE-GRANDE [18] experiments.

The three main features of an EAS array are the surface it covers, the spacing between the detectors and its altitude:

- The surface of the array is chosen depending on the flux of cosmic rays at the energy of interest: tens to hundreds of thousands of m^2 are enough to detect showers from primaries of energy in the knee region (10^{15} eV). To detect cosmic rays at energies above 10^{19} eV , where the flux is less than 1 particle per km^2 per century, surfaces of thousands of km^2 are necessary. The surface of the array then determines the maximum energy accessible by the detector on a reasonable timescale.
- The spacing of the array is chosen depending on the size of the showers generated by the cosmic rays of interest: a denser array will be able to detect smaller showers, produced by lower energy primaries. A spacing of the order of 1 km allows the detection of cosmic rays of about 10^{18} eV . The density of the array then determines the threshold energy accessible by the detector.
- The altitude of the array is also chosen based on the energy region of interest: as we have shown in the previous sections, the depth of the shower maximum in the atmosphere is dependent on the primary energy. Detecting a shower near its maximum is preferable, since the larger number of particles that the detector intercepts allows for a better reconstruction. For the highest energy cosmic rays, the optimal altitudes are between 1000 and 1500 m. The ideal altitude is higher for lower-energy cosmic ray detection.

The choice of detectors depends on the shower component we are interested in: scintillators sample mainly the electromagnetic component; detectors under absorbers (usu-

ally scintillators or Resistive Plate Chambers (RPCs), under lead, iron, concrete or else buried) can be used to select the muonic component, more penetrating. Water Čerenkov detectors are, on the other hand, sensitive to both muonic and e.m. components. From measurements of EAS particles, the following information on the primary cosmic ray can be obtained:

- The arrival direction is measured based on the timing information. Through a fit of the arrival times of the shower in the array detectors, it is possible to reconstruct the shower axis and hence the cosmic ray direction. The timing resolution of the detectors is then crucial for obtaining high precision on the arrival direction.
- The energy: to obtain an energy estimator the signal at a fixed optimal distance r_{opt} from the shower core is used. This distance depends on the array spacing. The signal $S(r_{opt})$ is obtained by fitting the so-called Lateral Density Function (LDF), which describes the signal in the single detectors as a function of the distance from the shower core. The attenuation of the signal due to the zenith angle also needs to be taken in account. The correlation between the estimator and the absolute energy can be done via other detectors (e.g, fluorescence telescope, as will be described in detail in section 4.5) or via comparison with Monte Carlo simulations.
- The mass: ground arrays do not have a direct access to the position of the shower maximum as they sample the shower at fixed depth. However, other observables related to X_{max} and to the number of muons in the shower, i.e., mass-sensitive, are accessible. This will be discussed at length in chapter 7.

2.2.2 Light detectors

Charged particles in the EAS emit light as they travel through the atmosphere: fluorescence light and Čerenkov light. Both of them can be used to detect showers. However, the light emitted is faint and can be observed only during dark nights in good atmospheric conditions. The duty cycle is around 10-15%.

Čerenkov emission is concentrated around the shower axis. Then, to obtain information about the EAS, the detector needs to be near the shower core. Multiple Čerenkov detectors can be used, forming an array as in the TUNKA experiment [76]. This technique is used also in high energy γ astronomy. Telescopes such as MAGIC [77], VERITAS [78], H.E.S.S. [79] and the forthcoming CTA [80] use this technique. However, in order to observe rare high energy events ($E > 10^{17}$ eV), Čerenkov light is inappropriate because it would require an unachievable number of detectors in order to have the required effective area.

In turn, the fluorescence emission is isotropic. The fluorescence yield in the atmosphere is about 4-5 photons per electron per meter at ground level pressure (this value represents the largest uncertainty of this technique). This means that only the highest energy events can induce enough fluorescence light for it to be detected. If atmospheric conditions are good, through a telescope with an aperture of ≈ 1 m² with sensitive detectors (e.g. photomultipliers), a shower from an UHECR can be detected from as far away as 20 km from the shower axis. This means that the effective area is large enough for observing events of the highest energy.

The fluorescence detection technique was first developed by Greisen and his team in the 60's [81], but the small telescopes and bad atmospheric conditions did not lead to the detection of EAS. The first successful detector, called Fly's Eye [82], was installed in the late 70's in Utah and took data from 1981 to 1993. It was then upgraded into the High resolution Fly's Eye or, more colloquially, HiRes, that was operational from 1993 to 2006.

Both through Čerenkov and fluorescence detection one can measure the shower longitudinal profile. From this, a calorimetric estimate of the primary energy and a direct measurement of the position of the shower maximum are possible. A number of corrections should be applied in order to obtain precise measurements, taking into account, for example, absorption and scattering of the light in the atmosphere. For this reason a good control of the atmospheric conditions is crucial for such detectors, i.e, those two techniques require the inclusion in the experiment of atmosphere monitoring instruments.

2.2.3 Other techniques

The two methods described above are the two main ones used for detecting EAS. However alternative techniques have been proposed and are currently under test.

Detecting radio emission from the shower particles in the MHz-GHz frequency range is a promising technique that is being currently studied, since it offers the possibility of having the capabilities of a fluorescence detector with about 100% duty cycle. Showers emit radiation in the radio range via various mechanisms, in particular the geo-synchrotron emission of the electrons and positrons of the electromagnetic cascade . The radio signal associated to cosmic rays was detected for the first time in 1965 [83], but it is only in the last few years that this technique has been studied and developed on a large scale. More recent developments are based on initial studies performed by the LOPES [84] and the CODALEMA [85] collaborations and the LOFAR radio telescope [86]. Since 2009, an array of radio detectors (AERA) has been installed at the Pierre Auger Observatory site.

Recent results of a test beam experiment at SLAC [87] showed that it could be possible to use microwave radiation to detect extensive air showers. This radiation, expected to be isotropic and broad in frequency, is interpreted as molecular bremsstrahlung (MBR) produced by the scattering of low energy electrons in the weakly ionized plasma produced by the shower with the neutral molecules of the atmosphere. Progres in the understanding of such a phenomenon has been made both with cosmic rays experiments (AMBER, MIDAS, EASIER experiments at the Auger site [88], CROME in the KASCADE one) and in beam experiments that produce artificial showers and record their emissions (Gorham[87], MAYBE[89], AMY[90]).

An additional method is the radar technique, in which a ground-based radio transmitter irradiates the ionization trail left behind the shower front and another ground-based antenna receives the scattered radio signal, as is done, for instance, for measuring meteor showers. A radar detector is being built at the Telescope Array site (TARA experiment [91]). However, this technique seems to be suitable only for very vertical showers above 10^{20} eV [92], thus limiting the interest in developing it on large scale. Observing the same

shower with two or more of the methods mentioned above allows cross-calibrations and helps to reduce systematics on the measurements. The Pierre Auger Observatory [93], that is going to be described in detail in chapter 3, uses both fluorescence telescopes and an

array of water Čerenkov detectors to detect EAS. Another hybrid experiment, Telescope Array (TA) [94] is in the USA. It uses scintillators as surface detectors overlooked by 3 fluorescence telescopes stations.

Chapter 3

The Pierre Auger Observatory

*Sparato tra gli astri in pallone gonfio di musica
solo al richiamo più lontano
voglio la notte e la voglio senza luna*

Vinicio Capossela, *Bardamù* - from *Canzoni a Manovella* (2000)

The Pierre Auger Observatory [95] is the first experiment capable of observing ultra-high energy cosmic rays with both an array of particle detectors at the ground and fluorescence telescopes.

The Observatory is located in Argentina, in the province of Mendoza, in the so-called pampa amarilla, a large plateau at 1400 m above sea level. A map of the observatory is shown in figure 3.1. The observatory has been in operation since 1 January 2004, increasing its size from 154 surface detectors and one fluorescence site up to completion by June 2008.

In this chapter, we will describe at first the Observatory, describing its site (section 3.1) and then entering into the details of its two main parts: the Surface Detector (SD, section 3.2) and the Fluorescence Detector (FD, section 3.3). Of the first one, we will describe only the technical features, leaving for the forthcoming chapter 4 the description of the trigger, selection and reconstruction of events. We will briefly describe the different upgrades of the Observatory in section 3.4. Finally, in section 3.5, we will highlight the most important results obtained after 10 years of data taking¹. We will not go into details of the results of the search for small- and intermediate-scale anisotropies, that will be described in detail in chapter 6.

¹Not all the results presented here are updated to the ones presented at ICRC 2015. This is because the writing of this thesis took place before the conference and not all the results were ready at the moment of writing this chapter.

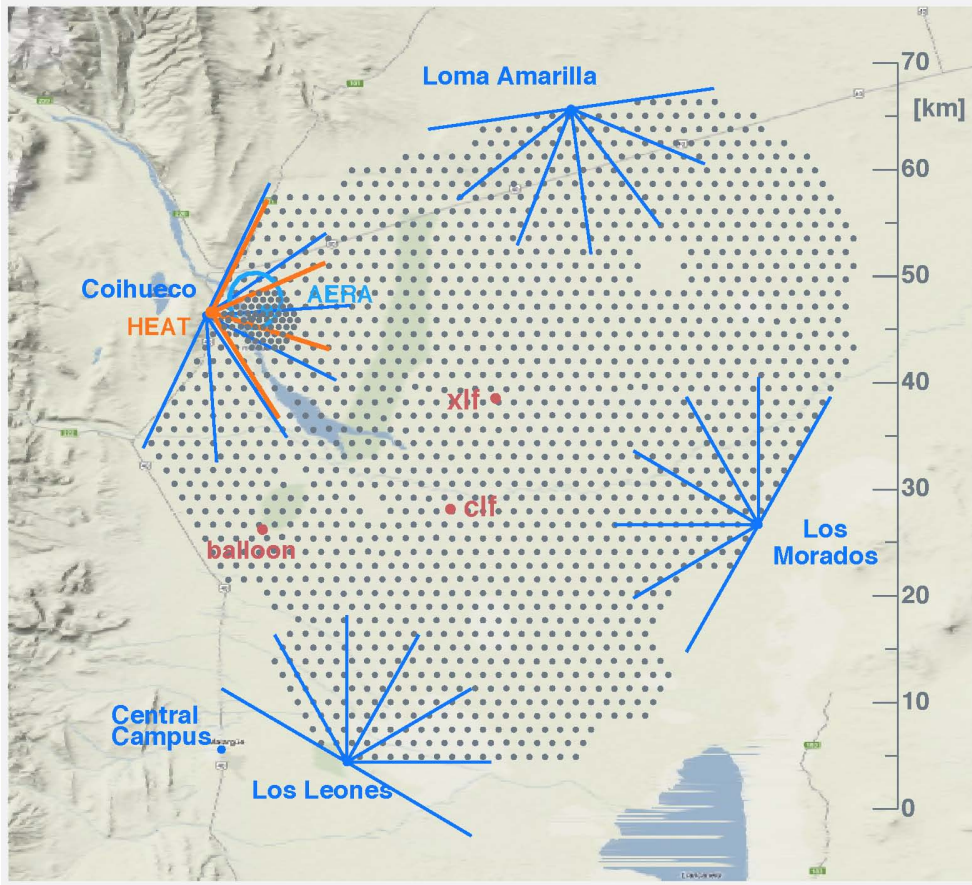


Figure 3.1: The Pierre Auger Observatory: dots represent stations of the surface detector (section 3.2) while blue lines represent the fields of view of the fluorescence detector telescopes (section 3.3). Orange lines represent the field of view of the HEAT telescopes (section 3.4.1). In front of them, the AMIGA detector (section 3.4.1) is visible as a denser array. The two red dots at the center of the array are the two laser facilities used to monitor the atmosphere (section 3.3.3). The blue circle near Coihueco is the location of the AERA array of radio detectors (section 3.4.2)

3.1 The site

The choice of site for building such a detector is crucial, in particular if we think that the Earth atmosphere can be considered as a part of the detector itself. The optimal altitude was chosen to be low enough ($< \sim 1500$ m above sea level) to allow the observation of the average shower maximum above the ground at the energies of interest. In turn, it is high enough ($> \sim 500$ m a.s.l.) for the shower maximum to be observed at ground (see section 3.3). In addition, because of the low fluxes at the energies of interest, a huge accessible surface was needed. The site also had to be relatively flat in order to avoid data-communication problems and ease the deployment of the detectors. Finally, good weather conditions, nearly at the level of the ones requested for optical telescopes, were needed for the operation of the fluorescence detectors.

The site latitude was evaluated considering as a final objective two sites for the observatory, one in the northern hemisphere and one in the southern. In this configuration, having two observatories at latitudes close to $\pm 35^\circ$ provided the most complete sky coverage. After evaluating several sites in Argentina, Australia and South Africa, the site in Argentina, near the city of Malargüe (Mendoza province, $35.1^\circ - 35.5^\circ$ S, $69,0^\circ - 69.6^\circ$ W at an altitude of 1300 - 1400 m asl) was selected for the southern observatory [93]. The northern site, which was planned to be built in Colorado, USA, was never realized due to lack of funding. In the city of Malargüe there is the office building that is the center of operation, where the local permanent staff works. The building is equipped with an assembly hall and a laboratory where broken parts of the detectors can be repaired. It also houses for the Central Data Acquisition System (CDAS).

3.2 The Surface Detector

The Surface Detector (SD) [96] of the Pierre Auger Observatory is composed of 1600 water Čerenkov detectors on a triangular grid. The spacing between the detectors is 1500 m resulting in a total surface of $\approx 3000 \text{ km}^2$. Each detector is represented as a dot in figure 3.1. The SD samples the particle components of extensive air showers (mainly muons, electrons, and photons) with a duty cycle of nearly 100%. It is fully efficient in the detection of showers generated by cosmic rays with energies above $\approx 3 \cdot 10^{18} \text{ eV}$ [97]. After the completion of the 1500 m-SD in 2008, part of it was filled with more detectors in order to obtain a denser array (so-called infill array) which is able to detect events with full efficiency down to $3 \cdot 10^{17} \text{ eV}$ [98, 99]. This infill detector, part of the AMIGA enhancement, will be described in section 3.4.

3.2.1 Čerenkov stations

Water containment

Each detector is composed of a roughly cylindrical polyethylene tank filled with 12000 liters of ultra-pure water, which ensure an high level of transparency, overlooked by three PMTs. A schematic view of each station is visible in figure 3.2. The water volume height is 1.2 m, chosen as a compromise between offering a large effective surface for horizontal showers (which is useful, among other things, in particular for UHE neutrino search) and logistic necessities. Inside the polyethylene structure, the water is contained in a circular cylinder, called liner, made of a flexible plastic material conforming approximately to the inside surface of the tank. The liner fulfills three functions: it contains the water, preventing contamination and loss of water and providing a barrier against any light that enters the closed tank; it diffusely reflects the light that is produced in the water; and it provides optical access to the water for the PMTs, such that PMTs can be replaced without exposing the water to the environment. The liner is composed of different layers, the innermost one being made of Tyvek [®]. This peculiar material was chosen due to its strength and excellent diffuse reflectivity for Čerenkov light in the near ultraviolet. Each part of the station has been designed for a lifetime of 20 years. The long-time performance stability is also one of the reasons for the water to be purified: it prevents the growth of micro-organisms that could affect the water transparency characteristics.

PMTs and electronics

The light produced by particles is detected by three PMTs looking downward in the water through windows of clear polyethylene. They are positioned at 1.2 m from the tank center axis in an equilateral triangle pattern. Signals are extracted from the last dynode and from the anode. The dynode signal (“high gain”) is amplified 32 times to extend the dynamic range, while the anode signal (“low gain”) is used for large signals when the dynode channel is saturated (see section 4.2 for further details). The six signals from each station (3x dynode and 3x anode) are read using front-end electronics having six 10 bit 40 MHz Fast Analog to Digital Converters (FADCs). The digitized signals are sent to a programmable logic device board, which is used to implement the various triggering decisions (see section 4.1 for a detailed description). In addition to the signals, a wide range of monitoring information, used for diagnostics, are collected from the stations. The time at which a local station triggers is crucial for the reconstruction of the events: it is measured by using a commercial Motorola GPS board (OnCore UT) with a precision of ~ 8 ns. This precision is achieved also thanks to the accurate measurement of the tank positions, performed during the deployment. Each station is powered by two 12V batteries that are recharged by two solar panels. For this reason, the electronics was designed in order to have a low power consumption (less than 10 W).

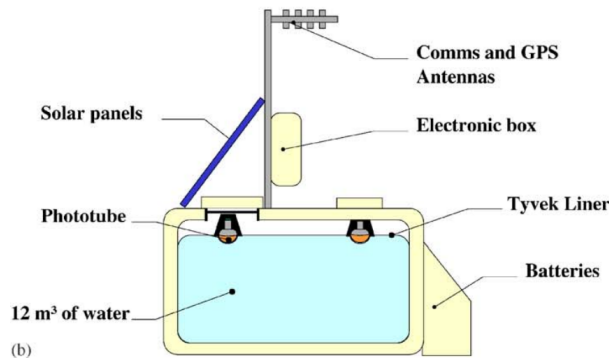


Figure 3.2: Schematic view of an SD station.

3.2.2 Data communication and acquisition

The data communication system of the Observatory consists of two radio networks. The first one, operating in the microwave band, is a high capacity network and provides communication from the FD sites to the central campus. It also retransmits data that are sent to and from the FD sites by individual SD stations via the second network, a wireless Local Area Network (LAN). All data are transmitted to the Central Data Acquisition System (CDAS), which is located in the central campus in Malargüe. The CDAS assembles the triggers from the surface array detectors (see section 4.1 for details), checks for FD-SD coincidences (so-called “hybrid” events, see section 3.3), gathers diagnostic data for these detectors and organizes the storage of data. The CDAS is programmed in C/C++ and is designed to run continuously with minimum intervention, being installed on a network of

six servers.

3.2.3 Maintenance

The maintenance of the SD is performed regularly by the local staff. The whole detector is constantly followed through a monitoring system that collects all diagnostic data sent by each tank (used for example to identify malfunctioning PMTs, see section 4.3). When a detector part shows abnormal functioning, a team is sent in the *pampa* to substitute it. In the case of PMTs, broken PMTs are taken back to the office building after substitution and repaired, if possible, so that they can serve as a spare part in the future. Thanks to this continuous maintenance, less than 2% (1%) of the PMTs (stations) are inactive on average over the whole array.

3.2.4 Aperture and exposure

The SD array trigger is fully efficient above $\sim 3 \cdot 10^{18}$ eV [97], meaning that above this energy, every cosmic ray will trigger SD independently of its energy, mass and arrival direction. The full-efficiency energy has been determined in independent ways using SD data, Hybrid data (i.e., events that are recorded both by SD and the fluorescence detector) and Monte Carlo simulations. Above this energy, the effective area coincides with the geometrical one and the aperture of the SD can be computed as the geometrical area integrated over solid angle. Thanks to the simple geometry of the array, the geometrical aperture is computed as the aperture of a hexagon composed of an active station surrounded by 6 active stations (elemental cell, see figure 3.3 for a sketch) multiplied by the number of such cells that are active at a certain moment in the array. The detection area per cell is 1.95 km^2 , so the corresponding aperture for showers with $\theta < 60^\circ$ is $a_{cell} \simeq 4.59 \text{ km}^2 \text{ sr}$. The number of active cells as a function of time is monitored second by second. To calculate the exposure over a given period of time, the above-defined aperture is then simply integrated over the number of live seconds. The main uncertainties in this calculation come from the determination of the N_{cell} , estimated to be around 1.5%, and from the estimation of the observation time. This was studied carefully using an empirical technique, based on the distribution of the arrival times of events, and it was found to be around 1%. The total uncertainty on the exposure is thus less than 3%.

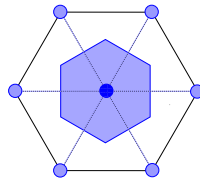


Figure 3.3: The effective area of each elementary cell (shaded blue area), corresponding to 1.95 km^2 .

3.3 Fluorescence Detector

The second component of the Pierre Auger Observatory, the fluorescence detector (FD), is designed to detect EAS in a complementary way with respect to the SD. The charged particles in the EAS excite atmospheric nitrogen molecules, causing them to emit fluorescence light in the 300-430 nm range. This faint light can be observed during clear nights by telescopes that can sample the longitudinal development profile of the EAS. As observation periods are limited to dark nights with good weather, the FD duty cycle is $\sim 15\%$. The integral of the longitudinal development profile ($\frac{dE}{dX}(X)$) gives the total energy dissipated electromagnetically. This energy accounts for $\sim 90\%$ of the total energy of the EAS, i.e. of the primary cosmic ray, the remaining $\sim 10\%$ being the so-called “invisible energy”, carried away by neutrinos and high energy muons that do not deposit all their energy in atmosphere.

3.3.1 The telescopes

The fluorescence detector (FD) [100] is composed of 24 telescopes distributed in four sites (Los Leones, Cohiueco, Loma Amarilla, and Los Morados) overlooking the surface array (see figure 3.1). Each telescope has a field of view of $30^\circ \times 30^\circ$ in elevation and azimuth: the combination of the 6 telescopes in each site allows for a 180° coverage in azimuth.

The FD telescopes design is based on Schmidt optics. Each of them includes a 3.6 m diameter mirror illuminating a camera composed of 440 PMTs. Each PMT (or pixel) has a field of view of $1.5^\circ \times 1.5^\circ$. A UV-transmitting filter is installed at the aperture to improve the signal-to-noise ratio in the fluorescence emission band. To reduce optical aberrations, corrector lenses are installed and arranged in a ring around the aperture diaphragm. A sketch of the FD optical system is shown in figure 3.4.

Triggering is done at hardware level at pixel (first) and camera (second) levels. A third-level trigger (TLT) is implemented in software mainly to reject lightning events and random alignments. TLT triggers are then processed at each FD site to merge all the telescope information. For energies above $\approx 10^{18}$ eV, the FD trigger is always accompanied by at least one triggered SD station, resulting in the so-called hybrid events. The hybrid system is thus fully efficient for the detection of cosmic rays at an energy significantly lower than the SD.

Once per year, an absolute calibration of the telescopes is performed with a uniform source positioned at the telescope aperture. This calibration ensures that the response of the PMTs is uniform but takes a long time to be performed. In addition to that, a relative calibration is done during each night of data-taking, both before and after each run. This is done by illuminating each camera in three different positions with optical fibers connected to light sources at different wavelengths. In this way, we can monitor the long-term time variations in the telescopes’ calibration, between successive absolute calibrations.

3.3.2 FD event reconstruction

An example of an FD event is shown in figure 3.5, where we can see the result of the merging of the outputs of the PMTs in two FD cameras that detected light from the

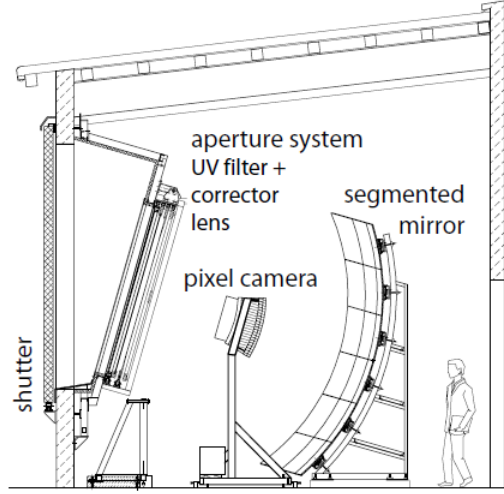


Figure 3.4: A schematic view of an FD telescope.

shower. The color code represents the timing of each pixel (early=red, late=violet). The signal and timing of each pixel is recorded and used in the event reconstruction.

The event reconstruction starts with the determination of the shower-detector plane (SDP). The SDP is the plane that includes the location of the eye and the line of the shower axis (see figure 3.6). Next, the timing information of the pixels is used for reconstructing the shower axis within the SDP. The geometry of the shower is reconstructed by fitting the formula

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

where R_p is the perpendicular distance from the camera to the axis, the angle χ_0 is the one between the track and the horizontal line in the SDP, and the angles χ_i are the ones between the same horizontal line and the pointing direction of each pixel (see figure 3.6). In some cases, in particular when the measured angular speed $d\chi/dt$ does not change much over the observed track length, there is a family of possible solutions for R_p and χ_0 and thus for the shower geometry. This degeneracy can be broken by using the timing information from the SD station which, as mentioned above, is always present in events above 1 EeV. This leads to a hybrid estimate of the shower geometry with a precision of typically 0.6° for the arrival direction of the primary cosmic ray [100]. The resolution of the geometry reconstruction can be improved if the same event is seen by 2 or more FD stations (so-called stereo, triple or quadruple event).

Once the geometry of the shower is known, the light collected as a function of time can be converted to energy deposit as a function of slant depth. For this purpose, the attenuation of the light from the shower track to the telescope aperture needs to be estimated and all contributing light sources need to be disentangled: fluorescence light, direct and scattered Čerenkov light from particles in the shower and multiple scattered-light. After that, the calorimetric energy of a shower is estimated by fitting the Gaisser-Hillas function [101] to the reconstructed energy deposit profile and by integrating it.

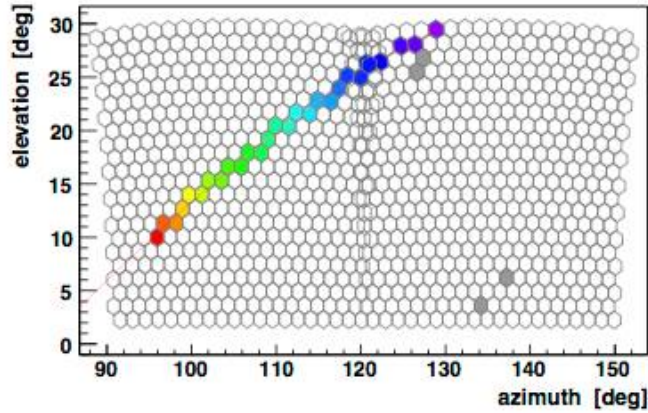


Figure 3.5: An example of FD trace. The color code indicates the timing of the trigger in each pixel (early=red, late=violet).

An example of this fit is shown in figure 3.7. A further correction for energy carried away by neutrinos and high energy muons (the so-called invisible energy). This correction has been estimated from hybrid data [102] and is of the order of 10% to 15%. The systematic uncertainty on the energy estimation is $\approx 14\%$ [103]. The main contributions are the uncertainties on the fluorescence yield estimate (3.6%), on the absolute calibration of the fluorescence telescopes ($\sim 9\%$), on the reconstruction method ($\sim 6.5\%$) and on the atmospheric monitoring (6.2%). From the same fit also the depth of the shower maximum (X_{\max}), the most important mass-sensitive variable (see paragraph 3.5.2), can be extracted.

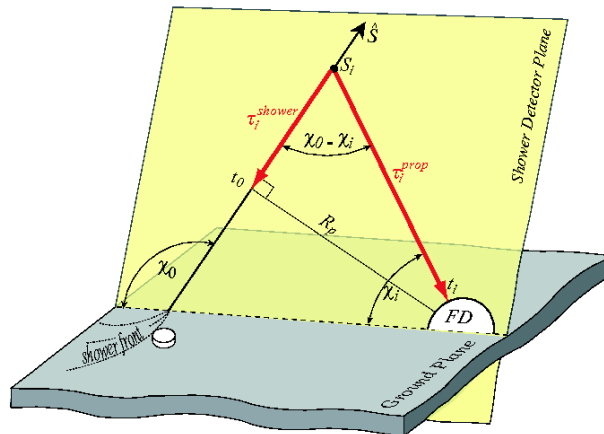


Figure 3.6: Schematic representation of the shower-detector plane.

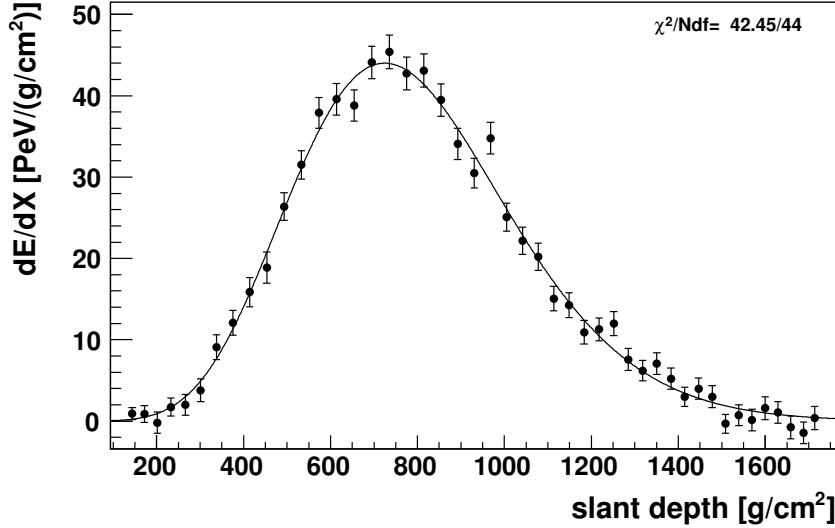


Figure 3.7: Energy deposit longitudinal profile (dots) and fit to Gaisser-Hillas function (line).

3.3.3 Atmospheric monitoring

An accurate reconstruction of the longitudinal profile with FD requires continuous monitoring of the atmospheric conditions. This is particularly important for estimating the attenuation of the light due to molecular and aerosol scattering as it travels from the shower to the telescope. An ensemble of instruments was designed and deployed to monitor the atmosphere at the Pierre Auger Observatory [104]. The aerosol monitoring uses two lasers placed near the center of the array (the Central Laser Facility (CLF) and the eXtreme Laser Facility (XLF), four elastic scattering lidar stations and two optical telescopes (Horizontal Attenuation Monitor, HAM, and Photometric Robotic Telescope for Atmospheric Monitoring, FRAM) that determine the wavelength dependence of the aerosol attenuation. Four infra-red cameras are also used for cloud detection.

The molecular properties of the atmosphere at the time of data taking are obtained as the 3-hourly data tables provided by the global data assimilation system (GDAS) [105], which have been extensively cross-checked with radio balloon soundings on site. The aerosol content of the atmosphere during data taking is continuously monitored. For this purpose, the vertical aerosol optical depth (VAOD) is measured on an hourly basis using laser shots provided by the CLF and the XLF and cross-checked by lidars operated at each FD site.

3.3.4 Hybrid exposure calculation

The exposure of the hybrid system is calculated using a time-dependent Monte Carlo simulation. The changing configurations of both fluorescence and surface detectors are taken into account for the determination of the on-time of the hybrid system. Within a time interval of 10 min, the status and efficiency of all detector components, down to the level of the single PMTs of the fluorescence detector, are determined. Moreover, all atmospheric

measurements as well as monitoring information are considered and used as input for the simulation. A detailed description can be found in [106]. The longitudinal profiles of the energy deposits are simulated with different hadronic interaction models. The influence of hadronic interaction models on the exposure calculation has been estimated to be lower than 2%. A 50% mixture of protons and iron nuclei has been assumed for the primaries. The quality cuts used for the event selection lead to only a small dependence of the exposure on the mass composition. The systematic uncertainty arising from the lack of knowledge of the mass composition is about 8% at 10^{18} eV and 1% at 10^{19} eV. The total systematic uncertainty of the derived exposure is estimated to be 10% at 10^{18} eV and 6% at 10^{19} eV.

3.4 Enhancements

3.4.1 Low-energy extensions

As we have shown in the previous sections, the Pierre Auger Observatory has been designed to detect cosmic rays in the EeV region and above. However, the energy range between 10^{17} and few 10^{18} eV is very important to understand the origin of ultra-high energy cosmic rays. Indeed, it is this region where the transition between galactic and extragalactic sources is expected to take place according to various models (see paragraph 3.5.1 for further details). In order to explore also this region two enhancements of the observatory were started in 2008, after the completion of the baseline SD and FD.

AMIGA

AMIGA (Auger Muons and Infill for the Ground Array) is the SD enhancement for low energy studies, and it consists of two different parts. The first part, which is now completed and taking data since 2010, is a denser sub-array of the SD built by adding 60 stations in a limited area of the detector (the one around the Coihueco FD building, see figure 3.1). This infill array has a spacing that is half the original one (750 m) and an area of 23.5 km^2 [107]. This spacing allows for a full-efficiency detection of events above $3 \cdot 10^{17}$ eV. The reconstruction of infill events is done in nearly the same way as for the 1500 m array that will be described in chapter 4, with small differences in the trigger implementation and the use of a slightly different energy estimator. In addition to the denser array, the AMIGA enhancement will include also muon detectors, consisting of 30 m^2 plastic scintillators buried at a depth of 2.3 m aside the SD station. The aim is the measurement of GeV muons to extract relevant information on the mass composition of cosmic rays. An engineering array of 7 modules made of 38 scintillators has been deployed by April 2015.

HEAT

HEAT (High-Elevation Auger Telescope) is the FD enhancement for low energy studies. It consists of three additional telescopes that are similar to the standard FD ones. They are installed at the Coihueco FD site, overlooking the AMIGA array (see figure 3.1). Their peculiarity is that while they can look at the atmosphere from ground (0°) to an elevation of 30° as standard FD telescopes, they can also tilt 29° upwards. In this tilted position,

they can observe the atmosphere at higher elevation, allowing the detection of lower energy showers, that develop higher in the atmosphere.

3.4.2 Radio and Microwave detection

In the last decades, a third way to detect cosmic rays other than surface arrays and fluorescence telescopes has started to be developed. An EAS traveling through the atmosphere induces also radio and microwave emissions. The potential of detecting in such way the EAS is high since this could allow the study of the shower longitudinal profile as a fluorescence detector does, but with a $\sim 100\%$ duty cycle as a surface array. Many different approaches to this technique are being developed in the Pierre Auger Observatory, with some promising results.

AERA

The Auger Engineering Radio Array (AERA) is composed of 153 radio detection stations, spread over an area of roughly 17 km^2 , in the same area as AMIGA (see figure 3.1). The detectors have various spacings, ranging from 150 m to 750 m. All stations feature two antennas, one aligned in the geomagnetic north-south direction and one in the east-west direction and are sensitive to the 30-80 MHz range. As of early March 2015, more than 5500 air showers were detected by three or more antenna stations in coincidence with the surface detector array (AERA+SD hybrid events). In addition, 324 air showers were simultaneously detected with the fluorescence detector (AERA+SD+FD hybrid events), 509 events in coincidence with the AMIGA muon counters (AERA+SD+AMIGA events) and 51 AERA+SD+FD+AMIGA hybrid events. The direction reconstruction is in good agreement with the surface detector reconstruction. A method for the energy reconstruction based on the measured energy density has been compared to the standard SD, showing a resolution of 17% [108].

Microwaves

Cosmic ray detection in the GHz band is being pursued at the Pierre Auger Observatory with two different approaches. MIDAS (MICrowave Detector of Air Showers) and AMBER (Air shower Microwave Bremsstrahlung Experimental Radiometer) are prototypes of an imaging parabolic dish detectors (FD-like approach). EASIER (Extensive Air Shower Identification with Electron Radiometer) is composed of radio receivers of wide angular coverage installed on the SD detectors. EASIER was able to perform the first detection ever of an air shower in microwaves in June 2011. More details can be found in [88].

3.4.3 Planned upgrades

Further upgrades have been planned for a new 10-year operation of the observatory, starting in 2016:

- A complementary measurement of the shower particles will be provided by deploying plastic scintillators (SSD) above the existing SD stations. This allows the sampling of the shower particles with two detectors having different responses to muons and electromagnetic particles.

- The SD stations will be upgraded with new electronics that will process both SD and SSD signals. Use of the new electronics also aims to increase the data quality (with faster sampling of ADC traces, better timing accuracy, increased dynamic range), to enhance the local trigger and processing capabilities and to improve calibration and monitoring capabilities of the surface detector stations.

With operation planned until 2024, event statistics will more than double compared with the existing Auger data set, with the critical added advantage that every event will have mass information. Obtaining additional composition-sensitive information will not only help to better reconstruct the properties of the primary particles at the highest energies, but also improve the measurements in the important energy range just above the ankle. Furthermore, measurements with the new detectors will help to reduce systematic uncertainties related to modeling hadronic showers.

3.5 Highlight of the Pierre Auger Observatory results

After more than 10 years of data taking, measurements of the Auger Observatory have advanced our understanding of ultra-high energy cosmic rays. Here we will summarize some of these measurements with the aim of explaining the scientific context where this thesis work has developed.

The Telescope Array

The Telescope Array experiment (TA) [94] is currently the largest UHECR observatory taking data in the northern hemisphere. It is located in Utah, USA and is a hybrid detector system consisting an array of 507 scintillation surface detectors and three fluorescence stations. In order to present a more complete scenario of the experimental status of ultra-high energy cosmic rays, we will briefly report in boxes like this one key results from this experiment. TA results will be presented aside the corresponding Auger results.

3.5.1 Energy Spectrum

The precise measurement of the flux of ultra-high energy cosmic rays has been one of the main motivation for building the Pierre Auger Observatory. Indeed, the measurements from AGASA and HiRes experiments led to controversial conclusions on the presence of a suppression at the highest energies. After ten years of operation, Auger data have allowed the measurement of the energy spectrum of cosmic rays above 0.1 EeV with unprecedented precision, combining different datasets from both FD and SD.

The principle of building the spectrum is rather simple: binning data in energy and dividing the number of events in each bin by the exposure, in order to obtain the flux. However, this has to be done carefully and uncertainties have to be deeply studied for both quantities. In particular, because of the steepness of the spectrum, uncertainties in energy can cause an important migration of events from lower to higher energy bins and this has to be taken into account. The uncertainty on the energy will be described in the next chapter, while that on the exposure was described in section 3.2.4.

The measurement described here is based on data collected by the FD and SD up to December 2012 [109]. It includes data from the 1500 m surface detector array (divided into two sets according to the shower zenith angle, 0-60° and 60-80°), from the 750 m infill array, and hybrid events, measured simultaneously with both the fluorescence detector and the surface detector. The spectrum resulting from the combination of these data is shown in figure 3.8, left. The different data sets cover different energy ranges: the 1500 m SD is fully efficient at $3 \cdot 10^{18}$ eV, while FD hybrid data lower this threshold to $1 \cdot 10^{18}$ eV. The full efficiency threshold for 60-80° events is slightly higher, $4 \cdot 10^{18}$ eV, while the 750 m array extends the measurement down to $3 \cdot 10^{17}$ eV. For all detectors and data sets, the upper energy limit is given only by the lack of events.

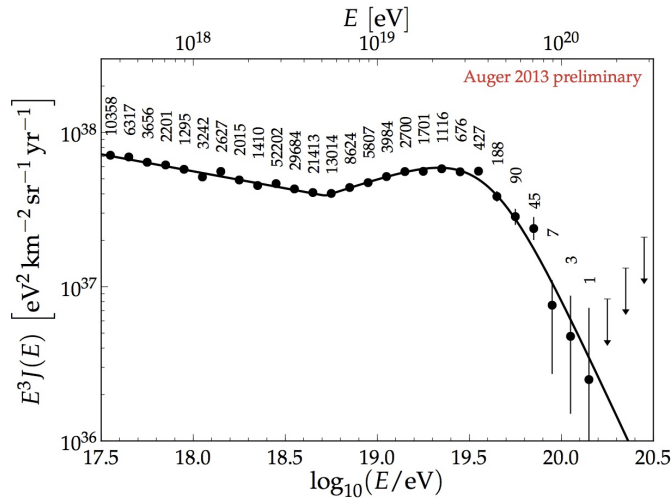


Figure 3.8: The spectrum of UHECRs as measured by Auger combining all data from SD vertical, SD horizontal, Infill and hybrid measurements, with its best fit. Fit results are in table 3.1. From [109].

The most important features that appear in the energy spectrum are the ankle (at about $5 \cdot 10^{18}$ eV) and the suppression at about $4 \cdot 10^{19}$ eV. The data are fitted with a formula including a power law before the ankle and a power law with smooth suppression above:

$$J(E) \propto \begin{cases} E^{-\gamma_1} & \text{if } E < E_{ankle} \\ E^{-\gamma_2} \cdot \left[1 + \exp\left(\frac{\log_{10} E - \log_{10} E_{1/2}}{\log_{10} W_c}\right)\right]^{-1} & \text{if } E > E_{ankle} \end{cases}$$

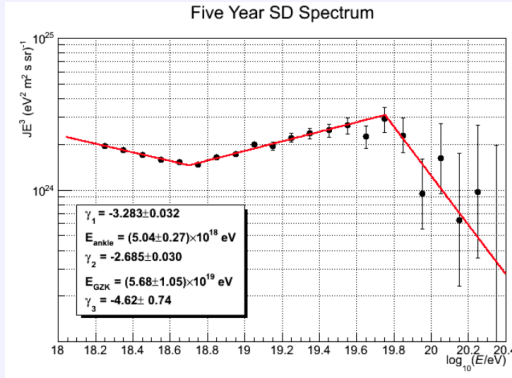
γ_1 and γ_2 are the spectral indices below and above the ankle. $E_{1/2}$ is the energy at which the flux has dropped to half of its value before the suppression, the steepness of which is described with $\log_{10} W_c$. Results of the fit are reported in table 3.1.

Parameter	Value ($\pm\sigma_{stat} \pm \sigma_{sys}$)
$\log_{10}(E_{ankle}/eV)$	$18.72 \pm 0.01 \pm 0.02$
γ_1	$3.23 \pm 0.01 \pm 0.07$
γ_2	$2.63 \pm 0.02 \pm 0.04$
$\log_{10}(E_{1/2}/eV)$	$19.63 \pm 0.01 \pm 0.01$
$\log_{10} W_c$	$0.15 \pm 0.01 \pm 0.02$

Table 3.1: Results of the fit of the UHECRs spectrum as measured by Auger, with its main features. See text for details. From [109].

Telescope Array and the energy spectrum

The Telescope Array collaboration published the spectrum of cosmic rays above $3 \cdot 10^{18}$ eV measured with data collected by their SD from May 2008 to May 2012 [111]. The spectrum was fitted with a broken power law with two breaking points: the ankle and the cutoff.



TA energy spectrum, fitted to a broken power law. Results of the fit are shown in the picture with γ_1 being the power law before the ankle position, γ_2 the power law between the ankle and the suppression (here reported at E_{GZK}) and γ_3 the power law after the suppression.

A comparison between the cosmic rays spectra measured by TA and Auger was made in [112]. The comparison of the energy spectra at the ankle region suggests that the energy scale of TA is about 13% higher than the one of the Pierre Auger Observatory. The spectra measured by the two observatories are however compatible within the uncertainties coming from both experiments.

3.5.2 Mass composition: depth of shower maximum from FD

Due to the low intensity of cosmic rays at the highest energies, the primary mass cannot be measured directly but is inferred from shower properties such as the longitudinal development or the particle densities at the ground (see paragraph 2.1), after making assumptions about the characteristics of high-energy particle interactions. The most accessible and reliable mass-related variable that can be observed at the Pierre Auger Observatory is the depth of shower maximum (X_{max}) through the FD telescopes. This is related to the logarithm of the primary mass, as it was described in paragraph 2.1. Due to fluctuations of the properties of the first few hadronic interactions in the cascade, the primary mass cannot be measured on an event-by-event basis but must be inferred statistically

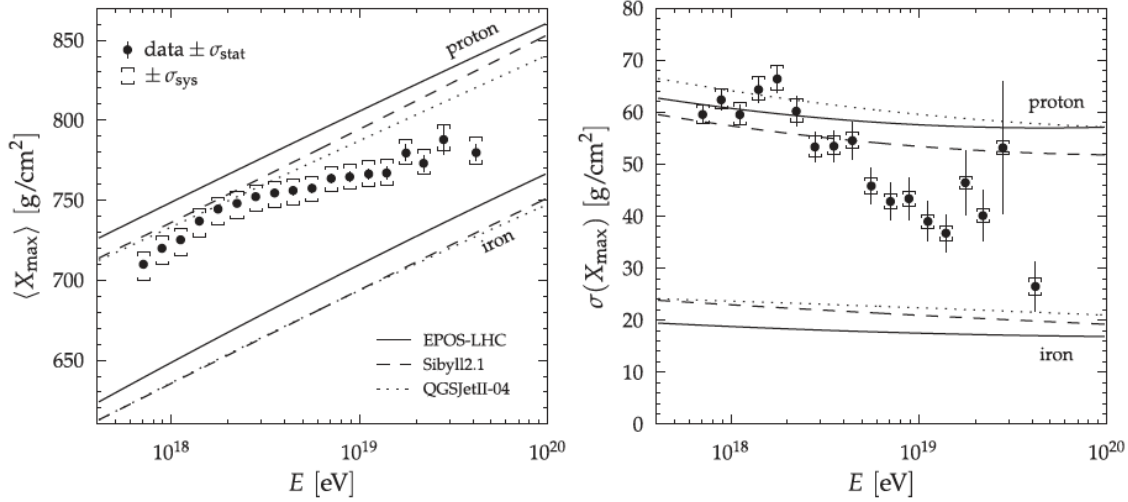


Figure 3.9: X_{\max} (left) and its RMS (right) as function of energy. Dots represent data measured by the Pierre Auger Observatory while lines show prediction from simulations based on different hadronic interaction models (see section 2.1.3 for further details). From [113].

from the distribution of shower maxima of an ensemble of air showers. The recent Auger publication of X_{\max} data [113] includes the study of the mean and standard deviation of the X_{\max} distributions as a function of energy. It also presents for the first time the X_{\max} distributions. Whereas with the first two moments of the distributions the overall features of primary cosmic-ray composition can be studied, only the distribution itself contains the full information on composition and hadronic interactions.

Hybrid data from 1st December 2004 to 31st December 2012 were used. Strict quality and fiducial cuts (see [113] for details) were applied, the latter ones to avoid an undersampling of the tails of the distributions due to the effective field of view of the telescopes. The resolution in the X_{\max} measurement is better than 26 g cm^{-2} (15 g cm^{-2}) at $\sim 6 \cdot 10^{17} \text{ eV}$ ($2 \cdot 10^{19} \text{ eV}$). The resulting measurements of the first two moments ($\langle X_{\max} \rangle$ and $\sigma(X_{\max})$) of the X_{\max} distributions (based on ~ 20000 events) are shown in figure 3.9, left and right panel respectively, as a function of energy. They are compared with predictions for proton and iron primaries for three different hadronic models: Epos-LHC, QGSJETII-04 and Sybill 2.1. The first two ones were tuned to match recent LHC data². Comparing the energy evolution for data and simulations it can be seen that the slope of the data is different than what would be expected for either a pure-proton or a pure-iron composition. We note an evolution of the average composition of cosmic rays towards lighter nuclei up to energies of $10^{18.27} \text{ eV}$. Above this energy, the trend reverses and the composition becomes heavier.

For a more quantitative study of the evolution of the composition, $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$ are converted to the first two moments of the logarithm of atomic mass number ($\ln A$) distribution via an analytical method [114]. The mean and variance of $\ln A$ are shown in figure 3.10 for the same three hadronic models as above. We can see that all three models agree about the composition being the lightest around $\sim 2 \cdot 10^{18} \text{ eV}$. The interpretation

²See section 2.1.3 for further details on hadronic interaction models.

with Epos-LHC leads to the heaviest average composition that is compatible with nitrogen at the highest energies while for the other two models it is closer to helium. Note that the interpretation with QGSJETII-04 leads to unphysical variances above $10^{18.4}$ eV and therefore this model is disfavored.

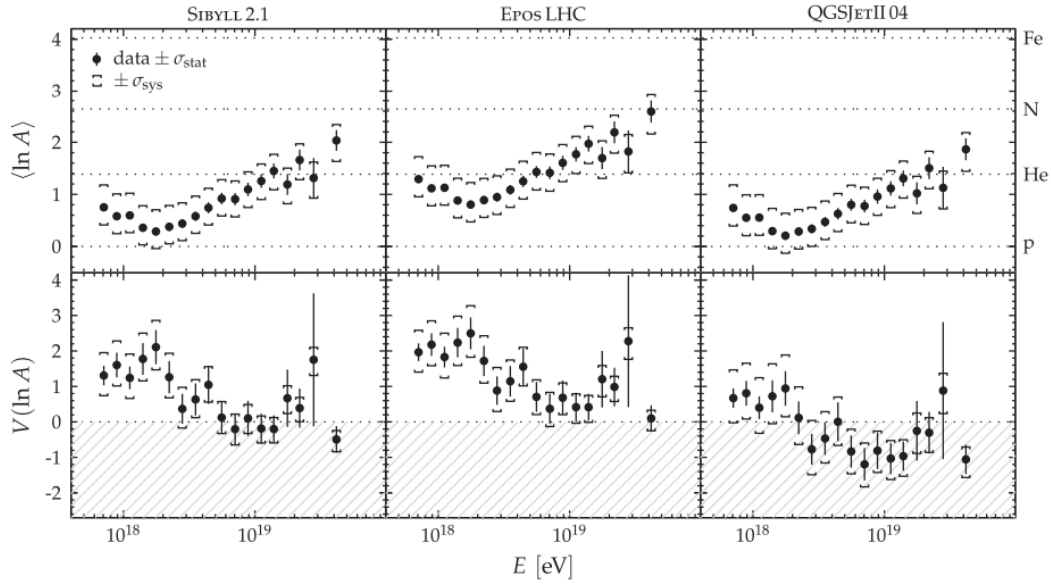


Figure 3.10: Predictions on the average composition as a function of energy based on X_{\max} measurements (top plots), and its variance (bottom plots), using different hadronic interaction models. See text for details. From [113].

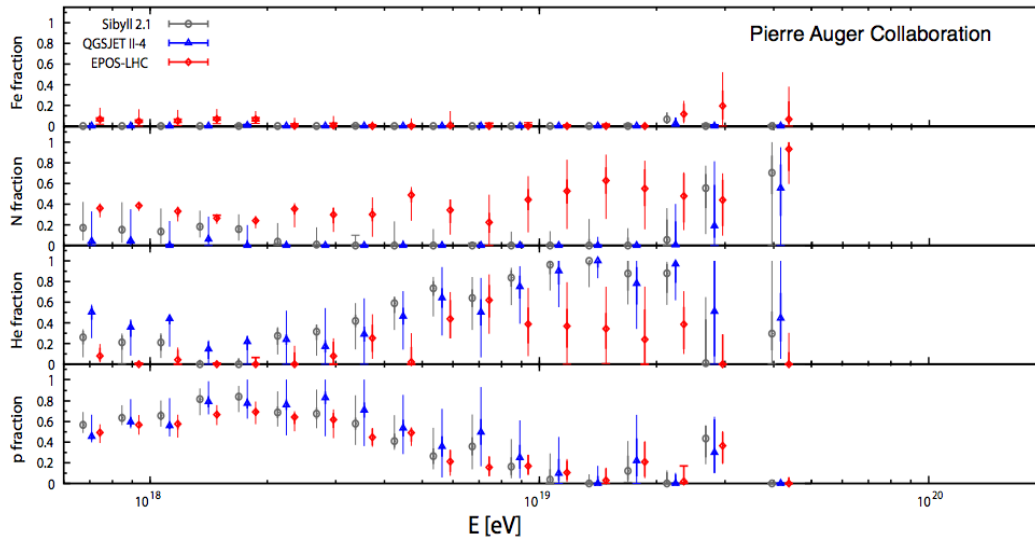
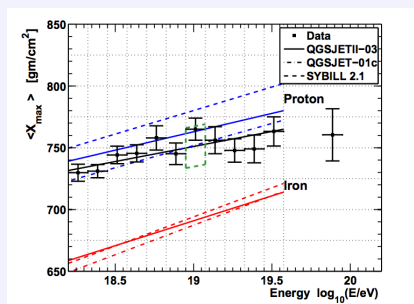


Figure 3.11: Estimate of the composition of ultra-high energy cosmic rays from X_{\max} distributions fit (see text for details). The estimated fraction for each primary considered in the fit is plotted as a function of energy. From [115].

A further step in mass composition interpretation can be taken by exploiting the full X_{\max} distributions. For a given hadronic interaction model, the X_{\max} distributions have been compared to predictions from simulations using several nuclear fractions: p, He, N and Fe. A binned maximum-likelihood discriminator is used to determine the best-fit fractions [115]. At first, a mixture of the two most stable types of particles, protons, and iron nuclei was considered, and then the fits were extended to include extra components. Specifically, helium and nitrogen nuclei are included as representatives of the intermediate range of nuclear masses. While a simple mixture of protons and iron primaries does not describe well data, with any hadronic interaction model, adding intermediate components greatly improves the fits for all of them. Figure 3.11 shows the best-fit fractions as a function of energy. The models appear to be in remarkable agreement regarding their predictions of the protons and iron nuclei contributions, despite large differences in the intermediate components predictions. Overall, a substantial change in the proton fractions is observed across the entire energy range, which rises to over 60% around the ankle region and subsequently drops to near zero just above 10 EeV with a possible resurgence at higher energies.

Telescope Array and X_{\max} measurements.

The Telescope Array collaboration measured as well X_{\max} using fluorescence telescopes. Their results indicate a lighter composition than the one measured by Auger when compared to pre-LHC simulations, in particular at the highest energies [116]. However, a recent joint study concluded, that, by applying appropriate corrections, the two measurements are in excellent agreement [117].



Data are the black points with error bars. The solid black line is a fit to the data. Colored lines are fits to MC. Blue is proton and red is iron. The green hashed box indicates the total systematic error.

3.5.3 Mass-sensitive observables from SD and hadronic models

Due to the duty cycle of 15% of the fluorescence telescopes, the data on the depth of shower maximum extend only up to the flux suppression region, i.e., $4 \cdot 10^{19}$ eV. Thus it is worth trying to access mass-sensitive observables through the 100% duty cycle SD to increase the number of events at the highest energies. Mass composition studies with SD data have been made within the Collaboration by using observables such as the azimuthal asymmetry of rise time in the station signals [12], the muon production depth of the showers [118] or the muon content at ground [119]. While the energy evolution of those observables is similar to that found by X_{\max} studies, the SD measurements, combined

with the measurement of the depth of shower maximum, provides important insights into the consistency of hadronic interaction models. We summarize here the results of muon measurements as they are sensitive to the details of the hadronic component of air showers.

Depth of maximum muon production, X_{max}^μ

Muons mainly originate from pions and kaons and are produced with a characteristic transverse momentum distribution inside a narrow cylinder around the shower axis. Since the radiation length of typical muons is much larger than the whole atmospheric depth, even in the case of inclined events, multiple scattering effects are negligible. Muon trajectories can thus be taken as straight lines from the point where they are produced. The time structure of the muon component reaching the ground can be exploited to obtain the distribution of muon production distances along the shower axis. The muon production depth (MPD), i.e. the total amount of traversed matter in g/cm^2 , is obtained by integrating the atmospheric density over the range of production distances.

The reconstruction of the MPD distribution requires the removal of the electromagnetic contribution from the total signal³. SD high-energy events ($E > 20 \text{ EeV}$) with zenith angle in the range 55° - 56° are used, as at these angles only a residual electromagnetic contamination is contributing to the total signal. As the number of muons increases with energy, the high-energy threshold allows us to reconstruct the MPD distributions. From the timing in the traces measured in each SD station it is possible to extract an estimate of the arrival time of muons at the ground. The arrival times of the muons are used for the reconstruction of their geometrical production heights along the shower axis and hence for the reconstruction of the muon production depth and in particular of its maximum X_{max}^μ . This is related to the hadronic component of the shower that is highly mass-sensitive.

In [118], X_{max}^μ was measured for 481 events in the cited energy and angular region, observed between 1 January 2004 and 31 December 2012. The evolution of $\langle X_{max}^\mu \rangle$ thus obtained is shown in figure 3.12, compared to predictions from Epos-LHC and QGSJETII-04 hadronic models. The two models predict a similar evolution of X_{max}^μ for proton and iron but a considerable difference in its absolute value. While Auger data are bracketed by QGSJETII-04, they fall below the Epos-LHC prediction for iron, thus demonstrating the power of the MPD analysis to constrain hadronic interaction models. Like $\langle X_{max} \rangle$, also $\langle X_{max}^\mu \rangle$ can be translated into $\langle \ln A \rangle$ using the predictions of the different hadronic models. The results of this conversion for two different hadronic models are shown in figure 3.13 compared to the conversion obtained by using X_{max} measurements. With QGSJETII-04, we obtain compatible values for $\ln A$ within 1.5σ , while in the case of Epos-LHC the result of the X_{max}^μ conversion indicates primaries heavier than iron, the measurements being incompatible at a level of at least 6σ . However, as described in the previous section, QGSJETII-04 has problems describing in a consistent way the first two moments of the $\ln A$ distribution obtained from the X_{max} measurements. In conclusion, the described SD measurement, in correlation with the information from the shower profile as observed by FD, sets valuable constraints on the most recent LHC-tuned interaction models, QGSJETII-04 and Epos-LHC. In particular, none of the interaction models provides a consistent description of both the electromagnetic and muonic shower profiles as measured

³In chapter 7 a detailed description of an algorithm that separates the muon and electromagnetic components in the FADC traces will be given.

by the Pierre Auger Observatory.

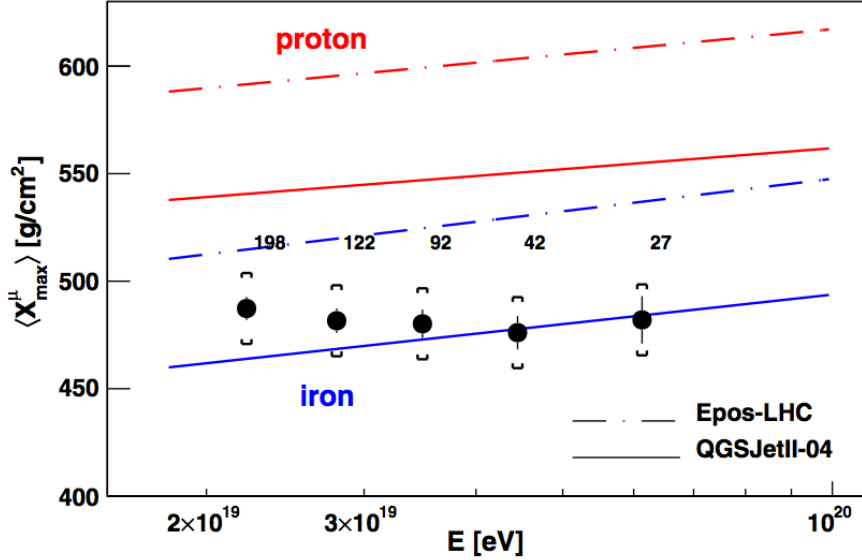


Figure 3.12: Average X_{max}^μ as a function of energy, compared to prediction from proton (red lines) and iron (blue lines) simulations made with QGSJETII-04 (full lines) and EPOS-LHC (dashed lines) hadronic interaction models. From [120].

Number of muons

Extensive air showers with zenith angles exceeding 60° are characterized at the ground by the dominance of secondary energetic muons, since the electromagnetic component has been largely absorbed in the large atmospheric depth crossed by the shower. The reconstruction of such inclined showers in Auger provides a direct measurement of the muon number at the ground. The muon number in inclined air showers is measured by using the relative scale factor N_{19} (see also next chapter) which relates the observed muon densities at the ground to the average muon density profile of simulated proton-induced air showers of fixed energy, 10^{19} eV. The latter has a shape that is practically independent of energy, composition or hadronic interaction models. The reliability of N_{19} as an estimator of the number of muons has been extensively tested with simulations comparing N_{19} to $R_\mu^{\text{MC}} = N_\mu/N_{\mu,19}$, the total number of muons in the simulated shower relative to the number of muons in the reference function. The average difference between N_{19} and R_μ^{MC} for proton or iron simulated using different hadronic models (including QGSJETII-04 and Epos-LHC) is always below 5%. We have used R_μ , defined as N_{19} corrected for the average bias of all the simulations, to study the muon content as a function of energy. The analysis is applied to hybrid events with zenith angles $62^\circ < \theta < 80^\circ$ for which a simultaneous measurement of muon number and of energy is possible. Only events with energy above $4 \cdot 10^{18}$ eV are used to ensure a 100% SD trigger probability.

The average values of R_μ , divided by the energy, are plotted for five energy bins in figure 3.14 (left). Data points are compared to simulations for proton and iron showers,

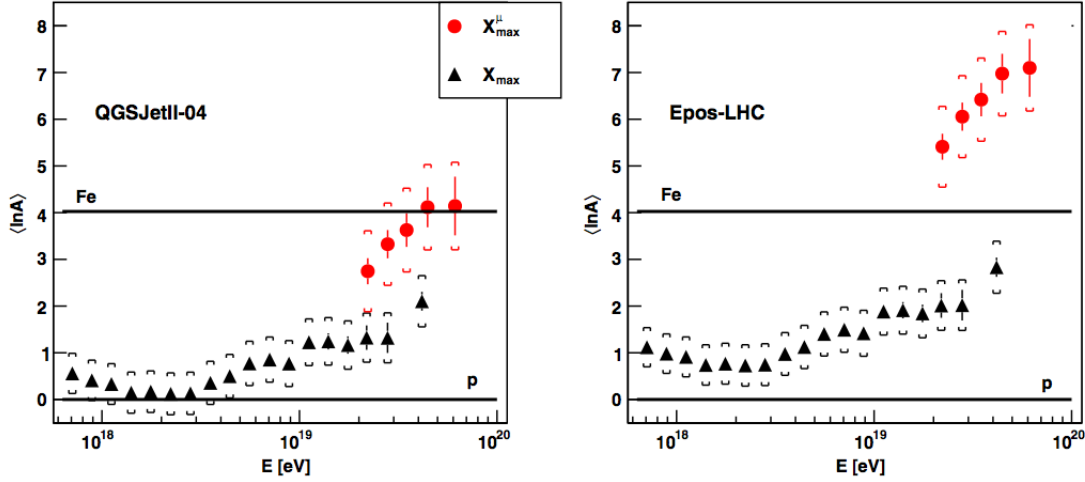


Figure 3.13: Predictions on the average composition as a function of energy based on X_{\max}^{μ} measurements (red dots), compared to those derived from X_{\max} (black triangles). In the left plot the conversion is done through the QGSJETII-04 hadronic interaction model, while in the right plot it is done through the EPOS-LHC one. See text for details. From [120].

performed with QGSJETII-04 and Epos-LHC hadronic models at $\langle \theta_{\text{data}} \rangle = 67^\circ$. The predictions for protons and iron nuclei are well separated, illustrating the power of $\langle R_{\mu} \rangle$ as a composition estimator. Our data strongly disfavor light composition in this energy range, in agreement with the X_{\max} measurements. In turn, the measured muon number is close or even higher than that expected in pure iron showers, a result not in agreement with studies based on X_{\max} which, as we have seen, indicate an average logarithmic mass $\langle \ln A \rangle$ being between proton and iron in the same energy range. Similarly to what was observed in the MPD studies, also in this case there is a tension between the X_{\max} and R_{μ} measurements with respect to the expectations from different models. Such a tension can be appreciated further in figure 3.14 (right) which displays Auger data at 10^{19} eV compared to different models and fixed composition in X_{\max} and R_{μ} space. The discrepancy between data and model predictions is shown by a lack of overlap of the data point with any of the model lines. The four hadronic interaction models fall short in matching our measurement of the mean logarithmic muon content. The discrepancy is smaller for the post-LHC models, and Epos-LHC performs slightly better than QGSJETII-04. As already concluded from the MPD studies, none of the models is covered by the total uncertainty interval of our data. The minimum deviation is 1.4σ . To reproduce the higher signal intensity in data, the mean muon number around 10^{19} eV in simulations would have to be increased by 30% to 80^{+17}_{-20} (sys.)%.

3.5.4 Arrival directions

The third important characteristic of the observed cosmic rays, besides their flux and their composition, is their arrival directions. The search for anisotropies at different angular scales can indeed provide valuable insights on the sources of cosmic rays and on their propagation to the Earth. Information on the sources might in particular be contained

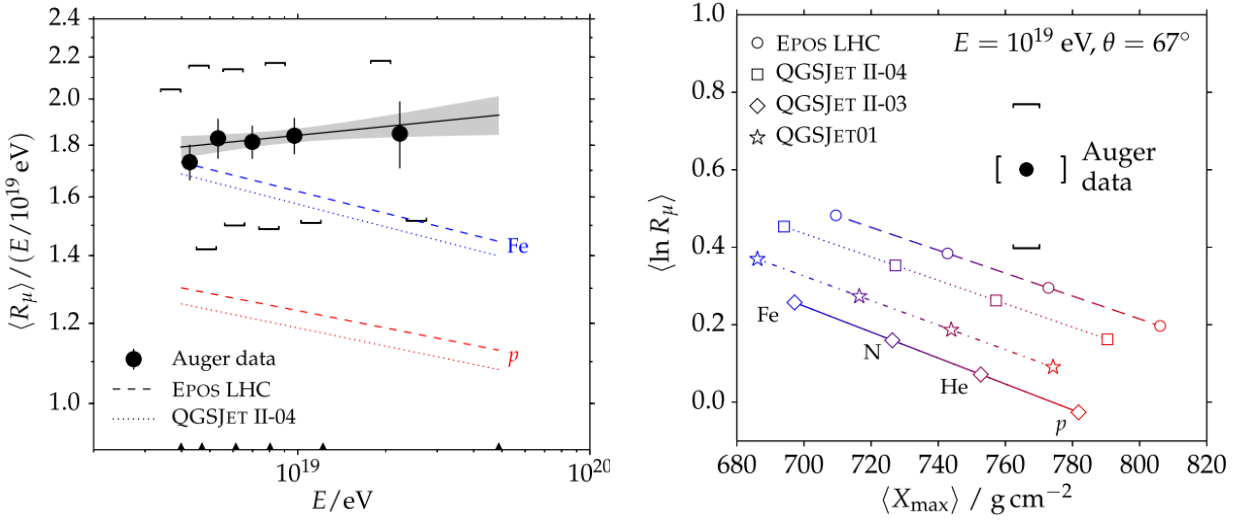


Figure 3.14: Left: Average muon content R_μ per shower energy E as a function of E . Square brackets indicate the systematic uncertainty of the measurement. The grey band indicates the statistical uncertainty of the fitted line. Shown for comparison are theoretical curves for proton and iron showers. Black triangles at the bottom show the energy bin edges. Right: R_μ as a function of the average shower depth in data (dot) compared to simulation predictions for different primaries and hadronic interaction models (lines). From [119].

in the small/intermediate-scale angular distribution. As this study is part of this thesis work, it will be presented in chapter 6. We will instead summarize here the results of the search for large-scale anisotropies performed in Auger at different energies.

Large scale anisotropies

As we have seen in section 1.1, for many theoretical models the transition between galactic and extragalactic sources of UHECRs is expected to happen at the ankle, in the EeV region. Large-scale anisotropies might thus be present at energies below the ankle if cosmic rays are of galactic origin, due to their escape from the Galaxy that might generate a dipolar large-scale pattern as seen from the Earth. On the other hand, above the ankle, if the sources are extragalactic, cosmic rays can come from cosmological distances, leading in principle to an isotropic distribution of their arrival directions. However, a dipolar flux could result from cosmic rays propagating diffusively in the extragalactic turbulent magnetic fields, in particular if the cosmic rays have a component with large charge Z . For these reasons, the search for large scale anisotropies, and their evolution versus energy, is of crucial importance for the determination of the origin of ultra-high energy cosmic rays.

Large-scale anisotropy studies take advantage of the large amount of data recorded by the SD arrays, both the 1500 m and the 750 m ones, to search for dipole or quadrupole patterns in the arrival directions of the detected cosmic rays [121]. The data set of the 1500 m array has been extended to include also events with zenith angle between 60° and 80° [122], leading to a 85% sky coverage and 30% more data. The method consists of fitting the counting rate as a function of right ascension with a sine wave. The harmonic analysis gives the amplitude and phase of the different harmonics (typically the first and second),

and the probability of detecting a signal due to fluctuations of an isotropic distribution with an amplitude equal to or larger than the observed one. Two different approaches are used depending on the energy range. If the events energy is above that corresponding to full trigger efficiency ($3 \cdot 10^{18}$ eV for the 1500 m SD and $3 \cdot 10^{17}$ eV for the 750 m one), the classical Rayleigh analysis is applied after correcting the counting rate for atmospheric effects and time-varying exposure. At energies where the arrays are not fully efficient, instead of applying the harmonic analysis to the total counting rate, the differences between Eastward and Westward directions [123] are used. As the two sectors are equally affected by detector instabilities and weather conditions, the use of their difference cancels these effects, allowing us to search for anisotropy in right ascension without requiring any evaluation of the detection efficiency.

The use of data from the two arrays has allowed us to study the amplitude and phase of the first harmonic over almost four orders of magnitude in energy, from 10^{16} eV up to the highest energies. Figure 3.15 summarizes the obtained results (in red, those obtained with the 750 m array, in black those with the 1500 m one), by showing the upper limits or measured amplitudes (left) and the phase (right) of the dipole equatorial component. There are several indications of anisotropies. As can be seen in the right panel, an alignment of phases below 1 EeV is observed: even if the amplitudes are not significant, a constancy of phases in ordered energy bins is expected to appear with a smaller number of events than necessary for the amplitudes to stand significantly above the background noise. The observed alignment is suggestive of real underlying dipolar-like anisotropies oriented in the direction of the right ascension of the Galactic center. A change to a phase in the opposite direction is observed at higher energies, where amplitudes starts to stand above the background noise in particular above 8 EeV [122]. At this energy, a first harmonic in right ascension with an amplitude $r_1^\alpha = (4.4 \pm 1.0) \cdot 10^{-2}$ with a chance probability of $P(\geq r_\alpha) = 6.4 \cdot 10^{-5}$ has been found. Under the assumption that the only significant contribution to the anisotropy is from the dipolar component, this observation would correspond to a dipole of amplitude $d = 0.073 \pm 0.015$ pointing to $(\alpha, \delta) = (95^\circ \pm 13^\circ, -39^\circ \pm 13^\circ)$. This direction suggests that the anisotropy imprinted on the CR arrival directions starts to be dominated by the flux of extragalactic CRs.

Telescope Array and large scale anisotropies

The Auger and TA collaborations have performed a joint search for large-scale anisotropies [124]. Given the location of the two experiments, the search for large-scale anisotropies using a joint data set allows for a full-sky coverage that enhance the performance of the analysis allowing one to measure the full set of spherical harmonic coefficients a_{lm} [125]. The dipole amplitude is observed to be $d = 0.065 \pm 0.015$ pointing to $(\alpha, \delta) = (93^\circ \pm 13^\circ, -46^\circ \pm 13^\circ)$. These results are in agreement with the ones reported with Auger only data, and moreover, due to the nature of the joint analysis, they do not rely on any assumption on the underlying flux of UHECRs.

3.5.5 Search for neutral particles

We summarize here the results of the search for primary photons, neutrinos and neutrons. Protons and nuclei interacting with the universal low energy photon background (CMB) are expected to produce a flux of UHE photons, that can propagate for a few tens of

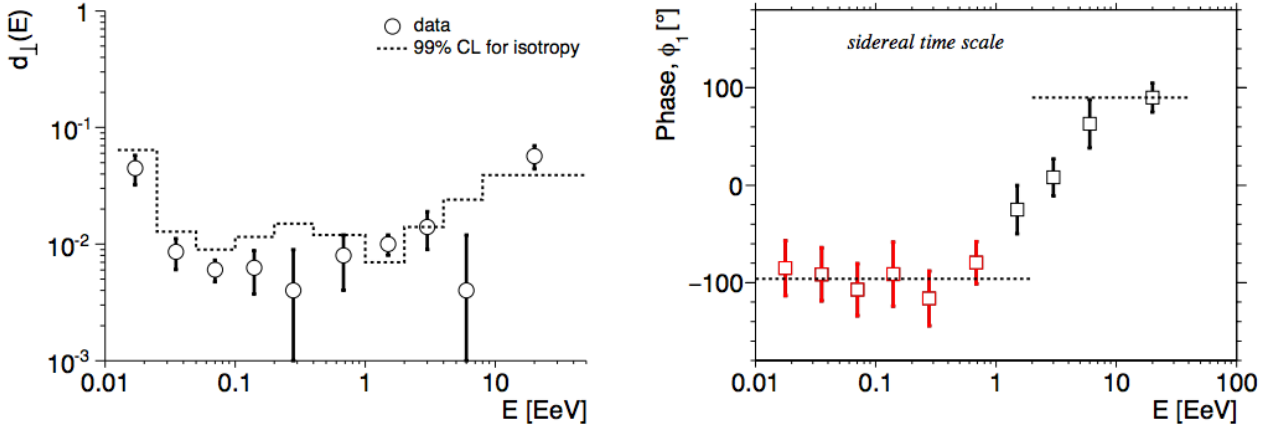


Figure 3.15: Left: measured equatorial dipole amplitude compared to prediction from isotropy at 99% CL. Right: measured phase. Red dots are coming from measurements done with the 750 m array, black dots with the 1500 m array. See text from details. From [126].

Mpc without being absorbed, and neutrinos (from the decay of charged pions, muons and neutrons), that can travel to the observer with no interaction or deflection. The expected cosmogenic fluxes depend on the composition and maximum energy of cosmic rays at the sources, emissivity, distribution and cosmological evolution of the acceleration sites. Thus, observing UHE photons or neutrinos can pose constraints on the UHECR origin and properties of the sources. In turn, the search for neutrons at EeV energies is relevant to test the origin, either galactic or extra-galactic, of EeV cosmic rays. Given the evidence, as we have seen, that at EeV energies the cosmic ray composition includes a strong proton component, a neutron flux should be in principle detectable if the sources of those protons are in the Galaxy and continuously emitting protons in all directions.

Neutrons

Neutrons are unstable particles with a lifetime of ~ 15 minutes. At the energy of our concern, EeV, this means that neutrons coming from the Galaxy can survive⁴ and be observed at the Earth. Neutron-induced showers are undistinguishable from the ones generated by protons, but, since neutrons are not deflected by magnetic fields, an excess of events from a given direction within the angular resolution of the observatory would be a signature of a neutron flux coming from that direction. The search for such an excess has been performed by the Auger collaboration, with two approaches. One was a “blind” search [127], scanning the whole sky in the search for excesses, and the other a “targeted” one, on a set of astrophysically interesting directions. The latter allows for higher statistical significance since the trial penalization is lower. For the “targetted” search, candidate galactic sources, like different kind of pulsars (millisecond, γ -ray), magnetars, microquasars, x-ray binaries and sources detected by HESS were considered, together with the Galactic center and the Galactic plane [128]. None of the two approaches found evidence for a EeV neutron flux,

⁴Above 1 EeV, the Galactic center is within the mean decay length, and above 2 EeV most of the Galactic disk is within range.

thus allowing us to put severe limits to such fluxes. The absence of any detectable neutron flux might suggest that the sources are extragalactic, or transient, or individually weak but densely distributed. The flux limits on neutrons couples with the absence of detectable amplitudes in large-scale anisotropy at EeV energies to constrain models in which EeV protons are produced by a low density of strong sources in the Galaxy.

Photons

Showers induced by photons develop deeper in the atmosphere compared to hadron showers, so the depth of shower maximum, X_{\max} , is larger for photon-induced events. In addition, they are also characterized by a smaller number of secondary muons and their signal in the SD has peculiar characteristics, related to the curvature of the shower front and to the risetime of the signals. Using these properties of photo-induced showers, also combined into multi-variate analyses, stringent upper limits on the photon fraction and flux were published by the Pierre Auger Collaboration using data both from the FD and the SD [129].

UHE photons could be generated by cosmic rays both by interaction of cosmic rays with the CMB (GZK effect, see section 1.2.2) and by pion photoproduction or inelastic nuclear collisions near their sources. In such way, a source within the Galaxy could then be identified by a flux of photons arriving from a single direction. Also, a substantial photon flux could be interpreted as a signature of “top-down” models, which predict that UHECRs come from the decay of exotic particles at the energies observed (see section 1.3.3 for further details on those models). Up to now no photon has been observed by Auger, resulting in strong flux limits on the photon flux. Such limits (due to both FD and SD data) are summarized in figure 3.16 compared to predictions from some top-down models and GZK production expected from proton or iron primaries. We can see that several of top-down models are ruled out. Moreover, the limits are starting to access the region where the observation of GZK photons from protons is expected. As the flux of such photons depend on the cosmic ray mass, a detection (or the lack of it) is another way of constraining the mass composition of UHECRs.

Neutrinos

Above $5 \cdot 10^{19}$ eV cosmic-ray protons interact with CMB photons and produce ultrahigh energy cosmogenic neutrinos of energies typically 1/20 of the proton energy. Their fluxes are uncertain and at EeV energies they depend mostly on the evolution with redshift z of the unknown sources of UHECRs, and on their spectral features at injection. Protons typically produce more neutrinos than heavier primaries do, so the measurement of the neutrino flux gives information on the nature of the primaries. UHE neutrinos are also expected to be produced in the decay of charged pions created in the interactions of cosmic rays with matter and/or radiation at their potential sources. In fact, at tens of EeV, neutrinos may be the only direct probe of the sources of UHECRs at distances farther than ≈ 100 Mpc.

Back in the 70s [130], it was suggested that array of particle detectors could detect UHE neutrinos by observing highly inclined showers. The idea is that neutrinos, having very small cross-sections, can interact at any point along their trajectories, while protons,

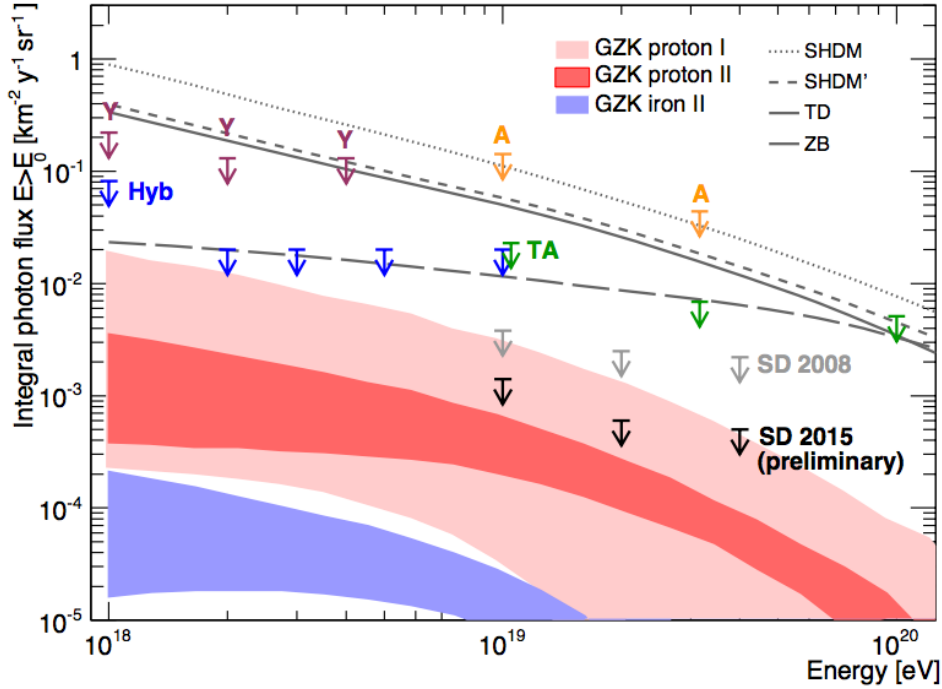


Figure 3.16: Photon limits with hybrid (blue arrows) and SD (black and grey arrows) data compared to Telescope Array limits (green arrows) and to predictions from top-down models (dashed lines) and GZK emission (fill colors: red for proton GZK, blue for iron). From [126].

nuclei, or photons interact shortly after entering the atmosphere. The signature for neutrino events is therefore inclined showers that interact deep in the atmosphere, thus having a considerable amount of electromagnetic component (“young” showers). In Auger neutrinos of energy above ~ 0.1 EeV can be detected in two different channels. The first one, called “downward-going”, concerns all neutrino flavors, and requires that the ν interacts in the atmosphere and induces an extensive air shower close to the ground. The second one, called “Earth-skimming”, is specific for τ flavored neutrinos, and the detected shower is the one induced by the decay products of an emerging τ lepton, after the propagation and interaction of an upward-going ν_τ inside the Earth. The identification of potential neutrino-induced showers is based on selecting those events that arrive in rather inclined directions, and then selecting among them those with FADC traces that are spread in time, indicative of an electro-magnetic signal and thus of the early stage of development of the shower. The exposure for neutrinos has been carefully computed, being higher for Earth-skimming neutrinos than for downward-going ones. No neutrino candidate was found, and the most recent results obtained with data between 1 January 2004 and 20 June 2013 (see [131]) led to the neutrino diffuse flux limits shown in figure 3.17. The Auger limits are shown together with those from neutrino experiments, i.e., Anita and IceCube, as well as with expectations for several cosmogenic models and with the Waxmann-Bahcall bound on neutrino production in optically thin sources. We can see that Auger is at the moment the detector with the highest sensitivity for neutrinos in the EeV region, giving the most stringent limits. Such limits are now starting to disfavor different models based on pure

proton composition at the highest energy, in particular the ones with a strong evolution of the sources (FR II-type).

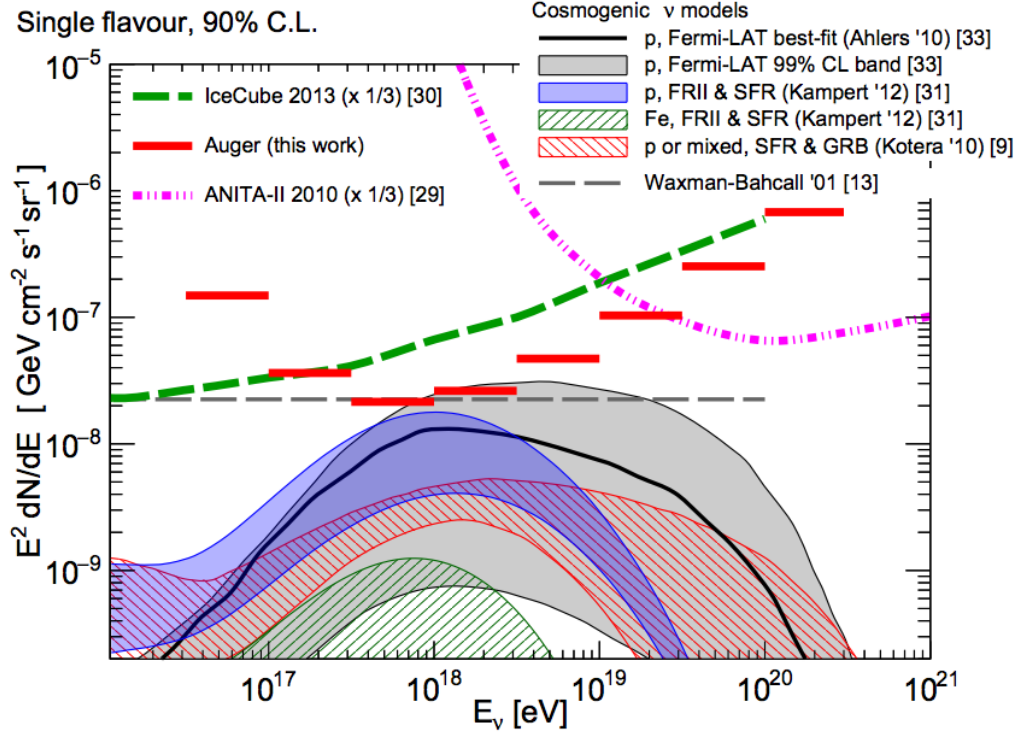


Figure 3.17: Upper limit to the normalization of the diffuse flux of UHE neutrinos (at 90 % C.L.) compared to the one coming from ANITA-II (dot-dashed line) and Ice Cube (dashed line) experiments (with appropriate normalizations to take into account the energy bin width, and to convert to single flavor) and to predictions for several cosmogenic neutrino models and the Waxman-Bahcall bound. From [131]

We have outlined in this chapter the main characteristics of the Auger Observatory. We have also presented a selection of the results published by the Auger collaboration. The rationale behind this selection has been that of setting the stage for the thesis work that will be presented in the following chapters.

The stringent limits to the flux of primary photons at UHE largely constrain top-down models for the origin of UHECRs and hence favor astrophysical objects as accelerators. Also, the indication of a large scale anisotropy at UHE in a direction that points outside the Galaxy suggests that the origin of such cosmic rays is extra-galactic, this being hinted at also by the non-detection of a Galactic neutron flux, a possible tracer of Galactic UHE-CRs sources. A suppression in the flux of UHECRs at energies above 40 EeV has been established beyond any doubt. While the energy at which the spectrum steepens is in accordance with that expected from the GZK effect, the spectrum alone does not allow

one to conclude whether the observed feature is due to propagation effects or to source properties, i.e. the maximum energy achievable in the acceleration process. Whatever the origin of the suppression, UHECRs are expected to come from sources relatively close to the Earth due to energy losses in their interactions with photons of the Cosmic Microwave Background. Indeed, these interactions limit the distance from which a source can contribute to the flux at Earth, i.e. less than ~ 200 Mpc for protons or Fe nuclei with energies above 60 EeV, and even smaller for intermediate mass nuclei. Thus, the number of candidate sources which could contribute to the measured fluxes at the highest energies is significantly reduced, and moreover, at such distances, the galaxies are distributed non uniformly. The measurement of the cosmic ray composition has been addressed through the measurement of the depth of shower maximum, X_{\max} . Interpretations of Auger data through the most updated models of hadronic interactions indicate that the fraction of heavy nuclei increases above the energy of the ankle (the spectral hardening taking place at $E \simeq 5$ EeV) up to the highest energies. While the small number of events detected by FD does not allow one to probe the primary mass evolution in detail at energies in excess of 40 EeV, interpretation of X_{\max} data does not exclude a possible resurgence of the proton component in the ultra-high energy range. Even if low-charge particles were to contribute only a fraction of the primary cosmic rays, anisotropic signals on small angular scales might show up for a larger enough number of events. In turn, should the UHECRs be significantly deflected, either due to their large charge or due to the presence of strong intervening magnetic fields, directional excesses might still be found at intermediate angular scales.

Such a scenario is the reason for the search for small- and intermediate scale anisotropies in the distribution of the arrival directions of UHECRs, both intrinsic and in correlations with “close-by” astrophysical objects, that will be described in chapter 6, as part of this thesis work. It is also the reason for our attempt to identify proton-candidates in the UHECR sample and using it to enhance the sensitivity of the search, as will be shown in chapter 7. The first step is in turn the building of the UHECR sample, which will be detailed in chapter 5, after explaining the reconstruction of the events in chapter 4.

Chapter 4

Reconstruction of surface detector events

*L'Argentina, ah l'Argentina che tensione!
quella croce del Sud nel cielo terso
la capovolta ambiguità d'Orione
e l'orizzonte sembra perverso*

Francesco Guccini - *Argentina* - from *Metropolis* (1981)

As anticipated in the previous chapter, in this one we will describe the procedures of recording and reconstructing events using the Auger surface detector (SD). We will focus on vertical events ($\theta < 60^\circ$) and we will describe step by step the reconstruction procedure performed by one of the two official reconstruction softwares, the so-called Herald. In parallel, in dedicated boxes, we will also outline the main characteristics of the reconstruction of horizontal events ($60^\circ < \theta < 80^\circ$) and of the other software, the so-called Offline. When an event is selected by the triggers (section 4.1), its reconstruction procedure is divided in 3 main parts: calibration of signals (section 4.2), selection of good stations (section 4.3) and, finally, the reconstruction itself (section 4.4). The final section (4.5) describes the energy calibration that is done by exploiting the fluorescence detector and the hybrid nature of the Pierre Auger Observatory.

In the Offline software

The text will follow the reconstruction procedure implemented in the Herald software. In the Auger Collaboration, another reconstruction software is used: the Offline, that follows a procedure that is conceptually identical. However, the two reconstruction algorithms differs in some crucial points. These points will be outlined in boxes like this one.

In the HAS reconstruction

The text will describe the reconstruction procedure used for events with $\theta < 60^\circ$, called vertical. However, a different kind of reconstruction has been developed in order to reconstruct also Horizontal Air Showers (HAS, $60^\circ < \theta < 80^\circ$). Although we will not go into the details of this reconstruction (detailed information can be found in [132]), the key points will be outlined in boxes like this one, in parallel with the corresponding vertical reconstruction section.

4.1 SD Triggers

Each of the three PMTs in the SD water Čerenkov detectors records the light produced by particles interacting in the 12 tonnes of water that the tank contains. Most of those particles are surviving muons from low energy showers that were absorbed in the high atmosphere, together with low energy γ and e^\pm . These particles are the background for our studies: in order to select high energy showers over this background, a set of hierarchic conditions has been imposed. The first two levels (T1 and T2) in the hierarchy are applied online locally at each detector. The third level (T3) is formed online at the CDAS, based on timing and position of stations satisfying the prior two levels; if the required conditions at this level are met, data acquisition and storage are started. The last two levels (T4 and T5) are, on the other hand, offline conditions applied on recorded data, and will be described in section 4.3. More details on the SD triggers can be found in [96].

4.1.1 Station trigger: T1 & T2

The first two level of triggers are applied locally in the SD stations. The first trigger, called T1, is required for the station to store in memory the traces of the three PMTs. Two kind of conditions may be satisfied in order to classify the station as T1: a simple threshold condition, called Th-T1, requires all three PMTs to have a signal higher than 1.75 VEM (Vertical Equivalent Muon, for further details see section 4.2) in at least one FADC bin¹. This trigger selects mainly atmospheric muons and reduce the trigger rate from ~ 3 kHz to ~ 100 Hz. The second condition, called Time Over Threshold (ToT-T1) requires on the other hand at least 13 bins (i.e. > 325 ns) in a sliding window of 120 bins (i.e. $3 \mu\text{s}$) to be above a threshold value of 0.2 VEM. This trigger is intended to select sequences of small signals spread over time. It is optimized for the detection of near-by, low energy showers, dominated by the electromagnetic component, or for high-energy showers where the core is distant from that detector. The ToT rate at each detector is < 2 Hz and is mainly due to the occurrence of two muons arriving within the duration of the sliding window.

A second, more stringent, local station trigger is the T2 trigger, that is applied to reduce the rate of events per detector to ~ 20 Hz. This is done because of the band-width

¹If only 2 PMTs are active, the threshold is raised to 2 VEM. If only one PMT is active, it is raised to 2.8 VEM.

limitation in the communication system between the detectors and the central campus. All ToT-T1 triggers are automatically promoted to ToT-T2 (or simply ToT), while a higher threshold value of 3.20 VEM, again with a 3-fold coincidence of the PMTs², is imposed to Th-T1 in order to be promoted to Th-T2.

Two new station triggers have been implemented in the array more recently (July-2013): the so-called deconvolved ToT (ToTd) and the Multiplicity of Positive Steps (MoPS). The first one is basically the same as the ToT but a deconvolution is first applied on the trace to account for the exponential decay of the signals. The second one is an algorithm that counts how many positive steps above a certain threshold are present in the trace. These triggers reduce the threshold for signals acquisition, in particular for those dominated by the electromagnetic component. As a result, there will be more stations contributing to the typical shower footprint. Thus, the implementation of the new triggers improves the event reconstruction and enables several physics analyses. In particular, the search for photon primaries at EeV energies. For the time being, they are not used in the standard event selection and reconstruction for the sake of uniformity with respect to the preceding years. For this reason, we will not deal with them in the following.

4.1.2 T3: array trigger

All T2-triggered stations notify the CDAS of their status. The CDAS then searches for real showers by looking at coincidences of T2 triggers in nearby stations. Two configurations are allowed. The first one, called ToT2C1&3C, asks for a three-fold coincidence of ToT stations among which one of them must have one of the other two in the first “crown” of stations (the 6 surrounding stations) and the other within the second “crown” (see figure 4.1, left). The rate of events that pass this condition is around 1600 per day, among which 90% are physics events. This configuration is mostly efficient for vertical events, up to 60°. The second configuration is more permissive. It requires 4 stations from any T2 in a less compact pattern: one station must have one of the others within the first “crown”, another within the second one and the last one can be as far as in the fourth “crown”. This configuration is called 2C1&3C2&4C4 and is sketched in figure 4.1, right. This configuration is more efficient for horizontal showers, that are much richer in muons. Around 1200 events per day are selected by this configuration, but only $\sim 10\%$ of them are real showers. In addition to these geometrical requests, for both configurations it is also requested that the timing of each station must be within $(6 + 5 C_n) \mu s$ from the first one, C_n being the crown where it lies. If any of the T3 conditions is satisfied, the CDAS records the event by saving not only the traces of the stations participating to the T3 trigger itself, but also the ones coming from the nearby stations that reported a T1 or T2 in coincidence with the T3 trigger within 30 μs .

²The threshold is raised to 3.8 (4.5) VEM for stations with only 2 (1) PMTs active.

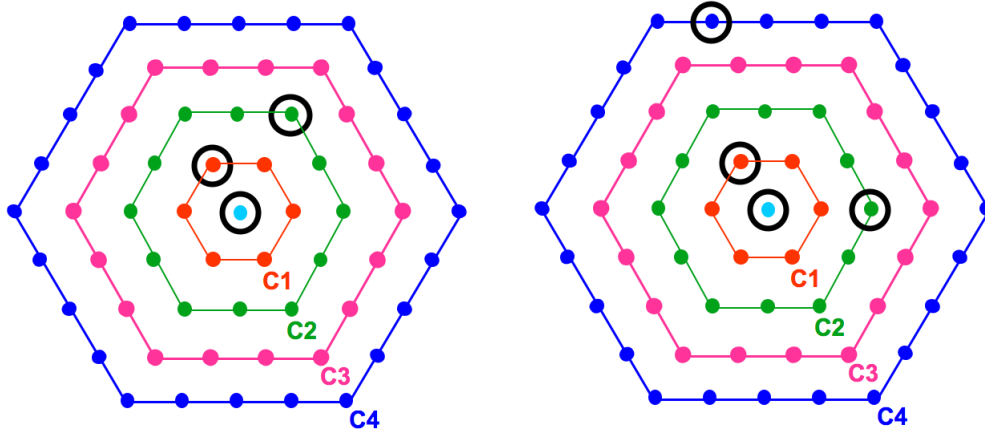


Figure 4.1: Examples of T3 Configurations for ToT2C1&3C2 (left) and 2C1&3C2&4C4 (right) from [96]

In the HAS reconstruction

Up to the T3 level, no difference is made between vertical and horizontal showers (HAS). Differences start at T4 level, when a first reconstruction of the arrival direction is available. For HAS the station selection logic is rather different, to account for the sparse ground topology characteristic of very horizontal showers. The procedure is called top-down as no seed is used. For trigger configurations with more than 3 stations, criteria that are applied are based on the start times of the traces (compatibility with a shower front) with tolerances depending on the estimated zenith angle. The initial configuration of N preselected stations is tested, and then successive trials with $N-1$, $N-2$,... stations are performed until a satisfactory configuration is found. For 3-fold topologies, a specific procedure is applied, taking into account the shape of the FADC signals. The top-down procedure includes also a processing of the FADC traces, in terms of integrated signal and signal/peak. Isolated stations are rejected, similarly to what is done for the vertical showers, but with a larger tolerance. Also, an event needs to have at least four selected stations before being passed to the reconstruction algorithm. Finally, T5 condition for HAS requires that the station that is nearest to the reconstructed core and its six adjacent stations are all operational.

4.2 Signal treatment

The conversion of the photoelectron (PE) signal as recorded by the PMTs into a calibrated variable is crucial for the correct reconstruction of each EAS event. A first calibration is done directly in the local station since it is fundamental for the uniformity of the trigger condition all over the array. However, a more refined calibration is computed offline, and the online calibration is used only in particular cases, that will be outlined in the following sections. After all signals in each station are calibrated, a list of calibrated stations is returned that will then undergo the station selection.

4.2.1 The Vertical Equivalent Muon (VEM)

Due to possible variations of the detector response versus time, both due to atmospheric and detector effects, the stations are calibrated offline on an event-by-event basis with a common source. To do this, we use the average charge collected by a PMT from the Čerenkov light produced by a vertical and central through-going muon (VEM) which can be used as a reference signal. A water Čerenkov detector cannot however select only vertical muons. Dedicated measurements using scintillators above and below test tanks to select vertical muons proved that the peaks of the distributions of the charge deposit and the pulse height distributions for omni-directional atmospheric muons are proportional to the values for a vertical through-going muon, Q_{VEM}^{peak} and I_{VEM}^{peak} , respectively.

Figure 4.2 shows histograms of the two distributions: the two smaller and broader peaks correspond to omni-directional muons. The dashed histograms represent the distributions for vertical muons, selected by test scintillators. Charge histograms, as well as other monitoring informations (overall called “calibration block”), are recorded by the CDAS for each station participating in a trigger, along with the signal traces: such histograms include the charges deposited by crossing muons each minute before the event. Recording the values before rather than after ensures that no residual tail of the event-generated trace will be present. Among other variables, the local station sends to the CDAS also the online calibration results, the ones that were used for checking the trigger condition. These values are usually not used but can be exploited in the case that the charge histogram is defective, as it may happen, for example, in the case of communication problems.

First the reconstruction software reads the calibration block and extracts the charge histogram. This histogram, if present, is then fitted in order to extract the position of the so-called muon hump. A correction is then applied to the fitted muon hump position to be converted into a good estimate of the real VEM. This correction has been measured [133] to be $Q_{VEM}^{est} = (0.96 \pm 0.03)Q_{VEM}^{peak}$.

4.2.2 Conversion of the FADC trace into VEM units

Before converting the FADC trace to VEM units, the baseline value is estimated and subtracted from the trace. This is done in the anode trace or in the dynode one, depending on the latter being or not saturated. An algorithm then selects the signal regions on the dynode channel even if these regions are separated in different pieces (as for example in small, long signals) and then the selected signals from the three PMTs are averaged into one. From this averaged trace the software extracts the start time to determine the station timing. The averaged trace is then integrated and converted to VEM to determine the station signal, S . The uncertainties on the signal value are Poissonian, i.e., they are taken as \sqrt{S} . Note that the errors introduced by the calibration have been proven to be negligible with respect to statistical fluctuations.

4.2.3 Saturation recovery

If the dynode trace is saturated (i.e., at least one bin is over 1024), the anode trace is used to evaluate the signal and its timing. If also the anode trace is saturated, a procedure is

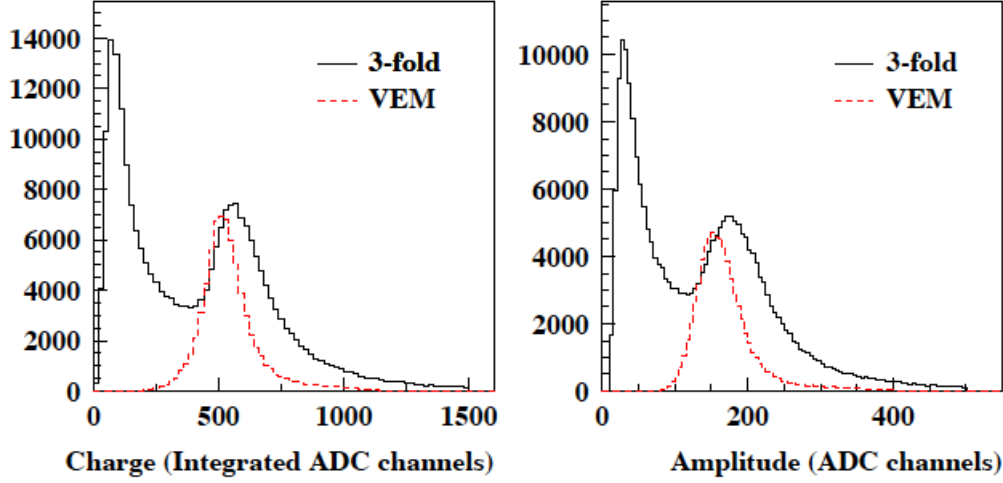


Figure 4.2: The distributions of the charge (left panel) and amplitude (right panel) measured by the stations for a 3-fold PMTs coincidence (black histogram). In red, the distributions of the charge (left panel) and amplitude (right panel) vertical muons, selected through scintillators. From [133].

performed to estimate the true signal, which a simple integral of the trace would underestimate. In the Herald software this is done by measuring the undershoot that is produced after the saturated signal both in the anode and in the dynode channels. In fact, since the readout chain is a simple RC stage, it is expected for the two quantities to be correlated [134]. The undershoot is measured as the difference between the baseline measured before the signal and the one measured after. In principle, the dynode undershoot allows for a better estimate. But the baseline after the signal may reach the “zero level for large signals (where the anode undershoot is greater than 1 FADC channel). The dynode undershoot cannot be measured in these cases, and the anode undershoot is used to determine the saturated signal. The relations between the undershoots and the “real” anode charge have been carefully studied in [135], and the saturated signal is then recovered as:

$$Q_a^{rec} = \begin{cases} 770 \cdot (U_d - 1.35 \cdot 10^{-4} Q_d + 0.055 k_d) - 222 & \text{if } U_a < 1 \text{ FADC channel} \\ 1669 + 14775 \cdot U_a & \text{if } U_a > 1 \text{ FADC channel} \end{cases}$$

where Q_a^{rec} is here expressed in FADC channel and should then be converted in VEM via the usual calibration, U_d and U_a are respectively the measured dynode and anode undershoots, Q_d is the measured dynode charge and k_d is the number of saturated dynode readings. The uncertainty is $\sim 10\%$ in the case of $U_a < 1$ FADC channel and $\sim 16\%$ otherwise.

4.3 Selection and quality

After the initial treatment of the FADC traces, each station is characterized by two main parameters that, together with its position, will be used in the reconstruction procedure:

the integrated signal and its start-time. The second step of the reconstruction chain builds on these parameters and consists in the selection of PMTs and stations to be used to define an event. As the study of the former is part of this thesis work, it will be detailed in section 5.1.1. Here we focus on stations and event selection.

4.3.1 Station selection and T4 physics trigger

After all PMTs are checked, the list of calibrated station is inspected and a selection is made. Indeed, a selection of physics events and of detectors belonging to each event is necessary as a large number of chance coincidence events is expected due to the large number of possible combinations among the single detectors. This is done via the T4 or “physics trigger,” which consists in selecting events that fulfill the timing and geometry expectations of a real EAS. Two different conditions are applied. The first one is the so-called 3ToT. It requires the presence of 3 nearby ToT stations in a triangular pattern. Thanks to the low rate of ToT, less than 1 chance coincidence is expected per day. The second condition is the so-called 4C1 where 4 nearby stations are required, with no specific T2 trigger request. In both case, it is required that the timing of the stations is coherent with a plane shower front moving at the speed of light. Due to their compactness, events with zenith angles below 60° are selected with high efficiency, i.e., more than 98%. The 4C1 works particularly well above $\sim 55^\circ$, where the efficiency of the 3ToT starts decreasing. It thus brings to 100% the efficiency for showers below 60° .

Besides disentangling accidental events, there is also the need to identify, and reject, accidental detectors in real events, i.e. detectors whose signals are by chance in time with the others, but that in fact are not part of the event. To this aim, we define a “seed” made by three neighbouring detectors in a non-aligned configuration. If there is more than one triangle of stations, the seed with the highest total signal is chosen. Once the triangle has been determined, the arrival direction is estimated by fitting the arrival times of the signals to a plane shower front moving with the speed of light. Subsequently, all other detectors are examined, and are defined as random stations, if their time delay with respect to the front plane is outside defined time windows [96]. Detectors that have no triggered neighbours within 3 km are always removed from the list of selected stations that is passed to the reconstruction module. If an event results to have only aligned stations, it is not used in the reconstruction since the arrival direction cannot be reconstructed accurately.

As we will see later, also stations that are active, but not triggered at the time of the event are used in the reconstruction, if they are close enough to the shower core position. This is due to the fact that the local trigger is fully efficient for signals above 10 VEM. Stations might be non-triggered just because the signal was sub-threshold. The definition of a non-triggered station relies on the knowledge of its activity at the time of the event. The activity of the stations and their live time are included in a database that is accessed by the reconstruction software. Non-triggered active stations in the proximity of the event (less than 5 km away) are stored in a dedicated silent stations list that is passed to the reconstruction module. Further information on this and on how mis-classified (so-called “fake non-triggered”) stations can impact on the event reconstruction will be given in section 5.1.2.

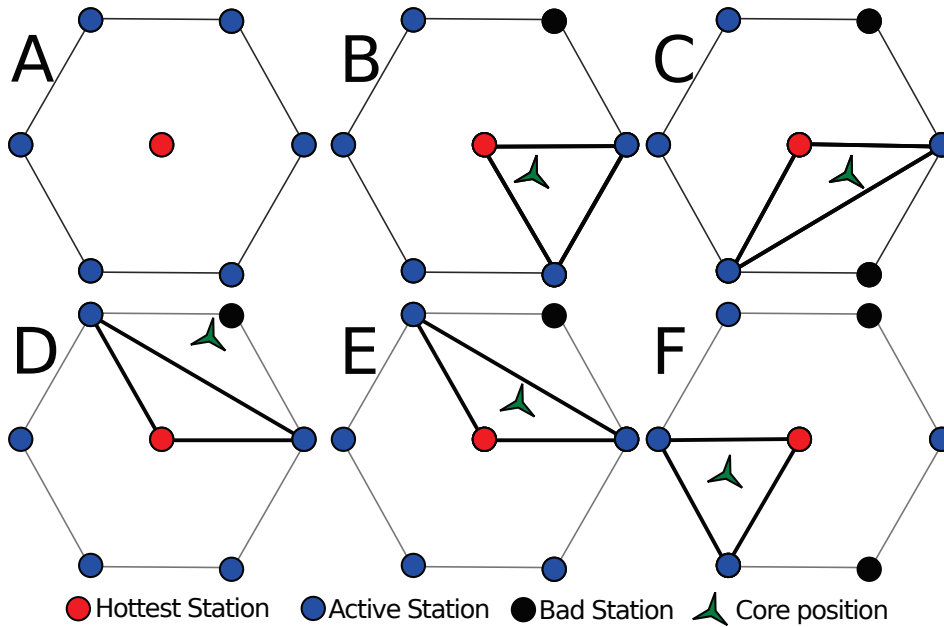


Figure 4.3: Examples of T5 Configurations.

A: a 6T5 condition (no matter where the core position is inside the hexagon).

B: an event satisfying a 5T5 - POS condition.

C: an event satisfying a 4T5 - POS2 condition.

D: a 4T5 event not satisfying POS or POS2 conditions.

E: a 5T5 event not satisfying a POS condition but rather a POS2 one.

F: a 4T5 event satisfying a POS condition.

If a 6T5 selection is applied, only A would pass. If a 5T5 POS selection is applied, only A and B would pass. If a 4T5 POS2 selection is applied A, B, C, E and F would pass. Condition D would be rejected in all Auger analysis so far.

4.3.2 T5: fiducial trigger

If an EAS falls on the border of the array, a part of the shower footprint may be missing. Such events could have wrong estimated core positions, and consequently, incorrect energies. The main task of the fiducial trigger is thus to select only events well contained in the array, ensuring that the shower core is properly reconstructed. The T5 trigger level requires that all the 6 stations surrounding the station with the highest signal (commonly referred to as the “hottest” station) have to be active at the moment of the event, even if they are not triggered (so-called 6T5 condition, see figure 4.3, A). The 6T5 selection is the standard one used in Auger for all physics analyses.

The 6T5 condition reduces the effective area of the array of $\sim 10\%$ on average, as events falling in the border region of the array, or close to inactive tanks, are not used. For this reason, a less strict condition is used for the highest energy events (above 40 EeV). In such events, the footprint on the ground is large enough to ensure a good sampling even without a whole active hexagon around the hottest station. This condition, called 5T5 POS requires only 5 of the stations surrounding the hottest to be active (5T5) and, in addition, it requires the reconstructed core position to be within an equilateral triangle of

active stations (T5 POS condition, see figure 4.3, B). This is the selection criterion used e.g., in point-sources search studies (such as [136], [137] and [138]).

A further even less strict condition has been devised to select the highest energy events, called 4T5 POS2. It requires only 4 active stations around the hottest one and the core to be within an isosceles triangle of active stations (see figure 4.3, C). The study of this criterion is part of this work and will be discussed in paragraph 5.1.3.

4.4 Reconstruction

Through the reconstruction procedure, the stations signals and their timing are exploited to extract the information on the primary cosmic ray: its arrival direction (zenith and azimuth angle, θ and ϕ), the position of the impact on the array (core), and its energy. As anticipated, we will describe here the official reconstruction procedure as it is performed in the Herald software.

4.4.1 Time fit: first estimate of the arrival direction

The first step of the reconstruction does not exploit the values of the station signals but only their timing. If we imagine a plane shower front arriving on the ground from the direction indicated by the vector \vec{a} at the speed of light c , stations at a position \vec{x} should be triggered at a time

$$ct(\vec{x}) = ct_0 - (\vec{x} - \vec{b}) \cdot \vec{a} \quad (4.1)$$

where t_0 is the impact time of the shower core at the ground, and \vec{b} points at the position of this impact point (see figure 4.4 for an illustrative sketch). If we assume that the position of each tank is known with infinite precision, the only uncertainty comes from the timing of each station, σ_t . We can then build a χ^2 function from the former formula:

$$\chi^2 = \frac{1}{\sigma_t} \sum_i (t_i - t(\vec{x}_i^{\vec{a}}))^2 = \frac{1}{c^2 \sigma_t^2} \sum_i (ct_i - ct_0 + \vec{x}_i^{\vec{a}} \cdot \vec{a})^2 \quad (4.2)$$

where $\vec{x}_i^{\vec{a}} = \vec{x}_i - \vec{b}$ is the position of the i -th station as measured from the shower core position. By minimizing this function, the vector \vec{a} can be evaluated. This vector translates into the local zenith and azimuth angle (θ and ϕ)³ The plane front is however only an approximation: EAS are in fact characterized by a curved shower front. Introducing this curved shape in the fit requires the knowledge of the shower core position and thus the curved fit is performed as a second step.

4.4.2 Lateral distribution function and global fit

After the first time fit has been performed, an estimate of the arrival direction of the primary is available. This is the point where the vertical and horizontal reconstruction starts to be different: if a shower is found to arrive with a zenith angle $\theta < 60^\circ$, then it will go through the vertical reconstruction procedure. In this case, the estimation of the

³Because of the huge spacial extension of the array, the Earth curvature should be taken into account. When comparing different θ and ϕ , we always refer to a corrected local value that gives the site value of each event to a common position in the array.

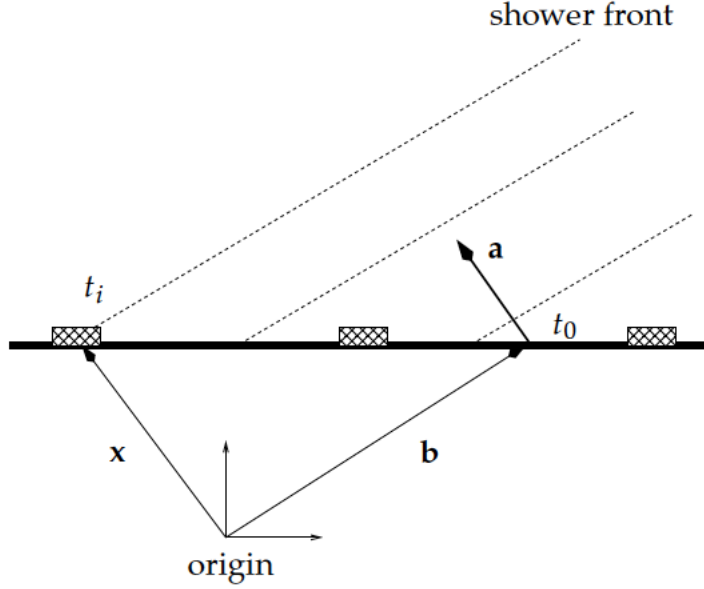


Figure 4.4: A sketch of the timing of a shower hitting a ground-based detector array. Courtesy of Hugo Rivera.

energy of the EAS is done by fitting an empirical Lateral Distribution Function (LDF) to the observed distribution of signals as a function of distance to the shower axis. The LDF has a normalization parameter that is the signal at a reference distance, S_{ref} . The optimal reference distance depends on the array characteristics (in particular on its spacing) and it is chosen to minimize the shower-to-shower fluctuations with respect to different LDF formulations and different primary masses. For the 1500 m array, the optimal distance was found to be 1000 m [139]. This reference signal at 1000 m, called S_{1000} , is correlated to the energy of the primary and is the energy estimator that will be used in the energy calibration procedure described in the following sections.

Lateral distribution function

Among the various possible formulations of the LDF parametrization, in the Herald reconstruction the predicted signal in a station at a distance r from the shower axis is parametrized as:

$$S(r) = \begin{cases} S_{1000} \left(\frac{r}{1000 \text{ m}} \right)^{(\beta + \gamma \cdot \log(r/1000 \text{ m}))} & \text{if } r > 300 \text{ m} \\ S_{1000} \left(\frac{r}{300 \text{ m}} \right)^{(\beta + 2 \cdot \gamma \cdot \log(3000 \text{ m}/1000 \text{ m}))} \cdot \left(\frac{3000 \text{ m}}{1000 \text{ m}} \right)^{(\beta + \gamma \cdot \log(3000 \text{ m}/1000 \text{ m}))} & \text{if } r \leq 300 \text{ m} \end{cases} \quad (4.3)$$

where S_{1000} represents the signal that would be observed at 1000 m from the shower core and β is related to the slope of the LDF. The β parameter is fitted only if there are enough degrees of freedom. Otherwise, it is fixed to a value that is based on a data-driven parametrization that depends on the energy and the arrival direction of the shower. The γ parameter has a fixed value and accounts for the small LDF curvature.

The distance r is computed with respect to the shower axis, not to the shower core. For this reason it depends both on the shower core position and on the arrival direction. As the arrival direction is already determined by the time fit, the LDF fit have then three free variables to be determined: S_{1000} , X_{core} and Y_{core} .

In the Offline software

The LDF treatment in the Offline software is different: the fitting function is a formula derived from the NKG function [56, 57] that is theoretically derived for purely em showers:

$$S(r) = S_{1000} \left(\frac{r}{1000 m} \right)^\beta \left(\frac{7000 m + r}{1000 m + 7000 m} \right)^{\beta+\gamma}$$

Moreover, the β parameter is always parametrized and never fitted even in events with large station multiplicity. The fit is made through a Maximum Likelihood method, rather than the χ^2 minimization used in Herald. This allows to differentiate the error form for different signals: poissonian errors are associated to small signals and gaussian errors to larger signals.

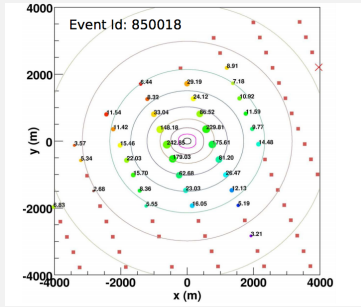
As a first step, a rough estimate of the core position is made. This is done by computing the barycenter of the shower footprint as the average of the coordinates of the stations, weighted by the root mean square of the corresponding signal. Then, a first LDF fit is performed and an estimation of S_{1000} is obtained. This being done, a second fit of the arrival direction, including a curved shower geometry, is performed. After this, a further selection for out-of-time stations is performed: if a station is found to be incompatible with the fitted shower front beyond uncertainties, it is removed from the list of selected stations and added to the random list. Up to 3 stations can be removed this way. Finally, a Global fit of both timing and LDF is performed, where all parameters (core position, LDF parameters, and arrival direction) are fitted at the same time in the same minimization process.⁴ It has to be noted that the Global fit takes into account both saturated and non-triggered ("silent") stations. The signals of saturated stations, recovered by using the procedure described earlier, are considered in the fit as lower bounds. The error associated to the recovered signals is taken as proportional to the signal rather than to its square root as in non-saturated stations. In turn, signals from non-triggered stations, are treated in the fit as upper bounds. This bound is fixed at 3 VEM, based on the trigger probability of the local station as a function of the signal in the station. Non-triggered stations are considered in the fit only if they are within 5000 m from the shower core. In figure 4.5 an example of a reconstructed Herald event is shown. The LDF fit is shown in the bottom panel.

⁴Further refinement can be made before the final fit: for example, in the default reconstruction option, the signal of each station is corrected for the azimuthal asymmetry. Namely, based on the arrival direction of the shower, the different response of early (or "upstream") and late (or "downstream") stations is corrected. The response of the detector is thus reported to a perfectly symmetrical one as if the shower fell vertically on the array.

In the HAS reconstruction

Horizontal showers require specific reconstruction methods since they are dominated by muons displaying complex density patterns at the ground level. The method used for the reconstruction in this case is based on a fit of the measured signals to the expected pattern, and requires modeling of a two-dimensional distribution of the muon number densities at the ground, the response of the detectors to the passage of muons, and the treatment of the electromagnetic component of the signal in the detectors. In particular, the expected pattern of a proton shower at 10^{19} eV is taken as a reference. What is then used as an energy estimator is N_{19} , the measure of the shower size, which is the relative normalization with respect to the reference muon distribution, namely:

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



An exemplary horizontal event projected onto the shower plane with the contour plot of the fitted distribution superimposed, indicating the signal measured in the 37 triggered stations and the position of the reconstructed core. From [132].

4.4.3 Uncertainty on the arrival direction

The angular accuracy of the surface detector is driven by the precision in the measurement of the arrival time of the shower front at each station (T_s). The variance of T_s has been carefully parametrized under the assumption that the particle arrival time in the shower front can be described as a Poisson process over a time interval T (see [140] for details), and can be written as:

$$V[T_s] = a^2 \left(\frac{2T_{50}}{n} \right)^2 \frac{n-1}{n+1} + b^2 \quad (4.4)$$

where b is the detector clock precision, T_{50} is the time between the start time of the trace and the time where it reaches 50% of the total signal and n is the number of particles that crossed the station in the time interval T . n is computed assuming that muons are the particles that contributes most to the time measurement: since muons arrive mainly parallel to the shower axis, the number of particles is defined as the ratio between the total signal and the track-length in the station, computed on the basis of the arrival direction. The parameter a is a scale factor that contains all the assumptions considered in the model and the treatment done to the FADC traces. To fit a and b ⁵ a study was performed using “doublets”, i.e., pairs of stations 10 meters apart. Doublets allow us to have two independent stations sampling the same area of the EAS. Through these stations, the time

⁵ b should be given by the GPS clock accuracy (about 10 ns) and the FADC trace resolution ($25/\sqrt{12}$ ns), that is $b \sim 12$ ns.

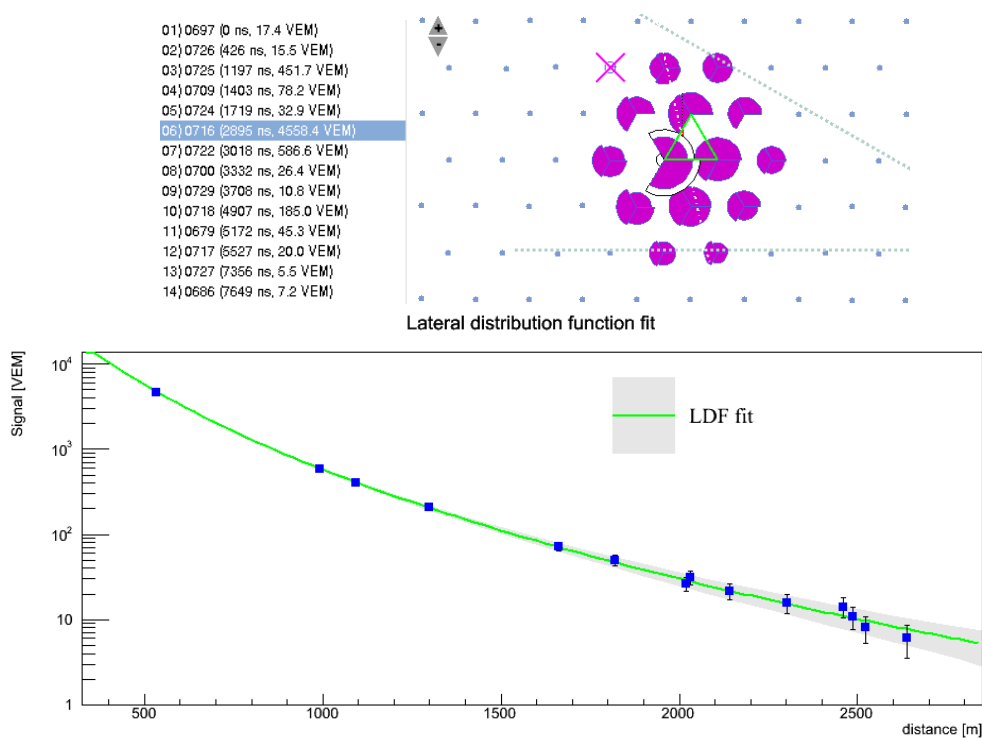


Figure 4.5: Reconstruction of event 201102605906 as shown in the Herald Event Display. On the top left, the list of stations with their measured signal, converted in VEM, is given. On the top right, the topology of the selected stations is shown. Each tank is represented by a circle divided into three parts, one for each PMT, whose size is proportional to the detected signal. 4 of the tanks have only two active PMTs. Stations marked with a cross are the one that participated to the trigger but were later rejected as random stations (see section 4.3). The bottom graph shows the LDF fit superimposed to the station signals as a function of their distances to the reconstructed shower core.

variance model has been validated. Then, given the time variance model, we can calculate the angular resolution on an event-by-event basis out of the minimization procedure. In figure 4.6 the angular resolution as a function of zenith angle for events with different station multiplicities is plotted. As we will see in the next chapters, for the high energy events it is always better than 1° , since the multiplicity for these events is always larger than 6.

4.4.4 Uncertainty on S_{1000}

The reconstruction accuracy of $S(1000)$ $\sigma_{S(1000)}$ is composed of three contributions: a statistical uncertainty due to the finite number of particles intercepted by a given SD station and the limited dynamic range of the signal detection; a systematic uncertainty due to assumptions on the shape of the lateral distribution function; and an uncertainty due to shower-to-shower fluctuations. The latter are found through Monte Carlo simulations: their magnitude is $\sim 10\%$, independently from energy.

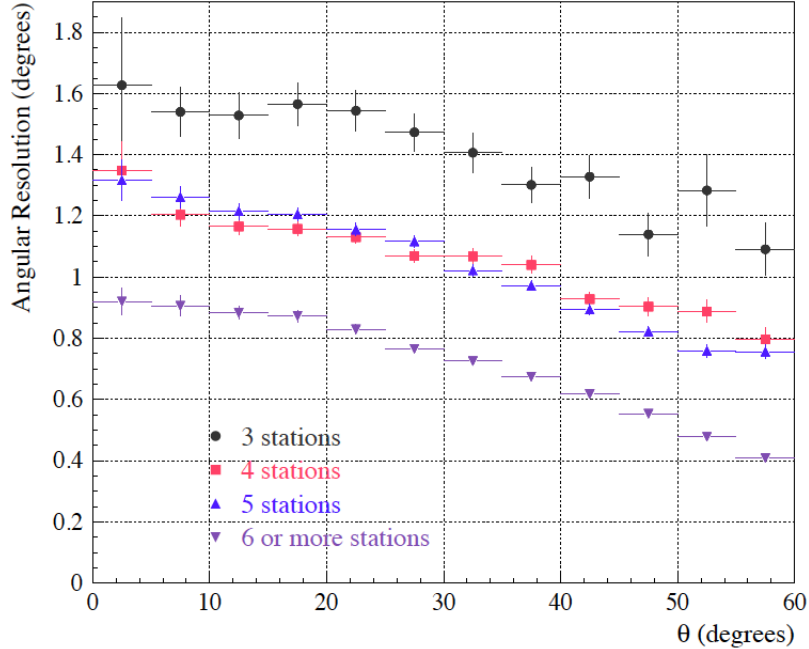


Figure 4.6: Angular resolution as a function of zenith angle θ for events above 3 EeV with different station multiplicities. From [140].

The statistical error in S_{1000} comes from the fitting one as a consequence of the uncertainty of the observed signals due to the finite dimension of the detectors. The average statistical errors vs S_{1000} are shown (black points) in Figure 4.7. For events with no saturated stations, it ranges between $\sim 10\%$ and $\sim 2\%$ depending on S_{1000} . For events with saturated stations, due to the larger uncertainty in the measured signal, the uncertainty in S_{1000} is systematically 5% higher. The second source of error in S_{1000} is a systematic effect arising from the lack of knowledge of the true LDF shape for a particular event. The choice of $r_{opt} = 1000$ m minimizes this effect, but r_{opt} changes from event-to-event. If r_{opt} of a given event is close to 1000 m, the fitted S_{1000} is independent of the value of β assumed. When it is not, fluctuations in the event by event β give rise to a systematic error. The contribution of the variations of r_{opt} to the uncertainty of S_{1000} was obtained by reconstructing the same event with different β values. The resulting systematic uncertainties are shown in figure 4.7: they amount to about 5% (10%) for events without (with) saturated stations.

4.5 From S_{1000} to primary energy

The final part of the reconstruction procedure is the energy calibration of the estimator S_{1000} . The hybrid feature of the Pierre Auger Observatory comes in hand here since the fluorescence detector has access to the primary calorimetric energy (an overview on how the energy is computed in the FD reconstruction was sketched in section 3.3.2). The

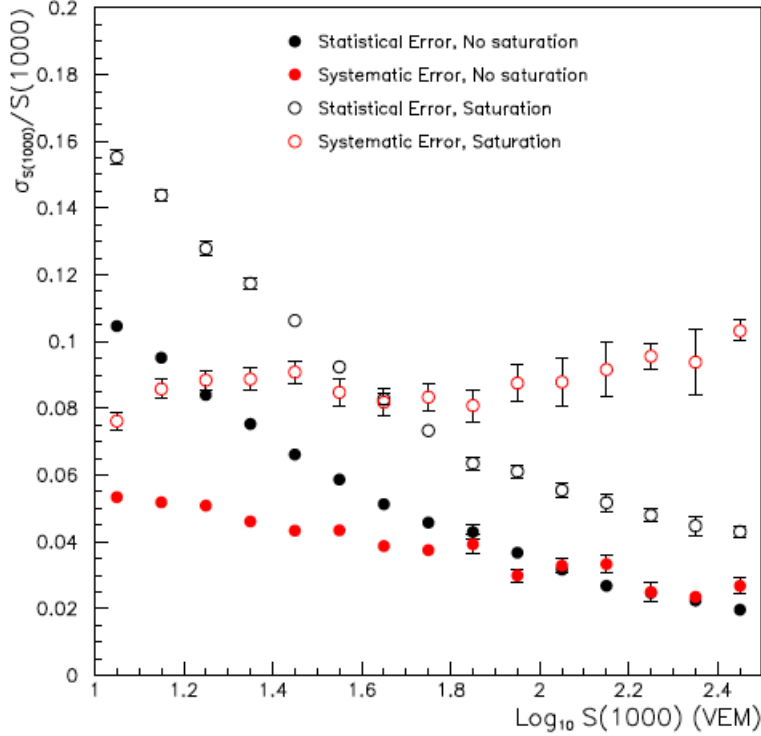


Figure 4.7: The average statistical (black) and systematic (red) errors as a function of $\log(S_{1000})$ for events with or without a saturated station. From [141].

cross-calibration between the SD energy estimator and FD calorimetric energy using a hybrid events sample is the method that allows us to derive a conversion formula that can be used for all events measured by the SD.

4.5.1 From S_{1000} to S_{38} : the constant intensity cut

For a given energy, the value of S_{1000} decreases with the zenith angle θ , due to the atmospheric attenuation of the shower particles and geometrical effects⁶. For this reason, values of S_{1000} for showers coming from different θ have to be corrected and reported to a common reference angle, here chosen to be the median zenith angle $\theta = 38^\circ$, before being correlated with the FD energy. A method to estimate the θ dependence of S_{1000} without using Monte Carlo simulations is the so-called Constant Intensity Cut (CIC). Given the intensity defined as the number of events with S_{1000} greater than a given value per unit $\sin^2 \theta$:

$$I(S_{1000}^*, \theta) = \frac{dN(S_{1000} > S_{1000}^*)}{d \sin^2 \theta} \quad (4.5)$$

⁶Indeed, the S_{1000} value depends also on the weather conditions. This effect is quite small and will not be considered in the following analysis. However, a proper set of weather corrections were implemented to correct this effect for particularly delicate analysis such as the search for large scale anisotropies.

assuming that the flux of cosmic rays is isotropic and given that the acceptance of the detector is known as a function of zenith angle, we can state that equal intensities at two different zenith angles correspond to the same lower energy cut. So, if we plot I as a function of S_{1000}^* for different angles, we obtain the curves shown in figure 4.8. A horizontal line in this graph, a line of constant intensity, corresponds then to a line of constant energy. If we choose a fixed value of the intensity, we can find the attenuation curve values for S_{1000} as a function of θ . Different attenuation curves can be obtained, depending on the chosen value of the constant intensity. In Auger, the intensity used is the one corresponding to an energy of around 10 EeV, well above 3 EeV, thus guaranteeing the full efficiency, at the same time allowing us to use an adequate number of events. The attenuation curve obtained is then fitted by a 3rd-degree polynomial function:

$$S_{1000} = S_{38}(a + b(\cos^2 \theta - \cos^2 38^\circ) + c(\cos^2 \theta - \cos^2 38^\circ)^2 + d(\cos^2 \theta - \cos^2 38^\circ)^3) \quad (4.6)$$

The a , b , c , d parameters for the attenuation curve in use in the Herald reconstruction are:

$$a = 1.00 \pm 0.01 \quad b = 1.08 \pm 0.05 \quad c = -1.26 \pm 0.09 \quad d = -0.69 \pm 0.47 \quad (4.7)$$

For each event, given S_{1000} and θ , we can then obtain the S_{38} value, that is to say the S_{1000} value that this shower would have if it had come from $\theta = 38^\circ$.

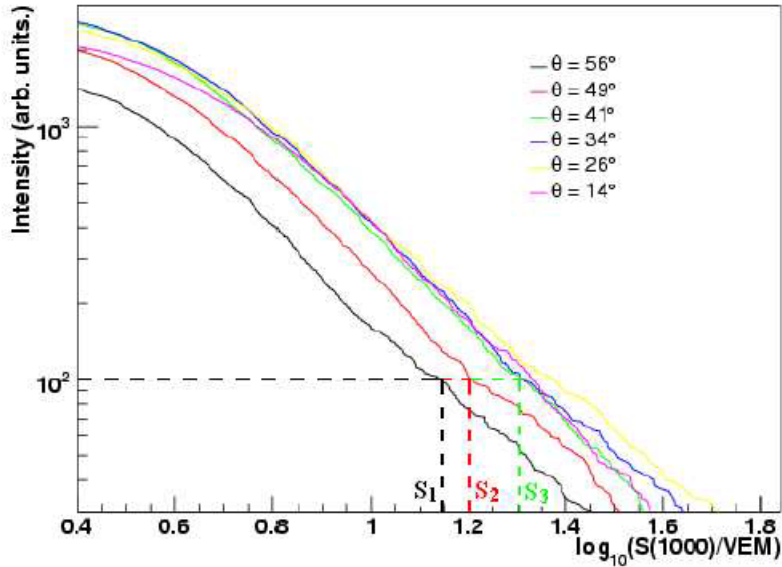


Figure 4.8: Intensity as a function of S_{1000} . For clarity, only 6 zenith bins are shown in this plot. Courtesy of Claudio di Giulio.

4.5.2 From S_{38} to energy: the “golden” hybrid events

To minimize the possible systematics on the energy determination, only carefully selected events are used for the calibration fit.

First of all, only events that have independently triggered the SD array and FD telescopes (so-called golden hybrid events) are used for the energy calibration of SD data. Then, on the SD side, a large bias could come from events below SD full efficiency (3 EeV) due to upward fluctuations of S_{1000} . These events are rejected through a maximum-likelihood method that takes into account the evolution of uncertainties with energy, as well as event migrations due to the finite resolution of the SD [142]. Moreover, only 6T5 events with the “hottest” station closer than 750 m to the shower core are selected.

Severe selection criteria are applied to the FD-side of these hybrid events. The reduced χ^2 of the longitudinal fit to the Gaisser-Hillas function must be less than 2.5. The χ^2 of a linear fit to the profile has to exceed the one from the Gaisser-Hillas fit by at least 4. The depth of shower maximum, X_{\max} , must be within the field of view of the telescopes and the fraction of the signal detected by the FD and attributed to Čerenkov light must be less than 50%. The limited field of view of the FD and the requirement to observe the EAS maximum can introduce a dependency in the event selection on the primary mass. To avoid this effect, a fiducial cut on the slant-depth range observed by the telescopes is applied, in order to ensure that the field of view is large enough to observe all plausible values of X_{\max} . It is furthermore required for the uncertainties on E_{FD} and on X_{\max} to be less than 20% and 40 g/cm², respectively. Further requirements are made on the atmospheric conditions at the event time: the vertical aerosol optical-depth profile, measured using laser shots generated by the central laser facility, is required to be smaller than 0.1 in the hour when the event took place. Furthermore, the cloud fraction in the field of view, measured from the information provided by the LIDAR systems of the observatory, is required to be less than 25%. Applying those cuts, 1469 events survived at the time of the energy calibration used in this thesis work, using data up to December 2012 [109].

4.5.3 From S_{38} to energy

After the selection of the hybrid events, the calibration can finally be performed. The relation between S_{38} and E_{FD} is well described by a single power-law function:

$$E_{FD} = A S_{38}^B$$

This formula is fitted to the data and the parameters A and B are extracted (see figure 4.9). The values for these parameters in the current Herald reconstruction are: $A = 0.1942 \pm 0.0046$ EeV $B = 1.005 \pm 0.0068$. The fit takes into account the resolutions of both the SD energy estimator (as previously described) and of the FD energy. The resolution of the latter include (as seen in Chapter 3) contributions from the knowledge of the atmosphere, from the FD calibration and from the FD reconstruction, including the invisible energy [103].

The resolution in the SD energy is inferred from the distribution of the ratio E_{SD}/E_{FD} , being $\sim 16\%$ for energies above 3 EeV and $\sim 12\%$ for energies above 10 EeV. The obtained power-law function is finally used for all SD events to extract the SD energy E_{SD} . The total systematic uncertainty on the SD energy comes from that on the FD energy scale: it is about 14% [103]. It includes contributions from the absolute fluorescence yield (3.6%), calibration of the fluorescence telescopes (9.9%), the invisible energy correction (3%), systematics in the reconstruction method used to calculate the shower longitudinal profile (6.5%), and atmospheric effects (6%).

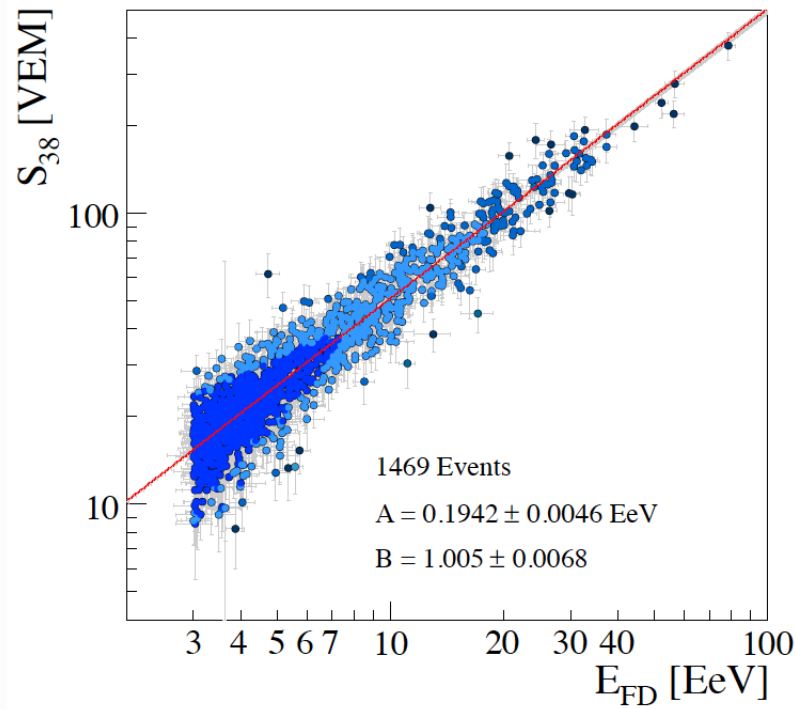


Figure 4.9: The calibration plot: S_{38} and E_{FD} for selected events are plotted. The line shows the fit that is used as official calibration to obtain the energy of all SD events. Courtesy of Claudio di Giulio.

In the Offline software

In the Offline software, the calibration procedure is exactly the same. However, since the upstream reconstruction is different, the S_{1000} values that enter the calibration and the selected calibration events might be slightly different. For this reason, the resulting values for A and B are different, namely $A = 0.190 \pm 0.005 EeV$ $B = 1.025 \pm 0.007$ based on the fit of 1475 events in the energy calibration currently in use in the Offline software[109].

In the HAS reconstruction

The principle of calibration for horizontal showers is the same as for the vertical ones, with N_{19} instead of S_{38} used as energy estimator, as pointed out in the previous box. However, due to the different characteristics of the HAS showers, the quality cuts, though with the same principles, impose different requests. The fit is then made with the function $N_{19} = A(E_{FD}/10^{19} EeV)^B$ and the latest results for the parameters are $A = 1.723 \pm 0.023$ and $B = 0.984 \pm 0.020$. The full list of imposed cuts and further details can be found in [132].

In this chapter we have described in detail how the reconstruction of an event is performed in the Herald software. In the next chapter we will use this software to reconstruct the highest energy events. The study of those events will help us also to ameliorate some steps of the reconstruction described in this chapter.

Chapter 5

Auger dataset of extremely high energy cosmic rays

*È tutto quel che posso
Quello che ho.*

Gianmaria Testa *Lampo* (2000)

The purpose of this thesis work is to study the arrival directions of the cosmic rays events at the highest energies, to extract information on their possible sources performing anisotropy tests on all the events and on a proton-enriched subsample. The first part of this study, that will be developed in the next chapters, is the selection of events that will be used in the analyses. In this chapter we describe the selection of such dataset, that consists of 454 vertical ($0^\circ \leq \theta \leq 60^\circ$) events with $E \geq 40$ EeV¹, that is the largest dataset of extremely high energy cosmic rays. In the following, we will refer to this as the 454-Evts dataset. The energy threshold of 40 EeV corresponds to the beginning of the flux suppression that is observed in the CR energy spectrum (see chapter 1.2.2).

In chapter 4, the principles of how an event that is recorded by the surface detector is reconstructed were outlined. An important part was dedicated to the selection of PMTs, stations and events. During this work, by inspecting the highest energy events, we were able to validate and improve such selections, ameliorating the event reconstruction and increasing the data sample. These new selections of PMTs, stations and events will be described in sections 5.1.1, 5.1.2 and 5.1.3, respectively. As the selected events and their reconstruction were settled, we performed further tests on the dataset, checking the

¹We note that this dataset is the one used in the recent publication [143] of the Pierre Auger Collaboration, to which this thesis work has contributed.

stability of the reconstruction over time (section 5.2) as well as that of the energy and angular resolution of the events (section 5.3). The work described in this chapter focuses mainly on vertical events reconstructed with the Herald software. A cross-check with the other reconstruction software, the `Offline`, was performed and will be discussed in section 5.3.2. Inclined ($60^\circ < \theta \leq 80^\circ$) events included in the dataset will be briefly discussed, together with a summary of the characteristics of the final dataset, in section 5.4. This final dataset, that includes both vertical and inclined events, will be dubbed in the following as 602-Evts dataset.

Herald software versions

In this chapter, different versions of the Herald software are named. As a reference, we show here a summary of the main differences between the versions mentioned in the text.

Version	Energy Scale	comments
v4r8	pre-2013	No PMT quality cuts (see par. 5.1.1) No non-triggered stations selection (see par. 5.1.2)
v5r0	2013	Daily quality cuts (see par. 5.1.1) No non-triggered stations selection (see par. 5.1.2)
v5r2	2013	PMT quality cuts (see par. 5.1.1) Non-triggered stations selection (see par. 5.1.2)

5.1 Studies on the selection criteria

5.1.1 PMT selection

Each station in the Auger SD array is equipped with 3 PMTs. After years in the field, for many technical reasons, some PMTs may not be working properly (we commonly refer to them as “bad” PMTs). A long-term commissioning study has been made before the beginning of this thesis work, in order to find a way to identify bad PMTs based on the diagnostic parameters that the station sends continuously to the CDAS, as mentioned in section 4.3. Details can be found in the internal note [144]: we report in table 5.1, a summary of the parameters and relative cuts that were proposed to identify bad PMTs. Those cuts (in the following “daily cuts”) are based on the RMS of the baselines distributions on a daily basis in the dynode (high gain) and anode (low gain) channels, or on the dynode/anode ratio. This ratio represents the amplification factor between the two channels and in principle is set to 32. The daily cuts may affect the whole PMT or the anode channel only (i.e., the PMT is turned off only if the dynode channel is saturated). Also, the average value over one day is used for the computation of some diagnostic parameters.

At the beginning of this work these cuts were not applied before the reconstruction. In particular, all Herald versions up to v4r8 did not implement them. In this work, we have exploited the 123 events with energy $E \geq 54$ EeV recorded from 1st January 2004 to 31st December 2009² to validate the use of the listed PMTs cuts. We have examined

²This was the dataset used for anisotropy studies in [138]. The energy given is obtained in the Herald v4r8 reconstruction.

Hierarchy	Parameter	Is bad if :	Kind of cut
1	Dynode baseline	RMS > 1.75 ADC channel	Full PMT
2	Dynode/anode ratio	MEAN < 25 ADC channel	Anode only
3	Anode baseline	RMS > 1.75 ADC channel	Anode only
4	Dynode/anode ratio	RMS > 5 ADC channel	Anode only
5	Dynode/anode ratio	MEAN > 45 ADC channel	Anode only

Table 5.1: List of the PMTs parameters that are checked in order to decide whether the PMT is to be removed or not. These cuts are applied on the daily average or RMS of each parameter. From [144].

the events one by one, station by station, PMT by PMT ³. PMTs that were identified by eye as problematic were manually removed from the reconstruction to see how the final reconstructed parameters (arrival direction, core position and energy) changed after their removal.

Out of the 123 considered events, only 10 had a malfunctioning PMT. In 7 of these, the differences in the reconstructed energy and arrival direction due to the removal of the malfunctioning PMTs were well within the respective uncertainties. However, in three cases, the presence of a “bad” PMT affected the reconstruction beyond uncertainties. In figure 5.1 an exemplary case is presented: in this event the PMT 3 of the station with the third largest signal (station ID: 1387) was malfunctioning. Indeed, the PMT was correctly identified as “bad” by the selection proposed, since it had too low a dynode/anode ratio value in the considered day (~ 7 , which is much lower than the proposed cut at 25, as indicated in row 2 of table 5.1. See top left panel for its daily behavior). The signal in the considered PMT is about 6 times the mean of the other two in the same station (see top right panel). When removing this PMT, the reconstruction changes: there is a variation in the core position estimate by 200 m, to be compared with the nominal uncertainty of ~ 50 m. The LDF shape changes (see bottom panels: the top one shows the LDF with the bad PMT and the bottom one without) and consequently S_{1000} changes too. This causes a variation well beyond the uncertainties in the reconstructed energy ($95.8 \rightarrow 65.6$ EeV, i.e. $\sim 30\%$, compared to the nominal 12% uncertainty) and in the reconstructed arrival direction (R.A.: $57.4^\circ \rightarrow 56.4^\circ$, Dec: $67.2^\circ \rightarrow -67.9^\circ$, i.e., 0.8° angular difference, compared to $\sim 0.5^\circ$ uncertainty)⁴. The PMT selection described here has been incorporated into the two official Auger reconstruction algorithms.

We also found some malfunctioning PMTs that were not identified from the cuts listed in table 5.1. Although only 2 events were affected in the 123 considered events and it had nearly no effect on the reconstruction, we implemented a more stringent identification of bad PMTs, based on new cuts, that are summarized in table 5.2: we reduced the allowed range for the dynode/anode ratio and added a new cut on the VEM peak value and on the dynode baseline. The previous cuts were based on the daily mean of the parameters.

³To quickly identify potential problematic PMTs, we modified the Herald Event Display in order to show the signal of each PMT instead of their average. “Bad” PMTs often show much lower or higher signal than the others and consequently such PMTs can strongly affect the reconstruction.

⁴The blue triangle at ~ 1000 m is another issue: it is a misidentified silent station. This will be discussed in paragraph 5.1.2.

That could cause PMTs with a “good” daily average of the diagnostic parameters not to be removed even if, at the very moment of the event, they were indeed malfunctioning. For this reason, we decided that the new cuts would be applied on the instantaneous values that the local station sends to the CDAS together with the information of the triggered event.

Parameter	Is bad if :	Kind of cut
VEM Peak	$< 20 > 90$	Full PMT
Dynode/Anode	$< 25 > 40$	Full PMT
Dynode Baseline	> 12 ADC channel	Full PMT

Table 5.2: List of the new PMTs parameters introduced by this work. They are checked on an event-by-event basis.

During this study we found only one case of mis-classification, i.e., properly working PMTs that were mis-classified as bad ones: in event 201210900870 all three PMTs of the station with the largest signal were removed by the daily quality cuts, although the parameters were fine at the moment of the event. While we believe that in general it is safe and conservative to always remove a PMT that has a “bad” daily average of some diagnostic parameters, we have proposed to release this criterium in the case where all the 3 PMTs of a station are removed at the same time, that would result in the removal of the whole station. This kind of situation is very rare and it is probably due to a problem of the transmission of the monitoring data more than to the PMTs themselves.

Both the new instantaneous cuts and the check for 3-PMTs-based rejection were implemented in the Herald software, version v5r2, the one used for the 454-Evts dataset. In this dataset, among the 454 events, in which nearly 20000 PMTs were involved, 58 have at least one PMT removed from the reconstruction by either the new or the old cuts. In particular, 49 PMTs were removed using the daily cuts and 14 using the instantaneous ones.

5.1.2 Station selection: misidentified non-triggered stations

As mentioned in chapter 4, some of the stations that are active at the moment of the event, but that report no trigger (non-triggered stations), are used in the LDF fit, namely the ones that are within 5000 m from the shower axis.

The use of non-triggered stations in the reconstruction implies that we need to know if a station was active and properly working even if it did not participate directly to the event. However, while a triggered station is certainly active, to assess if a non-triggered one is properly working, we need to access further informations. In particular, we look at the rate of T2 triggers (see paragraph 4.1) that is sent to the CDAS every minute. The rate is stored in so-called “T2 Files”. By looking at these files around the time of the event, we can assess if each non-triggered station was indeed active at that time. At the beginning of this work, no checks of this kind were done on the non-triggered stations used in the reconstruction. We have exploited the high energy events in the 454-Evts dataset also to verify the need of this check, by looking for misidentified non-triggered stations and, if present, by studying their impact on the reconstruction.

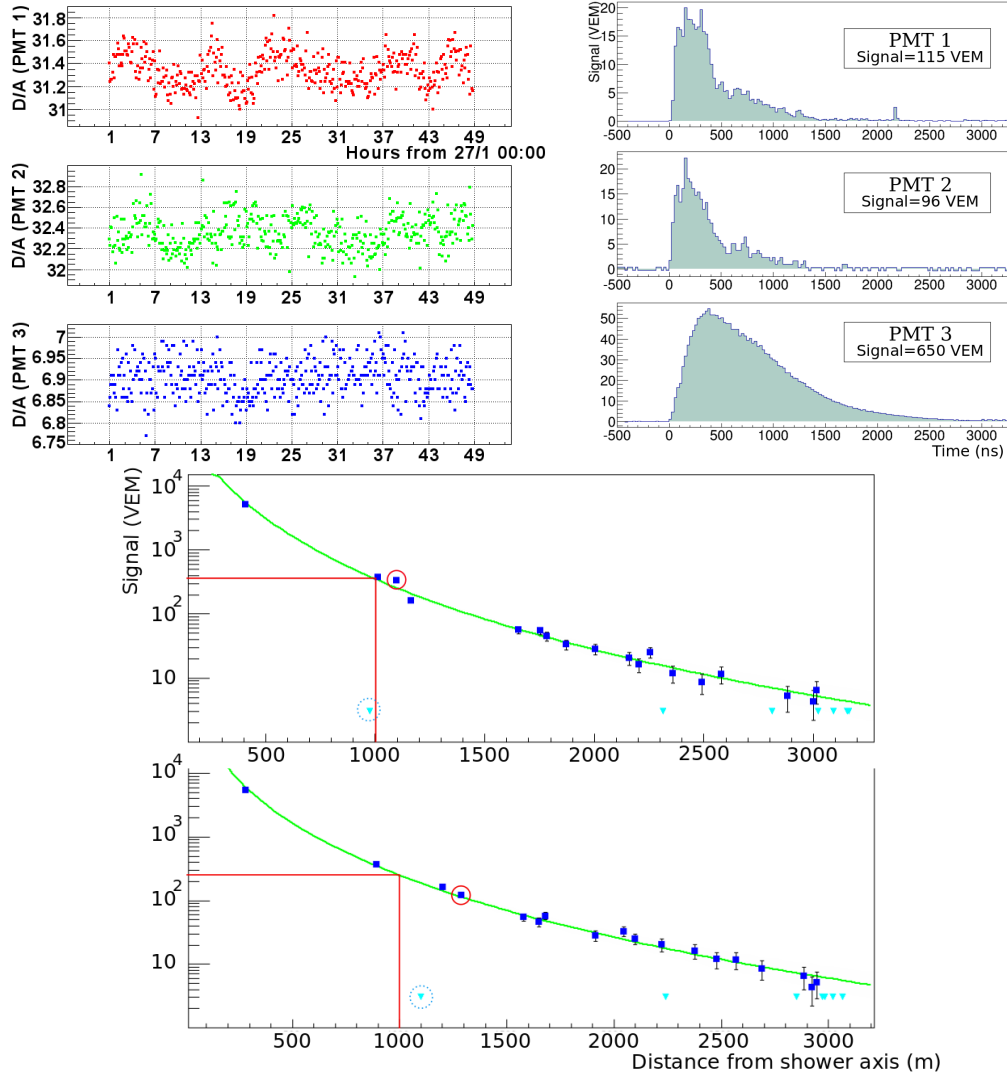


Figure 5.1: Event 201302701934. In the top, left panel, the D/A values at the time of the event for the three PMTs are reported. In the top, right panel, the PMT traces of the third hottest station show a clear problem in PMT3. The two bottom plots show the LDF fit before (top) and after (bottom) the bad PMT has been removed. See text for details.

We found indeed a few events where a station close to the core was classified as non-triggered, namely 4 events with a non-triggered station closer than 1500 m from the core. An example is shown in figure 5.1. In this event, besides the already mentioned PMT problems, a station at a distance ~ 1000 m is classified as non-triggered, whereas the signal expected from the LDF fit at 1000 m is ~ 100 VEM. We can see in figure 5.2, where the single station trigger efficiency is shown as a function of the detected signal⁵, that

⁵The trigger probability is obtained from the measured events as the ratio $N_T(S)/N_{ON}(S)$ where N_T is the number of triggered station and N_{ON} the total number of active stations with an expected signal S according to the LDF. Also, the trigger probability has been evaluated by means of “triplets”. Triplets are

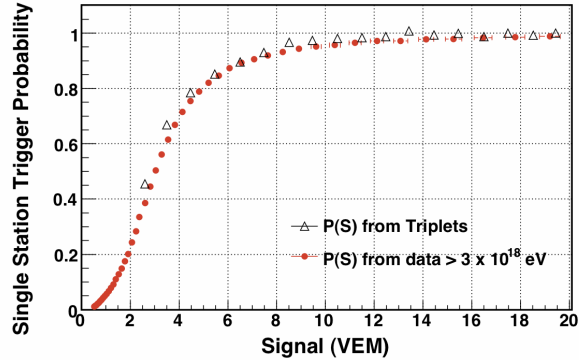


Figure 5.2: The single-station trigger probability, $P(S)$, as a function of the signal, measured with data (red dots) and with triplet stations (white triangles). From [96].

for signal higher than ~ 10 VEM the trigger probability is $> 95\%$. This means that the station under discussion cannot be active and not triggered as it should have measured a signal of ~ 100 VEM. The T2 files of the station show indeed that it was inactive at the moment of the event. By removing this station, the reconstructed energy changes from 66.7 to 70.4 EeV.

We thus found that the implementation of the check of the activity of non-triggered stations is important for a correct reconstruction of the events. Based on this study, this check, called CheckT2, was implemented in the official reconstruction in the Herald software v5r2⁶. We then re-examined the data set event by event to verify if the use of this selection was enough to identify and reject all inactive stations. We found that all the misidentified non-triggered stations were found, with only one exception: event 201133304681. In this event, even after the application of the CheckT2 selection, station 203 was reported as active but not triggered at ~ 1000 m from the shower axis, where the signal predicted by the LDF fit was well above 100 VEM. Such event is a particular case. It was recorded during a period of FD data taking. Station 203 is connected to the Central Laser Facility, one of the instruments used to monitor the atmosphere during the nights of FD operation (see section 3.3). The station was indeed active at the time of the event, but its communication system was probably overloaded due to the laser firing, so that it failed to transmit to the CDAS the recorded EAS signal. A correction for this problem is being developed and should be introduced in the next Herald release. For the time being, we manually forced the removal of that station in event 201133304681 that is part of the 454-Evts dataset. The energy of this event changes from 42.6 to 51.5 EeV after the correction, i.e., $\sim 20\%$.

positions in the array where additional stations (“slaves”) are added at 11 m from each other. For each slave station, the trigger probability versus recorded signal S is derived from the ratio between the number of events where both slaves have triggered and the number of events where the third one has triggered.

⁶The CheckT2 correction was introduced also in the latest version of the Offline software (v10r0).

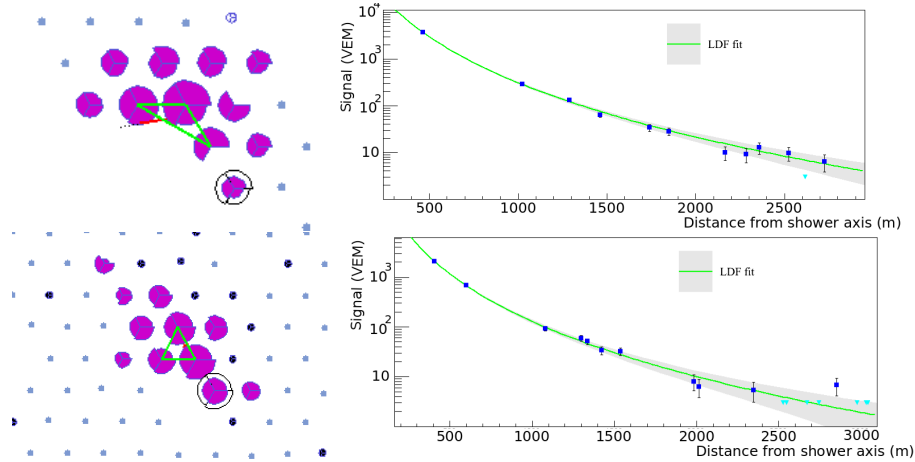


Figure 5.3: Two examples of 4T5 events. Event 201130905216 (top) fell at the border of the array, while event 201017300252 (bottom) fell in an area where a few stations were inactive.

5.1.3 Event selection

As shown in section 4.3.2, the last step of the event selection is the so-called “T5” fiducial criterion. In that section we described how the most stringent condition, 6T5, used in most of Auger analysis, is relaxed to a less stringent one, 5T5-POS, for anisotropy studies at the highest energies. We described as well a furtherly relaxed condition, 4T5-POS2. A first test of this criterion, to look for a possible degradation of the reconstruction accuracy, was performed on an early sample of 6T5 events (data taken between 2004 and 2010) with energies above 40 EeV. Active stations were removed around the one with the highest signal, so to simulate 4T5 events. The modified events were then reconstructed and both the arrival direction and the energy were compared with the original ones. The RMS of the distributions of the differences in angle and energies were found to be 0.3° and $\approx 3\%$, respectively, i.e., well within the statistical uncertainties.

By the time this work started, the 4T5-POS2 was officially adopted by the Collaboration in view of the new release of the UHECR data set. In the 454-Evts dataset, 58 events are 4T5-POS2, while 396 are 6T5 or 5T5-POS. The less strict condition yields an increase in the number of events of about 14%. In this work we verified the accuracy of the geometry and energy reconstruction for 4T5-POS2 events (also called in the following Relaxed T5 events, RT5). In figures 5.9 and 5.10 we show the angular and S_{1000} resolutions for events in the 454-Evts dataset with different T5 conditions. We can see that, though slightly deteriorated with respect to 6T5 events, the angular resolution for 4T5-POS2 ones is still below 0.5° and the S_{1000} resolution is better than 6%. A more complete discussion of the average resolutions will be provided in sections 5.3.1 and 5.3.2. To conclude, we show, as an illustration, two examples of 4T5 events in figure 5.3: either events at the border of array (top panel) or close to inactive stations (bottom panel) are recovered.

5.2 Detector aging and its impact on event reconstruction

The Pierre Auger Observatory started taking data in 2004, with ~ 150 stations, some of which were already active in the field since few years earlier. The deployment of the array went on until 2008, meaning that the age of the detectors is not uniform and some stations are much older than others. In figure 5.4 the age of each station is shown on the map of the array. The age of a station is an important characteristic because we have observed in fact a slow and regular variation of the response of the detectors over time. In particular, one of the features that have been observed to vary versus time is the Decay Time (DT, or width) of the elementary signal, i.e., the one induced by a single high energy muon used for the calibration of the SD stations (see section 4.2). The energy deposit of a muon is faster than the time width of one FADC bin but, due to reflection and absorption inside the detector, it is not actually recorded as a single-bin signal, but rather as a peak with an exponentially decreasing tail, characterized by a decay time. The spread of the elementary signal can be also quantified by looking at the integral of the signal (named Area) and the height of its peak (named Peak). If the decay time is higher, the signal pulse is wider and the ratio Area/Peak (A/P) is larger. The decrease in time of this quantity was already observed few years after the completion of the array, as we can see in figure 5.5 [145]. In this plot we can see the A/P evolution for the three PMTs in a given station (in this case, station 437) as a function of years, fitted by the product of a sinusoidal and a decreasing function to take into account the seasonal variations (introduced by the electronics, that is sensitive to the temperature), and the global decrease respectively. The origin of the global decrease, be it due to changes in water, in the reflectivity of the liners or in the electronics, is still under investigation.

Another observed variation is the decrease of the rate of events as a function of time below the full-efficiency energy. This is shown in figure 5.8, blue dots. In this work we studied the possible relationship between the decrease of the signal decay time and the loss of events. Moreover, we checked the effect of the detector aging on the reconstruction of the high energy events.

5.2.1 Shape of the signal and effect on the trigger

To obtain the value of the decay time, we fit the so-called shape histogram. It is produced by the station averaging random muon signals around 1 VEM that are measured during 60 seconds before every event. The histogram gives the shape of the elementary signal with a good statistical precision. The value of the decay time can then be extracted from an exponential fit of the shape histogram: common values are between 40 and 80 ns. In Figure 5.6 it is possible to see the distribution of the decay time of all the PMTs that were active in two days, well separated in time: 1st June 2007 (red, dashed histogram) and 10th January 2014 (blue, filled histogram). The double-peaked feature of these histograms is because of the presence, since the first years of operations of the detector, of two populations of PMTs. The reason for this original separation has not been discovered yet and is still under study by the Collaboration. A clear shift of the two distributions is visible: although the shape is the same, the more recent one is shifted by ~ 10 ns to the left. This means that, with the aging of the detector, values of DT for the same PMT get lower; so, as time passes, signals become narrower and narrower. This, in time, has an effect on the trigger.

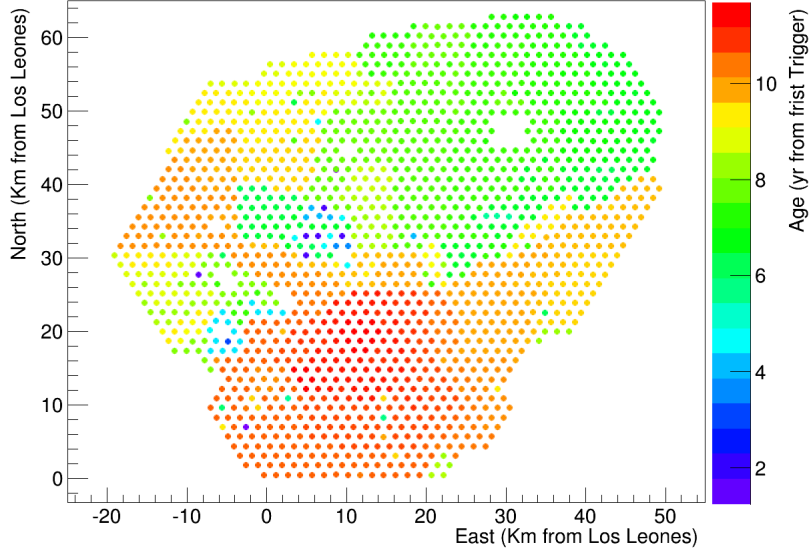


Figure 5.4: Map of the array indicating for each station the time since its first T2 trigger (in years), as of the end of 2014. The origin is at the FD Los Leones site.

As we have seen in section 4.1, two station triggers are used in Auger: Threshold (Th) and Time over Threshold (ToT). Since the Th trigger is related to the peak value of the signals, the changes in the width of the elementary signals does not affect it. On the other hand, the ToT trigger is very sensitive to changes in the shape of the elementary signals, due to its extension in time. For this work, we focused on studying the effects of the DT aging on the ToT and their consequences on the event rate. To study the effect of the DT aging on the event rate we converted all traces via a linear filter shape transformation (see [146] for further details), so that their DT is artificially set to 50 ns. This value is chosen as it is nearly the lowest one for the currently active PMTs, as one can see in Figure 5.6, blue histogram. In figure 5.7, an example of the effects of this transformation is shown. In the top plot, one can see the traces from a simulated shower measured in a station with decay time 75 ns (solid, blue line) and 50 ns (dashed, red line): signals are narrower in the 50 ns case. In the bottom plot, the results of the conversion of the 75 ns trace into a 50 ns one are shown: we can see that the transformed trace nicely reproduces the original 50 ns one (dashed red line in the top plot).

We included the transformation in a modified version of the Herald v5r0 software. The decay time is extracted for each PMT from a fit of the shape histogram. The PMT trace is then transformed from the measured DT to 50 ns and the resulting trace is stored in a dedicated variable, while the original trace is not overwritten. If the shape histogram is not available or the fit fails, no transformation is applied.

We then applied the ToT trigger to the modified traces and we check if the station still passes the trigger. If it does not, and the station is not triggered otherwise, the station is removed and classified as non-triggered. We finally reconstruct the event with the new list of stations. The signal used in the reconstruction is the one coming from the original

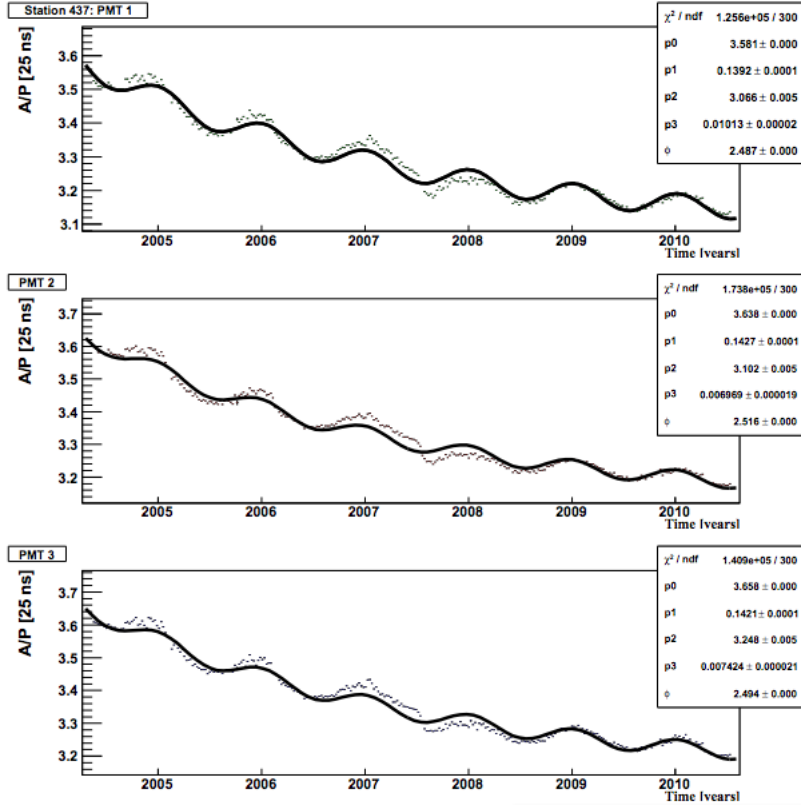


Figure 5.5: A/P as a function of time for station 437. The dots are the average of the A/P over 7 day and the continuous line is a fit with the convolution of a sinusoidal to take account of the seasonal variation and an exponential decay. From [145]

trace, not from the transformed one. We performed two different studies: the first one, on all the events recorded by SD, consisted in counting how many events were lost (i.e., were no longer T4 events) after the transformation. The second one, performed only on the highest energy events, consisted in studying the effects on the reconstruction when some of the stations participating in the event are lost after the transformation.

5.2.2 Effects on the event rate

We studied the effects of the DT transformation on all events recorded by the SD from 1st January 2004 to 18th June 2014. We checked for each event if the T4 trigger condition (see paragraph 4.3.1) held after the removal of the stations that did not pass the ToT trigger after the transformation. In table 5.3 (left side) we show the number of events recorded in each year and how many of them were no longer T4 after the DT transformation. We notice a relevant loss of events, and a strong time dependence. The data loss is lower for the last years because the average DT values are closer to 50 ns. We measured also the effect of the DT transformation for the events above full trigger efficiency, i.e., for energies above 1 EeV (right side of the table): these events are only marginally affected ($\lesssim 1\%$ of the events lost).

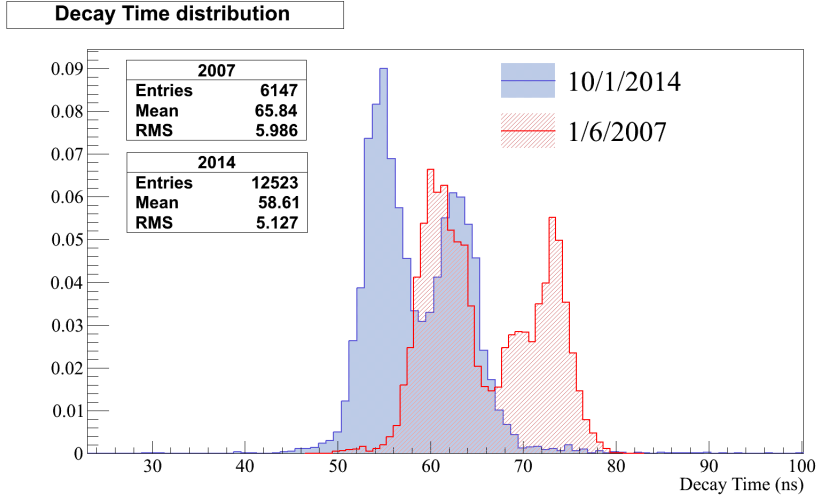


Figure 5.6: Distributions of the elementary signal decay time for all stations participating in events the 1st of June 2007 (red, dashed) and the 10th of October 2014 (blue, filled).

To better visualize the effects of the DT transformation on the event rate, we show in figure 5.8 the daily rate of all events as a function of time (blue dots) and compare it to the rate of events that survived after the transformation (red dots). While in the original data a decreasing trend is visible in the last years, after the correction this effect is removed.

5.2.3 Effects on the highest energy dataset

As shown above, the aging of the detector does not have an impact on the number of events above the full efficiency of SD (3 EeV). However, we studied the aging of the detector via the same procedure described before, also in the 454-Evts dataset. In this case, as expected, all events still passed the triggers and event selection criteria after the DT transformation, but in some of the events one or more stations resulted non-triggered. Among the 454 events considered, 246 lost at least 1 station after the transformation, with one event losing 6 stations out the original 27. The average multiplicity went from 14.6 to 13.83. All events still had a multiplicity greater than 6, thus yielding an angular resolution better than 1° (see section 4.4.3). Note that, as the ToT trigger is dominant in case of small spread signals, the lost stations in this dataset resulted to be those far away from the core.

We checked how the simulated loss of these stations impacted on the reconstruction: in case of a significant deterioration, the presence of an aging-related systematic should be taken into account. To this aim, we performed a new reconstruction of the events, classifying as non-triggered the stations that after the DT convolution did not trigger anymore. Considering the 246 events that lost at least one station after the convolution, the average change in the reconstructed energy is less than 0.05% with an RMS of less than 2%. Only two events show a change in energy of more than 10%. In those cases, the effect is larger due to the particular configuration of the footprint that enhances the

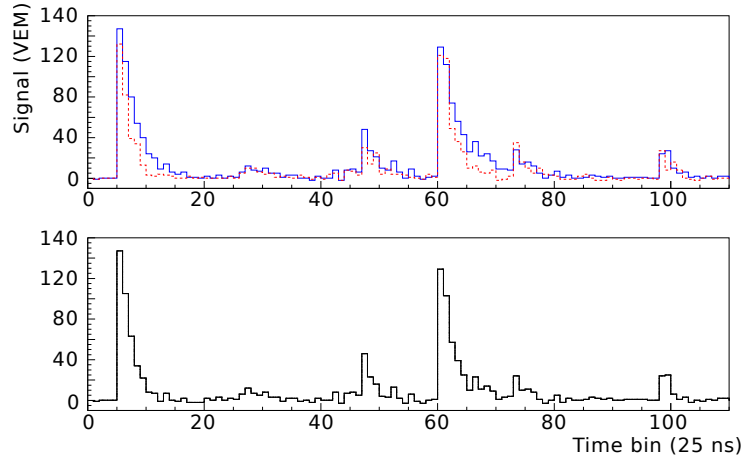


Figure 5.7: Top: PMT traces from a simulated shower corresponding to an elementary signal decay time 75 ns (solid, blue) and 50 ns (dashed, red). Bottom: result of the conversion of the 75 ns trace into a 50 ns one.

importance of stations far from the core. Overall, we can conclude that the detector aging has a negligible effect on the highest energies. Further tests of the long-term stability of the reconstruction of this dataset will be presented in the following.

5.3 Energy and angular resolutions

We have described in the previous chapter how the uncertainties on energy and arrival direction are estimated. In this section we will study the resolutions for the 454-Evts dataset, including checks versus different energy calibrations and event reconstructions.

5.3.1 Angular resolution

The angular resolution is an important parameter for the search for point sources and small scale anisotropies that is part of this thesis work. We have shown in section 4.4.3, in figure 4.6, that in general for events with more than 6 stations the angular resolution is better than 1° . Here we have studied specifically the angular resolution for the events in the 454-Evts dataset, in particular to verify if it is degraded by the relaxed selection of events used. In figure 5.9, top panel, we show the angular resolution obtained from the “global” fit (see chapter 4) errors as a function of zenith angle for 6T5 , 5T5-POS only and 4T5-POS2 only events in the 454-Evts dataset, and, as a reference, for all 6T5 events with multiplicity higher than 6 stations. We can see that the resolution for all the different selections of events in the 454-Evts dataset is better than 0.5° . Their average angular resolution is $0.24^\circ \pm 0.08^\circ$. With relaxed selection criteria, the angular resolution is indeed degraded, but only slightly. The average angular resolution for 6T5 events is $0.23^\circ \pm 0.08^\circ$,⁷ that raises to $0.25^\circ \pm 0.08^\circ$ for only 5T5 POS events and to $0.29^\circ \pm 0.10^\circ$

⁷Here the standard deviation is quoted as uncertainty rather than the standard error, to show the spread of the distribution.

Year	All			$E > 3 \text{ EeV}$		
	events	not T4	%	events	not T4	%
2004	67182	29042	43.23	1898	19	1.00
2005	148502	67806	45.66	3913	42	1.07
2006	210013	94675	45.08	5610	61	1.09
2007	310908	140430	45.17	8474	85	1.00
2008	379974	167460	44.07	10478	115	1.10
2009	198869	85684	43.09	5619	59	1.05
2010	352548	135436	38.42	10983	87	0.79
2011	396273	137247	34.63	12933	87	0.67
2012	382283	121386	31.75	12781	73	0.57
2013	384523	113237	29.45	12574	57	0.45
2014*	192210	53833	28.01	6020	23	0.38
total	3023285	1146236	37.91	91286	708	0.78

Table 5.3: Data loss after DT transformation from 1st January 2004 to 18th June 2014 (* 2014 is thus only partial). 6T5 events with $\theta < 60^\circ$ are considered.

for only the 4T5-POS2 ones. We note that such values represent only the statistical uncertainties coming from the fit. Systematic uncertainties, based on the study of hybrid events, are around 0.5° [140].

Given the extension in time of the dataset, we have also studied the stability of the angular resolution as a function of time. We divided the time period in 10 equal-exposure ones and evaluated the average angular resolution in each period. This is shown in figure 5.9, bottom panel: we can see that the average angular resolution is constant over all periods within the uncertainties. We performed a fit to a constant value (shown as a red line) and we found a value of $0.23^\circ \pm 0.02^\circ$, compatible with the average resolution of the whole dataset.

5.3.2 Energy resolution

Resolution of the energy estimator

As described in chapter 4, the energy estimate is based on the signal at 1000 m, S_{1000} , obtained via the LDF fit. The resolution of S_{1000} is the first source of uncertainty on the energy estimate. We show in figure 5.10, left, the relative uncertainty on S_{1000} , dS_{1000}/S_{1000} , obtained by the fit, as a function of energy for 6T5, 5T5 POS and 4T5-POS2 events. We can see that the uncertainty on S_{1000} decreases with energy for all events. The values for 6T5 and 5T5 POS events are comparable and always below 4%, while the uncertainty is a bit larger for 4T5-POS2 events, going up to $\sim 5\%$. In particular, the average uncertainty for all 6T5 events is $3.3 \pm 1.2\%$, for 5T5 POS events it is $3.5 \pm 1.2\%$ and for 4T5-POS2 events it is $4.9 \pm 1.9\%$. We checked also the stability in time of the uncertainties on S_{1000} as we did for the angular resolution, using the same equal-exposure time periods. As one can see from figure 5.10, right, the stability of the uncertainties is remarkable. The line is the fit to a constant value, C, that gives the average uncertainty for all the dataset; with $C=3.5 \pm 0.1\%$.

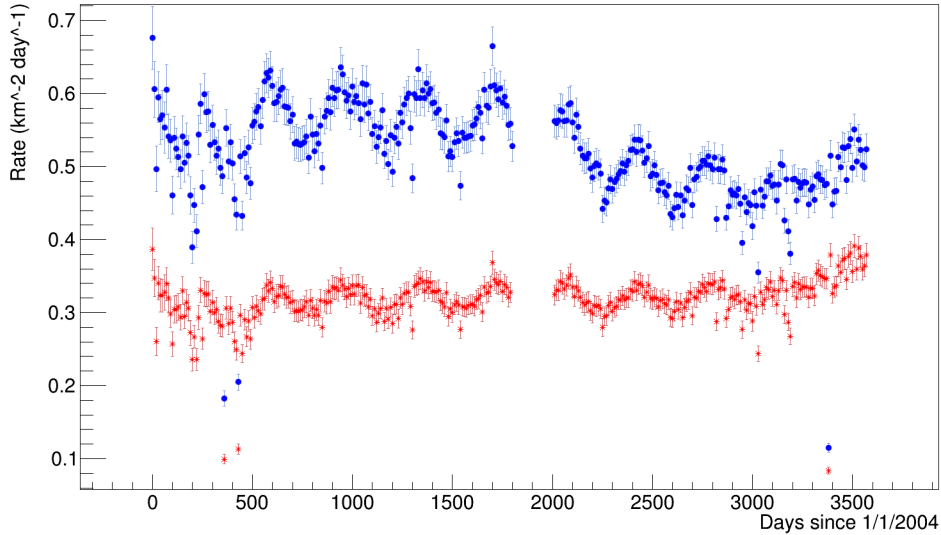


Figure 5.8: Daily event rate in 10 days binning as a function of the number of days since the start of the data taking before (blue dots) and after (red dots) the DT transformation (see text).

Stability of the energy estimation versus different reconstruction and energy scales

As seen in section 4.5, the energy calibration for SD events is done comparing the signal at 1000 m from the shower axis, corrected for the zenith dependence through the constant intensity cut, with the calorimetric energy measurements of FD in a selected sample of hybrid events. As more data are available, the calibration is updated periodically. This may change the energy estimates not only due the increasing statistics, but also because of updates in the FD reconstruction and calibration as well as in the SD reconstruction, like those that were introduced after the improved selections described in the previous sections.

An important change in the energy estimation was made in 2013 when the Auger energy scale was updated. The FD energy determination was updated taking advantage of more precise measurements of the fluorescence yield, of a better estimate of the invisible energy, and of a deeper understanding of the detector with consequently improved event reconstruction⁸. The overall effect of the introduction of the updated energy scale was the reduction of the systematic uncertainty on the energy, from 22% to 14%.

We have studied the changes in the energy estimation due to new energy scale for all events with $E > 3$ EeV. We show in figure 5.11, left, the differences in energy between the latest Herald reconstruction (v5r2, 2013 energy scale) and version v4r8, (pre-2013 energy scale) as a function of energy. Error bars represent the RMS. We can see that the energy differences are of the order of 14% at the lowest energies, becoming negligible

⁸For further information about the energy scale update see [103].

above ~ 20 EeV ($10^{19.3}$ eV). The 14% difference at the lowest energies is compatible with the systematic uncertainties related to the FD energy scale (22%). The distribution of the differences for the events in the 454-Evts dataset are shown in the right panel: the average difference is of the order of 2% with an RMS of 8%. Only 7 events show a difference of more than 22%: 4 of them have in fact problems of malfunctioning PMTs and 3 have misidentified non-triggered stations close to the core. These problems were not corrected in version v4r8 yet, i.e. these differences are not due to the change in the energy scale. These events account also for the large spread in the penultimate bin in the left panel of figure 5.11.

As we already mentioned in chapter 4, two independent reconstruction algorithms are used in Auger: Offline and Herald. Although based on the same strategy, the two reconstructions differ in a few key points (outlined in the boxes in chapter 4). Also, the energy estimators reconstructed by the two codes are calibrated independently. We have checked the differences in the estimated energy between the latest Herald (v5r2) and Offline reconstruction (v10r0), both using the 2013 energy scale. We show them in figure 5.12, left, as a function of the energy for all events with $E_{Herald} \geq 3$ EeV. The differences in the energy estimation are less than 5% on average over the whole energy range, i.e., well within the uncertainties of the energy scale. There is a slight trend in energy” at the lowest end the Herald estimation is higher, and lower at the highest energies. In the last energy bin the difference is 9% because of the few events that are in the tails of the distribution of the events in the 454-Evts dataset, shown in the right panel. The average energy difference is 1%, with only 6 events showing a difference in energy higher than the quoted statistical uncertainty of 12%. We carefully inspected all these events. In 4 cases the differences were due to the signal saturation recovery in the station with the highest signal, as this procedure is different in the two reconstructions. In one case, the difference was due to the different LDF fits, namely the slope of the LDF is fitted in the Herald, while it is fixed in the Offline (see chapter 4 for details). Finally, in one particular case, the Herald correctly removed a station that reported an incorrect position in the array, while the Offline did not. Excluding this last event, in the other 5 cases 3 of them showed a reduced χ^2 close to 1 for the reconstruction fit in for both reconstructions. In one case the Offline showed a $\chi^2 \sim 1$ while the Herald value was ~ 3.5 and in the remaining case the Herald χ^2 was close to 1 while the Offline one was ~ 8 . These differences are still under study by the collaboration. However, we remind here that these particular events represents only the $\sim 1\%$ of the 454-Evts dataset.

Stability of energy estimation versus time

The last test concerns the stability of the energy estimator as a function of time, given the duration of the data set. Due to the energy resolution indeed, low energy events can spill-in above the imposed energy threshold especially due to the steepness of the energy spectrum. This effect becomes larger as the uncertainty on the energy estimate grows, and this means that, if the energy resolution is degrading with time, we might observe an increasing flux. Under the reasonable assumption that the flux of cosmic rays and the trigger efficiency of the detector at the highest energies are constant over time, we have studied the behavior of the flux versus time to constrain the stability of the energy resolution.

The behavior of the observed flux in every 6 months of data taking for events with energy larger than 10 EeV (6T5 only) and 40 EeV (the 454-Evts dataset) is shown in figure 5.13 (top and bottom plot, respectively). We can see in both cases a slight increase. We estimated its magnitude using a linear fit. The result of the fit gives a slope $A=0.0032 \pm 0.0012$ per year for events above 10 EeV, and $A=0.00024 \pm 0.00015$ per year for events above 40 EeV. This corresponds to a drift of $1\% \pm 0.7\%$ per year for the events above 10 EeV and $3.3 \pm 2.9\%$ per year for the events above 40 EeV. We can see that both estimates are actually compatible with a constant rate of events.

Although overall the drift is not significant, we have used it to constrain the long-term stability of the energy resolution. Given the spectrum as measured by Auger (see section 3.8), the resulting maximum drift of the energy resolution in 10 years estimated to be a few percents in both the considered energy ranges. The possible effects are thus small and consistent with the 5% systematic uncertainty associated to the stability of the energy scale [103].

5.4 The final dataset and consistency checks

We described in the previous sections the studies and checks that we performed on the 454-Evts dataset of events with energy above 40 EeV and zenith angle $< 60^\circ$ recorded by Auger between the 1st January 2004 and 30th March, 2014. To complete the dataset we include here inclined events, with zenith angles $60^\circ < \theta < 80^\circ$. We sketched how the reconstruction is performed for horizontal showers in dedicated boxes in chapter 4. We selected inclined data fulfilling the 5T5 condition. Inclined showers can have a particularly elongated footprint. For this reason, the use of the 4T5 condition is not recommended in this case because very large portion of the shower footprint could be missing (e.g., if the event falls at the border of the array). The inclined events were selected in the same time period and with the same energy threshold as for the vertical ones. The PMT selection described in section 5.1.1 was applied to inclined events too. In total 150 events were selected this way.

Two of the 150 inclined events were already part of the vertical dataset. These two events, 200813701781 and 201218311437, were selected in both datasets because of small differences in the zenith angle estimations: for event 200813701781, the zenith angle estimated via the vertical reconstruction was $\theta_{Herald} = 59.7^\circ$, while with the horizontal one it was $\theta_{Inclined} = 60.3^\circ$, i.e., 0.6° difference. For event 201218311437, $\theta_{Herald} = 59.8^\circ$ and $\theta_{Inclined} = 60.1^\circ$, with a difference of 0.3° . Regarding the reconstructed energy, for the event 200813701781 $E_{Herald} = 46.7$ EeV, while $E_{Inclined} = 42.6$ EeV, i.e., 9.9% energy difference. For the event 201218311437 $E_{Herald} = 61.8$ EeV, while $E_{Inclined} = 54.5$ EeV, i.e., 12.6% energy difference. Considering that the two energy reconstruction are based on different methods and calibrated with different hybrid events, the results are in a good agreement and anyway compatible within statistical and systematical uncertainties. In the following we use for those events the energy and arrival direction as reconstructed by the vertical reconstruction.

The final dataset consists of 602 events, more than 3 times larger than the one used in previous studies [138]. Illustrative plots for the vertical and horizontal datasets are shown in figure 5.14, top and bottom panels, respectively, showing the distributions of zenith

angle, azimuth angle, multiplicity and energy (from top left to bottom left, clockwise). In table 5.4 the main characteristics of this dataset are presented; besides the number of events in each sub-dataset (row 1) we show the average station multiplicity (row 2): as expected, it is much higher for horizontal events than for vertical ones, and a little lower for 4T5-POS2 events than for 6T5 and 5T5-POS ones, as expected since these events are missing stations in the footprint. In row 3 we show the average energy which is constant within errors for all sub-datasets. In particular, the close values for all vertical events (52.7 EeV) and horizontal ones (51.2 EeV) indicates that the two energy scales are on average consistent. The average values for the statistical uncertainties in the energy estimators and angular resolutions (row 4 and 5) summarize the values shown in the previous paragraphs. The uncertainty on N_{19} is slightly higher than the one on S_{1000} , while, on the other hand, the arrival direction reconstructed for horizontal shower have a better resolution, this being due to the higher station multiplicity of those events.

The last row in the table shows the exposure for each of the sub-datasets. We verified the consistency between the numbers of events and exposures. Considering Poissonian uncertainties on the event numbers and 3% uncertainties on the exposure estimations, for vertical events the ratio of the number of events $N(6T5)/N(5T5POS) = 6.76 \pm 1.03$ is consistent with the ratio of exposures $\text{Exp}(6T5)/\text{Exp}(5T5POS) = 6.98 \pm 0.30$, and the ratio $N(4T5POS2)/N(5T5POS) = 1.14 \pm 0.22$ is consistent with $\text{Exp}(4T5POS2)/\text{Exp}(5T5POS) = 1.03 \pm 0.04$.

Finally, we performed a check of the compatibility of the vertical and inclined datasets finding that $N(\text{Vertical})/N(\text{inclined}) = 3.07 \pm 0.29$ and $\text{Exp}(\text{Vertical})/\text{Exp}(\text{inclined}) = 3.52 \pm 0.15$ are consistent account taken of the Poissonian fluctuations. On the other hand, the 14% difference between these ratios could also result from a $\sim 4\%$ mismatch between the vertical and inclined energy calibrations, which is compatible with the uncorrelated systematic uncertainties on the energy scale.

The final plot of this chapter (figure 5.15) shows the distribution of the position of the 602 events in the sky, in equatorial coordinates. As one can see, the inclined events extend the FOV up to $\delta = +45^\circ$, while previous datasets covered the sky only up to $\delta = +25^\circ$. Consideration on the events arrival directions, their anisotropy and their possible correlation with astrophysical objects will be the topic of the next chapter.

Value	# All vertical	6T5	5T5-POS	4T5-POS2	Horizontal	Total
Number of events	454	345	51	58	148	602
Average multiplicity (RMS)	14.6 ± 4.3	15.1 ± 4.1	14.1 ± 5.6	12.6 ± 3.5	33.4 ± 9.2	-
Average energy (RMS)	52.7 ± 13.3	52.1 ± 12.9	55.5 ± 15.8	53.4 ± 13.5	51.2 ± 10.6	-
$\langle dS_{1000}/S_{1000} \rangle$ ($\langle dN_{19}/N_{19} \rangle$)	$3.5 \pm 1.4\%$	$3.3 \pm 1.2\%$	$3.5 \pm 1.2\%$	$4.9 \pm 1.9\%$	4.5%	-
Angular resolution ⁹	$0.24^\circ \pm 0.08^\circ$	$0.23^\circ \pm 0.08^\circ$	$0.25^\circ \pm 0.08^\circ$	$0.29^\circ \pm 0.10^\circ$	0.14°	0.22°
Exposure (km ² sr yr) ($\pm 3\%$)	51753	40063	5740	5950	14699	66452

Table 5.4: Key informations on the final dataset. See text for details.

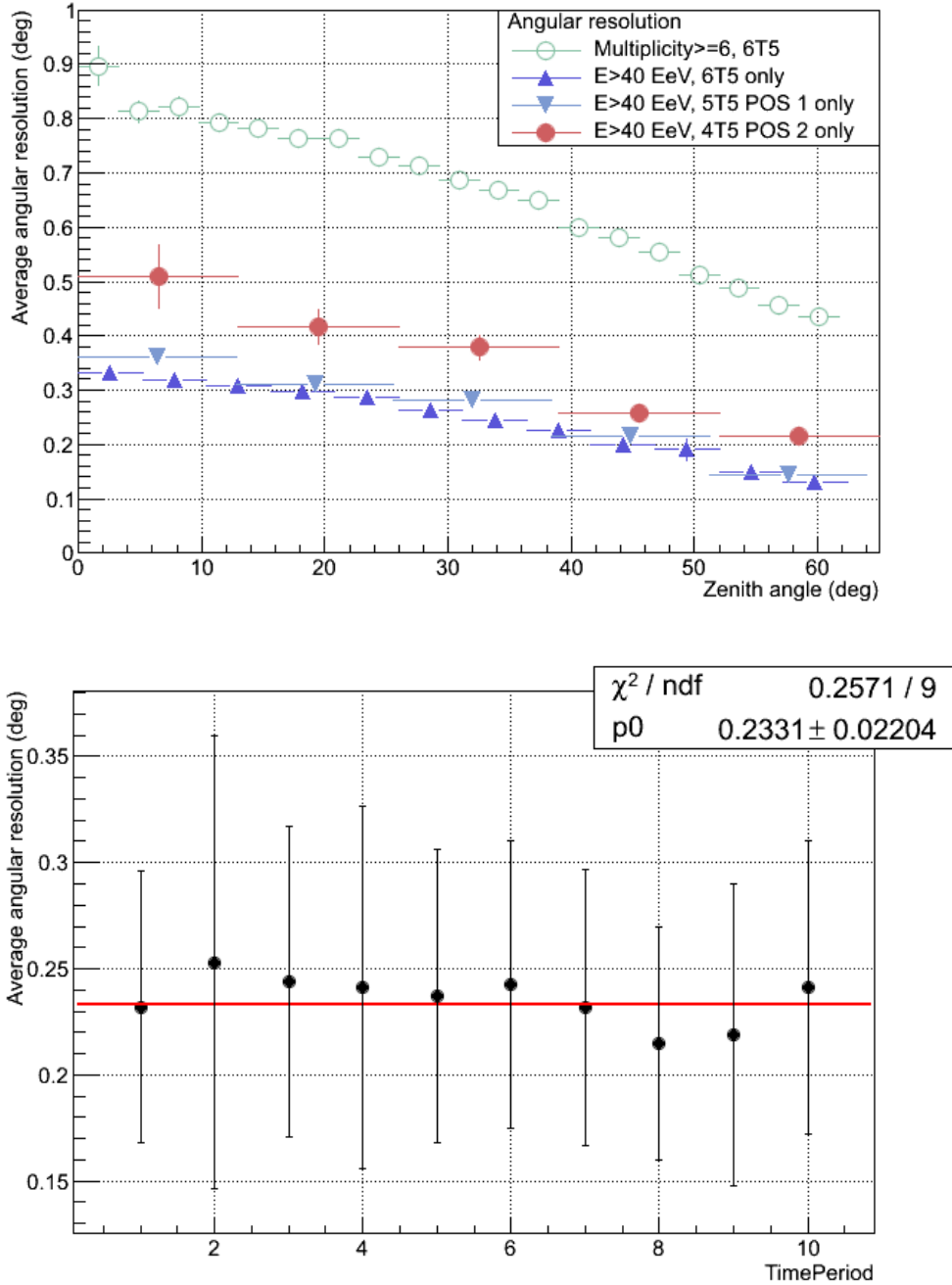


Figure 5.9: Top: average angular resolution estimated from the reconstruction fit as a function of zenith angle: 6T5 events with ≥ 6 stations (green empty dots) compared to 6T5 (blue upward triangles), 5T5 POS (light blue downward triangles) and 4T5 or POS 2 (red full dots) events in the 454-Evts dataset.

Bottom: average angular resolution estimated from the reconstruction fit as a function of time. The time, ranging from 1/1/2004 to 31/3/2014, is divided in 10 equal-exposure periods. The red line is a constant fit, with $C=0.23$.

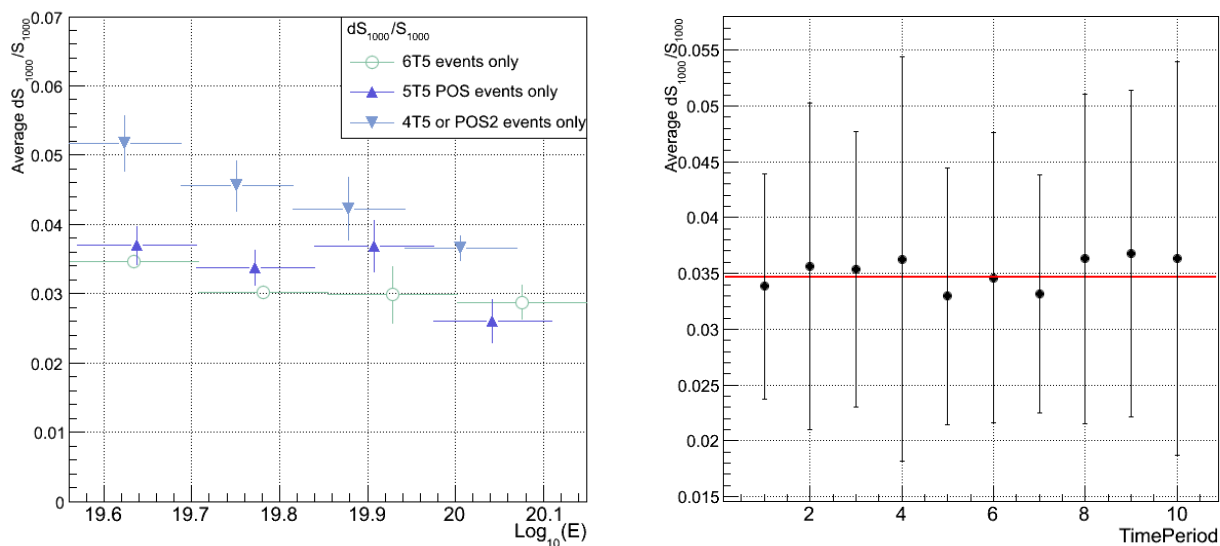


Figure 5.10: Left: average uncertainty on S_{1000} estimated from the reconstruction fit as a function of the energy: 6T5 events (green empty dots) compared to 5T5-POS (blue upward triangles) and 4T5-POS2 (light blue downward triangles) events in the 454-Evts dataset. Right: average uncertainty on S_{1000} estimated from the reconstruction fit as a function of time. The time, ranging from 1/1/2004 to 31/3/2014, is divided in 10 equal-exposure periods. The red line is a constant fit with $C=3.5\%$.

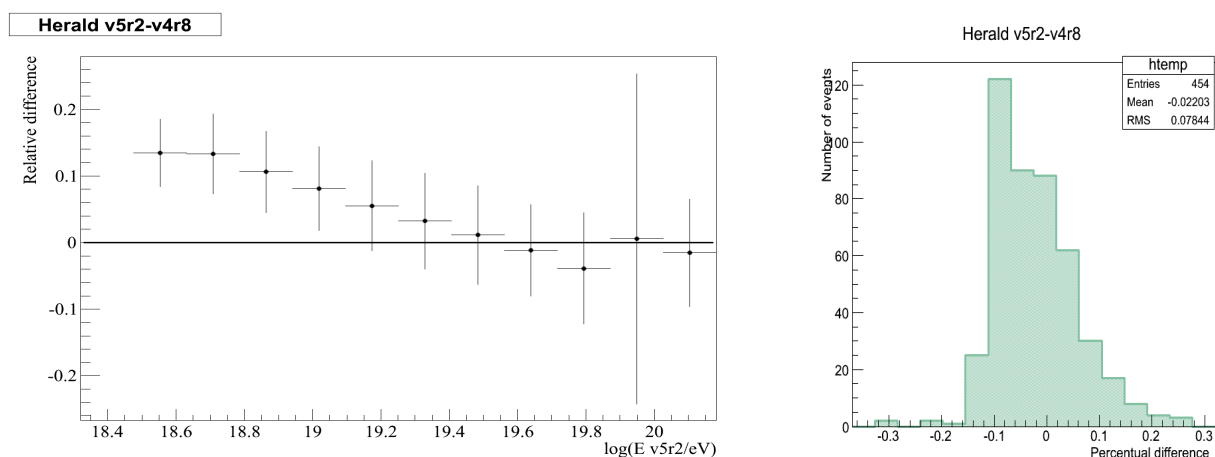


Figure 5.11: Left: relative energy difference Herald v5r2-v4r8 for events with $E(\text{v5r2}) > 3 \text{ EeV}$ up to March 2013, as a function of $E(\text{v5r2})$. The error bars show the spread of every distribution. Right: the same difference for the events in the 454Evts dataset.

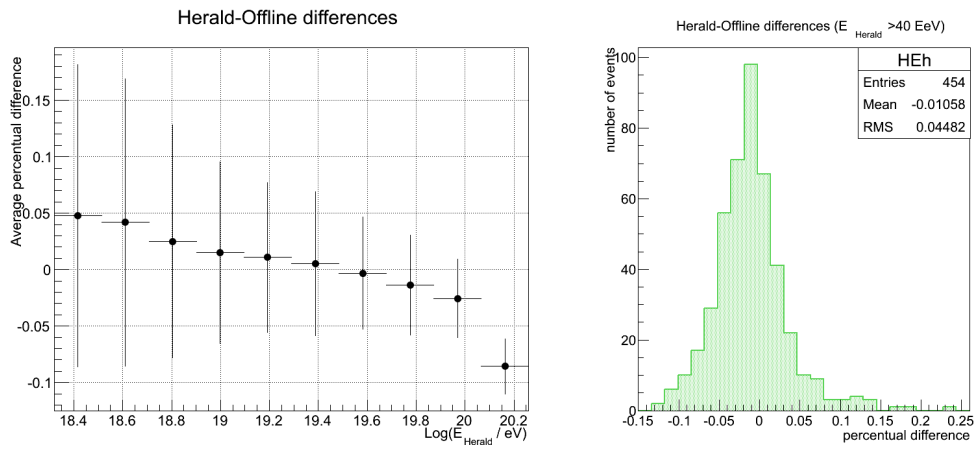


Figure 5.12: Left: relative energy difference Herald software v5r2- $\overline{\text{Offline}}$ v10r0 for events with $E_{\text{Herald}} \geq 3 \text{ EeV}$ up to the end of 2014, as a function of Herald energy. The error bars show the spread of the distribution in each energy bin. Right: relative energy difference for the 454-Evts dataset.

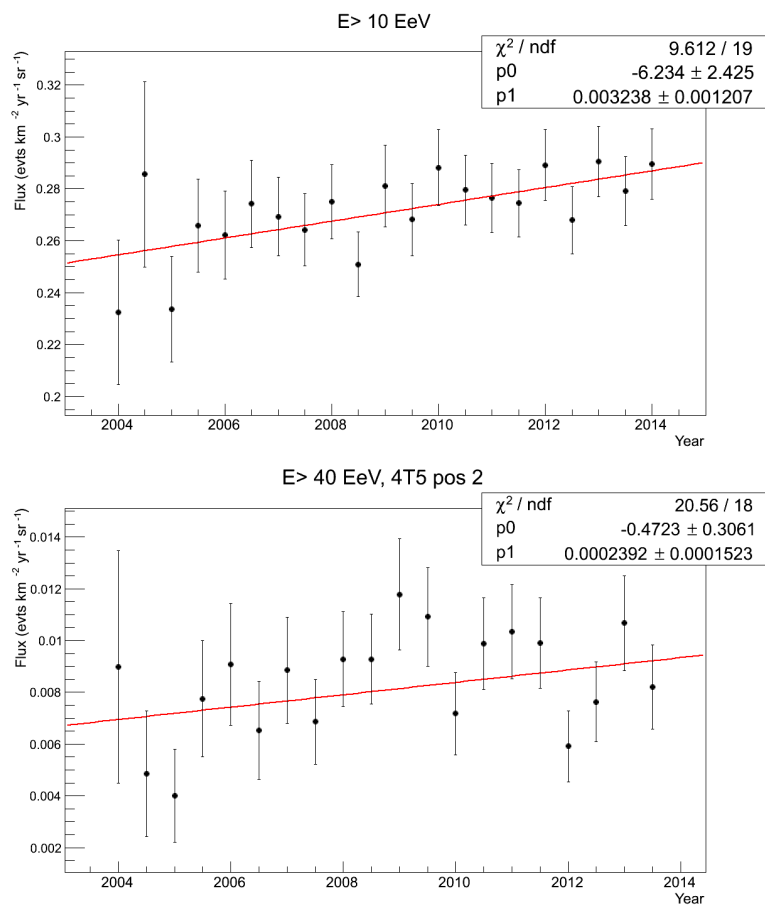


Figure 5.13: Flux of 6T5 events as a function of time for 2 different energy thresholds: 10 (top) and 40 EeV (bottom). The red line is the result of a linear fit.

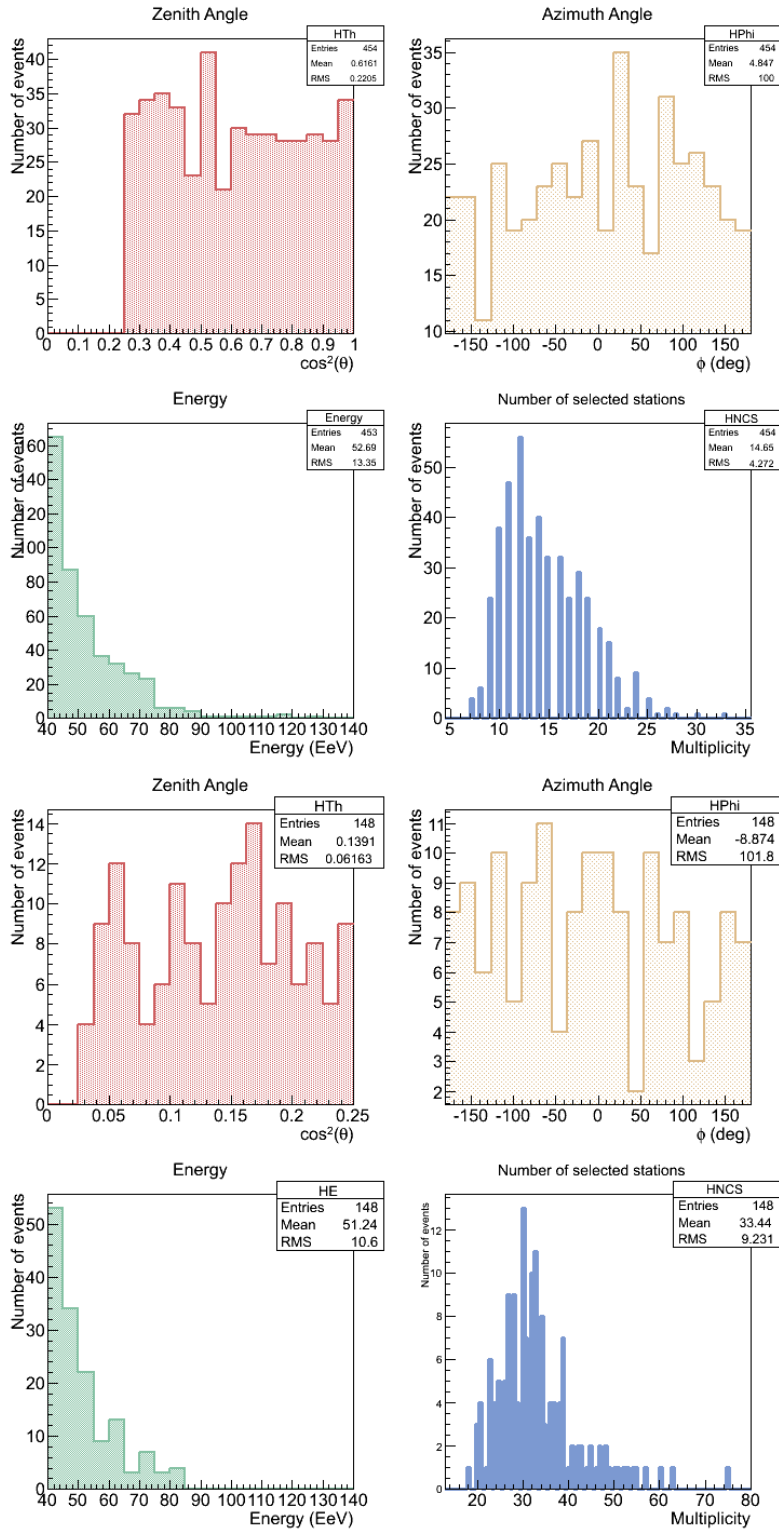


Figure 5.14: Zenith (red), azimuth (yellow), energy (green), and number of stations (blue) distributions for the 454 vertical events (top 4 panels) and for the for the 148 horizontal events (bottom 4 panels).

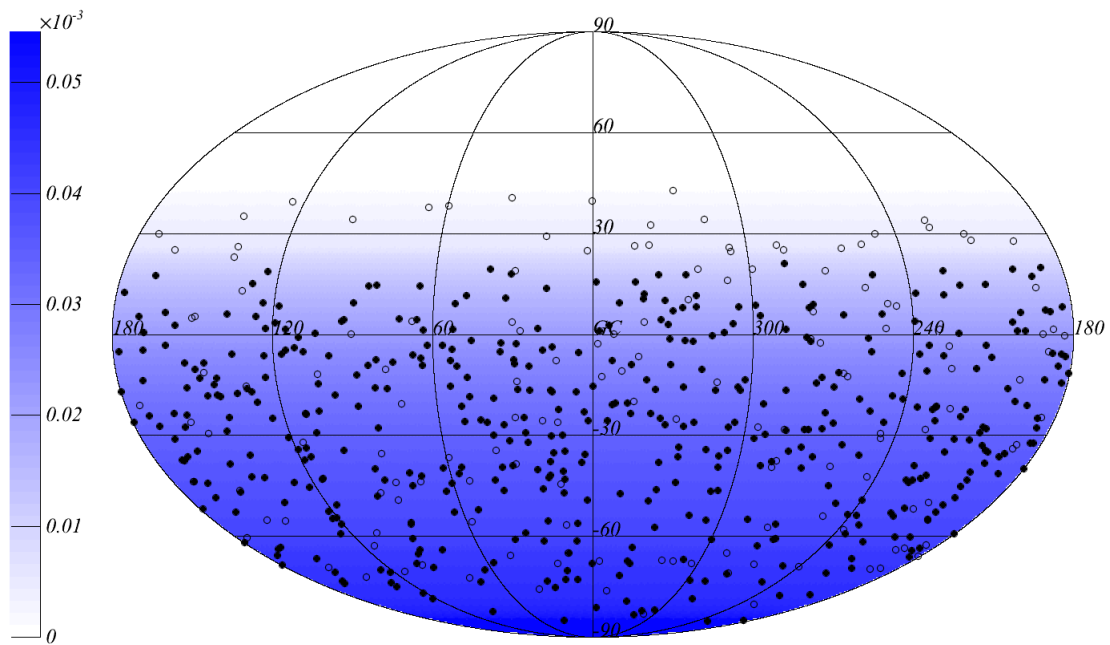


Figure 5.15: Arrival directions of the 602 events in equatorial coordinates: full dots represent vertical events, while empty dots inclined ones. Different shades of blue indicate the relative exposure in each point of the sky.

Chapter 6

Anisotropy studies

*"And the stars up above
directionless and drifting"*

Josh Ritter *Change of time* (2010)

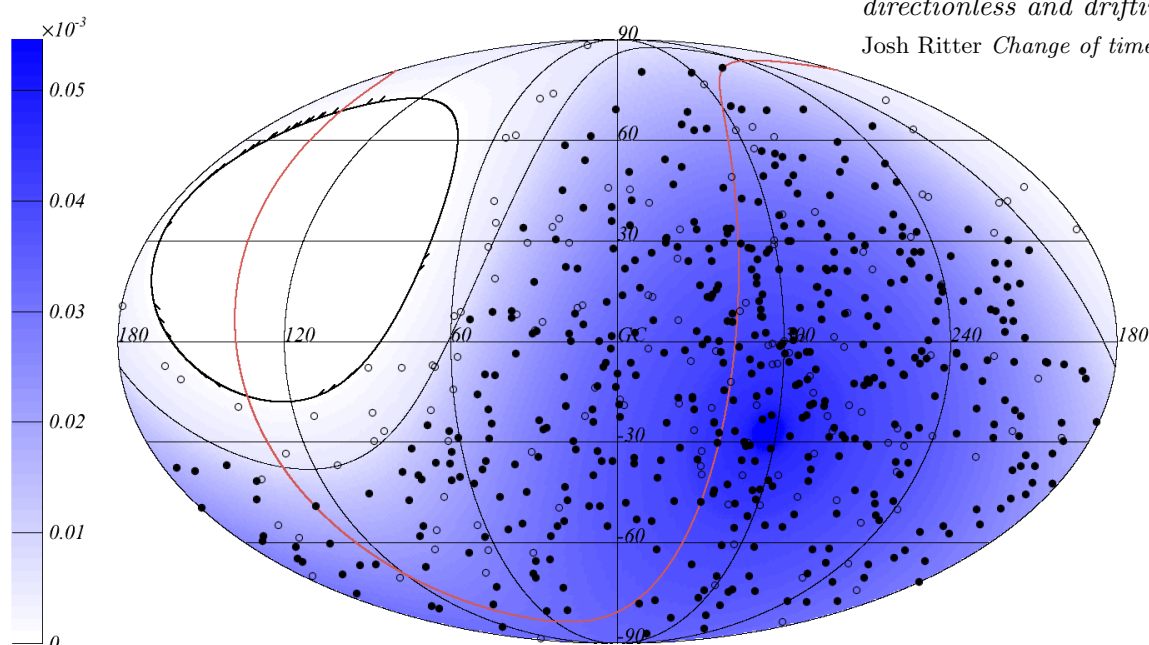


Figure 6.1: Arrival direction of all the 602 events described in chapter 5, in Galactic coordinates: full dots represent vertical events, while empty dots represent inclined ones. Different shades of blue indicate the relative exposure in each point of the sky. The red line represents the supergalactic plane. The thick and thin black lines demarcate the field-of-view limits for all events and for vertical ones only, respectively.

The study of the arrival directions of cosmic rays is perhaps the most natural way to search for their sources. In 2015 the Pierre Auger Observatory published [143] results of the search for anisotropy in the arrival directions of the highest energy cosmic rays, using the dataset described in the previous chapter 5. Figure 6.1 shows their distribution in the sky in Galactic coordinates. In that publication, to which this thesis work contributed, a search for intrinsic anisotropy and for correlations with astrophysical catalogues was

performed. After depicting in section 6.1 the general context in which this work took place, in this chapter we will show the most interesting results published in the cited paper, together with the cross-checks that we performed using, where possible, a different analysis, described in section 6.2. In detail, we will search for auto-correlations in the dataset (section 6.3), for possible correlations with astrophysical catalogues (section 6.4) and for correlation with the region of Centaurus A, the closest radio galaxy (section 6.5). Finally, we will show, in section 6.6, the results of the search for correlations of cosmic rays observed by both TA and Auger with a catalogue of high-energy neutrinos as observed by the IceCube experiment, for which we studied the deflections of cosmic rays in the Galactic magnetic field. Such studies were used in the analyses developed by a joint working group, composed of members of the three collaborations.

6.1 Context of the study

In chapter 3 we have shown some key results obtained by the Pierre Auger Observatory. A few of them are important to understand the analysis strategy exploited in this chapter.

The spectrum measured by Auger (paragraph 3.8) shows a suppression at around 40 EeV, this energy being compatible with the prediction from the GZK effect (see paragraph 1.2.2). However, this alone does not allow one to conclude whether the observed feature is due to propagation effects or to source properties, i.e., to the maximum energy achievable in the acceleration process (see chapter 1). In any case, the interactions of cosmic rays with photon backgrounds limit the distance from which a source can contribute to the flux at Earth at these energies. For instance, this distance has to be less than ~ 200 Mpc for protons or Fe nuclei with energies above 60 EeV, and even smaller for intermediate mass nuclei. The horizon for the 602 events considered here is thus limited.

The primary mass composition is inferred in Auger by FD measurements (see paragraph 3.5.2), that are available for only a fraction of the events. They indicate that the fraction of heavy nuclei increases towards the highest energies. However, the small number of events does not allow one to probe the primary mass evolution in detail at energies in excess of 40 EeV (only ~ 10 events are available in FD in this energy range).

Besides the uncertainty on the mass composition at the highest energies, also the knowledge of magnetic fields (that deflect cosmic rays, the deflection being proportional to their charge and inversely proportional to their energy) is limited. An overview of the Galactic and extra-Galactic magnetic fields was given in section 1.2.1 and we will go further in details about cosmic rays deflections at the highest energies in the forthcoming section 6.6.2. Here we remind that such deflections can range from few degrees to tens of degrees even for protons at the highest energies, according to recent models for the Galactic magnetic field.

A further element that sets the stage for the analyses presented in this chapter is a historical summary of the most relevant search for small-scale anisotropy performed in Auger. In 2006 the Pierre Auger Collaboration performed a first study to search for small scale anisotropy using events with $E \geq 40$ EeV and $\theta < 60^\circ$ collected from 1st January, 2004 to 26th May 2006 (so-called period I). On this dataset, an exploratory scan was performed over the energy threshold of the events, their angular separation from AGNs included in the 2006 edition of the Véron-Cetty and Véron (VCV) catalogue of active

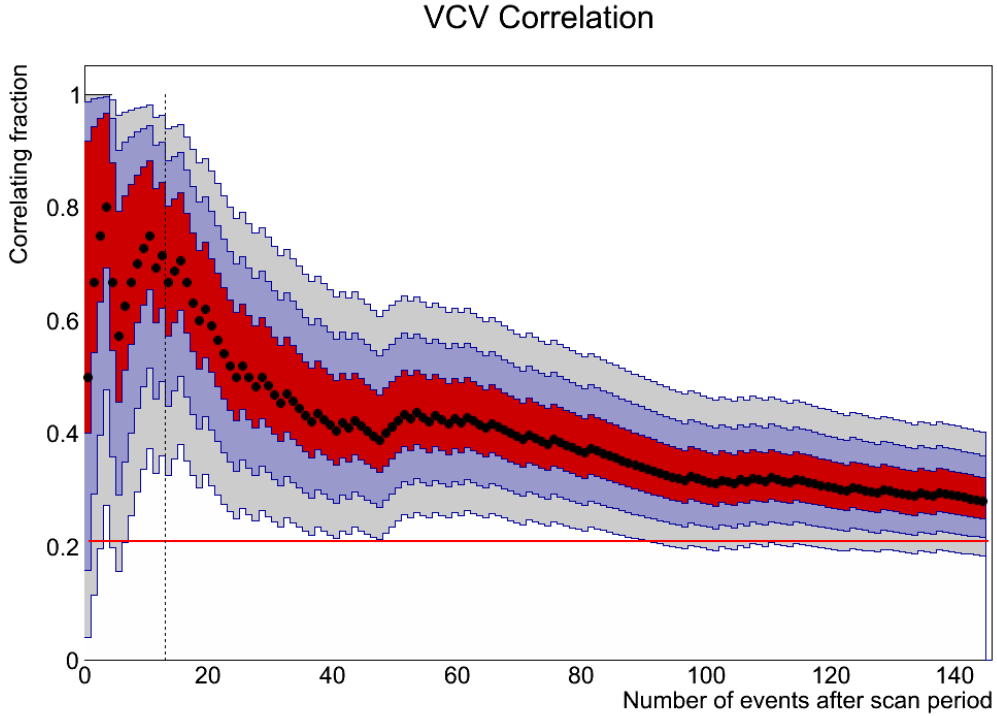


Figure 6.2: Fraction of events correlating with the VCV catalogue as a function of the number of events, ordered in time. The red, blue and gray shades are the 1,2 and 3 – σ dispersions of the observed fraction. The red line represents the isotropic expectation of 21% while the black dashed line delimits the events included in period II (see text).

galactic nuclei [147], and the maximum AGN redshift. The most significant excess was found for events with energy above 57 EeV with arrival directions within 3.1° of AGNs closer than 75 Mpc. A prescription was set on these parameters: the Auger search protocol was designed as a sequence of tests to be applied after the observation of each new event with energy above 57 EeV. The total probability of incorrectly rejecting the isotropy hypothesis along the sequence was set to a maximum of 1%. The prescription was fulfilled with the 13 events collected up to 31st August 2007 (so-called period II): among them, 8 events correlated, while 2.7 events (i.e., 21% of the total) were expected to correlate by chance in case of an isotropic distribution of arrival directions. This finding had a probability of $1.7 \cdot 10^{-3}$ of happening by chance [136, 137]. Subsequent analyses with enlarged statistics yielded a correlation (38_{-6}^{+7})% with data up to the end of 2009 [138] and (33 ± 5)% with data up to the end of June 2011 [148].

Preliminary to the analyses presented in the following, we updated the VCV analysis by using the vertical data set described in the previous chapter and, for internal coherence, the same VCV catalogue used in [137], although a newer version was published in 2010. Excluding Period I, there are 146 events above 53 EeV¹: 41 events correlate with VCV AGNs, with the angular and distance parameters fixed by the exploratory scan. The

¹Given the update of the energy scale (see paragraph 5.3.2) 53 EeV corresponds to 57 EeV in the energy scale in use in [136, 137].

updated fraction of correlations is then $(28.1_{-3.6}^{+3.8})\%$, which is 2 standard deviations above the isotropic expectation of 21%. The evolution of the correlation fraction with time is shown in figure 6.2. We can see that the correlation is dominated by the excess seen in period II (delimited by a dashed line). We note here that since the VCV correlation involves many different regions of the sky (besides the fact that cosmic rays with different energies have significant time delays), an explanation of the reduced correlation found after 2007 in terms of a transient nature of the signal would not be natural. Hence, the high level of correlation found initially was probably affected by a statistical fluctuation. We then concluded that this particular test does not yield a significant indication of anisotropy with the present data set.

The context delineated here, where the origin of the UHECR flux suppression has not yet been understood, their mass composition is not precisely known, the predictions of their deflections in magnetic fields are uncertain and the VCV test does not provide anymore a significant indication of anisotropy, led us to the analysis strategy described in the following section.

6.2 Analysis strategy

Given the scenario outlined before, in 2014 the Pierre Auger Collaboration decided to perform a new set of anisotropy tests exploring the arrival directions of all cosmic rays observed with energies above 40 EeV, i.e., the 602-Evts dataset described in chapter 5. Since this energy corresponds to the onset of the suppression in the observed flux, a limited number of contributing sources is expected above such a threshold. A large range of energy thresholds were considered in fact, from 40 EeV up to 80 EeV, the upper value still allowing to have a sizeable number of events (22) in the analysis. Also, given the large range of expected angular deflections caused by intervening magnetic fields and the unknown mass composition at these energies, it seemed sensible to explore a wide range of angular windows, between 1° and 30° . The angular range was motivated, at the lower end, by the angular resolution of the measurement of the arrival directions (see chapter 5) and, at the higher end, by the large deflections expected if cosmic rays are high- Z nuclei. The search for anisotropies was performed through various tests. First we studied “intrinsic” anisotropies, that can be revealed by the search for localized excesses of events over the exposed sky and by the analysis of the autocorrelation of arrival directions. Then we searched for correlations with known astrophysical structures, such as the Galactic and Super-Galactic Planes and the Galactic Center. Also, we studied the cross-correlation with astrophysical objects that could be considered as plausible candidates for UHECR sources. Specifically, we exploited flux-limited catalogs of galaxies (2MRS), of AGNs observed in X-rays (Swift BAT-70) and of radio galaxies with jets. For the last two samples, we performed an additional study, considering different thresholds in the AGN intrinsic luminosity. Finally, we focused on the distribution of events around the direction of Centaurus A.

All the correlations were searched, as said, scanning in energy threshold, E_{th} , from 40 EeV to 80 EeV in step of 1 EeV, and in the angular scale of the correlation, Ψ , from 1° up to 30° , in steps of 0.25° up to 5° , and of 1° for larger angles. For the correlations with catalogues, a scan was performed also in the maximum distance of the candidate sources,

D , from 10 Mpc up to 200 Mpc, in steps of 10 Mpc. To identify an excess with respect to isotropy, the observed number of correlating events was compared with that expected from an isotropic distribution having the same number of events above the corresponding energy threshold. This value was computed by simulating 10000 sets of isotropic arrival directions containing the same number of events as the dataset. The fraction of isotropic simulations, f , having an equal number to, or more, correlating events (pairs) than the data was computed: the excess corresponding to the minimum value of f , f_{\min} , was classified as the most significant one. The significance of this excess was evaluated by computing the post-trial probability, P , defined as the fraction of isotropic simulations which under a similar scan would lead to a value of f_{\min} smaller than the one obtained with the data. In the following this procedure will be dubbed as “published method”.

In this thesis work we performed a cross-check of the results of the analyses. In particular, we focused on the 2-points correlations (the autocorrelation and the correlations with astronomical objects). For the autocorrelation, we performed an independent analysis using the same approach as described before. For the other ones, we used a different approach (dubbed in the following the “analytical method”) to search for the most significant excess. Rather than generating isotropic events, we computed analytically $C_{iso}(\Psi, D, L_{min})$, i.e., the fraction of events that was expected to correlate for each catalogue, for each value of Ψ and D_{max} (and L_{min} if the luminosity scan was also performed). C_{iso} was computed based on the distribution of the candidate sources, the angle Ψ and the Auger directional exposure. Given C_{iso} , and the correlating fraction actually observed for each E_{th} , $C_{data}(E_{th}, \Psi, D, L_{min})$, we computed the cumulative binomial probability p to measure a correlating fraction C_{data} given C_{iso} . If we have n events above E_{th} and out of them k correlates (so that $C_{data} = k/n$):

$$p = \sum_{i=k}^n \binom{n}{i} \cdot C_{iso}^i \cdot (1 - C_{iso})^{(n-i)} \quad (6.1)$$

Note that the two methods treat in a different way the case of overlapping sources, i.e., when a cosmic ray is within Ψ degrees from more than one source: the published method counts all the pairs, while the analytical methods counts it as only one correlating event. This difference is important in particular for large catalogs, such as the 2MRS (see section 6.4). The purpose of this cross-check was to verify the stability of the minima found with the published method. For this reason, and since it would have been too much CPU-time consuming, we did not cross-check the post-trial penalized probability.

Note that, with the analytical method, we performed the scans using slightly different steps with respect to the published analysis: for the energy steps, we added the events, ordered by energy, one by one, rather than using fixed-energy steps. The number of events with $40 \text{ EeV} < E < 80 \text{ EeV}$ is 580, so that is the number of steps we performed in energy. Regarding the scan in the maximum distance D of the sources, we rather did a scan in the maximum observed redshift z of the source, ranging from $z=0.0025$ to $z=0.045$ with a step of 0.0025, which corresponds to $\sim 10.7 \text{ Mpc}^2$.

In the following sections we will present a selection of the published results of the analysis together with the cross-checked results performed in this work. In particular, as

²We adopt hereafter a Hubble constant of $H_0 = 70 \text{ km/s/Mpc}$ and the effective distances considered are taken as $D \equiv zc/H_0$, with z the source redshift obtained from the catalogue.

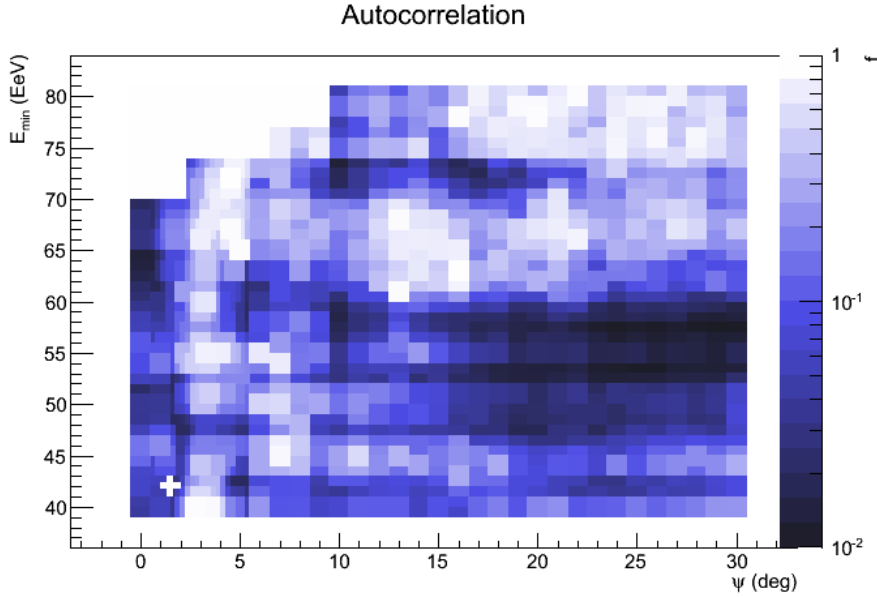


Figure 6.3: Fraction f obtained in the autocorrelation of events versus ψ and E_{th} . The white cross indicates the parameters for which the minimum value of f is obtained.

anticipated, we will focus on the results of the 2-points correlations.

6.3 Autocorrelation

A simple way to test the clustering of arrival directions is through a 2-points autocorrelation analysis, which is particularly useful to highlight scenarios where several sources lead to excesses around them on a similar angular scale.

With this method we looked for excesses in the number of pairs of events, i.e., excesses of “self-clustering,” by looking at the number of pairs separated by less than Ψ degrees, scanning in ψ and E_{th} . In this case, f is the fraction of isotropic simulations having an higher or equal number of pairs than the data, $f(\psi, E_{\text{th}})$.

The results of our analysis are shown in figure 6.3, where f is plotted as a function of the angular distance and the energy threshold. The color code indicates the values obtained for f . The white cross corresponds to the parameter values leading to the minimum value of this fraction, $f_{\min} = 0.025$, found at $\psi = 1.5^\circ$ and $E_{\text{th}} = 42$ EeV. These values are in agreement with the published ones. For these parameters, 30 pairs were expected on average for isotropic simulations while 41 were observed in the data. The post-trial probability for this excess, P , is $\simeq 70\%$: this means that the observed autocorrelation is compatible with the expectations from an isotropic distribution of the arrival directions of the events.

6.4 Cross correlation with astrophysical catalogues

6.4.1 Catalogues

In this section we search for correlations between the set of arrival directions and the celestial distribution of potential nearby cosmic ray sources. We choose approximately uniform and complete catalogues, namely the 2MRS catalogue of galaxies [149], the Swift-BAT [150] X-ray catalogue of AGNs³ and a recently compiled catalogue of radio galaxies with jets [151].

The three samples are quite complementary in identifying possible cosmic ray sources. The normal galaxies that dominate the 2MRS catalogue may trace the locations of gamma ray bursts and/or fast spinning newborn pulsars, whereas X-rays observed by Swift identify AGNs hosted mainly by spiral galaxies, while the radio emission catalogue selects extended jets and radio lobes of AGNs hosted mainly by elliptical galaxies.

2MRS

The 2MASS Redshift Survey (2MRS) [149] maps the distribution of galaxies in the nearby universe. It covers 91% of the sky, except for Galactic latitudes $|b| < 5^\circ$ (and $|b| < 8^\circ$ for longitudes within 30° of the Galactic center). In the region covered, it is essentially complete (at 97.6%) for magnitudes brighter than $K_s = 11.75$. It contains 43,533 galaxies with measured redshift: 37,209 of them are within 200 Mpc and 16,422 within 100 Mpc. About 90% of its objects have a redshift $z < 0.05$, which is the range of distances of interest for UHECR correlation studies due to the effects of the GZK horizon.

Swift-BAT

The Swift-BAT 70-month X-ray catalogue [150] includes sources detected in 70 months of observation of the BAT hard X-ray detector on the Swift gamma-ray burst observatory. It contains a total of 1210 objects: 705 of them are AGN-like (Seyfert I and II, other AGNs, blazars and QSOs) with measured redshift. The catalogue is complete over 90% of the sky for fluxes $> 13.4 \cdot 10^{-12}$ erg/(s cm²), measured in the X-ray band from 14 to 195 keV. We use this cut in flux hereafter to have a more uniform sample of nearby AGNs. 489 AGN-like objects survive the cut: 296 of them are within 200 Mpc and 160 are within 100 Mpc.

Radiogalaxies

The third catalogue we use is a compilation of radio galaxies produced in [151]. This is a combination of catalogues of observations at 1.4 GHz (NRAO VLA Sky Survey [152]) and 843 MHz (Sydney University Molonglo Sky Survey [153]), with redshifts of associated objects taken from 2MRS. A flux limit of 213 mJy (289 mJy) at 1.4 GHz (843 MHz) is imposed to the objects from each respective catalogue, which would correspond to the flux of Cen A as seen from a distance of about 200 Mpc. We select from this catalogue the radio galaxies having jets, which constitute a set of attractive candidates for UHECR

³The 2MRS and Swift catalogues have been updated after the earlier Auger study of correlations in [136, 148].

sources. There are in total 407 such jetted radio galaxies: 205 are within 200 Mpc and 56 are within 100 Mpc (for this catalogue we compute the distance using the redshift corrected for peculiar velocities that are also provided).

6.4.2 Cross-correlation with flux-limited samples

We have first studied the cross-correlation with the three flux-limited catalogues listed above (with the flux limits just described), including objects up to different maximum distances. This selection is based on the apparent luminosity, and is motivated by the fact that nearby sources may contribute significantly to the fluxes (in their corresponding electromagnetic band as well as in cosmic rays) even if they are intrinsically fainter than far away sources.

The basis of the cross-correlation technique is a counting of the number of pairs between the UHECRs and the objects in the chosen catalogue. We remind scanned over the minimum energy of the events, E_{th} , the maximum angular separation between the objects and the events ψ , and the maximum distance of the candidate sources D (maximum redshift, z , for the analytical method). Here we will show the results obtained in the published analysis together with the ones obtained with the cross-check performed for this work.

Figure 6.4 displays the results for the case of the 2MRS catalogue. The top panel shows f_{min} from the published analysis (blue circles), p_{min} for the analytical method (black asterisks) and the post-trial probability, P (red triangles), obtained for every distance D .

The minimum values for f are observed for $D = 90$ Mpc, for which $f_{\text{min}} \simeq 1.5 \cdot 10^{-3}$ and $P \simeq 8$ %. The bottom-left panel in the figure shows the distribution of $f(\psi, E_{\text{th}})$ as a function of energy threshold and angle for the value $D = 90$ Mpc giving rise to the minimum probability. The local minimum (indicated with a cross) is observed for $\psi = 9^\circ$ and $E_{\text{th}} = 52$ EeV. For these values, 20,042 pairs were expected on average from isotropic realizations while 23,101 were observed in the data. Considering the penalization due to the scan in D performed, the probability \mathcal{P} to obtain a value of P smaller than 8% from isotropic distributions for any value of D is $\mathcal{P} \simeq 24\%$.

For the analytical method, the minimum is found at 96 Mpc (in figure 6.4, top panel, black asterisks): the values of p for this value of D are shown in the bottom-right panel. The minimum is at $\Psi = 6^\circ$ and $E_{\text{th}} = 70$ EeV, with a $p_{\text{min}} = 1 \cdot 10^{-4}$. With the analytical method, the minimum obtained with the published analysis is found to have a value of $p = 2 \cdot 10^{-1}$. We can see that in this case the analytical method leads to results different the published one. This is probably due to the different response of the two methods in the case of overlapping sources, which is important with large catalogs as the 2MRS (see section 6.2). However, we can see that the behavior of p as a function of D in the top panel of figure 6.4 is similar to the one of f .

Similar plots to those presented above are included in Figure 6.5 for the case of the Swift-BAT catalogue. As shown in the top panel of the figure, the minimum values are reached here for $D = 80$ Mpc, where $f_{\text{min}} \simeq 6 \cdot 10^{-5}$ and $P \simeq 1\%$. Correspondingly, the bottom-left panel in the figure shows $f(\psi, E_{\text{th}})$ as a function of energy and angle at $D = 80$ Mpc. The local minimum (indicated with a cross) is at $\psi = 1^\circ$ and $E_{\text{th}} = 58$ EeV, where 9 pairs were observed while 1.6 were expected on average. After accounting for the penalization due to the scan performed in D , the probability of obtaining a value of P

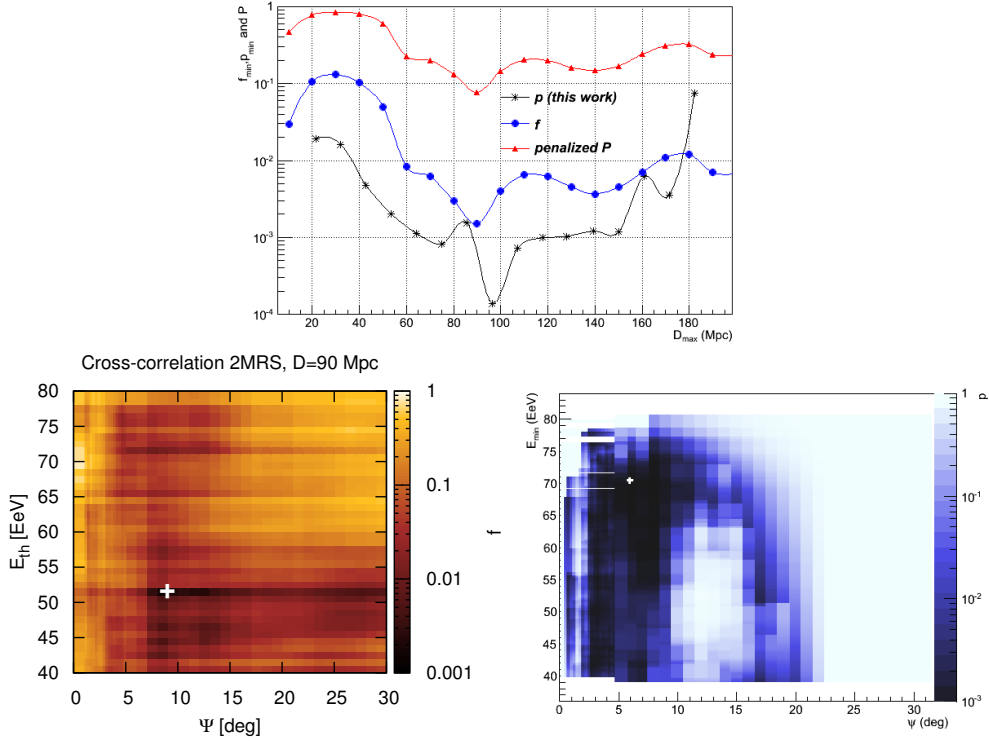


Figure 6.4: Cross-correlation of events with the galaxies in the 2MRS catalogue. The top panel shows the values of f_{\min} (blue circles), p_{\min} (black asterisks) and P (red triangles) as a function of the maximum distance D to the galaxies considered. The bottom-left panel shows the results of the scan in ψ and E_{th} for the value $D = 90$ Mpc corresponding to the minimum values of f and P in the top panel. The bottom-right panel shows the results of the scan with the analytical method in ψ and E_{th} for the value $D = 96$ Mpc corresponding to the minimum values of p in the top panel.

smaller than 1% from isotropic distributions for any value of D is $\mathcal{P} \simeq 6\%$.

The cross-check with the analytical method leads to a minimum compatible to the one found before, at $D = 86$ Mpc. In the bottom-right plot in figure 6.5, we can see that this minimum is found at $\Psi = 1^\circ$ and $E_{\text{th}} = 56$ EeV, where $p_{\min} = 6 \cdot 10^{-4}$.

The results of the cross-correlation with jetted radio galaxies are shown in figure 6.6. The minimum value $f_{\min} \simeq 2 \cdot 10^{-4}$, with $P \simeq 1.4\%$, is obtained for $D = 10$ Mpc (see top panel). The only object included in this catalogue within such a distance is the Centaurus A galaxy. Since the correlation with Cen A is discussed separately in the next section, we consider here the second minimum, which is found for $D = 90$ Mpc. This minimum corresponds to $f_{\min} \simeq 4 \cdot 10^{-4}$ and $P \simeq 3.4\%$. The bottom-left panel in the figure thus shows the results of the scan in angle and energy for $D = 90$ Mpc. The minimum occurred for $\psi = 4.75^\circ$ and $E_{\text{th}} = 72$ EeV, where 13 pairs were observed in the data and 3.2 were expected on average. The chance probability for getting $P \leq 1.4\%$ (corresponding to the absolute minimum found) for any value of D is $\mathcal{P} \simeq 8\%$.

The cross-check with the analytical method led to three similar minimum with $p_{\min} \sim 10^{-3}$, the first two being at $D = 75$ Mpc and $D = 118$ Mpc, both with $\Psi = 3^\circ$ and

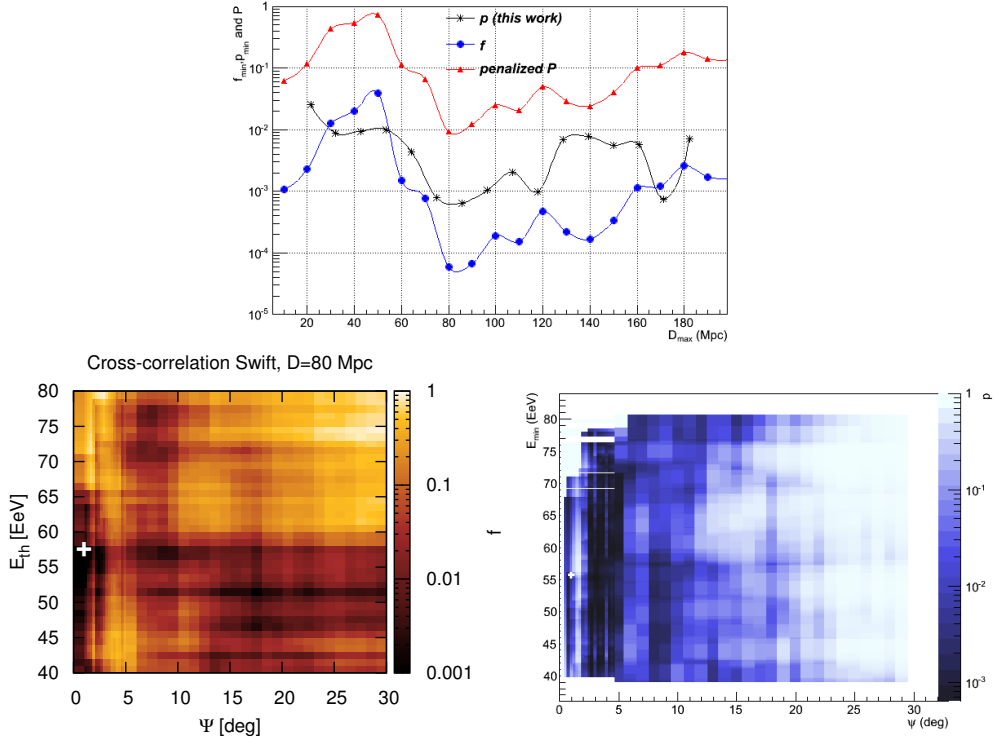


Figure 6.5: Cross-correlation of events with the galaxies in the Swift-BAT catalogue. The top panel shows the values of f_{\min} (blue circles), p_{\min} (black asterisks) and P (red triangles) as a function of the maximum distance D to the galaxies considered. The bottom-left panel shows the results of the scan in ψ and E_{th} for the value $D = 80$ Mpc corresponding to the minimum values of f and P in the top panel. The bottom-right panel shows the results of the scan with the analytical method in ψ and E_{th} for the value $D = 86$ Mpc corresponding to the minimum values of p in the top panel.

$E_{\text{th}} \sim 42$ EeV. The third one is compatible to the one found with the previous method, standing at $D = 96$ Mpc, $\Psi = 5^\circ$ and $E_{\text{th}} \sim 69$ EeV. The bottom-right plot in figure 6.6 shows this third minimum.

6.4.3 Cross-Correlation with Bright AGNs

In the case of the AGNs in the Swift catalogue and of the radio-galaxy one it was chosen to scan also on the measured intrinsic luminosity of the objects. This is motivated by the fact that the maximum cosmic ray energy, E_{max} , achievable at the sources may be linked to their intrinsic electromagnetic bolometric luminosity \mathcal{L} . In particular, one could expect that $(E_{\text{max}}/Z)^2 \propto \mathcal{L}$, if the energy density in the magnetic field is in equipartition with the energy in synchrotron emitting electrons in the acceleration region (see, e.g., [154]). Hence, it might happen that only sources intrinsically brighter than some given luminosity are able to accelerate cosmic rays above the threshold energies considered in this work. On the other hand, for the radio galaxies the luminosity is also correlated with the Fanaroff-Riley class, with FRII galaxies being generally brighter than FRI ones (see

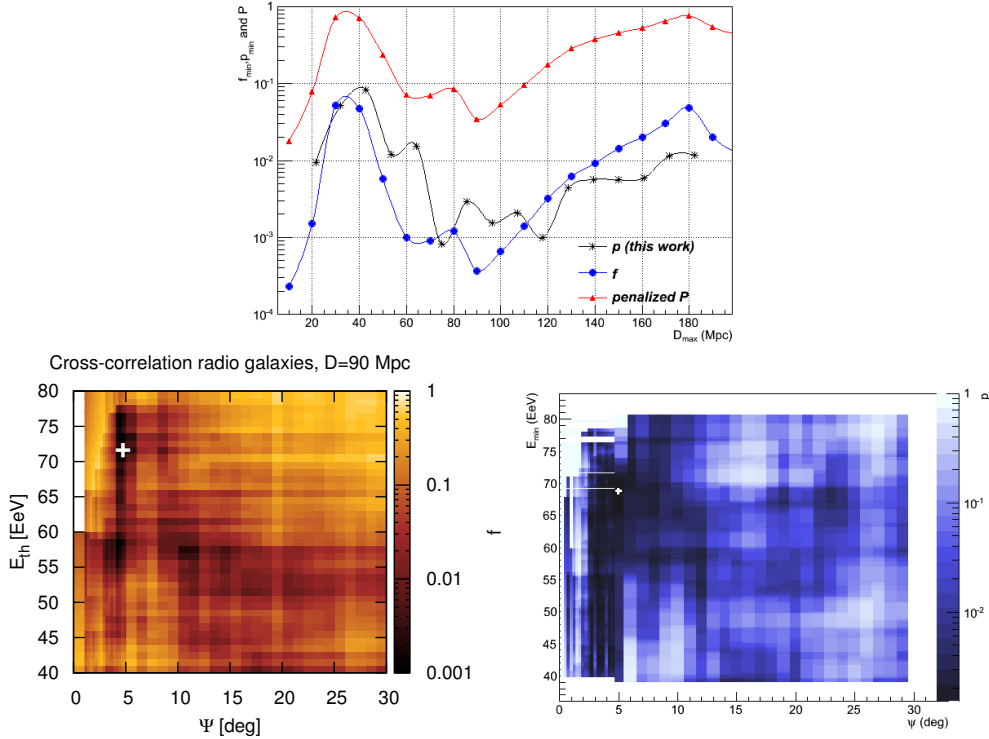


Figure 6.6: Cross-correlation of events with the radio galaxies. The left panel shows the values of f_{\min} (blue circles), p_{\min} (black asterisks) and P (red triangles) as a function of the maximum distance D to the galaxies considered. The center panel shows the results of the scan in ψ and E_{th} for the value $D = 90$ Mpc corresponding to the minimum values of f and P in the top panel. The right panel shows the results of the scan with the analytical method in ψ and E_{th} for the value $D = 96$ Mpc corresponding to the third minimum values of p in the top panel.

section 1.3.2 for details).

We present here the results of the scan over the minimum source luminosities performed, considering for the Swift AGNs the reported X-ray band luminosity \mathcal{L}_X and for the radio-galaxy sample the reported radio luminosity \mathcal{L}_R , computed per logarithmic energy bin at 1.1 GHz. For Swift, the scan was performed over the threshold values for \mathcal{L}_X from 10^{42} erg/s up to 10^{44} erg/s, while for the radio galaxies it was performed over the threshold values for \mathcal{L}_R from 10^{39} erg/s up to 10^{41} erg/s, considering three logarithmic steps per decade, for a total of 7 luminosity values in each case. These luminosity values cover most of the range spanned by the actual luminosities of the AGNs that are present in the catalogues (just 10 AGNs from the Swift sample have $\mathcal{L}_X < 10^{42}$ erg/s, while only 3 AGNs from the radio-galaxy sample have $\mathcal{L}_R < 10^{39}$ erg/s). Given the additional scan performed in \mathcal{L} , a slightly coarser scan was performed in D , using 20 Mpc steps to cover from 10 Mpc up to 190 Mpc.

Considering first the Swift catalogue, we show in the top-left panel of figure 6.7 the resulting values of f_{\min} as a function of the maximum AGN distance and the minimum adopted luminosity \mathcal{L}_{\min} in the respective bands (the white region in the top-left corner of

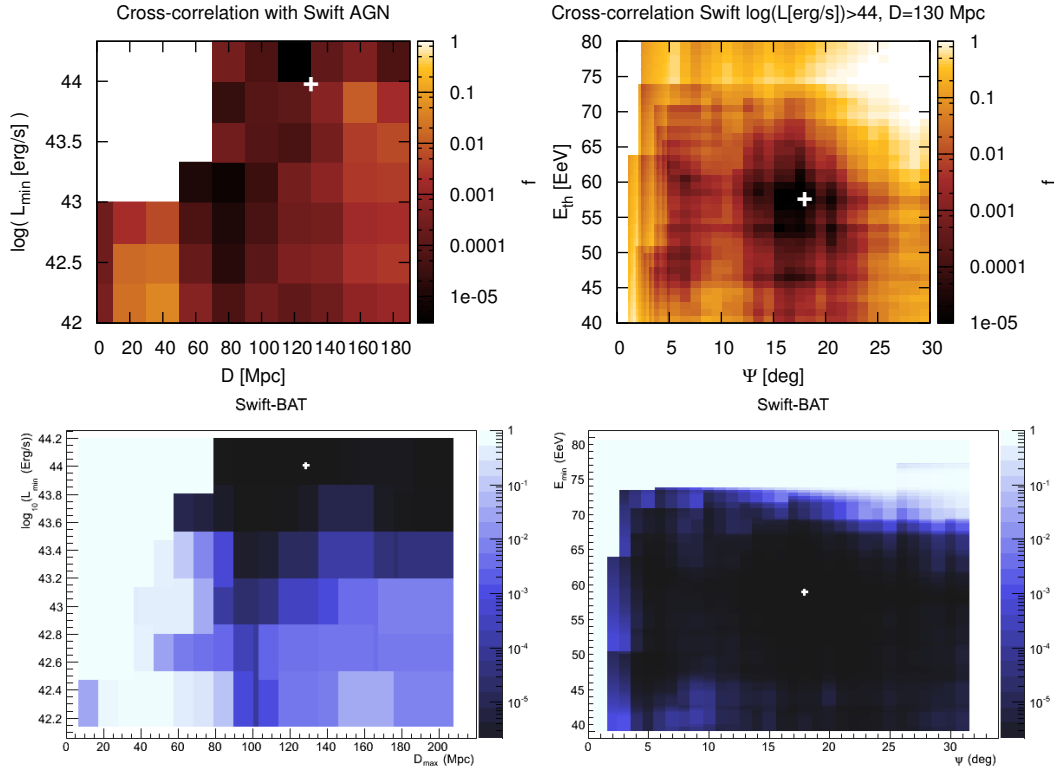


Figure 6.7: Cross-correlation of events with the AGNs in the Swift catalogue as a function of D and \mathcal{L}_{\min} (top-left panel) and detail of the scan in Ψ and E_{th} for the minimum found (top-right panel). The bottom plots are similar, but showing the p-value computed with the analytical method.

the plot is due to the absence of nearby objects above those luminosity thresholds). The values of f_{\min} are obtained after scanning on Ψ and E_{th} as in the previous subsection. The minimum value of $f_{\min} = 2 \cdot 10^{-6}$ is obtained for $D = 130$ Mpc and $\mathcal{L} > 10^{44}$ erg/s. The top-right panel shows the details of the scan in Ψ and E_{th} for $D = 130$ Mpc and $\mathcal{L} > 10^{44}$ erg/s. The minimum corresponds to the values $\Psi = 18^\circ$ and $E_{\text{th}} = 58$ EeV. For these parameters there are 10 AGNs and 155 events, and 62 pairs are obtained between them while the isotropic expectation is 32.8. The probability to find values $f_{\min} < 2 \cdot 10^{-6}$ in isotropic simulations after making the same scan on Ψ , E_{th} , \mathcal{L}_{\min} and D is $\mathcal{P} \simeq 1.3\%$.

The bottom plots shows the cross-check results with the analytical method, that found a minimum compatible with the one previously described, at $\mathcal{L} > 10^{44}$, $D > 130$ Mpc, $\Psi = 18^\circ$ and $E > 58.79$ EeV, with a value $p_{\min} = 2 \cdot 10^{-6}$.

Figure 6.8 is similar, but for the sample of radio galaxies. The scan in luminosity leads to two minima with very similar probabilities, both for $D = 90$ Mpc (see the top-left panel). The first one has $f_{\min} = 5.1 \cdot 10^{-5}$ and corresponds to $\mathcal{L} > 10^{39.33}$ erg/s, $\Psi = 4.75^\circ$ and $E_{\text{th}} = 72$ EeV, the angle and energy being equal to the parameters already obtained in the previous subsection (Figure 6.6). The main difference is that 32 AGNs remain within 90 Mpc once the luminosity cut is imposed, compared to the original sample of 39 AGNs in the flux-limited sample, so that the expected number of pairs becomes 2.4 while 13 are actually observed. The second minimum has $f_{\min} = 5.6 \cdot 10^{-5}$ and corresponds to

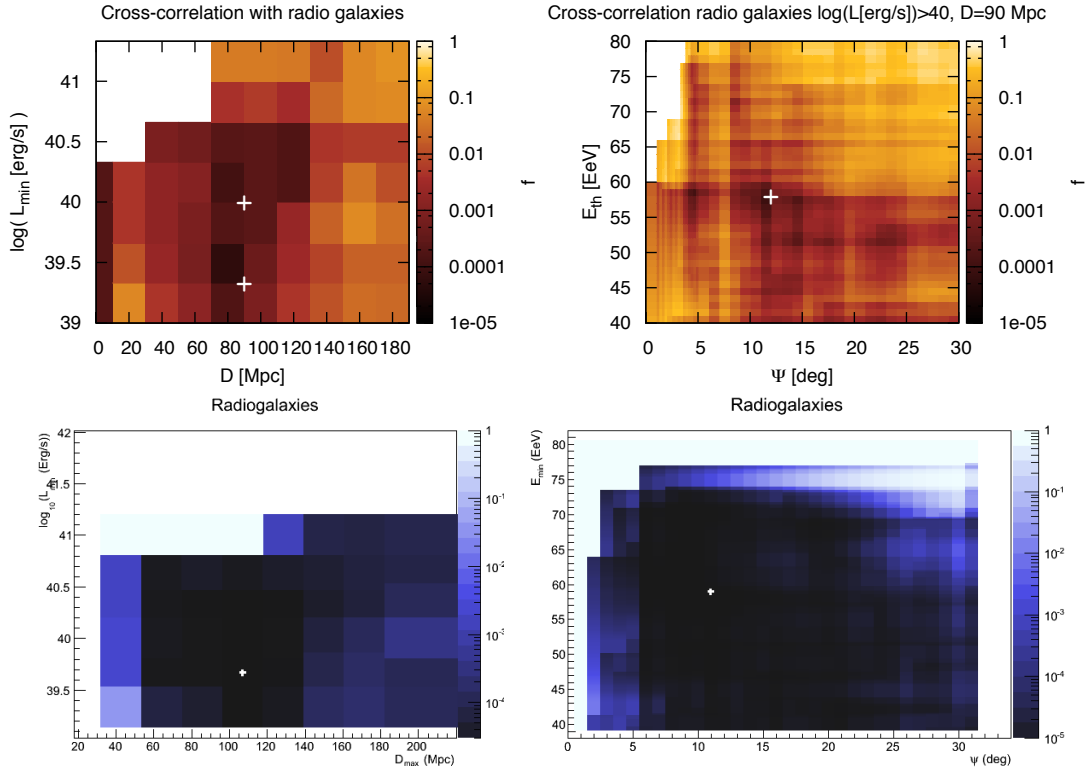


Figure 6.8: Cross-correlation of events with the radio galaxies as a function of D and \mathcal{L}_{\min} (top-left panel) and detail of the scan in Ψ and E_{th} for the second minimum found (top-right panel). The bottom plots are similar, but showing the p-value computed with the analytical method.

$\mathcal{L} > 10^{40}$ erg/s. The top-right panel shows the scan in Ψ and E_{th} for this minimum, which leads to $\Psi = 12^\circ$ and $E_{\text{th}} = 58$ EeV. The globally penalized probability of getting $f_{\min} < 5.1 \cdot 10^{-5}$ after a similar scan with the radio galaxies turns out to be in this case $\mathcal{P} \simeq 11\%$.

The bottom plots shows the cross-check results from the analytical method, that found a minimum at $\mathcal{L} > 10^{39.66}$, $D > 110$ Mpc, $\Psi = 11^\circ$ and $E > 58.9$ EeV. Both the minima found with the published method have in the analytical one a probability $p \sim 2 \cdot 10^{-3}$, while the p_{\min} found at the cited parameters has a value $p = 3 \cdot 10^{-4}$.

Overall, we may conclude that the analytical method found results closer to the ones from the published one when the value of \mathcal{P} was smaller, i.e., when the minimum was better defined.

6.5 Single-source correlation (CenA)

Centaurus A is the nearest radio-loud active galaxy, at a distance of less than 4 Mpc. In addition, the nearby Centaurus cluster is a large concentration of galaxies lying in approximately the same direction and at a distance of ~ 50 Mpc. It is interesting to note that the most significant region found in the search for localized excesses published in

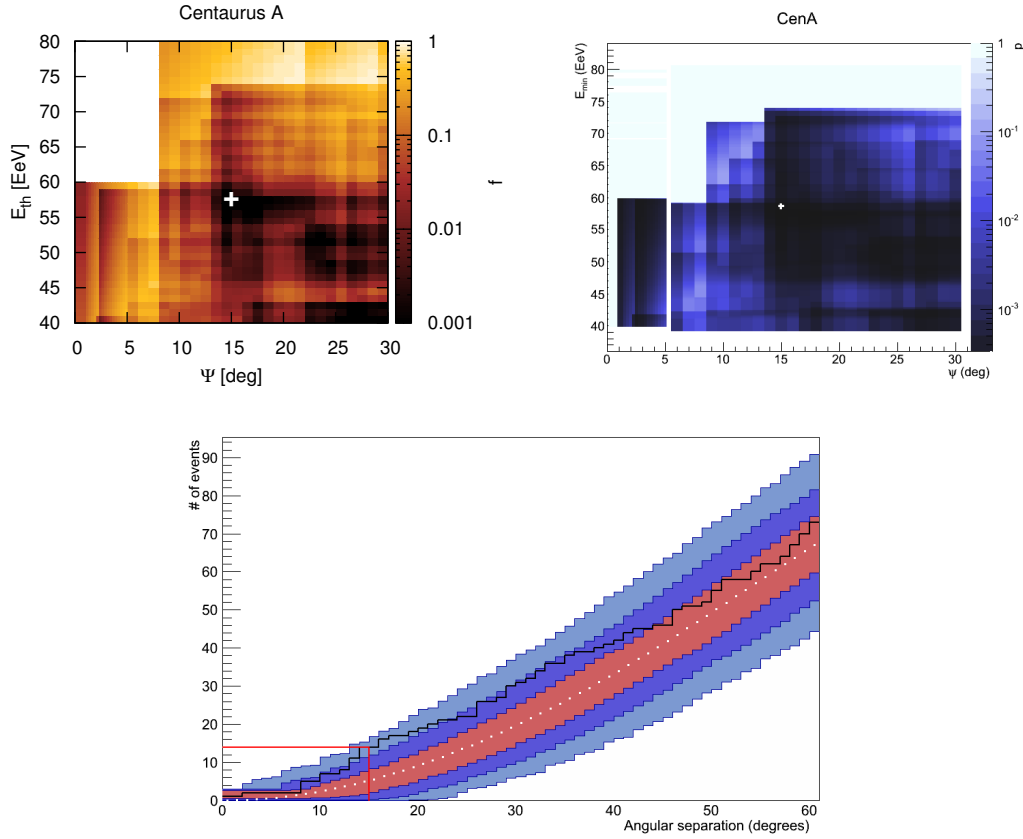


Figure 6.9: Correlation of events with the Cen A radio galaxy as a function of the angular distance and the energy threshold, E_{th} (top-left panel). The top-right panel shows the same results obtained with the analytical method. The bottom panel shows the cumulative number of events (black line) for the threshold $E_{\text{th}} = 58$ EeV, compared to the isotropic expectation (white points) and its 1, 2 and 3 σ dispersion, respectively in red, dark blue and light blue.

[143], though not statistically significant, was found for the events with $E \geq 54$ EeV in a 12° -radius windows centered at about 18° from the direction of Centaurus A (Cen A).

The search for correlation with the position of Cen A was performed by varying the energy threshold of events between 40 EeV and 80 EeV and by counting events in angular radii ranging from 1° to 30° . Figure 6.9 (top-left panel) shows the fraction f of isotropic simulations that yield a number of pair larger or equal than the data. The minimum value of f is $f_{\text{min}} = 2 \cdot 10^{-4}$, corresponding to $E_{\text{th}} = 58$ EeV and $\psi = 15^\circ$. There are 14 events (out of a total of 155) observed while 4.5 are expected on average from isotropic distributions. The fraction of isotropic simulated data sets that yield a smaller value of f_{min} under a similar scan is $\mathcal{P} \simeq 1.4\%$.

The analysis with the analytical method found a perfectly compatible minimum at $E_{\text{th}} = 58.58$ EeV and $\psi = 15^\circ$, as it is possible to see in figure 6.9 (top-right panel).

For completeness, we show in the bottom panel of the figure the number of events with energy above 58 EeV as a function of the angular distance from Cen A indicating also the 68, 95 and 99.7% ranges obtained with isotropic simulations.

6.6 Search for correlation with neutrino “catalogues”

Another possible approach to search for anisotropy in the highest energy cosmic rays distribution is to look for correlations with the arrival directions of ultra-high energy astrophysical neutrinos. UHECRs accelerated in astrophysical sources are naturally expected to produce high-energy photons and neutrinos in the interactions with the ambient matter and radiation (see paragraph 1.2.2 for details). While photons are expected to be attenuated by interactions with the CMB, neutrinos travel essentially unattenuated. The IceCube [155] collaboration has recently published a data-set of ultra-high energy neutrinos showing evidence of detection of neutrinos from astrophysical origin at energies > 100 TeV. A joint working group was created in 2014 between the IceCube, Pierre Auger and Telescope Array Collaborations to search for angular correlations between the arrival directions of high-energy neutrinos and UHECRs. Here we will provide a summary of the data used, analyses performed and main results; further details can be found in [156].

For Auger, the events with $E \geq 52$ EeV, contained in the data-set described in the previous chapter were used, resulting in 231 events. For TA, 87 events above 57 EeV, recorded in the period ranging from 11th May 2008 to 1st May 2014, were used.

IceCube is a cubic-kilometer neutrino detector installed in the ice at the geographic South Pole between depths of 1450 m and 2450 m. Neutrino reconstruction relies on the optical detection of Čerenkov radiation emitted by secondary particles produced in ν interactions in the surrounding ice or the nearby bedrock. Depending on the flavor of the interacting neutrino and the type of interaction, different signatures are expected in the detector: charge current ν_μ interactions produces track-signals as the produced μ traverses the detector. On the other hand, charged current $\nu_{e,\tau}$ and all neutral current interactions produce cascades. Track events have a better angular resolution ($\sim 1^\circ$) than cascade ones ($\sim 15^\circ$). The background coming from atmospheric neutrinos is expected to be dominated by track-like events. Two different data sets have been considered in this work. The first one (“high-energy cascades”) is a set of 39 cascades with reconstructed $E > 30$ TeV and where the interaction occurs within the detector (4-years HESE, High Energy Starting Events). The second set of events (“high-energy tracks”) is composed of two parts; the first part is made of the 7 tracks in the HESE sample that have energies and directions which make them more likely to be of extraterrestrial origin than the other track events in that sample. The second part is made of 9 muon tracks found in a search of a diffuse up-going ν_μ flux. These tracks, found in two years of data (May 2010- May 2012), belong to a high energy excess with respect to atmospheric predictions⁴.

6.6.1 Analysis strategy

Two analyses were performed to search for neutrinos-UHECRs correlations. The first is a cross-correlation analysis between the arrival directions of UHECRs and high energy cascades and tracks. The principles of this analysis are the same as described for the cross-correlation with astrophysical catalogues. In this case, the scan is done only on the

⁴A third data set was used in the working group: a point source sample obtained with 4 years of data. This sample is composed of well reconstructed, through-going muons that could be associated with charged-current muon neutrino interactions and is composed of 394,000 events. However, since the analyses related to this catalogue were not related to this thesis work, we will not go into its details, that can be found in [156].

angular separation, between 1° and 30° , with steps of 1° . A second analysis is based on the stacking method as this is a well known way of accumulating multiple weaker signals to enhance the discovery potential. Since neutrinos are not deflected on their way to Earth, the stacked sources are the set of neutrinos arrival directions. An unbinned likelihood method is used with the log of the likelihood function defined as:

$$\ln \mathcal{L}(n_s) = \sum_{i=1}^{N_{\text{Auger}}} \ln \left(\frac{n_s}{N_{\text{CR}}} S_{\text{Auger}}^i + \frac{N_{\text{CR}} - n_s}{N_{\text{CR}}} B_{\text{Auger}}^i \right) + \sum_{i=1}^{N_{\text{TA}}} \ln \left(\frac{n_s}{N_{\text{CR}}} S_{\text{TA}}^i + \frac{N_{\text{CR}} - n_s}{N_{\text{CR}}} B_{\text{TA}}^i \right) \quad (6.2)$$

where n_s , the number of signal events, is the only free parameter. $N_{\text{CR}} = N_{\text{Auger}} + N_{\text{TA}}$, is the total number of events. The signal PDFs (S_{Auger}^i and S_{TA}^i), in which the different neutrino positions are stacked, take into account the exposure and angular resolution of the CR observatories, the likelihood maps for the reconstruction of the neutrino arrival directions and the magnetic field deflection D . The background PDFs (B_{Auger}^i and B_{TA}^i) are the normalized exposures of the CR observatories.

6.6.2 Magnetic fields deflections

The likelihood method described before is a powerful yet CPU-time consuming method. It would have been impossible to scan on the magnetic deflections D as we scan on ψ for the 2-point cross correlation method. Our contribution to this working group was to perform a study on the most recent models for the Galactic Magnetic Field (GMF) in order to give an estimation for D values to be used in the likelihood analysis.

In section 1.2.1, we summarized the current knowledge of the Galactic and extragalactic magnetic fields. In this section, we will go in deeper detail of the effects of magnetic fields on cosmic rays propagation, studying the deflections predicted by different Galactic magnetic field (GMF) models for UHECRs. To measure the deflections we used a propagation code, called CRT [157]. This code allows one to forward-track and back-track cosmic rays in the GMF using Runge-Kutta integration techniques. The forward-tracking method consists in following the cosmic rays from the position of the source and through the GMF. Though it is the most natural one, this method has the problem of being particularly time-consuming since only a small fraction of the forward-tracked events actually hits the Earth. A better solution is to backtrack the cosmic rays from the Earth through the GMF and record their direction when they exit the Galaxy. This approach is particularly suitable to measure the average deflections over the sky and it is the one that will be used in the following. A full description of the code can be found in [157] and the code itself can be downloaded from <http://crt.mps.ohio-state.edu/index.html>. Various GMF models among the ones suggested in literature are included in the CRT code. Here we will show results obtained with two of the most recent GMF models: the Jansson-Farrar 2012 and the Pshirkov-Tinyakov one.

The Jansson-Farrar 2012 GMF model (JF) [33] is a complete model of the Galactic magnetic field for both its regular and turbulent components. The latter is further divided into a purely random component and a striated one which is aligned with the local large-scale field and has the same relative magnitude everywhere in the Galaxy. The model has 21 free parameters that are fitted to measurements coming from forty thousands extragalactic Faraday rotation measurements and from the polarized synchrotron emission as mapped by WMAP.

The Pshirkov-Tinyakov 2011 GMF model (PT) [158] is made of two components: one in the Galactic disk (spiral or ring) and the other one in the halo. The parameters of these components (4 for the disk and 5 for the halo) were determined by fitting different field geometries to the observed rotation measurements, in particular the ones obtained by NRAO VLA Sky Survey catalogue [152], and a compilation by the authors.

In order to see which effects we should expect on the cosmic rays arrival directions because of the GMF, we generated 100 000 proton events isotropically distributed on the sky. The energy of each event is randomly generated ranging from 40 to 150 EeV following the cosmic rays spectrum as measured by Auger (see section 3.8)⁵.

In figures 6.10 and 6.11, left panels, we show the map of expected deflections for the regular JF2012 and PT2011 models. The sky has been divided into 3072 equal-area pixels⁶ and the color code of each pixel indicates the average deflection for the cosmic rays generated with a departure direction in that pixel. In the right panels of each figure, the histograms show the distributions of the number of pixels with a certain average deflection. We can see that both models predict an average deflection of about 6° , with an RMS of about $3 - 4^\circ$, and that the maximum deflection can go in both cases up to $\sim 30^\circ$. Both distributions present a double-peaked feature. This is due to different average deflections in the northern and southern Galactic hemispheres. This can be better seen in the picture by looking at the red back-slashed and blue forward-slashed histograms, showing for both models the distributions of deflections for the northern and southern Galactic hemispheres, respectively.

In spite of these similarities, however, the two models show very different patterns in the sky, as visible from the deflection maps. In particular, the directions with the maximum deflections are localized around the Galactic center and in the regions north and south of it (longitude around 0°) for the JF model. They are in turn stronger on the Galactic plane, at around $\pm 50^\circ$ in longitude, for the PT model.

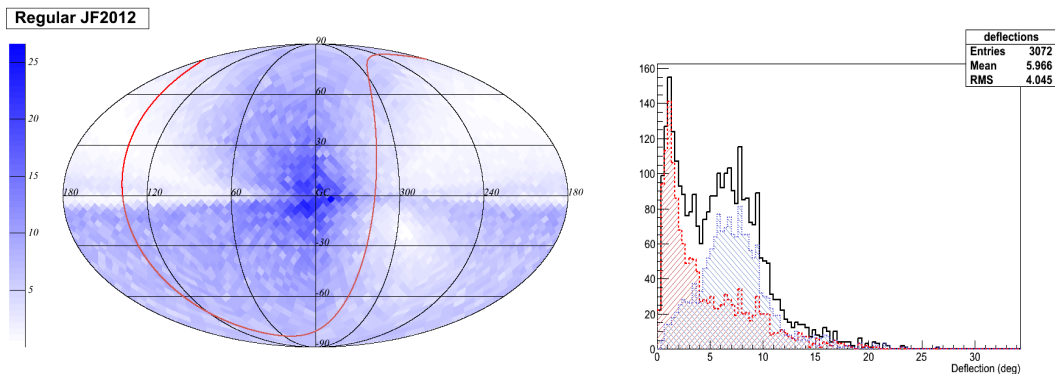


Figure 6.10: Left: Deflections for Jansson Farrar 2012 Regular GMF. The supergalactic plane is shown as a red line. Right: Number of pixels with a certain angular deflection for the JF2012 regular field in the northern (red backslashed) and southern (blue forward-slashed) Galactic hemisphere, and in the whole sky (black line).

We computed also the deflections predicted by the complete JF2012 magnetic field, su-

⁵The results does not change significantly if we use the spectrum as measured by TA.

⁶We used the NASA Healpix tool, described in <http://healpix.jpl.nasa.gov/>

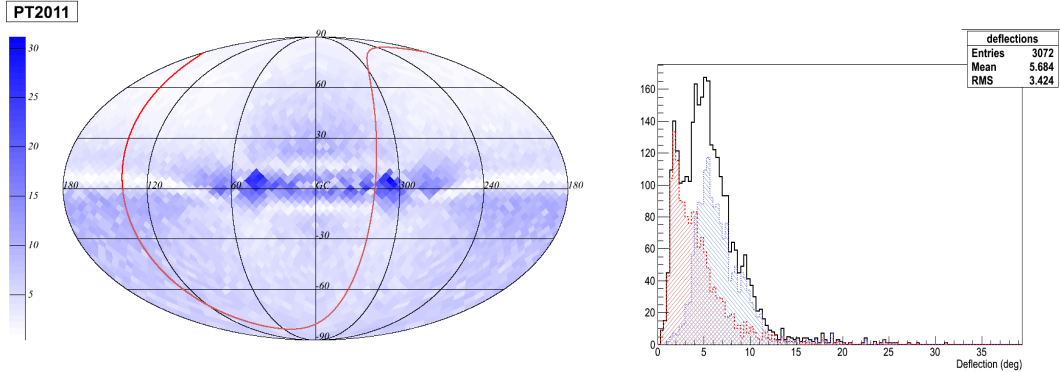


Figure 6.11: Same as in figure 6.10, but for Pshirkov-Tinyakov 2011 model.

perimposing the regular component with the striated and random ones. This computation is strongly CPU-time consuming so we limited our study to ~ 4000 events. The results are shown in figure 6.12, with the deflection maps in the left panel and the histogram of deflections in the right one, as for the previous figures. Comparing the histogram to what we obtained for the JF regular field alone (fig 6.10, right), we can see that the peak at low deflections has disappeared, whilst the mean value of the deflection did not change significantly. We can conclude that adding the random field affects significantly only the regions with small deflections.

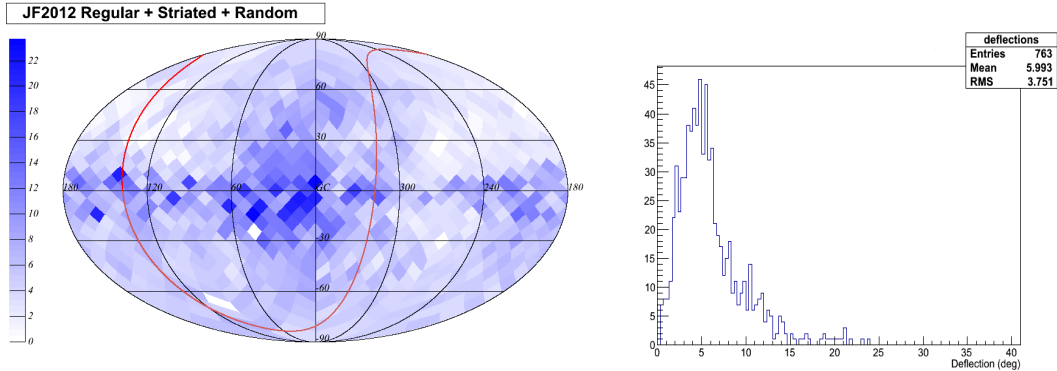


Figure 6.12: Same as in figure 6.10, but with JF regular+striated+random GMF model. In this case, due to the lower number of events, a larger pixelation was used.

The results presented above show an average deflection of the order of $\sim 6^\circ$ for protons events with average energy ~ 50 EeV. For this reason, the value for the average deflection at 100 EeV requested in the likelihood method was chosen to be $D = 3^\circ$. To take also into account the possible presence of cosmic rays heavier than protons we chose to use as well the values $D = 6^\circ$ and $D = 9^\circ$.

6.6.3 Preliminary results

Results for the cross-correlation analysis with the neutrino catalogues are shown in figure 6.13: in the two panels the relative excess of pairs with respect to isotropic expectations

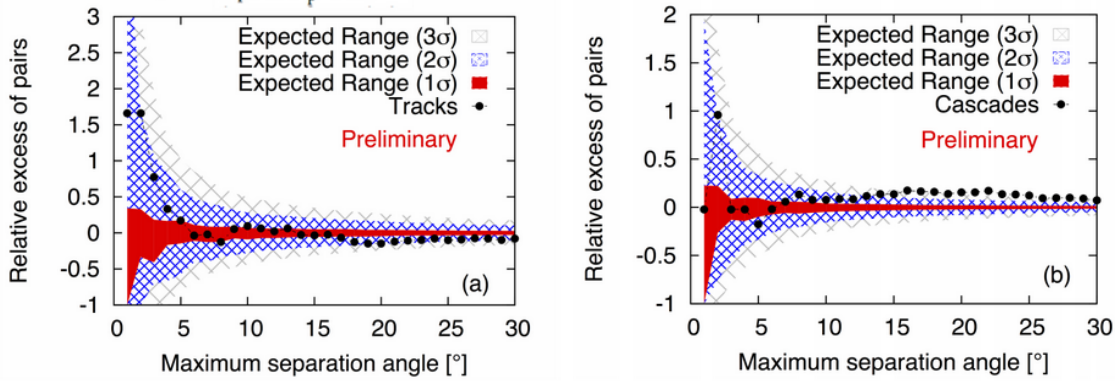


Figure 6.13: Relative excess of pairs, $[n_p^{obs}(\Psi)/\langle n_p^{iso}(\Psi) \rangle] - 1$ as a function of the maximum angular separation between the neutrino and UHECR pairs, for the analysis done with the high-energy tracks (left) and with the high-energy cascades (right). The $1-\sigma$, $2-\sigma$ and $3-\sigma$ fluctuations expected from an isotropic distribution of arrival directions of CRs are shown in red, blue and grey, respectively.

are plotted for the high energy tracks (left) and the high energy cascades (right).

For the tracks, the maximum departure from isotropy is found at 2° , where 4 pairs were observed while 1.5 were expected from isotropy. The penalized probability for finding a higher or equal number of pairs from an isotropic distribution of UHECRs is 34%.

For the cascades, the maximum departure from isotropy is found at 22° , where 575 pairs were observed while 490.3 were expected from isotropy. The corresponding penalized probability is $4.8 \cdot 10^{-4}$. This excess is driven by cascades in the excess region found by the two UHECRs experiments: the Cen A region excess in the Auger dataset and the TA hot spot⁷. To consider the effect of the UHECR distribution, the p-value has been also computed under the assumption of an isotropic distribution of neutrinos, obtaining a value of $8.5 \cdot 10^{-3}$. The results for the likelihood stacking method are shown in Table 6.1:

D	Tracks		cascades	
	n_{sCR}	pre-trial p-value	n_{sCR}	pre-trial p-value
3°	4.3	0.22	53.7	$2.1 \cdot 10^{-3}$
6°	0.5	0.48	85.7	$2.7 \cdot 10^{-4}$
9°	-	under-fluctuation	106.1	$3.8 \cdot 10^{-4}$

Table 6.1: Joint Auger-TA-IceCube Stacking analysis results.

the post-trial p-value for the most significant excess, found at $D = 6^\circ$ for cascade events, is $8 \cdot 10^{-4}$. We remind here that the value of the magnetic deflection D is computed at 100 EeV and so, taking into account also the angular resolution of the cascade events and of cosmic rays, respectively considered $\sim 15^\circ$ and 1° , this value and the 22° found in the cross-correlation analysis may indeed correspond to comparable magnetic deflections of

⁷The TA collaboration reported in [159] an indication for an excess of cosmic rays above 57 EeV in a circle of 20° around the position R.A. = 146.7° , Dec. = 43.2° . This “hot spot” has now a post-trial significance of 4σ [160].

cosmic rays.

Although none of these findings is significant, it will be interesting to see how these excesses evolve with new cosmic ray and neutrino data.

6.7 Conclusions

In this chapter we described a selection of the most recent results published by Auger on small/intermediate-scale anisotropies: the recent paper [143] and the preliminary results of the correlation with IceCube neutrinos [156].

We started by showing the results of the autocorrelation analysis that led to no significant excess with respect to isotropy. We cross-checked this result independently, and found the same result.

Then, we discussed the results of the searches for cross-correlations with astrophysical catalogues. We cross-checked these results using a different approach based on an analytical method. In table 6.2, we summarize the results from both the published method (left part) and from the analytical one (right part). We can see that for all the three studies presented here (cross-correlation with flux-limited catalogues, cross-correlation with bright AGNs and correlation with Cen A, respectively in rows 1-3,4-5 and 6) our cross-checks led to results compatible with the published ones in the case of the most significant findings (Swift AGNs, bright Swift AGNs and Cen A). It led to different minima in the search for correlations with 2MRS and the Radio-galaxies catalogue, as in fact those are not well-defined minima.

Finally, we presented the work done in collaboration with the IceCube and Telescope Array experiments, to search for correlation between cosmic rays and a catalogue of high-energy neutrinos. In particular, our contribution to the work was the study of magnetic fields deflections that were used in one of the performed analyses. The study led to interesting results when comparing the arrival direction of UHECRs above 52 EeV (57 EeV for the TA sample) with cascades neutrino events in the HESE IceCube dataset. This hint of correlation, found with both a cross-correlation and a likelihood stacked analysis, is a preliminary result: work is still in progress aimed at a joint publication.

Overall, none of the tests performed yields a statistically significant evidence of anisotropy in the distribution of UHECRs. It will be in any case interesting to follow with future data the evolution of the excesses found in the cross-correlation studies, particularly from Cen A and from the bright AGNs for cosmic ray energies in excess of 58 EeV. Also the search for correlation with neutrinos found some interesting hints of correlation that are worth being followed with future data. It is worth noting that all the most interesting excesses are found at angular scale of the order of $15^\circ - 20^\circ$.

The observed lack of anisotropy could be explained in different ways, for example, it could be due to a large number of sources. Another explanation, as already mentioned, is the possibility of having a large fraction of heavy cosmic rays, highly deflected by magnetic fields. In the following chapters we will try and develop a method in order to select light events in the dataset. We will also try to perform some of the tests presented in this chapter on such a subsample.

Objects	published analysis						Analytical analysis (this work)			
	E_{th} [EeV]	Ψ [$^{\circ}$]	D [Mpc]	\mathcal{L}_{min} [erg/s]	f_{min}	\mathcal{P}	E_{th} [EeV]	Ψ [$^{\circ}$]	D [Mpc]	\mathcal{L}_{min} [erg/s]
2MRS Galaxies	52	9	90	-	$1.5 \cdot 10^{-3}$	24%	70.3	6	96	-
Swift AGNs	58	1	80	-	$6 \cdot 10^{-5}$	6%	55.6	1	86	-
Radio galaxies	72	4.75	90	-	$2 \cdot 10^{-4}$	8%	42.1	3	75	-
Swift AGNs	58	18	130	10^{44}	$2 \cdot 10^{-6}$	1.3%	58.8	18	130	10^{44}
Radio galaxies	72	4.75	90	$10^{39.33}$	$5.1 \cdot 10^{-5}$	11%	58.9	11	110	$10^{39.66}$
Centaurus A	58	15	-	-	$2 \cdot 10^{-4}$	1.4%	58.6	15	-	-

Table 6.2: Summary of the parameters of the minima found in the cross-correlation analyses.

Chapter 7

Multi-variate analysis for mass discrimination

*As easy it was to tell black from white,
It was all that easy to tell wrong from right.
And our choices were few and the thought never hit
That the one road we traveled would ever shatter and split.*

Bob Dylan, *Bob Dylan's dream* - from *The Freewheelin' Bob Dylan* (1963)

We have seen in the previous chapter that no statistically significant anisotropy signal was found in the highest energy dataset. However, some interesting hints of an excess were found at $\sim 15^\circ$ when comparing the arrival directions of the events with energy $E \geq 58$ EeV with the positions of bright AGNs in the swift-BAT catalogs and of the closest AGN, Centaurus A. In paragraph 3.5.2, we pointed out how interpretations of Auger data through the most updated models of hadronic interactions indicate that the fraction of heavy nuclei increases above the energy of the ankle and up to the highest energies. This, with the expected magnetic deflections described in section 6.6.2, might suggest a scenario where cosmic rays, although possibly coming from anisotropically distributed sources, may be isotropized by magnetic fields. In this scenario, only light-nuclei primaries would keep information on their sources distribution.

In the light of this, we have devised a multi-variate analysis method to select candidate light nuclei events in the dataset described in chapter 5. To do this, we have used simulated showers where we have adopted, for the hadronic interactions, the two most updated, LHC-tuned, models, i.e., QGSJETII-04 and EPOS-LHC (see section 3.5.3 for details). In section 7.1 the sample of simulated events is described. In section 7.2 the choice of a set of mass-discriminating observables accessible by the Auger surface detector is presented, based on EAS properties described in chapter 2. We have then developed a Multi-Variate Analysis (MVA), to combine such observables in order to maximize the p-Fe discriminating power. We will describe the MVA, the methods used and their performances in section 7.3.

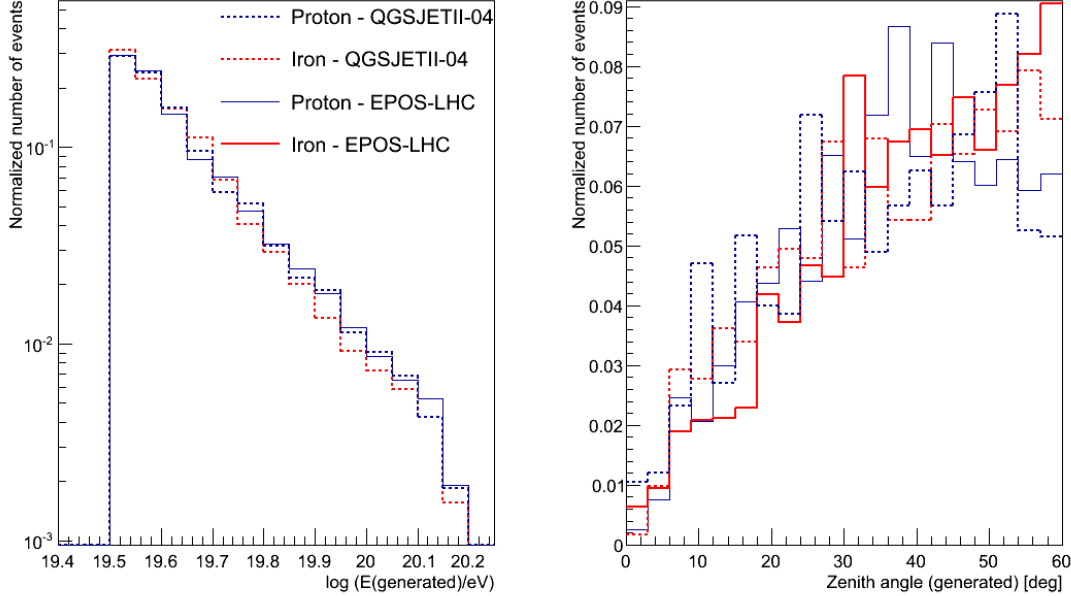


Figure 7.1: Generated energy (left panel) and zenith (right panel) distributions for the 4 sets of simulations used in this work. Histograms are normalized because of the different sizes of the samples.

We have pointed out in chapter 4 that vertical ($\theta < 60^\circ$) and horizontal showers are different in many aspects. Because of this, the reconstruction for the latter ones is different and also the mass-sensitive observables should be selected in a different way. In this work, we chose to focus on the vertical events: all the results obtained in this chapter, and in the forthcoming chapter 8, are based on the 454-Evts dataset described in chapter 5.

7.1 The sample of simulated events

In order to test the capability of separating light from heavy primaries based solely on observables of the surface detector, we have simulated events using the two most up-to-date, post-LHC, hadronic interaction models: EPOS-LHC and QGSJETII-04. We considered proton and iron primaries in this work, as a reference for light and heavy primaries. Showers have been generated through the CORSIKA software (see paragraph 2.1.3), following a $\propto E^{-1}$ energy spectrum and the angular distribution of the data. The simulated events have then been reweighed to match a power law $\propto E^{-4.3}$, that approximately describes the observed Auger spectrum above the suppression energy [161]. For illustration, the energy and zenith distributions of the simulated samples (red lines for iron, blue lines for proton, full lines for EPOS and dashed lines for QGSJET) are shown in figure 7.1, left and right panel respectively.

Each simulated shower is generated following its development from the first interaction

point to the ground. Each particle of the shower is followed, taking into account its interaction, decay and energy-loss processes until its energy is lower than a given energy threshold (thinning procedure). Such energy threshold is defined as a fraction of the primary energy, in our case $E_{thin} = 10^{-6} E_{primary}$. At lower energies, to reduce the number of particles followed during the cascade development, and thus the computing time, only a subsample of particles is processed; a weight is associated to them to properly take into account the fraction of particles that were dropped. As the simulated shower reaches the ground, the response of the detector is simulated from the Čerenkov light in the stations to the signal measured in the PMTs and finally through the electronics. Then, the same reconstruction used for real showers is performed. The full simulation of the Auger SD detector is available only in the Offline framework [162]. For this reason, in the following, this software is used, in particular the version v9r5. We remind here that the 454-Evts dataset described in chapter 5 was selected by applying a cut at 40 EeV on the energy as reconstructed by the Herald software, E_{herald} (see chapter 4 for details on the differences between the two reconstructions). For the sake of consistency with simulations, we reconstructed all those 454 events with the Offline software. We found that the energy estimated, $E_{offline}$, is greater than 32 EeV and lower than 150 EeV for all events. In the following we will use this energy boundaries to select simulated events, so that data are fully bracketed by simulations. In total, we used 5929 proton and 8517 iron events simulated with EPOS-LHC and 5046 proton and 7256 iron events simulated with QGSJETII-04.

7.2 Mass-sensitive observables

In chapter 2 we outlined the main properties of the EAS that are sensitive to the primary mass. As we have seen, heavier primaries produce showers that develop higher in the atmosphere and with a higher muon content than a lighter primary of the same energy. Neither the depth of the shower maximum nor the number of muons can be directly measured with the SD only. However, it is possible to develop techniques of analysis of SD data to obtain observables related to such characteristics of the showers. Many studies are on-going in the Auger Collaboration: we have examined these observables and selected for this work the most suitable ones for differentiating proton-induced showers from iron-induced ones at the highest energies. They are described below, following a schematic classification into observables related to the longitudinal development, to the lateral development or to the muon content of the shower.

7.2.1 Longitudinal development

The depth of the shower maximum, X_{max} , is the premiere observable used to study composition, as described in section 3.5.2. The longitudinal development of the shower is observed by the fluorescence detectors, but is not directly accessible with the surface detector. However, showers that develop at different depths show different characteristics that can be measured at ground by the SD array. The longitudinal development of a shower is shown in figure 7.2, from left to right. We can see that the characteristics of the shower vary if it is “young” or “old” (having traversed a small or large portion of

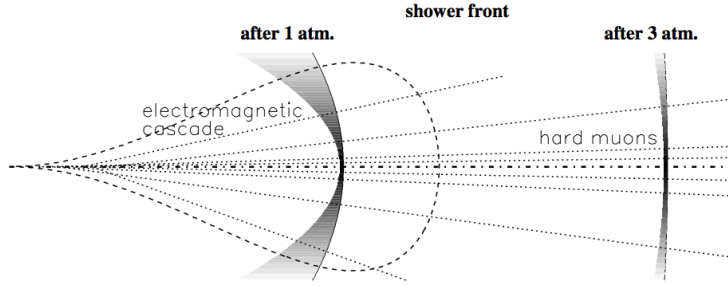


Figure 7.2: A schematic view of a shower developing in the atmosphere (from left to right). The dotted straight lines indicate the muons while the dashed lobes represent the electromagnetic component. The shaded regions represents the different shape of the shower front at two stages of the development of the shower. From [163].

atmosphere, respectively¹).

Shower front curvature

A shower in its early stages (e.g., after 1 atmosphere, as in the left part of figure 7.2) has a more curved front. In the case of a lighter primary, that interacts deeper in the atmosphere, the shower will reach the ground at an earlier stage than one induced by a heavier primary. For this reason, the radius of curvature of the shower front in a proton-induced shower is expected to be smaller than in an iron-induced one at the same energy and zenith angle. The radius of curvature of the shower front is fitted in the standard geometrical reconstruction, as was described in chapter 4, and it will be used as one of the discriminating observables.

The radius of curvature, R_c , is dependent on both zenith angle and energy in a non trivial way. A parametrization for correcting both dependences, obtained by fitting data with $E \geq 10$ EeV, was proposed in [164]. We will then use $\delta R_c \equiv R_c^{fit}/R_c$ as a p-Fe discriminating observable. Its distribution for data and for the 4 sets of simulated showers is shown in figure 7.3, left: we can see that the distributions for the proton and iron simulated showers are compatible between the two models and that a slight separation between proton and iron distributions is visible for both the QGSJET and the EPOS sets of simulated showers. Also, data points are bracketed by simulated distributions. The dependencies on energy and zenith angle are shown, for simulated events only, in the center and right panels of the same figure: we can see that, while the parametrization is successful in removing the dependence on energy, a dependence on the zenith angle is still present after the correction. This is probably due to the fact that the parametrization was made for events with $E > 10$ EeV. This effect will be reduced by binning in zenith angle the samples, as will be described in section 7.2.3.

Rise time azimuthal asymmetry (m/q)

The rise time of the detected signal, T_{50} , defined as the time the signal takes to rise from 10% to 50% of the total, can be measured by each SD station. While in highly

¹The value corresponding to 1 atm. is roughly 1000 g cm^{-2} .

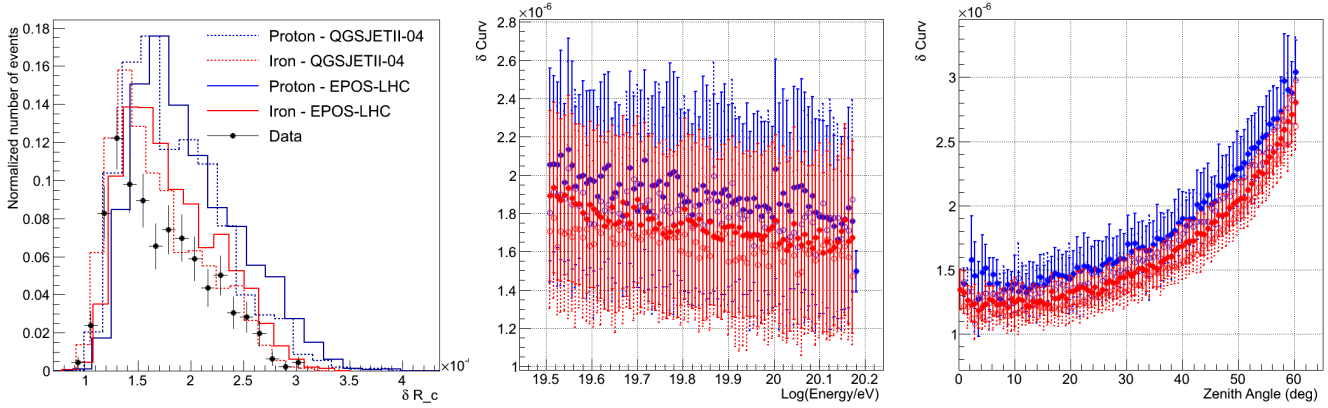


Figure 7.3: Left: distribution of δR_c (see text) for simulated events with $E > 32$ EeV (histograms) and data (dots). Center and right: its dependency on energy and zenith angle, respectively.

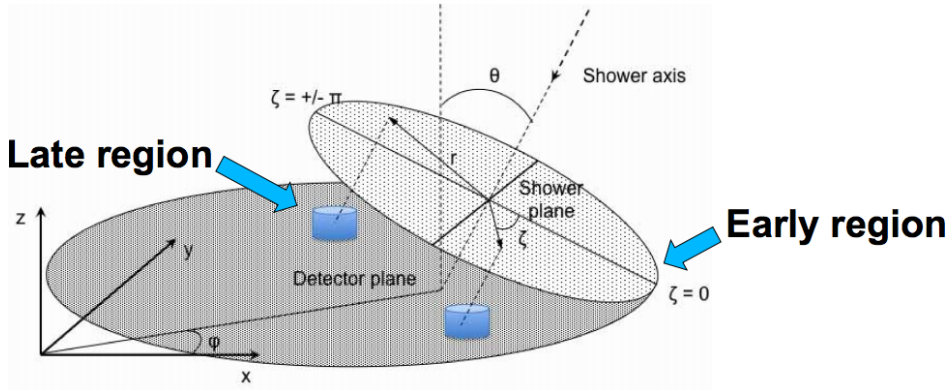


Figure 7.4: Schematic view of the shower geometry as it hits the ground. The incoming direction of the primary particle defines two regions: early ($\pi/2 < \zeta < \pi/2$) and late ($\pi/2 < \zeta < \pi/2$). Note the different amount of atmosphere traversed by the particles reaching the detectors in each region.

vertical showers the SD signals show a perfect circular symmetry around the polar angle ζ in the shower plane, in more inclined showers particles reaching late detectors ($\pi/2 < \zeta < \pi/2$, see figure 7.4) traverse longer atmospheric paths than those arriving to early detectors ($\pi/2 < \zeta < \pi/2$). Consequently, for inclined showers, both the magnitude and rise time of the signals depend on ζ . The observable described here is based on the above-mentioned rise time azimuthal asymmetry. This is due to two effects. On the one hand, a contribution comes from the absorption of the electromagnetic signal. Since the particles that reach late detectors traverse longer atmospheric paths, we expect a larger attenuation of electrons and photons as compared to early detectors. On the other hand, there are also contributions to the asymmetry from geometrical effects. In this case, not only is the electromagnetic component important, but muons also play a role. Although track length and solid angle effects are compensated by the near-cubical design of the detectors, the angular distributions of muons impinging on the detectors are different, as

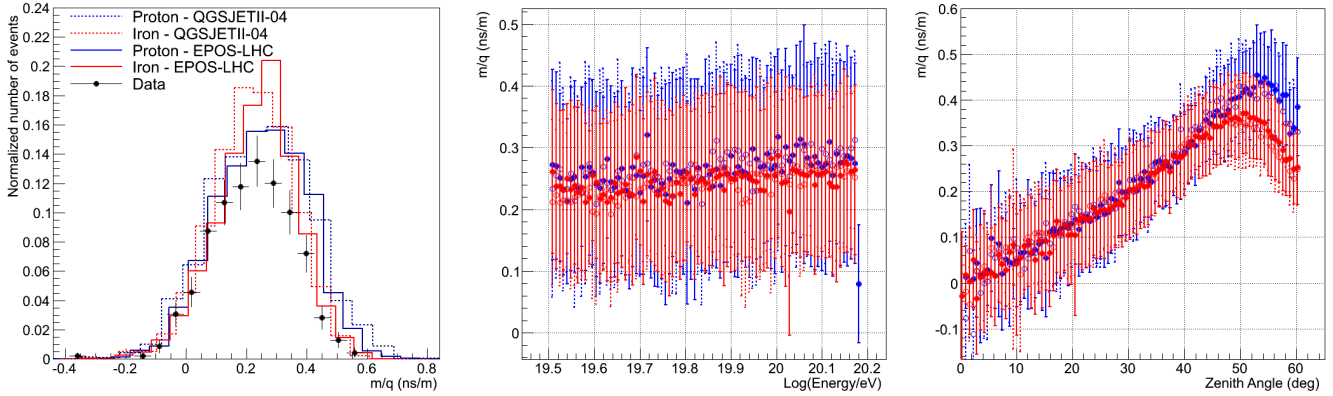


Figure 7.5: Left: distribution of m/q (see text) for simulated events with $E > 32$ EeV (histograms) and data (dots). Center and right: its dependency on energy and zenith angle, respectively.

late detectors record more muons emitted closer to the shower axis [165]. Geometrical effects predominate at low angles, while for showers with $\theta > 30^\circ$ attenuation effects are the main contribution. The rise time asymmetry dependence on θ is correlated with the longitudinal development of the shower [165]. As such it is sensitive to the primary mass.

We use the following parametrization of the asymmetry of the rise time [166]:

$$\frac{T_{50}}{R} = q + m \cdot \cos\zeta \quad (7.1)$$

where R is the station distance from the core, ζ its azimuthal angle on the shower plane (the plane perpendicular to the shower axis) and the parameters are fitted using the stations used in the reconstruction. The mass-sensitive observable that we will use is the ratio m/q .

The distribution of this observable is shown in figure 7.5, left panel for the simulated events (histograms) compared to data (dots): no clear p-Fe separation is visible in this case. The distributions for EPOS and QGSJET simulated events are compatible, and data points too follow the same distribution. This observable is independent on the energy while it shows a complex dependence on the zenith angle as anticipated (see figure 7.5, right and left panels, respectively). This dependence has not been studied nor corrected for this work, as in fact this observable achieves the best p-Fe discriminating power only for showers above $\theta = 30^\circ$, where it is due to the shower absorption. The effect of this dependence will be further reduced by the use of bins in zenith, as already mentioned.

7.2.2 Lateral development

Number of candidate stations

The number of candidate stations (NCS) used for the reconstruction of the event is an observable sensitive to the shower lateral development. The lateral trigger probability (LTP) as a function of distance from the shower core is higher for heavier primaries [167], as one can see in figure 7.6, where the LTP is shown for showers of 10 EeV in

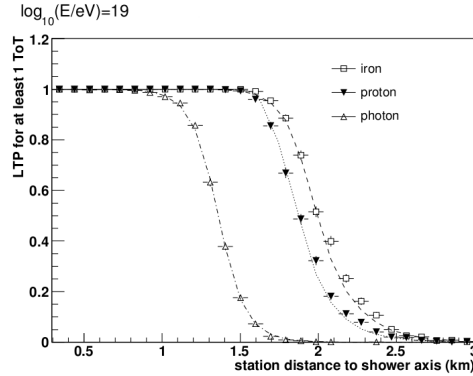


Figure 7.6: Probability for a cosmic-ray shower to trigger a station as a function of distance from the shower core (Lateral Trigger Probability). From [167].

case of different primaries. This effect is due to the fact that the station triggers are more sensible to the muonic component, which is larger in showers induced by heavier primaries. Thus, on average, an iron-induced shower will trigger a higher number of stations than a proton-induced one arriving at the same zenith angle with the same energy.

The number of candidate stations is directly dependent on the zenith angle, θ , because of the projection of the shower footprint on the ground plane. This dependence can be removed by using $\delta\text{NCS} \equiv \cos^2 \theta * \text{NCS}$, as we will do in the following. NCS is also expected to be dependent on the primary energy, in a less trivial way. The distribution of this observable is shown in figure 7.7, left panel, for the simulated events (histograms) compared to data (dots). We note that there are less candidate stations in proton-induced showers than in iron-induced ones, as expected, both in EPOS and QGSJET-based simulated events, that show coherent results. The data distribution is also nicely bracketed by simulations. The residual dependencies on zenith angle and energy are shown in the center and left panels, respectively. A clear dependence is visible in both plots. This effect will be reduced by the energy and zenith binning described in 7.2.3.

Slope of the lateral distribution function (β)

The detected signal as a function of distance from the shower core is parameterized by the LDF, as defined in paragraph 4.4.2, where β is the negative parameter representing the slope of the LDF. Muons survive at larger distance from the shower axis than the electromagnetic component. A proton-induced shower, that is poorer in muons, will be more concentrated around the shower axis and, therefore, its LDF will be steeper than the one for an iron-induced shower. In the standard Offline reconstruction, $\beta_F(E, \theta)$ is not fit, but parameterized as a function of energy and zenith angle. However, high energy events have enough triggered stations to fit the LDF with β as a free parameter. For the aim of our study, we thus re-perform the LDF fit, using the core position and arrival direction as obtained from the standard reconstruction, leaving S_{1000} and β as the only free parameters.

The slope of the LDF is dependent on both zenith angle and energy. This dependence,

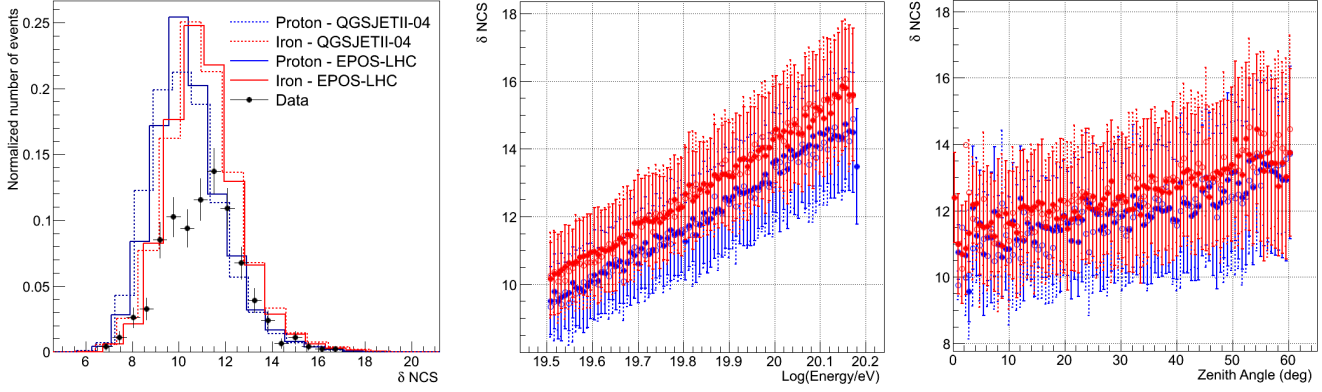


Figure 7.7: Left: distribution of δNCS (see text) for simulated events with $E > 32$ EeV (histograms) and data (dots). Center and right: its dependency on energy and zenith angle, respectively.

as mentioned before, has its own parametrization (β_F). We will then use in the following

$$\delta\beta = \frac{\beta - \beta_F}{\beta_F} \quad (7.2)$$

as p-Fe discriminating observable in the following. Its distribution for simulated and real events is shown in figure 7.8, left: also in this case simulated events from both the EPOS and QGSJET sample are consistent, with a slight separation between proton and iron distributions. Data points are well bracketed by simulations. From the plots in the middle and right panel, that show the energy and zenith dependence, respectively, we can see that the correction via the parametrization is effective in reducing both dependencies of this observable.

The parameter S_b

Another shower width-related observable proven to be mass sensitive [168] is a weighted sum of the signals in the stations defined as:

$$S_b = \sum_{i=1}^N \left[S_i \cdot \left(\frac{r_i}{r_0} \right)^b \right] \quad (7.3)$$

where $r_0 = 1000$ m, S_i is the signal in the i -th station, r_i its distance from the shower axis and the sum runs over all stations lying between 500 m and 2000 m from the shower core to avoid saturated signals. It was proven in [168] that, in order to achieve the best proton-iron discrimination power, the parameter b should be set to 3. The S_b parameter is sensitive to the global LDF shape.

By replacing in equation 7.3 the NKG parametrization of the LDF (described in section 4.4.2) for the signal S_i of each station we obtain:

$$S_b = \sum_{i=1}^N \left[S_{1000} \left(\frac{r_i}{1000 \text{ m}} \right)^\beta \left(\frac{7000 \text{ m} + r_i}{1000 \text{ m} + 7000 \text{ m}} \right)^{\beta+\gamma} \left(\frac{r_i}{1000 \text{ m}} \right)^b \right] \quad (7.4)$$

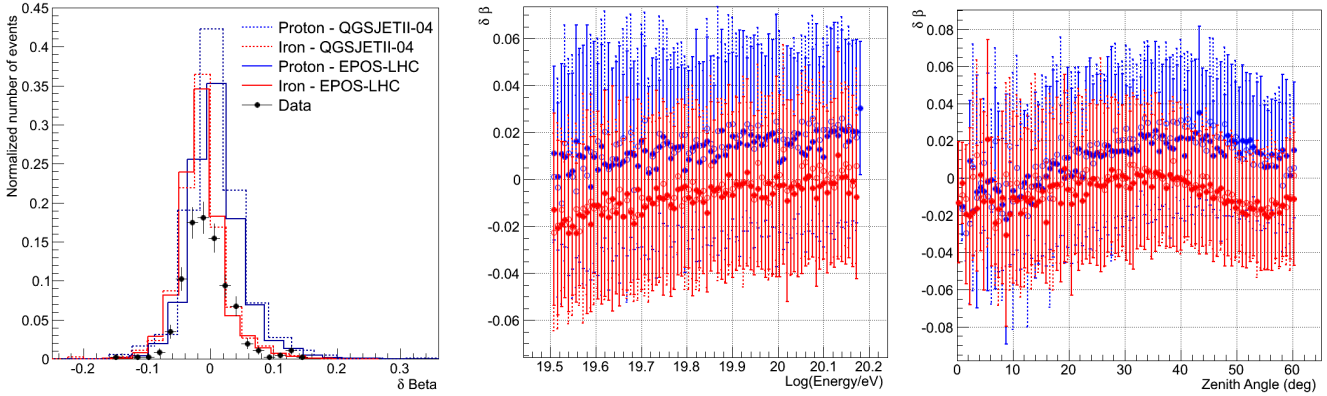


Figure 7.8: Left: distribution of $\delta\beta$ (see text) for simulated events with $E > 32$ EeV (histograms) and data (dots). Center and right: its dependency on energy and zenith angle, respectively.

After rearranging it, we get:

$$S_b = S_{1000} \sum_{i=1}^N \left[\left(\frac{r_i}{1000 m} \right)^{\beta+b} \left(\frac{7000 m + r_i}{1000 m + 7000 m} \right)^{\beta+\gamma} \right] \quad (7.5)$$

From equation 7.5 we see that S_b is directly dependent on S_{1000} and thus on energy. To remove this dependence, we will use $\frac{S_b}{S_{1000}}$ as a p-Fe discriminating observable in the following. An empirical correction for the zenith dependence can also be applied by further multiplying the observable for $\cos^2 \theta$. The distribution of the final observable,

$$\delta S_b \equiv \frac{S_b \cos^2 \theta}{S_{1000}} \quad (7.6)$$

is shown in figure 7.9, left for both EPOS and QGSJET simulated events and for data. We can see that the two proton and the two iron distributions are nicely superimposed while being slightly separated from each other. Data points are bracketed by simulations. In the middle and right plots we can see the dependencies on energy and zenith: the corrections we applied are effective and only small dependencies are left.

7.2.3 Muon content

Muonic signal at 1500 m

The trace measured in each station is different in the case of a muonic signal or an electromagnetic one. Muons tends to generate spike-like and short signals, while the electromagnetic signals are smoothed and elongated, as we can see in figure 7.10, where a PMT trace in a station hit by a simulated shower is shown. The blue trace is the muonic component, the red one is the electromagnetic one.

Different algorithms have been concieved inside the Collaboration to extract information on the muonic and/or electromagnetic signal in each station based on the temporal

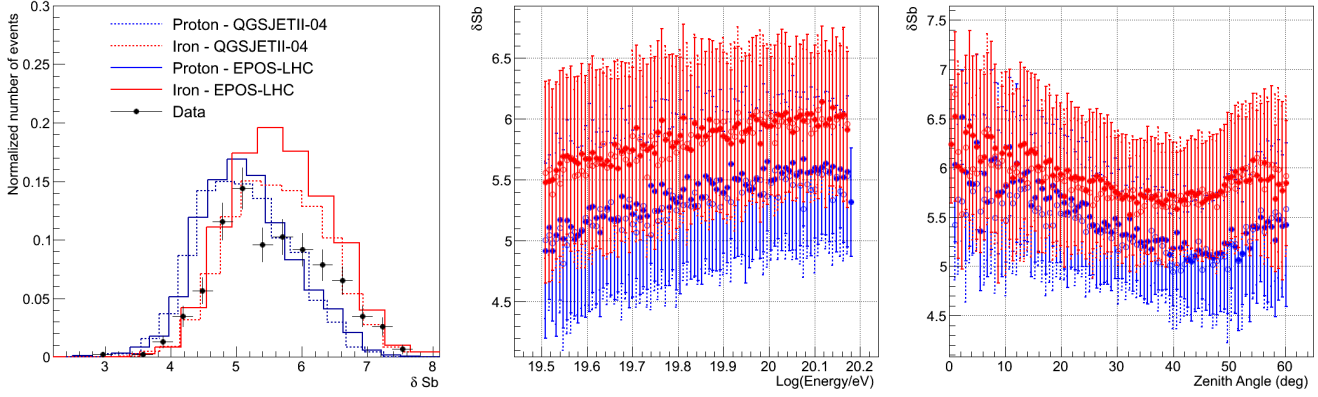


Figure 7.9: Left: distribution of δS_b (see text) for simulated events with $E > 32$ EeV (histograms) and data (dots). Center and right: its dependency on energy and zenith angle, respectively.

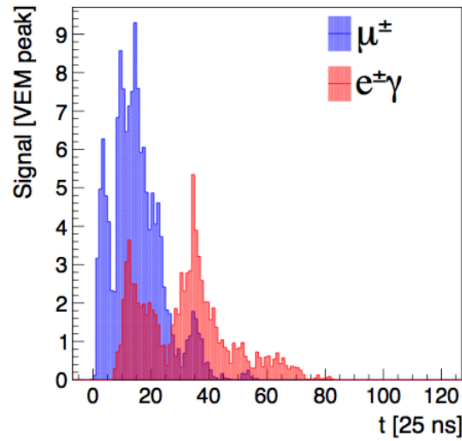


Figure 7.10: The trace generated in a station by the muonic (blue) and electromagnetic (red) components of a simulated shower. Courtesy of Laura Collica.

structure of the signal trace. For this work, we have used the so-called smoothing technique described in [169]: for each bin in the averaged trace of the three PMTs in each station, we compute the average signal in the N_{bin} bins before and after the considered bin, obtaining

$$S^{smoo}(i) = \sum_{k \in [i-N_{bin}, i+N_{bin}]} S(k) / (2N_{bin} + 1) \quad (7.7)$$

If $S^{smoo}(i) > S(i)$, we attribute the exceeding signal, $S_\mu(i) = S^{smoo}(i) - S(i)$, to the muonic component and we subtract it from the total signal of that bin. In this way, we obtain a “smoothed” trace, that will undergo again the same procedure N_{iter} times². Finally, we estimate as the electromagnetic signal of the station the integral of the final smoothed trace. The muonic signal of the station is obtained as the subtraction of the electromagnetic signal from the total one.

²In our case, $N_{iter} = 4$

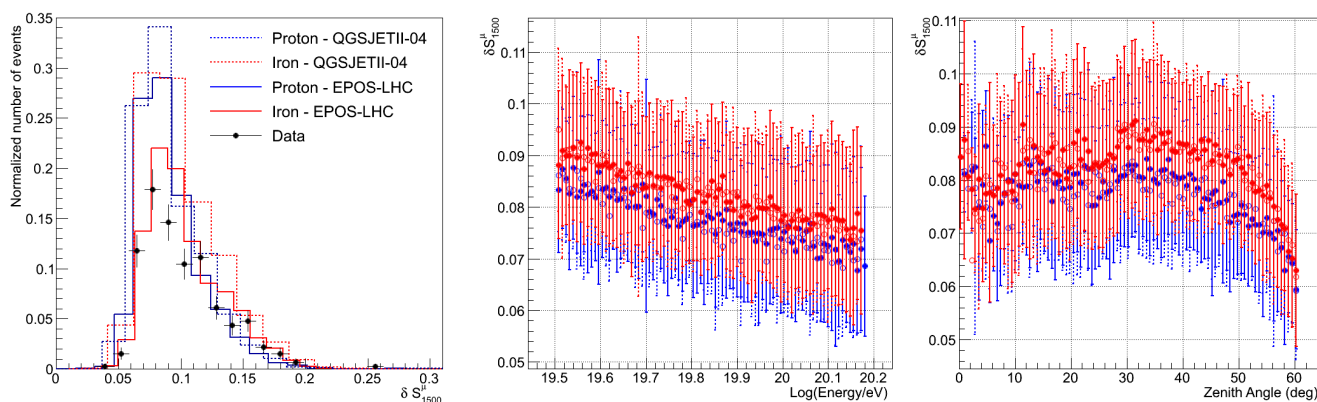


Figure 7.11: Left: distribution of δN_{1500}^μ (see text) for simulated events with $E > 32$ EeV (histograms) and data (dots). Center and right: its dependency on energy and zenith angle, respectively.

We use only non-saturated dynode (high-gain) signals, in a range of distances from the core depending on the primary energy. In our case, the optimal range was found to be from 800 to 3000 meters. Using the muonic signal in stations within this range, we fit a muonic LDF and extract the signal at a reference distance S_{1500}^μ as a proxy of the number of muons. We chose the reference distance to be 1500 m, well within the selected distance range.

S_{1500}^μ is related to the normalization parameter of the total LDF, S_{1000} , and thus on energy. We reduced this dependence through the ratio S_{1500}^μ/S_{1000} . S_{1500}^μ depends also on the zenith angle, as S_{1000} does. We corrected the θ dependence by applying the same correction applied on S_{1000} via the CIC procedure (see section 4.5 for details), obtaining the final corrected observable, δS_{1500}^μ . In figure 7.11, left panel, we show the distributions of δS_{1500}^μ for simulated showers (histograms) and real events (dots). The distributions for proton and iron obtained with the two hadronic interaction models are nicely superimposed, and they bracket the data points. From the center and right panels of the same figure, we can see that the applied corrections largely removed the energy and zenith angle dependency of this observable.

Rise time at a reference distance

We have already mentioned that the rise time of the signal, T_{50} , is a mass-sensitive observable. We expect that, at a given energy, the rise time at a reference distance should be different for proton and iron-induced showers. This is related to the different time shape of muon and electromagnetic-induced traces, as described in the previous paragraph: since the muons produce shorter signals, if the muonic component is larger, as in iron-induced showers, we expect a faster rise time than in muon-poorer showers, like proton-induced ones.

This observable is dependent on energy and zenith angle in a non-trivial way. A parametrization of such dependency, based on data, is provided in the Offline software.

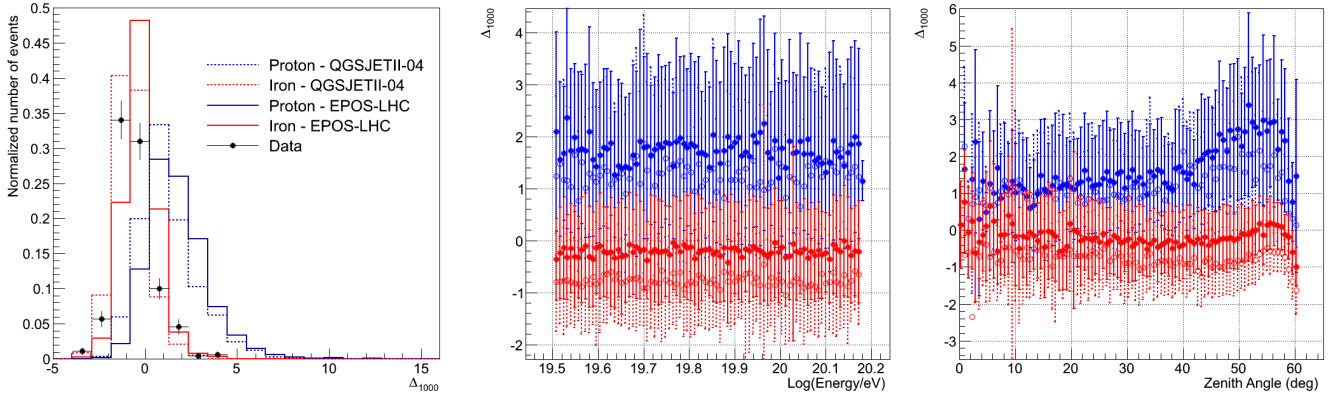


Figure 7.12: Left: distribution of Δ_{1000} (see text) for simulated events with $E > 32$ EeV (histograms) and data (dots). Center and right: its dependency on energy and zenith angle, respectively.

We will then use for our study the so-called Δ_{1000} observable, computed as

$$\Delta_{1000} = (T_{50}^{fit}(1000) - T_{50}^{par}(1000)(\theta, S_{1000})) / RMS(T_{50}^{par}(1000)(\theta, S_{1000})) \quad (7.8)$$

where $T_{50}^{fit}(1000)$ is the rise time at 1000 m, extracted from a fit over all the stations, and $T_{50}^{par}(1000)(\theta, S_{1000})$ is the parametrization value for the reconstructed θ and S_{1000} . More details can be found in [170]. In figure 7.12, left, we show that the two distributions for simulated proton events (blue histograms) are nicely superimposed, as iron ones (red histograms) are. The proton and iron distributions are, on the other hand, well separated, showing the effectiveness of this observable in separating proton from iron events. The data points (black dots) are well bracketed within the four simulation distributions. The center and right panels of the same figure show the energy and angular dependencies of Δ_{1000} , respectively. We can see that the used parametrization successfully removes the most part of the dependencies.

Zenith angle and energy binning

We have shown that observables may have dependences on the zenith angle, and that this dependence may be different for proton and iron events, so that the best mass-discriminating observables for very vertical showers might not be the same as for showers with zenith angle near 60° . For this reason, and to further reduce the zenith dependency of the observables, we binned our simulation and data sample in zenith bins and worked independently in each one of them. The binning was done in fixed steps in $\cos^2 \theta$ in order to have roughly the same number of events in each bin. After trying different binnings, we found that the best compromise between small bins, for better reducing the dependences, and a large enough number of simulated events in each bin, crucial for the multi-variate analysis, was the division of the datasets in 6 bins of size 0.125 in $\cos^2 \theta$.

The observables may show also an energy dependence. We thus divided the sample also according to the energy, in only 2 bins in this case, in order to have a large enough number of simulated events in each of them. The final purpose of this analysis is to re-perform

Bin:	1	2	3	4	5	6	Total
Zenith Range	[0°, 21°)	[21°, 30°)	[30°, 38°)	[38°, 45°)	[45°, 52°)	[52°, 60°]	[0°, 60°]
High Energy Bin ($E > 50$ EeV)							
EPOS Proton	819	824	667	712	725	665	4412
EPOS Iron	995	919	910	927	967	1039	5757
QGSJET Proton	650	639	618	620	546	617	3690
QGSJET Iron	830	835	730	851	774	1043	5063
Low Energy Bin ($32 < E < 50$ EeV)							
EPOS Proton	380	419	375	346	431	423	2347
EPOS Iron	380	415	510	535	539	772	3151
QGSJET Proton	475	475	358	352	378	384	2422
QGSJET Iron	475	509	470	391	569	605	3019

Table 7.1: Number of simulated events in the high energy bin (top 4 lines) and in the low energy bin (bottom 4 lines) in each zenith bin.

on proton-like events the two anisotropy tests that were the most significant ones in the analyses presented in chapter 6: the correlation with Centaurus A and with bright Swift AGNs. Both of these excesses were found for events with $E_{herald} \geq 58$ EeV: we found it convenient to divide the sample in order to have all of these events in the same energy bin. Since the binning is made according to the Offline energy, we use $E_{offline} = 50$ EeV as the energy to separate the two bins, being ~ 50 EeV the minimum $E_{offline}$ for the events with $E_{herald} \geq 58$ EeV. In the following, we will refer to the two bins as “low-energy bin” ($32 \text{ EeV} < E_{offline} \leq 50 \text{ EeV}$) and “high-energy bin” ($50 \text{ EeV} \leq E_{offline} \leq 150 \text{ EeV}$). Table 7.1 summarizes the number of simulations for proton and iron-induced showers generated with EPOS-LHC and QGSJETII-04 available in each zenith and energy bin.

Finally, we have checked the p-Fe separation capabilities for each observable, after applying the zenith and energy dependence corrections, and including the zenith and energy binning. To do that, we needed to quantify the separation between the distributions of each observable for proton and iron simulations, that, in our case, will be considered as “signal” and “background”, respectively. This can be done in different ways, each suitable for different applications. In our case, our aim is to obtain a selection of proton-like events in the sample, with as low iron-like contamination as possible. Defining the purity obtained for a given cut V_c on the observable value V as:

$$P = \frac{S(V > V_c)}{S(V > V_c) + B(V > V_c)} \quad (7.9)$$

where S and B are the number of signal and background events, and the efficiency as:

$$\mathcal{E} = \frac{S(V > V_c)}{S(\forall V)} \quad (7.10)$$

we quantify the separation as the efficiency at 80% purity. This means that for each observable we set a cut that selects a 80% pure sample, i.e., the sample selected by this cut is composed of 80% signal (proton) simulations and 20% background (iron) ones.

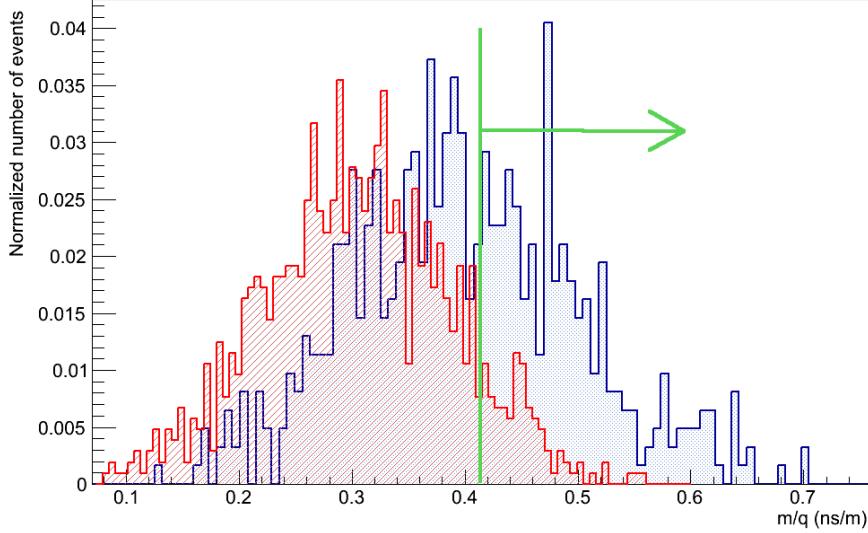


Figure 7.13: Distribution of the rise time azimuthal asymmetry in the zenith bin 6 ($52^\circ - 60^\circ$) and high energy bin ($E > 50$ EeV) for proton (blue) and iron (red) QGSJETII-04 simulated events. The green line represents the cut set to have 80% purity (see text for details).

Then, we evaluate the percentage of the selected signal sample with respect to the total one. An example is given in figure 7.13, where the distribution of the rise time azimuthal asymmetry in the zenith bin 6 and high energy bin are shown for proton (blue) and iron (red) QGSJETII-04 simulated events. The green line represents the cut to get 80% purity. The corresponding efficiency in this example is 41%, i.e., 41% of the total proton sample is selected by this cut.

We show in figure 7.14 the values for the efficiency at 80% purity ($\mathcal{E}(V_c^{80})$) for the unbinned zenith range (1-6) and as a function of the zenith angle bins, for the QGSJETII-04 (left plot) simulated events and for the EPOS-LHC ones (right plot). Upper and lower panels correspond to the high and low energy bins, respectively. We can see that Δ_{1000} is the best discriminator in every zenith and energy bin, with efficiency higher than 80% for the higher zenith values. In general, a better performance is obtained in the higher θ bins for all observables: this is due to the fact that, because of the $\cos^2 \theta$ -flat division, the size of the bins decreases with θ , the first bin being 21° large and the fifth one only 7° . As expected, the rise time azimuthal asymmetry (m/q) is effective only in the higher zenith bins, in particular in the 6th one where it reaches and efficiency $\sim 40\%$ in both simulated samples. No observable reaches $\mathcal{E}(V_c^{80}) > 50\%$ with the exception of the radius of curvature and Δ_{1000} . By comparing the efficiencies in the unbinned zenith range (left points, marked 1-6) and the ones in the single zenith bins, we can see that all the observables have in at least one zenith bin an efficiency higher than in the unbinned zenith range. A remarkable case is the radius of curvature, that shows in each zenith bin an higher efficiency than in the unbinned zenith range.

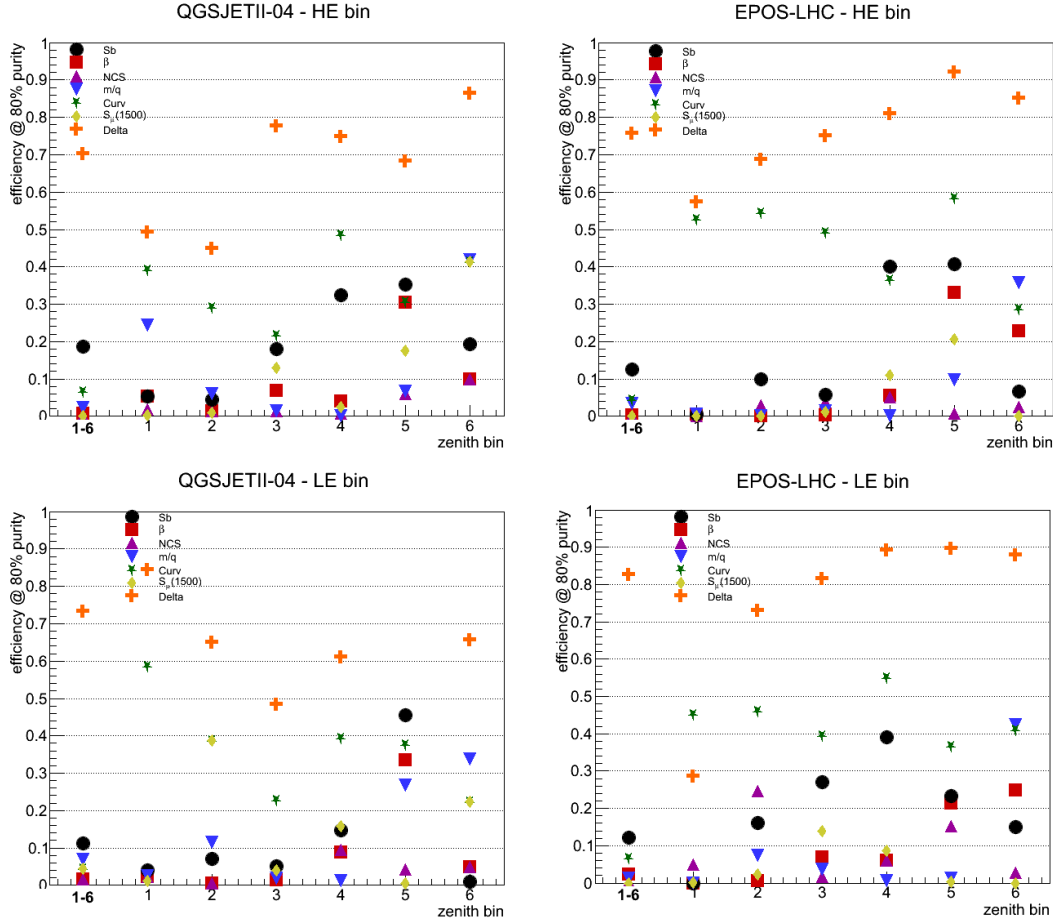


Figure 7.14: The values for the efficiency at 80% purity for the considered discriminating observables (see text for details), for the unbinned zenith range (1-6) and in single zenith bins for QGSJETII-04 (left) and EPOS-LHC (right) simulations, for the high energy (top) and low energy (bottom) bins.

7.3 Multi-variate analysis

The observables described above, all mass-sensitive, although with different efficiencies, can be combined to enhance the p-Fe discrimination capability. An appropriate tool to do so is the MultiVariate Analysis (MVA). MVA is a general term that denotes a variety of methods capable to train themselves on how to separate signal from background events (in our case, proton from iron-induced showers) once they are given a set of classified events (in our case, simulated events) and a list of discriminating observables (in our case, the mass-sensitive ones) available for each event.

The multi-variate statistical analysis has been developed starting from the last decades of the 20th century and is currently a very active research field. The need of processing large quantities of data and discriminating between classes of events which have very similar experimental signatures has brought MVA to be largely used in the physics community.

The most common package for MVA in the particle physics field is TMVA [171], which

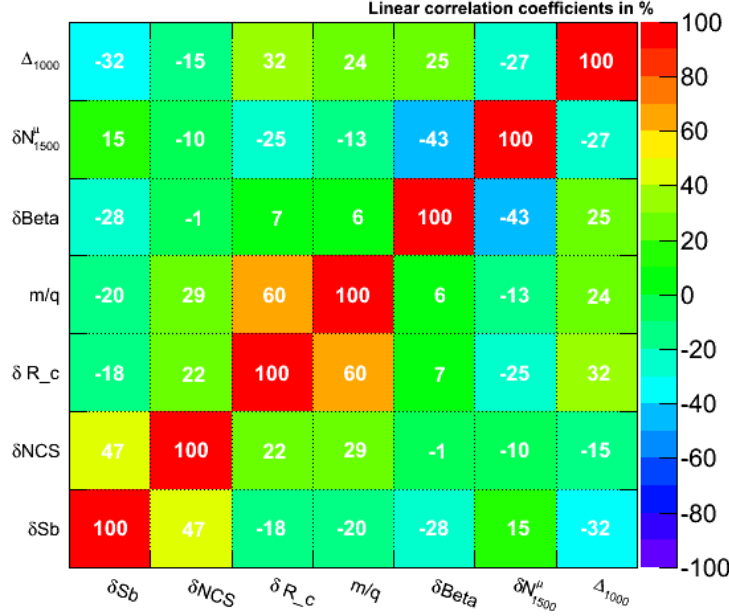


Figure 7.15: The linear correlation coefficients between all the selected observables, for EPOS simulated showers from a 50%p- 50%Fe mixture of primaries.

is part of CERN’s ROOT distributions. In particular, in the following, we will use ROOT 5.32/00 and TMVA Version 4.1.2.

Among the different MVA methods, Boosted Decision Trees (BDT) and Artificial Neural Networks (ANN, in particular TMVA’s Multi-Layer Perceptron, MLP) are widely used in high-energy astrophysics and also inside the Auger Collaboration. In the following, we will exploit them together with the Support Vector Machine (SVM) method, that we chose because it gave performances that are comparable to BDT and MLP methods in our case. All these methods are capable of dealing with correlated observables. Indeed, the selected observables are all related to intrinsically connected characteristics of the shower, and for this reason some of them may be correlated with others. We show in figure 7.15 the linear correlation coefficients between all the selected observables, for a 50% mixture of EPOS proton and iron simulated events, with no binning applied. As expected, one can see that some observables are highly correlated, such as the curvature with the rise time azimuthal asymmetry (60% correlation), S_b and the number of candidate stations (47%) or N_{1500}^μ and β (-43%). This observation imposes the use of MVA methods able to deal with correlated observables. The three methods described below are indeed of this kind.

Boosted decision trees

The simplest MVA method is the so-called cut selection: it scans on all the possible values of C_{min} and C_{max} for each observable, finding the cut values that select the region where

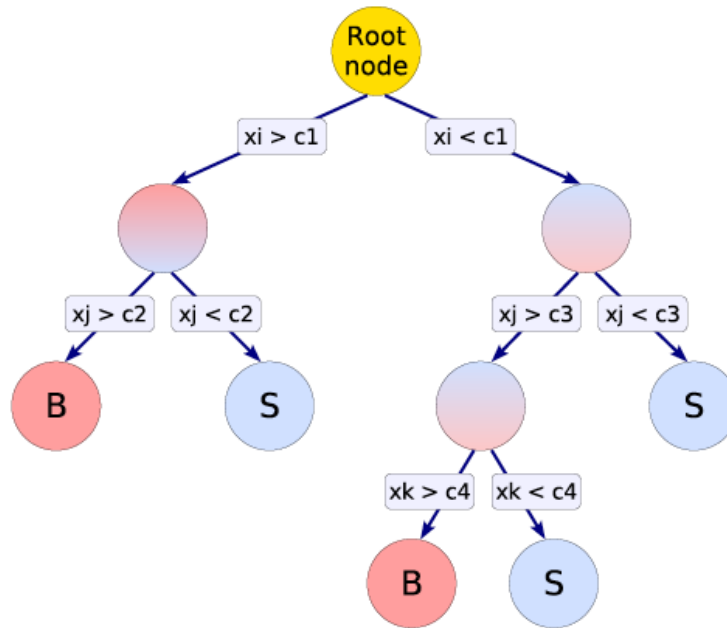


Figure 7.16: An example of a decision tree: starting from the root node, a sequence of binary splits using the discriminating observables x_i is applied to the data. Each split uses the observable that at this node gives the best separation between signal and background when being cut on. The same observable may thus be used at several nodes, while others might not be used at all. The leaf nodes at the bottom end of the tree are labeled S for signal and B for background depending on the majority of events that end up in the respective nodes. From [172].

the ratio of signal over background, usually quantified as $S/\sqrt{(S+B)}$ where S is the number of signal events and B the number of background ones, is the highest. The cut selection would select a single region which is an hyper-rectangle in n dimensions, where n is the number of observables. The cut approach can be generalized by trying to divide the whole space in more regions (hyper-rectangles) that can be labeled as “signal” or “background” regions. This is the principle of the so called decision trees methods. It is achieved by applying a succession of binary cuts (the tree) on the training sample, adjusting such cuts in order to maximize the signal-background separation in the final regions (often called leaves). The tree can then be used to test if an independent event falls in a signal or background leaf. An example of a decision tree is shown in figure 7.16. The main issue in using tree-based methods is the high variance, i.e. the possibility of having very different results with a small change in the training sample; this problem is solved using particular algorithms such as the so-called boosting.

Boosting is a procedure that can be, in principle, applied to all MVA methods: after a training, all the events are weighted so that during a second training procedure their effect is higher or lower with respect to the previous one. The most popular boosting algorithm is the so-called AdaBoost [173] (adaptive boost), where events that were misclassified during the training of a tree get a higher event weight in the training of the following

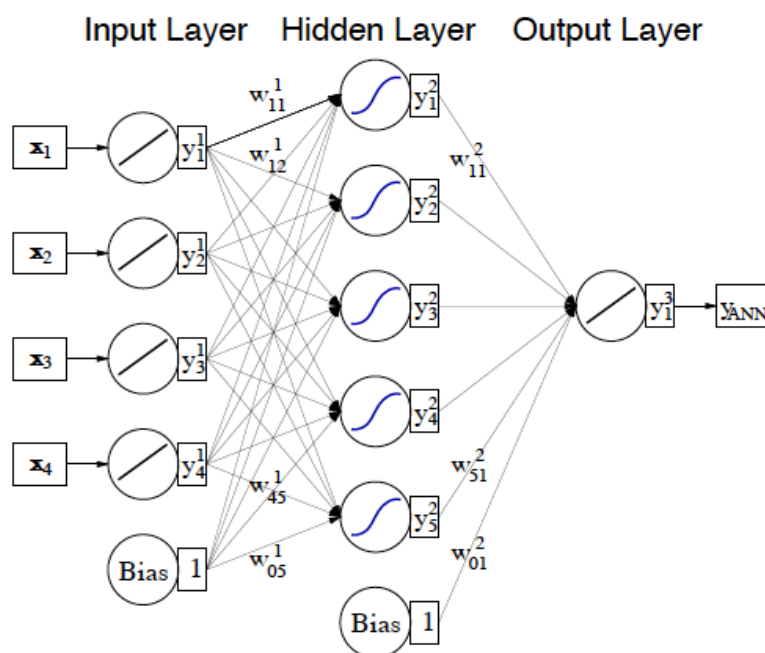


Figure 7.17: An example of a multilayer perceptron with one hidden layer. The first layer (input) is composed of four nodes, one for each input observables, the second layer (hidden) is the hidden one, composed of five neurons and the final layer (output) is the output one, with only the output observable y_{ANN} . Each neuron is connected with all subsequent neurons where upon each bonding has its own weight w as indicated. Two bias nodes are also introduced. They are connected with nodes in a higher layer node, but no node, regardless of type, connects to any bias node. They account for some special cases that cause the algorithm to break, namely when all node weights approach zero. From [172].

trees. This way, a forest of decision trees is produced: the events to be classified will then be processed by all the trees in the forest and the final output will be a proper average of results of each tree.

Multi-layer perceptrons

A more complex estimator that is also able to manage with highly correlated observables, or observables which have a very poor discrimination power, is the Artificial Neural Network (ANN). The ANN is a family of statistical learning models inspired by biological neural networks, first formulated by Warren McCulloch and Walter Pitts in 1943 [174]. ANN are composed by n nodes called neurons, that are linked to each other: each neuron produces a certain response at a given set of input signals. The connections have numeric weights that can be tuned via the training procedure. ANN are commonly used, for example, in handwriting or speech recognition softwares. To reduce the complexity of the structure, that would in principle scale as n^2 , a common approach, called Multilayer Perceptrons (MLP), is to organize the neurons in layers and only allow direct connections from a given layer to the following (see figure 7.17 for a sketch). The first layer is the input

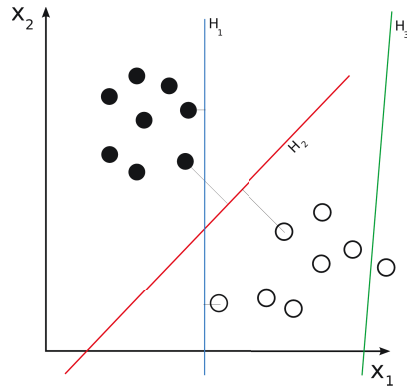


Figure 7.18: A description of the basic principle of SVM: while line H_3 does not separate the two populations of dots, both H_2 and H_1 do. However, only H_2 maximizes the distance between the line and the two populations.

layer, that holds the input observables $x_1, x_2, \dots, x_{nvar}$. Subsequent layers are hidden layers while the last one holds only the output observable, the neural net estimator M_{ANN} . The optimal configuration of hidden layers and number of neurons has to be determined for each classification problem individually.

The neural network starts its training nearly blindly, but it can tell how wrong it is based on how far off its answers are, and adjust the weights a little to make it more correct the next time. This is done for a fixed number of loops (selected by the user). It is important to note that one of the parameters that can be tuned in this method is the learning rate, η , i.e., how important are considered the mis-classification in each loop. If the learning rate is too high, the method risks to focus on unphysical fluctuations of the training sample (overtraining), while, if it is too low, the method may not converge. A common approach is to use a decaying learning rate, i.e., η is not fixed, but it is higher in the first loops, where the risk of overtraining is negligible, and then decreases.

Support vector machines

Support vector machines, or kernel machines, are a class of MVA methods developed by Vladimir Vapnik and his team in the 90s [175]. The main idea of the SVM is to build a hyperplane that separates signal and background vectors (events). The position of the hyperplane is obtained by maximizing both the separation between signal and background populations and also their “distance” to the plane. A sketch of how this works in 2 dimensions is proposed in figure 7.18: while line H_3 does not separate the two populations of dots, both H_2 and H_1 do. However, only H_2 maximizes the distance between the line and the populations. A detailed description of the SVM formalism can be found in [172].

7.3.1 Methods training and fine-tuning

The first step to deal with MVA, and probably the most important one, is the training of the methods. Given the results presented in section 7.2.3, the MVA training has been implemented independently in each zenith and energy bin. TMVA performs automatically

the training of the method and subsequently tests it with an independent simulation sample. The splitting of the given simulation sample in the training and test samples is done randomly by TMVA itself.

The use of an independent test sample is important to check if the methods suffer from the so-called overtraining. The overtraining happens when a method uses unphysical fluctuations in the training sample as discriminating features. We show as an example in figure 7.19 the MVA output distributions for the test and train samples for MLP and SVM (with the final fine-tuned parameters as described below) in zenith bin 6 of the high energy bin, trained on EPOS-LHC simulated events. The train-sample distributions (dots) for both signal and background are nicely superimposed to the correspondent test-sample distributions (histograms). This means that the method does not suffer from overtraining, since the MVA response is compatible for the test and train samples.

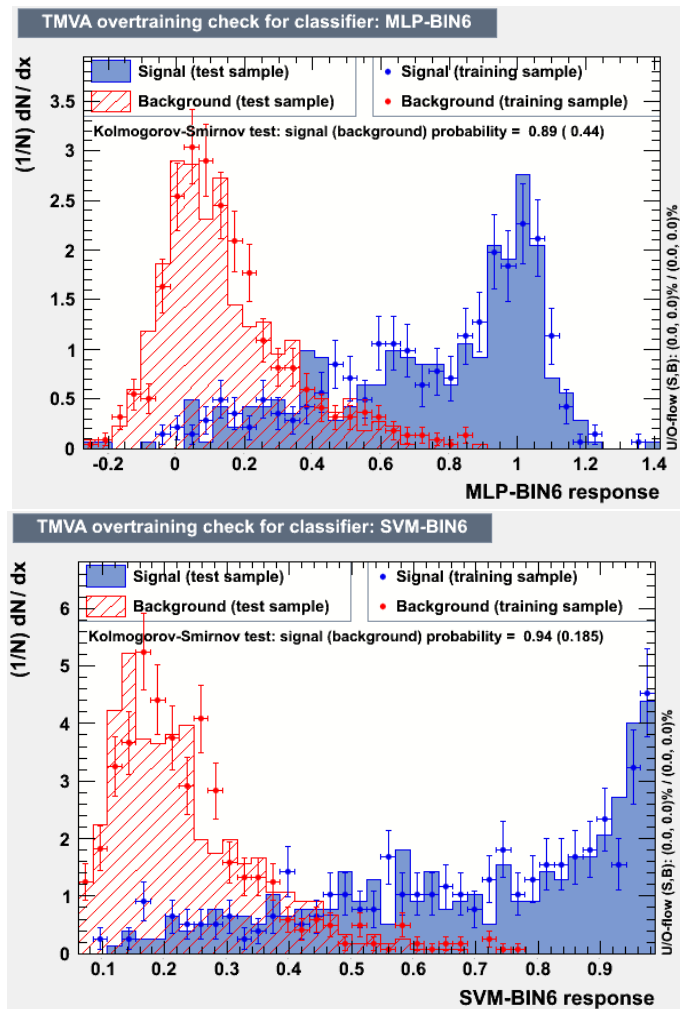


Figure 7.19: MVA output distributions for MLP (top) and SVM (bottom) trained with EPOS-LHC simulations in zenith bin 6 of the high energy bin. Train-dample: dots; test-sample: histograms.

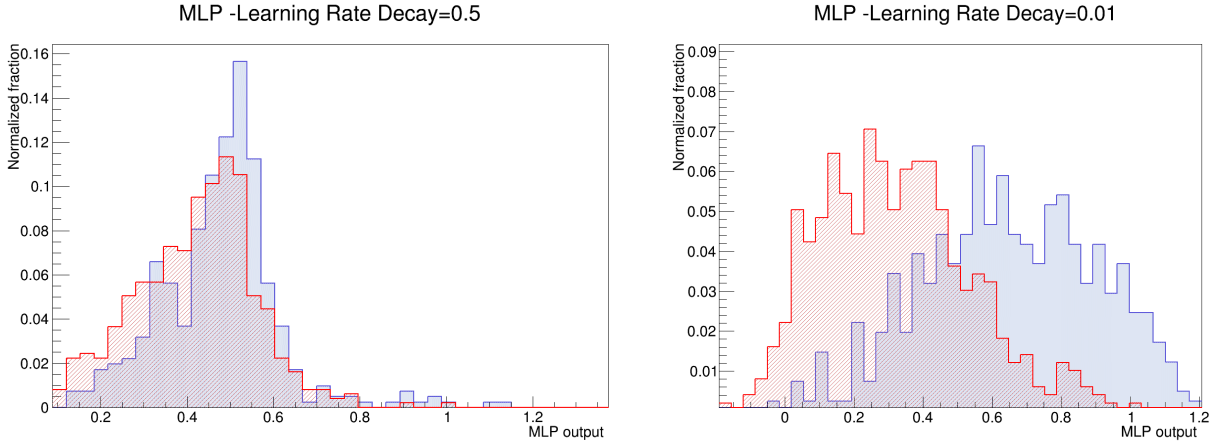


Figure 7.20: MVA MLP method output distributions for proton EPOS simulations (signal, blue dotted histogram) and iron ones (background, red dashed histogram). Left, obtained for an MLP method trained with a learning rate decay parameter of 0.5, right with the same parameter set to 0.01.

All the three methods we chose are complex structures with lots of possibility for fine-tuning. For example, for boosted decision trees it is possible to change the number of trees in the forest, the algorithm used for boosting, the purity requested for a leaf to be classified as signal and many others, allowing a huge number of possible configurations. The importance of the fine-tuning of these parameters can be seen by looking, for example, at figure 7.20, where the output of two MLP methods are shown. The only difference between the two methods is that the left one has been trained with a learning rate decay parameter 0.5, while the right one with the same parameter set at 0.01. We can see that in the left plot the separation achieved between the signal and background histograms (proton and iron simulations, respectively in blue and red) is much lower than the one in the right plot.

We thus performed a fine-tuning of the methods by varying their parameters and searching for the best signal-background separation. In order to quantify this separation, we used the same approach described in paragraph 7.2.3, computing the efficiency at a given purity. Here we requested a higher purity, 90%, since we expect a better separation with the MVA than with single variables. We then performed the fine-tuning of the methods by maximizing the efficiency at 90% purity $\mathcal{E}(M_c^{90})$ measured on the test sample. During the fine-tuning procedure, we also removed one by one all the observables, to see if the methods were misguided by poorly discriminating observables. This was not the case, since the methods showed always equal or lower $\mathcal{E}(M_c^{90})$ values when observables were removed. For this reason, we decided to keep all observables in use for all methods in all bins.

The results of the fine-tuning of the methods are summarized in figure 7.21, where the efficiency at 90% purity for the three fine-tuned methods in all zenith angle bins is shown, for QGSJETII-04 (left panels) and EPOS-LHC (right panels) simulated events in the high energy (top panels) and low energy (bottom panels) bins. The values of the fine-tuned parameters for the three methods can be found in appendix A. We can see

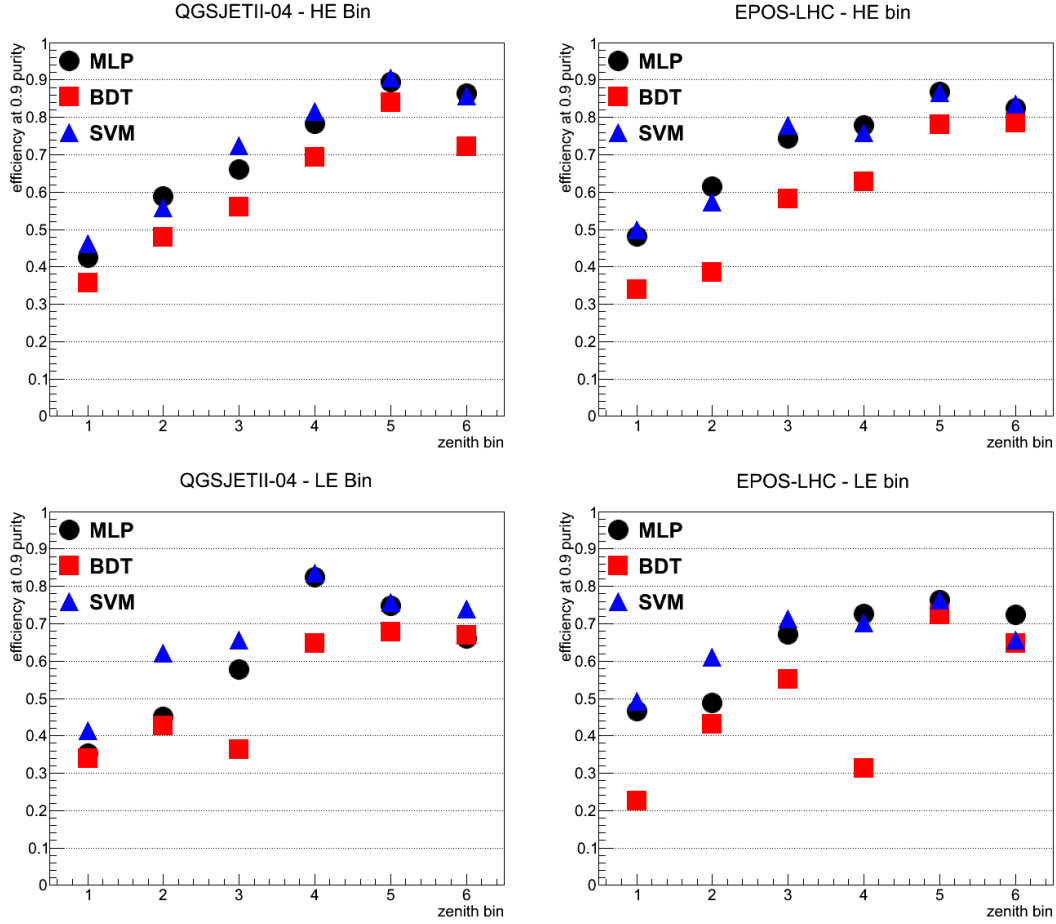


Figure 7.21: Performances of the fine-tuned methods, quantified as the efficiency at 90% purity (see text for details) for MLP (black circles), BDT (red squares) and SVM (blue triangles) methods trained with QGSJETII-04 (left panels) and EPOS-LHC (right panels) simulations in the high energy (top panels) and low energy (bottom panel) bins.

that the efficiency obtained by the BDT method is always lower than the ones obtained with the MLP and SVM ones; for this reason we decided to use only the latter two ones in the application of MVA on data, described in chapter 8. In general, the efficiencies obtained with the two hadronic interaction models are comparable; higher efficiencies are obtained in the high energy bin than in the low energy one. With the MVA we achieve in general a better discrimination than with the single observables: we obtain roughly the same efficiency as the Δ_{1000} alone (see figure 7.14), but imposing a 10% higher purity.

7.4 Conclusions

In this chapter we have identified a set of 7 mass-sensitive observables measured by the SD. We have studied these observables by using proton and iron events simulated with two different hadronic interaction models, namely EPOS-LHC and QGSJETII-04. We have shown that all observables measured in real data are well bracketed by the simulated

events. To reduce the dependencies of these observables on energy and zenith angle we have decided to divide the datasets in 6 zenith and 2 energy bins. In order to combine informations from each of the observables, we have used three multi-variate analysis methods: Boosted Decision Trees (BDT), Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM). We have trained each method independently in each energy and zenith bin, and then we have fine-tuned them. After the fine-tuning, we have found that the BDT method is always less performing than the other two, so we have decided to drop it. At the end of the work described in this chapter, we have thus built 2 MVA methods trained in each zenith and energy bin for each hadronic interaction models, for a total of 48 methods. In the next chapter, we will apply those methods to data, in order to select the most proton-like events that will be used for anisotropy tests.

Chapter 8

An application of MVA: correlations test with candidate proton events

"E vedere di nascosto l'effetto che fa"

Enzo Jannacci *Vengo anch'io. No tu no* (1968)

In this final chapter, we present an application of the MVA, discussed in the previous chapter, to data, aimed at performing anisotropy tests on a proton-enriched dataset. Indeed, we have shown in chapter 6 how magnetic fields can affect the arrival directions of cosmic rays, and we have concluded that only light events might preserve the directional information on their sources. In the same chapter, we have also reported that the two most significant deviations from isotropy are found for events with $E > 58$ EeV in correlation with Centaurus A and with luminous AGNs. In the following, we test the hypothesis that these excesses are due to the presence of proton-like events in the dataset.

We will describe in section 8.1 the application of the MVA to data, and we will explain how the subsamples of proton-like events are built. In section 8.1.1, we will show a comparison of the candidate protons, that we identified with the MVA, with those found with the so-called Universality method, a different approach to extract mass information from SD. Finally, we will use the proton-enhanced sample to test the two most significant excesses found in the anisotropy tests with the whole sample, and we will discuss our findings in section 8.2.

8.1 The proton-enhanced datasets

We have applied the MVA to data with $E_{\text{Herald}} \geq 58$ EeV (120 events). All these events are included in the high-energy bin, as defined in section 7.2.3. We therefore input the 120 events into the 24 MVA trained in the high-energy bin. Each event is processed by the four

methods (MLP and SVM trained with EPOS-LHC and QGSJETII-04 simulated events) trained in the relative zenith bin. This way, we obtained for each event four MVA outputs: $M_{MLP,SVM}^{EPOS,QGSJET}$. In principle, we could have then selected the proton-like events based on the cut at 90% purity, as described in the previous chapter. However, this would have been an arbitrary choice, based on simulated events. We thus chose to order the events by decreasing proton-likeness (i.e., decreasing M) and then select different subsamples containing the 5%, 10%, 15% ... 45% most proton-like events. The MVA outputs for different methods and bins are not comparable, so the sorting was made independently in the 6 zenith bins, for MLP and SVM and for QGSJET- and EPOS-trained methods. In the likely case that the chosen percentage, applied to the total number of events in each bin, did not lead to an integer number, we rounded it by excess. The number of events in each subsample, in each bin, are shown in the top 6 rows of table 8.1.

We merged the events in each zenith bin, for each proton-likeness fraction, and we ended with 9 mutually inclusive subsamples of candidate protons (the 5%-45% most proton-like in step of 5%) covering the whole zenith range, for each of the two methods MLP and SVM, for each of the two models, for a total of 36 subsamples. The number of events in each of the 9 subsamples is shown in the seventh line of table 8.1. Since the two methods, MLP and SVM, are expected to behave similarly as they are trained with the same set of simulated events, in order to reduce the fluctuations due to the methods, we decided to intersect the subsamples identified by MLP and SVM. We did so by selecting only the events that were chosen by both methods for a given proton-likeness fraction and for a given training sample (EPOS or QGSJET). On the other hand, the physics behind the two interaction models EPOS-LHC and QGSJETII-04 is different and so we decided not to mix the events selected on the basis of the two sets of simulated events. We ended up with 18 subsamples: 9 obtained from EPOS-trained methods and 9 obtained from QGSJET-trained ones. In the eighth and ninth rows of table 8.1, the number of events in each of these intersected subsamples are shown. By comparing the number of events in the intersected subsamples with the original ones in line 7, we can see that, in general, the two methods agree, with $\sim 60 - 90\%$ common events. This means that the MVA is robust and that the fluctuations associated to the different methods are limited.

We show the arrival directions of the events in the subsamples, in Galactic coordinates, in figure 8.1, top panel, for the events selected via QGSJET-trained MVA and in the bottom panel for the ones selected with EPOS-trained MVA. For the sake of legibility, we show only the arrival directions of odd subsamples, 5%, 15%... 45%, with larger dots for the smaller fractions: for example, the largest dots represent the events in the 5% most proton-like events subsamples.

Although in the following we will keep the candidates selected with the two hadronic interaction models separated, for completeness we counted also the number of events that are common in the EPOS and QGSJET subsamples. The number of common events is shown in the last line of table 8.1. We can see that in general there is a $\sim 60 - 80\%$ agreement between the two methods, with the exception of the 5% subsample. This is a remarkable agreement, in particular considering the differences in the two models described in section 3.5.3.

Most proton-like fraction:	5%	10%	15%	20%	25%	30%	35%	40%	45%	100%
Events with $E \geq 58$ EeV										
BIN 1	1	2	3	3	4	5	5	6	7	14
BIN 2	2	3	4	5	6	7	9	10	11	23
BIN 3	2	3	4	6	7	8	9	11	12	25
BIN 4	1	2	3	4	5	5	6	7	8	16
BIN 5	2	3	4	5	6	7	8	8	10	20
BIN 6	2	3	4	5	6	7	8	9	10	22
Total:	10	16	22	28	32	39	45	52	58	120
EPOS MLP-SVM Intersection:	7	10	17	23	29	31	37	43	50	-
QGSJET MLP-SVM Intersection:	6	12	18	23	26	33	38	45	49	-
EPOS/QGSJET intersection:	1	8	11	14	17	25	30	37	41	-

Table 8.1: Number of events in the final sub-data sets.

8.1.1 Cross-check with the Universality method

Within the Auger Collaboration, another method is under development to obtain mass-sensitive observable using SD data. It is the so-called Universality approach [176], based on the assumption that the shower produced by a cosmic ray has so many particles that, after a few interaction lengths, all its relevant distributions (particle densities, energy spectra, radial and longitudinal dependence) are function of a very reduced set of parameters: the energy in the electromagnetic cascade (E), its depth of the maximum (X_{\max}) and the number of muons (N_{μ}) [177]. Such an assumption results into a model that describes showers initiated by protons and nuclei, up to iron, as well as photon showers, using only these three parameters. This assumption was found to be valid in simulated events, almost independently from the adopted hadronic interaction model.

Consequently, based on a global fit of the integrated signal and its temporal structure measured in each SD stations, one can estimate E , X_{\max} and N_{μ} on an event-by-event basis. The main advantage of this method is that it combines all experimentally collected information of an air shower for deriving the physics observables. In this sense, a universality-based reconstruction can be considered as a very advanced multivariate analysis of shower data, that employs parameterized physics relations to combine the different measured quantities.

The main drawback of this approach is that the primary energy and the number of muons are very strongly correlated and cannot be determined independently from the data of the water-Čerenkov detectors alone. This problem can be solved with different approaches: in the “El Universal” reconstruction software, that we will use in the following, the number of muons is parametrized as a function of energy and of X_{\max} , based on hybrid events. Because of the small number of hybrid events at the highest energies (~ 10 for $E > 40$ EeV), the parametrization is extrapolated from measurements at lower energies. Moreover, although, in principle, independent on hadronic interaction models, this method actually uses parametrizations tuned on simulated showers.

We checked if the candidate light events selected with our method were classified as “light” also by the Universality method, i.e., if they showed a high value of X_{\max} . To do

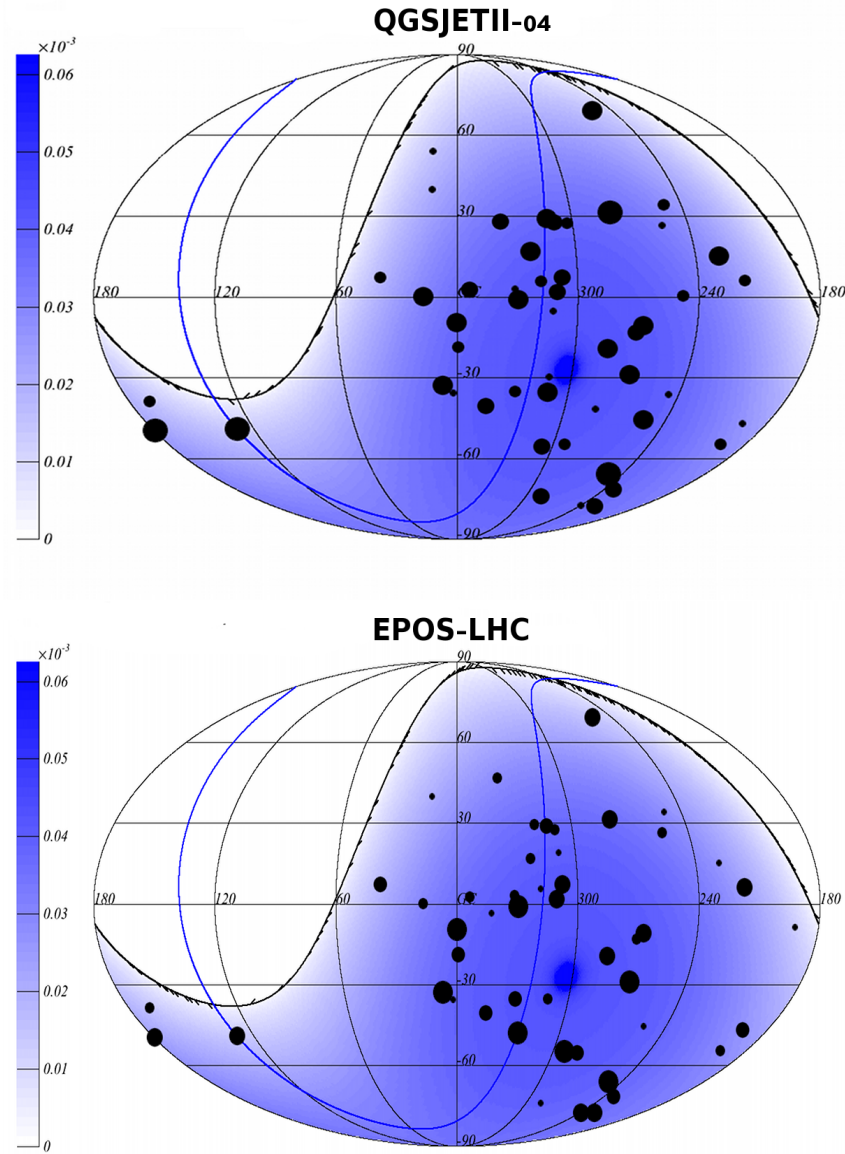


Figure 8.1: Map of the arrival directions of the most proton-like events as selected by QGSJETII-04-trained MVA (top panel) and EPOS-LHC-trained one (bottom panel). We plot the 5%,15%... 45% most proton-like events with decreasing size of the dots; for example, the largest dots represent the events in the 5% most proton-like subsample. The blue shade represents the exposure of Auger.

so, we computed the averaged X_{\max} in each of the 18 subsamples. Since X_{\max} is dependent on the energy, as described in chapter 2, we rescaled all the inferred X_{\max} to a common energy (10 EeV) so that they could be directly comparable. We did it by adopting a fixed elongation rate (i.e., the rate of change of X_{\max} as a function of energy) of 50 g/cm² per decade, i.e., we considered the quantity:

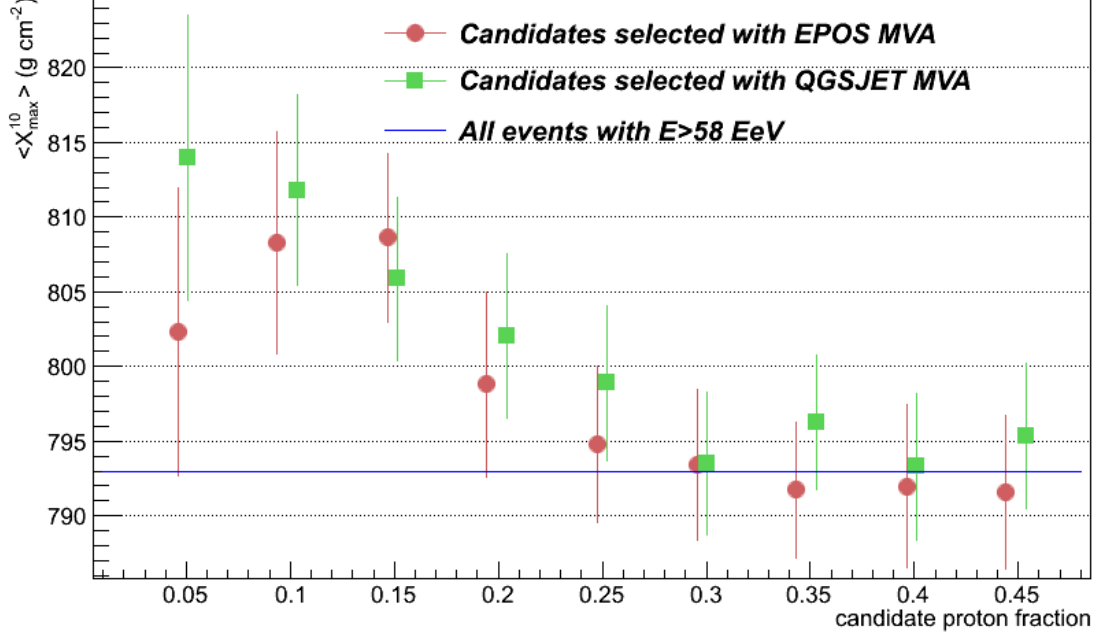


Figure 8.2: Average value of X_{max} estimated via the Universality method rescaled to 10 EeV, X_{max}^{10} , for each subsample of candidate protons selected with EPOS-trained MVA (red dots) and QGSJET-trained MVA (green squares) versus the fraction of proton candidates. The blue line represents the average X_{max}^{10} for all events with $E > 58$ EeV.

$$X_{max}^{10} \equiv X_{max} - 50 \frac{\text{g}}{\text{cm}^2} \log_{10} \left(\frac{E}{10 \text{ EeV}} \right) \quad (8.1)$$

We show in figure 8.2 the behavior of the average X_{max}^{10} value, in the subsamples selected by QGSJET- and EPOS-trained MVAs, as a function of the fraction of candidate protons. We can see that the behavior of $\langle X_{max}^{10} \rangle$ inferred by the Universality method is coherent with the decreasing proton-likeness of the events: $\langle X_{max}^{10} \rangle$ decreases as the selected fraction increases. As a reference, the blue line shows the average X_{max}^{10} of the 120 events.

8.2 Correlation tests

We finally used the 18 subsamples of events with $E \geq 58$ EeV, described in the previous section, to perform two tests on the two most significant excesses found in the analyses shown in chapter 6: the correlation with bright Swift-BAT galaxies and the correlation with Centaurus A. We fixed the parameters at the values corresponding to the minima found in the analyses with the whole dataset: $E_{th} = 58$ EeV and $\Psi = 15^\circ$ for Cen A; $E_{th} = 58$ EeV, $\Psi = 18^\circ$, $D_{max} = 130$ Mpc and $L_{min} = 10^{44}$ erg/s for Swift AGNs. We

then counted the fraction of correlating events (i.e., the fraction of events lying within an angle Ψ from Cen A or from any of the 10 selected AGNs) for each of the subsamples, and compared it to the probability of obtaining it from a randomly-extracted subsample. To build the latter, we extracted randomly 10000 subsamples of the 120 events with $E > 58$ EeV, of the same size of the MVA-selected ones, and we counted the number of correlating events in each of them.

We show in figure 8.3 the evolution of the correlating fraction as a function of the candidate-proton fraction in the case of the Swift AGNs (black dots), compared to the 1 and $2 - \sigma$ dispersion of the expected correlation from random-picked subsamples (filled red and blue areas, respectively), for candidates selected with QGSJET (top panel) and EPOS (bottom panel) -trained methods. The correlating fraction for the whole sample of 120 events is shown with a green line. We also show the correlating fraction expected from isotropy (blue line and dots), as a reference. We can see that all data points are compatible with a random-picked subsample. We conclude that this test did not yield any enhancement of the significance with respect to the full dataset.

Corresponding results for the correlation with Cen A are shown in figure 8.4. Also in this case, the results are compatible with the ones expected by randomly picking subsamples of the full dataset. Overall, the two tests did not yield any significant result.

We note here that a similar test was performed in [178], using candidate proton events selected with the Universality method. This test did not find either any significant improvement of the correlation with the light subsample.

8.3 Conclusions

The anisotropy tests performed on the proton-enriched samples did not yield any significant improvement of the correlation with Cen A and with bright Swift AGNs. In principle, three factors can contribute to this fact:

- The MVA method is ineffective: if there is indeed a light component in the data, it is not found by this method. This may be due to the method itself, or to the fact that the hadronic interaction models do not properly describe the showers, so that the MVA is wrongly trained. The cross-check with the Universality indicates that, on average, the two approaches agree. However, the Universality method also relies on the same hadronic interaction models, though less directly.
- The observed excesses are fluctuations: for this reason the candidate protons do not show a higher correlation than randomly picked events. In this case, one might consider performing further tests on different catalogues of candidate sources, using the complete dataset (down to 40 EeV) and selecting proton-enriched samples. We note, however, that if the number of actual sources is large, it may still be challenging to identify them, even by studying proton-enriched data.
- There is no proton component in our data, or it is too small to be detected, given the uncertainties on the mass-selecting methods. This is the worst-case scenario for cosmic-ray astronomy: if the events at the highest energies are all heavy nuclei, there is almost no chance that they can be used to track back the position of their sources.

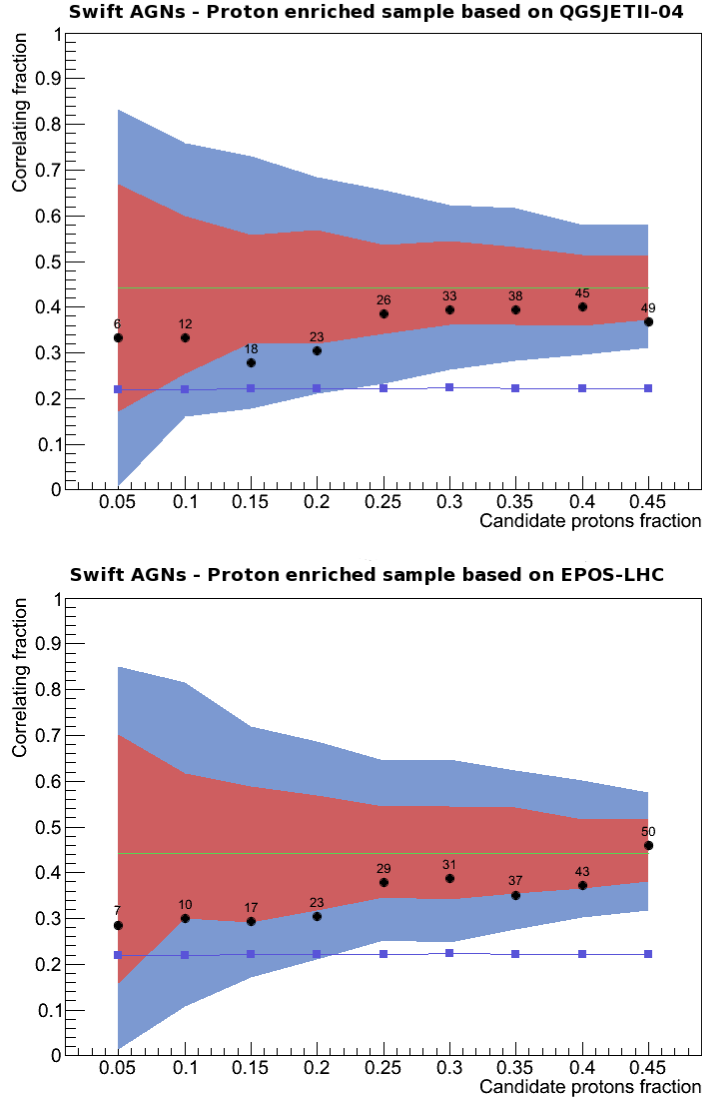


Figure 8.3: Fraction of events correlating with the positions of the AGNs in the Swift-BAT catalogue with $L > 10^{44}$ erg/s and $D < 130$ Mpc versus the fraction of candidate protons (black dots). The red and blue areas corresponds to 1 and 2 σ dispersions of randomly picking subsample of events from the data set, The blue dots and line represent the expectations from isotropy. The number of events in each point are indicated. The green line indicates the correlation for all the events with $E > 58$ EeV in the 454-Evts dataset. The top panel corresponds to candidate protons selected with QGSJET-trained MVA, and the bottom panel for EPOS-trained ones.

The actual scenario can contain a mixture of these three factors. Future measurements, in particular with the planned Auger upgrades aiming to enhance the mass sensitivity of the SD, can provide further elements to ameliorate the mass-discriminating methods allowing one to rely less and less on hadronic interaction models. Simultaneously, the new data recorded in the forthcoming years will help to confirm or disprove the anisotropy

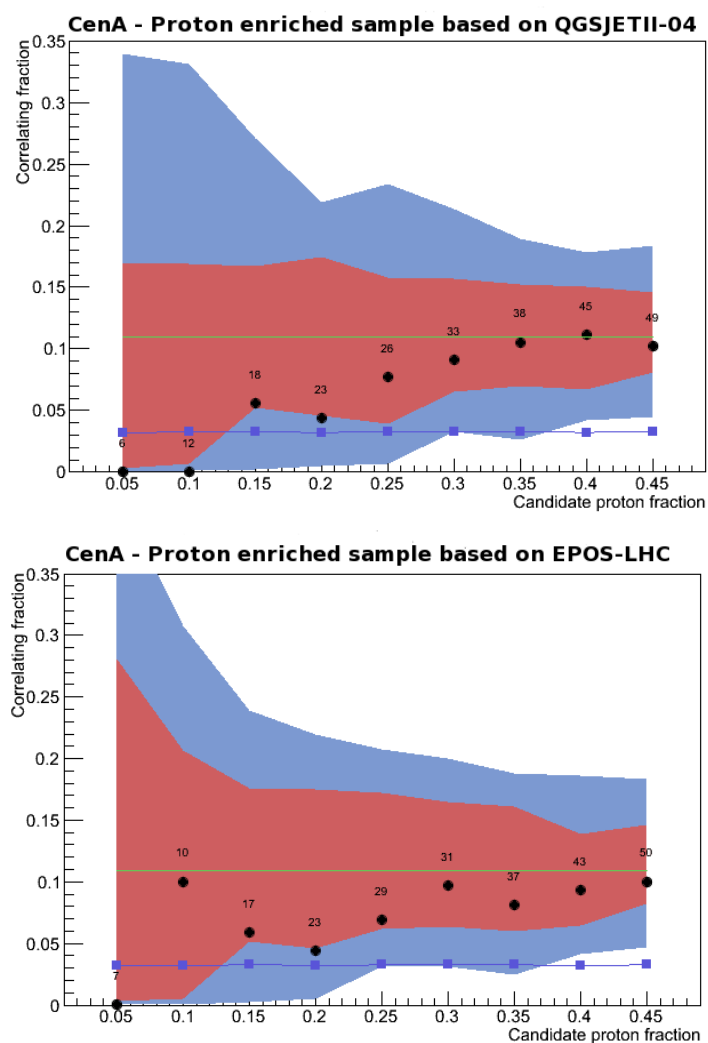


Figure 8.4: Fraction of events correlating with the positions of Cen A versus the fraction of candidate protons (black dots). The red and blue areas corresponds to 1 and 2 σ dispersions of randomly picking subsample of events from the data set, The blue dots and line represent the expectations from isotropy. The number of events in each point are indicated. The green line indicates the correlation for all the events with $E > 58$ EeV in the 454-Evts dataset. The top panel corresponds to candidate protons selected with QGSJET-trained MVA, and the bottom panel for EPOS-trained ones.

hints we observed. On the astrophysical side, new results could help, on the one hand, to better identify the candidate sources and, on the other hand, to better constrain the magnetic fields models, possibly allowing us to correct the arrival direction of cosmic rays to take into account the deflections, at least for proton-like events. The next decade will then be crucial for the future of cosmic-ray astronomy.

Conclusions and outlook

"E non c'è niente da capire"

Francesco De Gregori *Niente da capire* from Francesco De Gregori (1974)

The identification of ultra-high energy cosmic ray sources is one of the key topics of high energy astroparticle physics. Through this work, we tried to add some pieces of information to the current scenario, and we did so by studying the highest energy cosmic rays observed by the Pierre Auger Observatory. The work can be divided in three different steps: the first step was the selection of the largest ever dataset of cosmic rays with $E > 40$ EeV; the second step was to search for anisotropies in the arrival direction of these events; the final one was to devise a method to extract information on the mass of the events in order to perform anisotropy test only on the most proton-like ones.

The first part of this work was described in chapter 5. We selected the vertical ($\theta < 60^\circ$) events with $E > 40$ EeV and carefully look through all of them, performing checks on their reconstruction. Through these checks, we were able to identify problems in the PMT and stations selections and ameliorate them. These enhancements were then implemented the official reconstruction of Auger events. Also, we performed checks on the stability in time of the detector and reconstruction performances, given the 10-years extension of the dataset. In particular, we studied the change in time of the shape of the signal recorded in the SD stations, finding out that it does indeed impact on the detection and reconstruction of low-energy events, but it has nearly no impact on the highest energy ones. Further checks were performed on the angular and energy resolution of the events, finding out that they have been stable in time. Particular checks were performed on the events selected via relaxed quality cuts that were recently introduced in the Collaboration. As a final step, we added to the 454 selected vertical events 148 inclined ones ($60^\circ < \theta < 80^\circ$) in the same range of energy, performing test on the consistency of the two datasets.

The 602-events dataset was then used to search for anisotropies in their arrival directions (chapter 6). This was done by searching for intrinsic anisotropy, in particular through an autocorrelation test, and for correlation with catalogues of candidate sources. This work was published in [143], and this thesis contributed to the cross-check of the results. The chosen catalogues were those of 2MRS galaxies, mapping the distribution of nearby matter, of the Swift-BAT AGNs, observed in X-rays, and of radiogalaxies selected based on their morphology. We searched for correlations between the 60 events and these

catalogues by scanning on the minimum energy of cosmic rays, E_{th} , on the maximum separation angle between the cosmic rays and the sources, Ψ , and on the maximum distance of the sources, D_{max} . We obtained the combination of parameters that maximized the correlation with respect to what was expected from an isotropic distribution of cosmic rays. A further scan was also performed, only for the Swift and radiogalaxy catalogues, on the minimum luminosity of the sources, L_{min} . Finally, we searched for excesses with respect to the position of Centaurus A, the closest AGN. Overall, none of these analyses led to a significant evidence of anisotropy. The two most significant deviations from isotropy were found in the 4-parameters scan with Swift AGNs and in the correlation with Cen A, both yielding a penalized probability of chance coincidence of the order of $\sim 1\%$. In both cases, the minimum probability was found for $E_{th} = 58$ EeV and a rather large Ψ , $\sim 15^\circ - 18^\circ$. The minima obtained through our cross-checks were found to be completely consistent in these two tests, as well as in those that resulted in a chance probability of a few percent.

A further analysis was performed in a joint task force with the Ice Cube and Telescope Array experiments; it was aimed at searching for correlations between the arrival direction of UHECRs and of high-energy neutrinos. In that context, we performed a study on the predicted deflections of the cosmic rays arrival directions, caused by the Galactic magnetic field. We did so by simulating the field based on the most updated models for the GMF (Jensson-Farrar 2012 and Psirkov-Tinyakov 2011) and we found at 50 EeV, an average deflection of $\sim 6^\circ$ for proton cosmic rays for both models. This value was then adopted in one of the analyses, performed through a likelihood method. Also this test did not yield any significant evidence of anisotropy. Preliminary results with the 4-years catalog of neutrino cascade-events showed however a hint of correlation with chance coincidence probability of the order of $\sim 1\%$.

The lack of anisotropy found by all these tests is quite remarkable and can be explained in different ways, one of them being the presence of a large fraction of heavy nuclei at ultra-high energies. Such events would then be isotropized by magnetic fields, losing the directional information on their sources. This was the motivation to build a method to select light events in the dataset (chapter 7). To do this, we first selected 7 SD mass-sensitive observables. These observables are related to the characteristics of showers that are different in case of a proton or iron primary: the longitudinal and lateral profile of the shower and its muon content. The p-Fe discriminating capability of these variables was tested by means of two sets of simulated proton and iron events. The two sets were generated using two different hadronic interaction models: EPOS-LHC and QGSJETII-04. These models are the most updated ones, tuned on LHC data. Neither of them can describe coherently the muonic and the electromagnetic component of EAS observed in Auger. For this reason, we used them both, as a way to account for systematic effects due to the models. When studying the performances of the observables in simulated events, we found that they were dependent on the energy and zenith angle of the events. For this reason, we binned the simulated and real events in 2 energy and 6 zenith bins.

We then combined the observables, to enhance the p-Fe discriminating power into a the Multi Variate Analysis (MVA). MVA is a general term that denotes a variety of methods capable to train themselves on how to separate signal from background events (in our case, proton and iron-induced showers) once they are given a set of classified events (in our case, simulated events) and a list of discriminating observables available for each event (in our case the mass-sensitive observables). We have chosen three MVA methods:

Boosted Decision Trees, Multi-Layer Perceptron (MLP, a type of artificial neural network) and Support Vector Machine (SVM). We trained them on simulations, separately in each energy and zenith bin. After fine-tuning the methods, we checked their p-Fe discrimination power, and found that BDT were less performing than the other two in each energy and zenith bin. For this reason we decided to drop the BDT method and proceeded with the remaining ones.

The MLP and SVM methods, trained separately on events simulated on the basis of EPOS-LHC and QGSJETII-04 interaction models, were finally applied to the real data (chapter 8). We did not select a single subsample of candidate protons, but rather ordered the events by decreasing proton-likeness and then selected the 5%,10%...45% most proton-like ones. These mutually inclusive subsamples were selected independently using EPOS and QGSJET-trained MLP and SVM methods, resulting in 36 different sets of subsamples. Since the physics behind the two interaction models is different, we decided not to mix the events selected on the basis of the two. On the other hand, the two methods, MLP and SVM, are expected to behave the same way, as they are trained with the same set of simulated events. In order to reduce the fluctuations due to the methods, we decided to intersect the sub-datasets selected by MLP and SVM, selecting only the events chosen by both methods for a given proton-likeness fraction. We cross-checked the proton-likeness of the selected events by using a different technique of reconstruction of SD data, the so-called Universality. This technique differs from the standard reconstruction since it infers also the depth of shower maximum X_{\max} by fitting at the same time the signal in the station and its shape. We computed the average X_{\max} values for each of the subsamples and found that it decreases when selecting an higher percentage of proton-like data, as expected if we are adding each time heavier events.

We exploited the subsamples to test the two most significant excesses found with the total dataset, the one obtained with bright AGNs and the one obtained with Centaurus A. In particular, we checked the number of events that correlates in each of the subsamples, fixing the scan parameters found as chance probability minimum with the whole dataset. We found that the excesses were not driven by the proton-like events: the correlating fraction was for all subsamples consistent with what expected from randomly picked subsamples of the whole dataset.

From this work, we have seen that a remarkable isotropy is found in the high energy events. This unexpected result may arose from three effects: at first, the number of sources may be large and they might be spread isotropically in the sky. Secondly, the magnetic fields deflections may be larger than expected, even for protons. Thirdly, the highest energy cosmic rays may be primarily heavy nuclei and thus strongly deflected by magnetic fields.

In this scenario, the capability of selecting the low-Z component, if present, among all the highest energy cosmic ray events would be an important leap towards the achievability of detecting cosmic rays sources. As pointed out at the end of chapter 8, the lack of results obtained with our mass discrimination method may be due to a combination of different factors. At first, the method may not work properly. It can be a problem of the method itself or of the simulated events on which it is trained: with future refinements of hadronic interaction models, based on the new collisions in LHC at 14 TeV and on Auger data too, this method can be trained again with simulated events more similar to the real ones. Another possibility is that the astrophysical tests we performed on the candidate protons

did not yield any significant results because the hints of correlation found with Swift AGNs and Cen A are actually fluctuation with no statistical significance. In the future, the candidate protons should be tested for correlation with different candidate sources: we trained the MVA method to be usable on the whole high energy dataset from 40 EeV to the highest energies. It can be used to select candidate protons in the whole dataset and on these events we can perform new scans or different analyses. The final possibility is that cosmic rays at these energies do not have a light fraction, or, if present, it is particularly low. In the latter case, improvement in the mass-discriminating capabilities of the detector, as programmed in the Auger upgrade, could reduce the uncertainties on mass estimation and allow us to detect even a small fraction of light events.

With new detectors, new hadronic interaction models and new astrophysical results that will improve our knowledge of the candidate sources and of the magnetic fields, the next decade has the chance of being the turning point towards cosmic-ray astronomy.

Appendix A

Multi Variate analysis parameters and results plot

A.1 Fine-tuned parameters

In the following, the best values for the main multi-variate analysis parameters are presented.

<i>Feature name</i>	<i>value</i>
Number of Trees	400
Boost Type	AdaBoost
Node Purity Limit	0.55
Separation type	Gini Index
Prune Method	Cost Complexity
Prune strength	-1

Table A.1: Parameters chosen for BDT method. Entries in **bold** are different from the default ones in TMVA.

<i>Feature name</i>	<i>value</i>
Number of training cycles	100
Hidden layers	N+5
Neuron activation function type	tanh
Neuron input function type	sum
Training method	Back Propagation (BP)
Learning rate	0.02
Decay rate	0.01
Test rate	10

Table A.2: Parameters chosen for MLP method. Entries in **bold** are different from the default ones in TMVA.

<i>Feature name</i>	<i>value</i>
Cost parameter	1
RBF kernel parameter: Gamma	0.5
Tolerance parameter	0.1
Maximum number of training loops	1000
Number of training subsets	1

Table A.3: Parameters chosen for SVM method. Entries in **bold** are different from the default ones in TMVA.

Bibliography

- [1] E.-S. Seo, J. F. Ormes, R. E. Streitmatter, S. J. Stochaj, W. V. Jones, S. A. Stephens, and T. Bowen. Measurement of cosmic-ray proton and helium spectra during the 1987 solar minimum. *The Astrophysical Journal*, 378:763–772, 1991.
- [2] N. L. Grigorov, Y. V. Gubin, B. M. Jakovlev, I. D. Rapoport, I. A. Savenko, V. V. Akimov, and V. E. Nesterov. Energy Spectrum of Primary Cosmic Rays in the 1011-1015 eV According to the Data of Proton-4 Measurements. In *International Cosmic Ray Conference*, volume 1, page 170, 1971.
- [3] M. Nagano, M. Teshima, Y. Matsubara, H. Y. Dai, T. Hara, N. Hayashida, M. Honda, H. Ohoka, and S. Yoshida. Energy spectrum of primary cosmic rays above $10^{17.0}$ eV determined from extensive air shower experiments at Akeno. *Journal of Physics G Nuclear Physics*, 18(2):423–442, 1992.
- [4] KASCADE Collaboration. KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems. *Astropart. Phys.*, 24:39, 2005.
- [5] The Pierre Auger Collaboration. Observation of the suppression of the flux of cosmic rays above $4 \cdot 10^{19}$ eV. *Physical review letters*, 101(6):061101, August 2008.
- [6] S. Yoshida, N. Hayashida, K. Honda, M. Honda, S. Imaizumi, N. Inoue, K. Kadota, F. Kakimoto, K. Kamata, S. Kawaguchi, N. Kawasumi, Y. Matsubara, K. Murakami, M. Nagano, H. Ohoka, M. Teshima, I. Tsushima, and H. Yoshii. The cosmic ray energy spectrum above 3×10^{18} eV measured by the Akeno Giant Air Shower Array. *Astroparticle Physics*, 3(2):105–123, 1995.
- [7] The HiRes Collaboration. First Observation of the Greisen-Zatsepin-Kuzmin Suppression. *Phys. Rev. Lett.*, 100(10):101101, March 2008.
- [8] J. J. Beatty and S. Westerhoff. The highest-energy cosmic rays. *Annual Review of Nuclear and Particle Science*, 59:319–345, 2009.
- [9] K.-H. Kampert. Ultrahigh-Energy Cosmic Rays: Results and Prospects. *Braz.J.Phys.*, 43:375–382, 2013.
- [10] R. Aloisio, V. Berezhinsky, P. Blasi, and S. Ostapchenko. Signatures of the transition from galactic to extragalactic cosmic rays. *Physical Review D*, 706, 2008.
- [11] T. P. A. Collaboration. Measurement of the Depth of Maximum of Extensive Air Showers above 10^{18} eV. *Phys. Rev. Lett.*, 104:91101, 2010.

- [12] D. G. Pinto and the Pierre Auger Collaboration. The Pierre Auger Observatory II: Studies of Cosmic Ray Composition and Hadronic Interaction models. In *Proceedings, 32nd International Cosmic Ray Conference (ICRC 2011)*, 2011.
- [13] T. G. Guzik, J. H. Adams, H. S. Ahn, G. Bashindzhagyan, J. Chang, M. Christl, A. R. Fazely, O. Ganel, D. Granger, R. Gunasingha, and Others. The ATIC long duration balloon project. *Advances in Space Research*, 33(10):1763–1770, 2004.
- [14] E. S. Seo, H. S. Ahn, J. J. Beatty, S. Coutu, M. J. Choi, M. A. DuVernois, O. Ganel, T. G. Kang, K. C. Kim, M. H. Lee, and Others. Cosmic-ray energetics and mass (CREAM) balloon project. *Advances in Space Research*, 33(10):1777–1785, 2004.
- [15] S. Sposato. The Trans-Iron Galactic Element Recorder (TIGER): A Balloon-borne Cosmic-Ray Experiment. In *International Cosmic Ray Conference*, volume 5, page 29, 1999.
- [16] E. S. Seo, T. Anderson, D. Angelaszek, S. J. Baek, J. Baylon, M. Buénerd, M. Copley, S. Coutu, L. Derome, B. Fields, and Others. Cosmic ray energetics and mass for the international space station (ISS-CREAM). *Advances in Space Research*, 53(10):1451–1455, 2014.
- [17] F. Cadoux, F. Cervelli, V. Chambert-Hermel, G. Chen, H. Chen, G. Coignet, S. Di Falco, J. M. Dubois, E. Falchini, A. Franzoso, and Others. The AMS-02 electromagnetic calorimeter. *Nuclear Physics B-Proceedings Supplements*, 113(1):159–165, 2002.
- [18] A. Haungs, W. Apel, and F. Badae. Investigating the 2nd knee: The KASCADE-Grande experiment. *Journal of Physics: ...*, 2006.
- [19] T. Waldenmaier. IceTop - Cosmic ray physics with IceCube. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 588(1-2):130–134, 2008.
- [20] M. E. Bertaina. Cosmic rays from the knee to the ankle. *Comptes Rendus Physique*, 15(4):300–308, 2014.
- [21] R. Iuppa. Anisotropy in the cosmic radiation at TeV energy. *arXiv preprint arXiv:1302.7184*, 2013.
- [22] G. Di Sciacio and R. Iuppa. On the Observation of the Cosmic Ray Anisotropy below 10^{15} eV. *arXiv ePrint 1407.2144*.
- [23] M. Amenomori, S. Ayabe, X. J. Bi, D. Chen, S. W. Cui, L. K. Ding, X. H. Ding, C. F. Feng, Z. Feng, Z. Y. Feng, and Others. Anisotropy and corotation of galactic cosmic rays. *Science*, 314(5798):439–443, 2006.
- [24] Z. Cao. TeV gamma-ray survey of the northern sky using ARGO-YBJ experiment. *International Cosmic Ray Conference*, 7:212, 2011.

- [25] A. A. Abdo, B. T. Allen, T. Aune, D. Berley, S. Casanova, C. Chen, B. L. Dingus, R. W. Ellsworth, L. Fleysler, R. Fleysler, and Others. The large-scale cosmic-ray anisotropy as observed with Milagro. *The Astrophysical Journal*, 698(2):2121, 2009.
- [26] P. Desiati. Observation of TeV-PeV cosmic ray anisotropy with IceCube, IceTop and AMANDA. *Nucl.Instrum.Meth.*, A742:199–202, 2014.
- [27] A. U. Abeysekara and Others. Observation of Small-scale Anisotropy in the Arrival Direction Distribution of TeV Cosmic Rays with HAWC. 2014.
- [28] A. Chiavassa, W. D. Apel, J. C. Arteaga-Velázquez, K. Bekk, M. Bertaina, J. Blümer, H. Bozdog, I. M. Brancus, E. Cantoni, F. Cossavella, and Others. Studies of the cosmic ray spectrum and large scale anisotropies with the KASCADE-Grande experiment. In *Journal of Physics: Conference Series*, volume 531, page 12001. IOP Publishing, 2014.
- [29] J. L. Han, R. N. Manchester, A. G. Lyne, G. J. Qiao, and W. van Straten. Pulsar rotation measures and the large-scale structure of Galactic magnetic field. *Astrophys.J.*, 642:868–881, 2006.
- [30] G. R. Farrar. The Galactic Magnetic Field and Ultrahigh-Energy Cosmic Ray Deflections. *Comptes Rendus Physique*, 15:339–348, 2014.
- [31] N. Jarosik, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, M. Halpern, R. S. Hill, G. Hinshaw, A. Kogut, E. Komatsu, and Others. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: sky maps, systematic errors, and basic results. *The Astrophysical Journal Supplement Series*, 192(2):14, 2011.
- [32] F. R. Bouchet. The Planck satellite: status & perspectives. *Modern Physics Letters A*, 22(25n28):1857–1863, 2007.
- [33] R. Jansson and G. R. Farrar. A New Model of the Galactic Magnetic Field. *The Astrophysical Journal*, 757:14, 2012.
- [34] Y.-Y. Jiang, L. G. Hou, J. L. Han, X. H. Sun, and W. Wang. Do Ultrahigh Energy Cosmic Rays Come from Active Galactic Nuclei and Fermi γ -ray Sources? *Astrophys.J.*, 719:459–468, 2010.
- [35] A. Letessier-Selvon and T. Stanev. Ultrahigh energy cosmic rays. *Reviews of Modern Physics*, 83(3):907–942, September 2011.
- [36] K. Kotera and A. V. Olinto. The Astrophysics of Ultrahigh-Energy Cosmic Rays. *Annual Review of Astronomy and Astrophysics*, 49(1):119–153, 2011.
- [37] D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright. The Cosmic Microwave Background spectrum from the full COBE FIRAS data set. *The Astrophysical Journal*, 473(2):576, 1996.
- [38] K. Greisen. End to the Cosmic-Ray Spectrum? *Physical Review Letters*, 16(17):748–750, April 1966.

- [39] G. T. Zatsepin and V. A. Kuz'min. Upper limit of the cosmic-ray spectrum. *Zh. Eksp. Teor. Fiz., Pis'ma Red.*, 4(3):114–117, 1966.
- [40] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s., 1965.
- [41] D. Harari. The flux suppression at the highest energies. *Comptes Rendus Physique*, 15(4):376–383, 2014.
- [42] E. Fermi. On the origin of the cosmic radiation. *Physical Review*, 75(8):1169–1174, April 1949.
- [43] E. Amato. The origin of galactic cosmic rays. *Int. J. Mod. Phys.*, D23:1430013, 2014.
- [44] J. Paglione, C. Petrovic, L. Taillefer, S. R. Julian, P. M. C. Rourke, D. Aoki, J. Flouquet, D. Hall, K. Heuser, G. R. Stewart, S. State, C. Bergemann, T. M. Rice, D. T. Son, A. O. Starinets, S. A. Kivelson, B. Spivak, and M. Nicklas. Detection of the Characteristic Pion-Decay Signature in Supernova Remnants. *Science*, 193(February):807–811, 2013.
- [45] A. M. Hillas. The Origin of Ultrahigh-Energy Cosmic Rays. *Ann. Rev. Astron. Astrophys.*, 22:425–444, 1984.
- [46] V. L. Ginzburg and S. I. Syrovatskii. *The origin of cosmic rays*. Elsevier, 1964.
- [47] F. Halzen and E. Zas. Neutrino fluxes from active galaxies: A model-independent estimate. *The Astrophysical Journal*, 488(2):669, 1997.
- [48] C. M. Urry and P. Padovani. Unified Schemes for Radio-Loud Active Galactic Nuclei. *Publications of the Astronomical Society of the Pacific*, 107:803, September 1995.
- [49] B. L. Fanaroff and J. M. Riley. The morphology of extragalactic radio sources of high and low luminosity. *Monthly Notices of the Royal Astronomical Society*, 167(1):31P—36P, 1974.
- [50] G. Ghisellini, G. Ghirlanda, F. Tavecchio, F. Fraternali, and G. Pareschi. Ultra High Energy Cosmic Rays , Spiral Galaxies and Magnetars. 000(August), 2008.
- [51] G. B. Gelmini, O. E. Kalashev, and D. V. Semikoz. GZK photons as ultra-high-energy cosmic rays. *Journal of Experimental and Theoretical Physics*, 106(6):1061–1082, 2008.
- [52] J. Ellis, V. Mayes, and D. Nanopoulos. Ultrahigh-energy cosmic rays particle spectra from crypton decays. *Physical Review D*, 74(11):115003–+, December 2006.
- [53] P. K. Grieder. *Cosmic rays at Earth*. Gulf Professional Publishing, 2001.
- [54] W. Heitler. *Quantum Theory of Radiation*. Oxford University Press, 1944.
- [55] J. Matthews. A Heitler model of extensive air showers. *Astroparticle Physics*, 22(5-6):387–397, January 2005.

- [56] K. Greisen. *Progress in Cosmic-Ray Physics*, volume 3. Interscience Publishers, Inc., New York, 1956.
- [57] K. Kamata and J. Nishimura. The lateral and the angular structure functions of electron showers. *Progress of Theoretical Physics* . . . , 6:93–155, 1958.
- [58] D. Heck and F. (Karlsruhe). CORSIKA: A Monte Carlo code to simulate extensive air showers. *Report FZKA 6019*, 1998.
- [59] E.-J. Ahn, R. Engel, T. K. Gaisser, P. Lipari, and T. Stanev. Cosmic ray interaction event generator SIBYLL 2.1. *Phys. Rev. D*, 80(9):94003, November 2009.
- [60] N. N. Kalmykov and S. S. Ostapchenko. The nucleus-nucleus interaction, nuclear fragmentation, and fluctuations of extensive air showers. *Physics of Atomic Nuclei*, 56:346–353, 1993.
- [61] S. Ostapchenko. QGSJET-II: results for extensive air showers. *ArXiv Astrophysics e-prints*, 2004.
- [62] T. Pierog and K. Werner. Muon Production in Extended Air Shower Simulations. *Phys. Rev. Lett.*, 101(17):171101, 2008.
- [63] A. Fassò, A. Ferrari, J. Ranft, and B. E. Sala. FLUKA: a multi-particle transport code. *CERN-2005-10, INFN/TC_05/11, SLAC-R-773*.
- [64] G. Anelli, G. Antchev, P. Aspell, V. Avati, M. G. Bagliesi, V. Berardi, M. Berretti, V. Boccone, U. Bottigli, M. Bozzo, and Others. The TOTEM experiment at the CERN large hadron collider. *Journal of Instrumentation*, 3(08):S08007, 2008.
- [65] O. Adriani, L. Bonechi, M. Bongi, G. Castellini, R. DAlessandro, A. Faus, K. Fukatsu, M. Haguenaue, Y. Itow, K. Kasahara, D. Macina, T. Mase, K. Masuda, Y. Matsubara, H. Menjo, G. Mitsuka, Y. Muraki, M. Nakai, K. Noda, P. Papini, A.-L. Perrot, S. Ricciarini, T. Sako, K. Suzuki, T. Suzuki, Y. Shimizu, K. Taki, T. Tamura, S. Torii, A. Tricomi, J. Velasco, W. Turner, and K. Yoshida. Early results of the LHCf experiment and their contribution to ultra-high-energy cosmic ray physics. *Nuclear Physics B - Proceedings Supplements*, 212-213:270–276, March 2011.
- [66] S. Ostapchenko. Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model. *Physical Review D*, 83(1):14018, 2011.
- [67] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner. EPOS LHC: test of collective hadronization with LHC data. *arXiv preprint arXiv:1306.0121*, 2013.
- [68] N. N. Kalmykov and G. B. Khristiansen. Possible explanation for scaling violation in hadron interactions above 1000 TeV. *ZhETF Pisma Redaktsiiu*, 37:247–249, 1983.
- [69] T. Pierog. LHC results and high energy cosmic ray interaction models. In *Journal of Physics: Conference Series*, volume 409, page 12008. IOP Publishing, 2013.
- [70] B. Rossi and K. Greisen. Cosmic-ray theory. *Rev. Mod. Phys.*, 13:240–309, 1941.

- [71] P. Auger, P. Ehrenfest, R. Maze, J. Daudin, and R. Fréon. Extensive Cosmic-Ray Showers. *Reviews of Modern Physics*, 11(3-4):288–291, July 1939.
- [72] J. Linsley. Evidence for a Primary Cosmic-Ray Particle with Energy 10^{20} eV. *Physical Review Letters*, 10(4):146–148, February 1963.
- [73] AGASA Collaboration. Akeno Giant Air Shower Array (AGASA) covering 100 km² area. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 311(1–2):338–349, 1992.
- [74] M. Aglietta, G. Badino, L. Bergamasco, C. Castagnoli, A. Castellina, G. Cini, M. Dardo, B. D’Ettorre-Piazzoli, W. Fulgione, P. Galeotti, and Others. The EAS-TOP array at E0= 1014- 10 16eV: Stability and resolutions. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 277(1):23–28, 1989.
- [75] KASCADE Collaboration. The cosmic-ray experiment KASCADE. *Nuclear Instruments and Methods in Physics Research Section A: accelerators, spectrometers, detectors and associated equipment*, 513(3):490–510, 2003.
- [76] S. F. Berezhnev, D. Besson, N. M. Budnev, A. Chiavassa, O. A. Chvalaev, O. A. Gress, A. N. Dyachok, S. N. Epimakhov, A. Haungs, N. I. Karpov, and Others. The Tunka-133 EAS Cherenkov light array: status of 2011. *Nuclear Instruments and Methods in Physics Research Section A: accelerators, spectrometers, detectors and associated equipment*, 692:98–105, 2012.
- [77] The MAGIC collaboration. The MAGIC gamma-ray observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 553(1):274–281, 2005.
- [78] J. Holder, V. A. Acciari, E. Aliu, T. Arlen, M. Beilicke, W. Benbow, S. M. Bradbury, J. H. Buckley, V. Bugaev, Y. Butt, and Others. Status of the VERITAS Observatory. *arXiv preprint arXiv:0810.0474*, 2008.
- [79] J. A. Hinton, H. Collaboration, and Others. The status of the HESS project. *New Astronomy Reviews*, 48(5):331–337, 2004.
- [80] M. Actis, G. Agnetta, F. Aharonian, A. Akhperjanian, J. Aleksić, E. Aliu, D. Allan, I. Allekotte, F. Antico, L. A. Antonelli, and Others. Design concepts for the Cherenkov Telescope Array CTA: an advanced facility for ground-based high-energy gamma-ray astronomy. *Experimental Astronomy*, 32(3):193–316, 2011.
- [81] K. Greisen. Highlights in air shower studies, 1965. In *International Cosmic Ray Conference*, volume 2, page 609, 1965.
- [82] R. M. Baltrusaitis and R. Cady. The Utah Fly’s eye detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 240:410–428, 1985.
- [83] J. V. Jelley, J. H. Fruin, N. A. Porter, T. C. Weekes, F. G. Smith, and R. A. Porter. Radio pulses from extensive cosmic-ray air showers. 1965.

- [84] T. Huege for the LOPES Collaboration. Radio detection of cosmic ray air showers with LOPES. *BRAZ.J.PHYS.*, 36:1157, 2006.
- [85] D. Ardouin, A. Belltoile, D. Charrier, R. Dallier, L. Denis, and Others. Radioelectric Field Features of Extensive Air Showers Observed with CODALEMA. *Astropart.Phys.*, 26:341–350, 2006.
- [86] H. Falcke and P. Gorham. Detecting radio emission from cosmic ray air showers and neutrinos with a digital radio telescope. *Astroparticle Physics*, 19:477–494, 2003.
- [87] P. W. Gorham, N. G. Lehtinen, G. S. Varner, J. J. Beatty, A. Connolly, and Others. Observations of Microwave Continuum Emission from Air Shower Plasmas. *Phys.Rev.*, D78:32007, 2008.
- [88] R. Gaïor and The Pierre Auger Collaboration. Detection of cosmic rays using microwave radiation at the Pierre Auger Observatory. *Proc. 33rd ICRC, Rio de Janeiro*, 2013.
- [89] M. Monasor, M. Bohacova, C. Bonifazi, G. Cataldi, S. Chemerisov, J. R. T. Neto, P. Luis, B. Fox, P. W. Gorham, C. Hojvat, and Others. The Microwave Air Yield Beam Experiment (MAYBE): measurement of GHz radiation for Ultra-High Energy Cosmic Rays detection. *arXiv preprint arXiv:1108.6321*, 2011.
- [90] K. Louedec, J. Alvarez-Muñiz, M. Blanco, M. Boháčová, B. Buonomo, G. Cataldi, M. R. Coluccia, P. Creti, I. De Mitri, C. Di Giulio, and Others. The Air Microwave Yield (AMY) experiment-A laboratory measurement of the microwave emission from extensive air showers. *arXiv preprint arXiv:1310.4662*, 2013.
- [91] R. Abbasi, M. A. B. Othman, C. Allen, L. Beard, J. Belz, D. Besson, M. Byrne, B. Farhang-Boroujeny, A. Gardner, W. H. Gillman, W. Hanlon, J. Hanson, C. Jayanthmurthy, S. Kunwar, S. L. Larson, I. Myers, S. Prohira, K. Ratzlaff, P. Sokolsky, H. Takai, G. B. Thomson, and D. Von Maluski. Telescope Array Radar (TARA) observatory for Ultra-High Energy Cosmic Rays. *Nuclear Instruments and Methods in Physics Research A*, 767:322–338, 2014.
- [92] J. Stasielak, R. Engel, S. Baur, P. Neunteufel, R. Šmída, F. Werner, and H. Wilczyński. Feasibility of radar detection of extensive air showers. *ArXiv e-prints*, 2014.
- [93] The Pierre Auger Collaboration. Properties and performance of the prototype instrument for the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 523(1-2):50–95, May 2004.
- [94] The Telescope Array Collaboration. Telescope Array Experiment. *Nuclear Physics B - Proceedings Supplements*, 175–176(0):221–226, 2008.
- [95] The Pierre Auger Collaboration. The Pierre Auger Cosmic Ray Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2015.

- [96] The Pierre Auger Collaboration. The surface detector system of the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 586(3):409–420, 2008.
- [97] The Pierre Auger Collaboration. Trigger and aperture of the surface detector array of the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 613(1):29–39, 2010.
- [98] I. C. Maris and the Pierre Auger Collaboration. The AMIGA infill detector of the Pierre Auger Observatory: performance and first data. 2011.
- [99] F. Sanchez and the Pierre Auger Collaboration. The AMIGA detector of the Pierre Auger Observatory: an overview. 2011.
- [100] The Pierre Auger Collaboration. The fluorescence detector of the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 620(2–3):227–251, 2010.
- [101] T. K. Gaisser and A. M. Hillas. Reliability of the Method of Constant Intensity Cuts for Reconstructing the Average Development of Vertical Showers. volume 8 of *Proc. 15th ICRC, (Plovdiv)*, page 353, 1977.
- [102] M. J. Tueros. Estimate of the non-calorimetric energy of showers observed with the fluorescence and surface detectors of the Pierre Auger Observatory. In *Proceedings of the International Cosmic Ray Conference*, 2013.
- [103] V. Verzi and the Pierre Auger Collaboration. The energy scale of the Pierre Auger Observatory. *Proceedings of the 33rd ICRC, Rio de Janeiro, Brasil*, 2013.
- [104] The Pierre Auger Collaboration. A study of the effect of molecular and aerosol conditions in the atmosphere on air fluorescence measurements at the Pierre Auger Observatory. *Astroparticle Physics*, 33(2):108–129, 2010.
- [105] R. D. McPherson, K. H. Bergman, R. E. Kistler, G. E. Rasch, and D. S. Gordon. The NMC operational global data assimilation system. *Monthly Weather Review*, 107(11):1445–1461, 1979.
- [106] The Pierre Auger Collaboration. The exposure of the hybrid detector of the Pierre Auger Observatory. *Astroparticle Physics*, 34(6):368–381, 2011.
- [107] F. Suarez and the Pierre Auger Collaboration. The AMIGA muon detectors of the Pierre Auger Observatory: overview and status. In *Proc. 33rd Int. Cosmic Ray Conf. (Rio de Janeiro)*, 2013.
- [108] J. Schulz and The Pierre Auger Collaboration. the Status and Prospects of the Auger Engineering Radio Array. *PoS (ICRC 2015) 615*, 2015.

- [109] A. Schulz. The measurement of the energy spectrum of cosmic rays above 3×10^{17} eV with the Pierre Auger Observatory. *Proceedings of the 33rd ICRC, Rio de Janeiro, Brasil*, 2013.
- [110] K.-H. Kampert and P. Tinyakov. Cosmic rays from the ankle to the cutoff. *Comptes Rendus Physique*, 15(4):318–328, 2014.
- [111] T. Abu-Zayyad, R. Aida, M. Allen, R. Anderson, R. Azuma, E. Barcikowski, J. W. Belz, D. R. Bergman, S. A. Blake, R. Cady, and Others. The cosmic-ray energy spectrum observed with the surface detector of the telescope array experiment. *The Astrophysical Journal Letters*, 768(1):L1, 2013.
- [112] W. Tkaczyk, V. Rizi, R. Sato, D. Boncioli, C. C. H. Jui, M. Fukushima, J. Alvarez-Muniz, S. T. Roca, M. Minamino, J. Maller, and Others. Pierre Auger Observatory and Telescope Array: Joint Contributions to the 33rd International Cosmic Ray Conference (ICRC 2013). Technical report, 2013.
- [113] The Pierre Auger Collaboration. Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV. *Physical Review D*, 90(12):122005, 2014.
- [114] The Pierre Auger Collaboration. Interpretation of the depths of maximum of extensive air showers measured by the Pierre Auger Observatory. *Journal of Cosmology and Astroparticle Physics*, 2013(02):026–026, February 2013.
- [115] The Pierre Auger Collaboration. Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications. *Physical Review D*, 90(12):122006, 2014.
- [116] R. U. Abbasi, M. Abe, T. Abu-Zayyad, M. Allen, R. Anderson, R. Azuma, E. Barcikowski, J. W. Belz, D. R. Bergman, S. A. Blake, and Others. Study of Ultra-High Energy Cosmic Ray composition using Telescope Arrays Middle Drum detector and surface array in hybrid mode. *Astroparticle Physics*, 64:49–62, 2015.
- [117] R. Abbasi, J. Bellido, J. Belz, V. de Souza, W. Hanlon, D. Ikeda, J. P. Lundquist, P. Sokolsky, T. Stroman, Y. Tameda, and Others. Report of the Working Group on the Composition of Ultra High Energy Cosmic Rays. *arXiv preprint arXiv:1503.07540*, 2015.
- [118] The Pierre Auger Collaboration. Muons in air showers at the Pierre Auger Observatory: Measurement of atmospheric production depth. *Physical Review D*, 90(1):12012, 2014.
- [119] The Pierre Auger Collaboration. Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events. *Physical Review D*, 91(3):32003, 2015.
- [120] The Pierre Auger Collaboration. Erratum: Muons in air showers at the Pierre Auger Observatory: Measurement of atmospheric production depth [Phys. Rev. D 90 , 012012 (2014)]. *Phys. Rev. D*, 92(1):19903, 2015.

- [121] I. Sidelnik and The Pierre Auger Collaboration. Measurement of the first harmonic modulation in the right ascension distribution of cosmic rays detected at the Pierre Auger Observatory: towards the detection of dipolar anisotropies over a wide energy range. *arXiv preprint arXiv:1307.5059*, 2013.
- [122] V. M. Aranda, F. Arqueros Martínez, D. García Pinto, M. Flores, I. Andrés, J. Rosado, and J. R. Vázquez Peñas. Large scale distribution of ultra high energy cosmic rays detected at the Pierre Auger Observatory with zenith angles up to 80 degrees. *Astrophysical journal*, 802(2), 2015.
- [123] R. Bonino, V. V. Alekseenko, O. Deligny, P. L. Ghia, M. Grigat, A. Letessier-Selvon, H. Lyberis, S. Mollerach, S. Over, and E. Roulet. The East-West method: an exposure-independent method to search for large-scale anisotropies of cosmic rays. *The Astrophysical Journal*, 738(1):67, 2011.
- [124] The Pierre Auger Collaboration. Searches for Large-scale Anisotropy in the Arrival Directions of Cosmic Rays Detected above Energy of 1019 eV at the Pierre Auger Observatory and the Telescope Array. *The Astrophysical Journal*, 794(2):172, 2014.
- [125] P. Billoir and O. Deligny. Estimates of multipolar coefficients for searching for cosmic ray anisotropies with non-uniform or partial sky coverage. *Journal of Cosmology and Astroparticle Physics*, 2008(02):9, 2008.
- [126] The Pierre Auger Collaboration. The Pierre Auger Observatory Upgrade Preliminary Design Report. *Pierre Auger Collaboration Internal*.
- [127] The Pierre Auger Collaboration. A search for point sources of EeV neutrons. *The Astrophysical Journal*, 760(2):148, 2012.
- [128] The Pierre Auger Collaboration. A targeted search for point sources of EeV neutrons. *The Astrophysical Journal Letters*, 789(2):L34, 2014.
- [129] M. Settimo and The Pierre Auger Collaboration. An update on a search for ultra-high energy photons using the pierre auger observatory. page 17, 2011.
- [130] V. S. Berezhinsky and A. Y. Smirnov. Cosmic neutrinos of ultra-high energies and detection possibility. *Astrophysics and Space Science*, 32(2):461–482, 1975.
- [131] The Pierre Auger Collaboration. Improved limit to the diffuse flux of ultrahigh energy neutrinos from the Pierre Auger Observatory. *Physical Review D*, 91(9):92008, 2015.
- [132] The Pierre Auger Collaboration. Reconstruction of inclined air showers detected with the Pierre Auger Observatory. *JCAP*, 1408(08):19, 2014.
- [133] X. Bertou, P. S. Allison, C. Bonifazi, P. Bauleo, C. M. Grunfeld, M. Aglietta, F. Arneodo, D. Barnhill, J. J. Beatty, N. G. Busca, and Others. Calibration of the surface array of the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 568(2):839–846, 2006.

- [134] M. Aglietta, I. De Mitri, S. Maglio, S. Maldera, I. C. Maris, D. Martello, G. Navarra, and M. Roth. Recovery of saturated signals of the surface detector. *Pierre Auger Collaboration Technical Note*. GAP-2008-030.
- [135] F. Arneodo, S. Maldera, and G. Navarra. A procedure to derive the anode charges for saturated signals from the dynode and anode undershoots. *Pierre Auger Collaboration Technical Note*. GAP-2005-025.
- [136] The Pierre Auger Collaboration. Correlation of the highest-energy cosmic rays with the positions of nearby active galactic nuclei. *Astropart.Phys.*, 29:188–204, 2008.
- [137] The Pierre Auger Collaboration. Correlation of the highest-energy cosmic rays with nearby extragalactic objects. *Science*, 318(5852):938–943, 2007.
- [138] The Pierre Auger Collaboration. Update on the correlation of the highest energy cosmic rays with nearby extragalactic matter. *Astroparticle Physics*, 34(5):314–326, 2010.
- [139] D. Newton, J. Knapp, and A. A. Watson. The optimum distance at which to determine the size of a giant air shower. *Astroparticle Physics*, 26(6):414–419, 2007.
- [140] C. Bonifazi, A. Letessier-Selvon, E. M. Santos, P. A. Collaboration, and Others. A model for the time uncertainty measurements in the Auger surface detector array. *Astroparticle Physics*, 28(6):523–528, 2008.
- [141] M. Ave and The Pierre Auger Collaboration. Reconstruction accuracy of the surface detector array of the Pierre Auger Observatory. *arXiv preprint arXiv:0709.2125*, 2007.
- [142] R. Pesce and The Pierre Auger Collaboration. Energy calibration of data recorded with the surface detectors of the Pierre Auger Observatory: an update. 2011.
- [143] The Pierre Auger Collaboration. Searches for anisotropies in the arrival directions of the highest energy cosmic rays detected by the Pierre Auger observatory. *The Astrophysical Journal*, 804(1):15, 2015.
- [144] I. Lhenry-Yvon and P. L. Ghia. Implementation of PMTs Quality Cuts in Auger Sd Data. *Pierre Auger Collaboration Technical Note*. GAP-2009-080.
- [145] R. Sato. Long Term Performance of the Surface Detectors of the Pierre Auger Observatory. In *International Cosmic Ray Conference*, volume 3, page 204, 2011.
- [146] P. Billoir, L. Caccianiga, and P. L. Ghia. A trial to restore uniformity of SD response over space and time. *Pierre Auger Collaboration Technical Note*. GAP-2014-098.
- [147] M.-P. Véron-Cetty and P. Véron. A catalogue of quasars and active nuclei: 12th edition. *A&A*, 455(2):773–777, 2006.
- [148] K.-H. Kampert and The Pierre Auger Collaboration. Highlights from the Pierre Auger Observatory. *arXiv preprint arXiv:1207.4823*, 2012.

- [149] J. P. Huchra, L. M. Macri, K. L. Masters, T. H. Jarrett, P. Berlind, M. Calkins, A. C. Crook, R. Cutri, P. Erdo\ugdu, E. Falco, and Others. The 2MASS Redshift Surveydescription and data release. *The Astrophysical Journal Supplement Series*, 199(2):26, 2012.
- [150] W. H. Baumgartner, J. Tueller, C. B. Markwardt, G. K. Skinner, S. Barthelmy, R. F. Mushotzky, P. A. Evans, and N. Gehrels. The 70 month swift-BAT all-sky hard X-ray survey. *The Astrophysical Journal Supplement Series*, 207(2):19, 2013.
- [151] S. van Velzen, H. Falcke, P. Schellart, N. Nierstenhöfer, and K.-H. Kampert. Radio galaxies of the local universe-All-sky catalog, luminosity functions, and clustering. *Astronomy & Astrophysics*, 544:A18, 2012.
- [152] J. J. Condon, W. D. Cotton, E. W. Greisen, Q. F. Yin, R. A. Perley, G. B. Taylor, and J. J. Broderick. The NRAO VLA sky survey. *The Astronomical Journal*, 115(5):1693, 1998.
- [153] T. Mauch, T. Murphy, H. J. Buttery, J. Curran, R. W. Hunstead, B. Piestrzynski, J. G. Robertson, and E. M. Sadler. SUMSS: a wide-field radio imaging survey of the southern sky-II. The source catalogue. *Monthly Notices of the Royal Astronomical Society*, 342(4):1117–1130, 2003.
- [154] G. R. Farrar and A. Gruzinov. Giant AGN flares and cosmic ray bursts. *The Astrophysical Journal*, 693(1):329, 2009.
- [155] T. Deyoung. The IceCube neutrino observatory: Status and initial results. *Advances in Space Research*, 51(2):238–241, 2013.
- [156] The IceCube Collaboration, The Pierre Auger Collaboration, and The Telescope Array Collaboration. Search for a correlation between the UHECRs measured by the Pierre Auger Observatory and the Telescope Array and the neutrino candidate events from IceCube. *PoS (ICRC 2015) 1082*, 2015.
- [157] M. S. Sutherland, B. M. Baughman, and J. J. Beatty. CRT: A numerical tool for propagating ultra-high energy cosmic rays through Galactic magnetic field models. *Astroparticle Physics*, 34:198–204, 2010.
- [158] M. S. Pshirkov, P. G. Tinyakov, P. P. Kronberg, and K. J. Newton-McGee. Deriving the global structure of the galactic magnetic field from Faraday rotation measures of extragalactic sources. *The Astrophysical Journal*, 738(2):192, 2011.
- [159] R. U. Abbasi, M. Abe, T. Abu-Zayyad, M. Allen, R. Anderson, R. Azuma, E. Barcikowski, J. W. Belz, D. R. Bergman, S. A. Blake, and Others. Indications of Intermediate-Scale Anisotropy of Cosmic Rays with Energy Greater Than 57 EeV in the Northern Sky Measured with the Surface Detector of the Telescope Array Experiment. *The Astrophysical Journal Letters*, 790(2):L21, 2014.
- [160] M. Fukushima. Recent Results from Telescope Array. 2015.

- [161] F. Salamida and the Pierre Auger Collaboration. Update on the measurement of the CR energy spectrum above 1018 eV made using the Pierre Auger Observatory. *Proceedings of the 32nd ICRC, Beijing*, 2011.
- [162] S. Argiro, S. L. C. Barroso, J. Gonzalez, L. Nellen, T. Paul, T. A. Porter, L. Prado Jr., M. Roth, R. Ulrich, and D. Veberic. The Offline Software Framework of the Pierre Auger Observatory. *Nuclear Instruments and*, 707, July 2007.
- [163] M. Lemoine and G. Sigl. *Physics and astrophysics of ultra high energy cosmic rays*, volume 576. Springer Science & Business Media, 2001.
- [164] D. Harari, D. Melob, S. Mollerach, E. Roulet, and G. van Aar. On the curvature radius R_c , the effect of the time variance and the start-times, and its correlation with X_{max} . *Pierre Auger Collaboration Technical Note*. GAP-2012-145.
- [165] M. T. Dova, M. Manceido, A. G. A. Mariazzi, M. E. Mancenido, A. G. A. Mariazzi, H. Wahlberg, F. Arqueros, and Others. Time asymmetries in extensive air showers: a novel method to identify UHECR species. *Astropart.Phys.*, 31(March 2009):312–319, 2009.
- [166] D. G. Pinto and The Pierre Auger Collaboration. Measurements of the longitudinal development of air showers with the Pierre Auger Observatory. 2011.
- [167] The Pierre Auger Collaboration. The Lateral Trigger Probability function for the Ultra-High Energy Cosmic Ray showers detected by the Pierre Auger Observatory. *Astroparticle Physics*, 35(5):266–276, 2011.
- [168] G. Ros, A. D. Supanitsky, G. A. Medina-Tanco, L. Del Peral, J. C. D’Olivo, and M. D. Rodríguez Frías. A new composition-sensitive parameter for ultra-high energy cosmic rays. *Astroparticle Physics*, 35(3):140–151, 2011.
- [169] A. Castellina and G. Navarra. The muon/em ratio and the SD energy scale. *Pierre Auger Collaboration Technical Note*. GAP-2007-126.
- [170] C. Jarne, M. T. Dova, and H. Wahlberg. Update on the study of R1000 distributions. *Pierre Auger Collaboration Technical Note*. GAP-2014-116.
- [171] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS, ACAT:40*, 2007.
- [172] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss. TMVA 4 User Guide. *arXiv preprint physics/0703039*, 2007.
- [173] Y. Freund and R. E. Schapire. *J. of Comput. and Syst. Sci.*, 55:119–139, 1997.
- [174] A. K. O. Tiba, G. A. Medina-Tanco, and S. J. Sciutto. Neural networks as a composition diagnostic for ultrahigh energy cosmic rays. 2005.
- [175] V. Vapnik, I. Guyon, and T. Hastie. Support vector machines, 1995.
- [176] P. Lipari. Concepts of “age” and “universality” in cosmic ray showers. *Phys. Rev. D*, 79(6):63001, March 2009.

- [177] X. Bertou. Thoughts for a new year: the Universality - A new paradigm in Auger? *Pierre Auger Collaboration Technical Note*. GAP-2013-002.
- [178] D. Harari, S. Mollerach, and E. Roulet. Revisiting the anisotropy hints using cosmic ray composition indicators. *Pierre Auger Collaboration Technical Note*. GAP-2014-118.

Glossary

Offline One of the two softwares used to reconstruct Auger events, the other being the Herald.

X_{\max} Depth of shower maximum. A key characteristic of the shower, in particular mass-sensitive. See chapter 2 for details.

2MRS 2MASS Redshift Survey, a catalog of nearby galaxies observed in IR. See paragraph 6.4.1 for details.

454-Evts dataset The dataset selected in chapter 5, containing events with $E \geq 40$ EeV and $\theta < 60^\circ$. As the name says, it is composed of 454 events recorded from 1st January 2004 to 31st March 2014 .

602-Evts dataset The 454-Evts dataset added with 148 horizontal events with $E \geq 40$ EeV and $60^\circ < \theta < 80^\circ$. It is described at the end of chapter 5. As the name says, it is composed of 602 events recorded from 1st January 2004 to 31st March 2014 .

Auger The Pierre Auger Observatory, named after Pierre Victor Auger, french physicists. It is described in details in chapter 3.

Cen A Centaurus A (NGC 5128), the closest radiogalaxy. Located at R.A. 13h 25.5m and Dec. $-43^\circ 01'$, it is classified as FRI. See paragraph 6.5 for details.

EAS Extensive Air Shower, the shower induced by a cosmic ray hitting Earth's atmosphere.

EPOS An high energy hadronic interaction model, used for simulating EAS. Its most recent version is EPOS-LHC. See chapter 2 for details.

FD Fluorescence Detector, one of the two main component of the Auger detector. See section 3.3 for details.

FR I and II Fanaroff-Riley type I and II radiogalaxies. A classification of radiogalaxy based on the morphology of their lobes. See details in section 1.3.2.

GMF Galactic Magnetic Field, the magnetic field of our Galaxy.

GZK effect Effect of interaction of UHECRs with cosmic backgrounds. See paragraph 1.2.2 for details. It is named after the physicists who proposed it in the first place: Greisen, Zatsepin and Kuz'min. The **GZK horizon** is the limit on the distance of a candidate source of UHECRs derived from this effect.

Herald One of the two softwares used to reconstruct Auger events, the other being the Offline. Herald is the software used throughout this thesis except if differently specified.

QGSJET An high energy hadronic interaction model, used for simulating EAS. Its most recent version is QGSJETII-04. See chapter 2 for details.

SD Surface Detector, one of the two main component of the Auger detector. See section 3.2 for details.

Station One of the 1600 water Čerenkov detectors that form Auger SD. See paragraph 3.2 for details.

Swift-BAT Catalog of sources detected in 70 months of observation of the BAT hard X-ray detector on the Swift gamma-ray burst observatory. See paragraph 6.4.1 for details.

TA Telescope Array, an hybrid detector of UHECRs located in Utah, USA. it is described in dedicated boxes in chapter 3.

UHCER Ultra-High Energy Cosmic Rays, the highest energy cosmic rays, usually the convention is that UHECRs have $E > 10^{18}$ eV.

List of Figures

1.1	Cosmic Rays Flux	3
1.2	Composition above 10^{15} eV.	5
1.3	GMF deflections	8
1.4	GZK horizon.	10
1.5	Hillas Plot	13
1.6	AGN scheme	14
1.7	Fanaroff-Riley galaxy classification	15
2.1	Extensive Air Shower.	18
2.2	Heitler's model scheme	19
2.3	Elongation rate.	22
2.4	P-P cross section in models	23
3.1	The Pierre Auger Observatory	30
3.2	SD Station Schematic view	32
3.3	Elementary cell	33
3.4	FD camera schematic view	35
3.5	FD trace	36
3.6	Shower-detector plane	36
3.7	FD-observed shower profile	37
3.8	Auger spectrum	41
3.9	Auger X_{\max} measurements	43
3.10	$\ln A$ predictions from X_{\max} measurements	44
3.11	Composition fractions	44
3.12	Auger X_{\max}^{μ} measurements	47
3.13	$\ln A$ predictions from X_{\max}^{μ} measurements	48
3.14	Number of muons in the showers	49
3.15	Auger γ limits.	51
3.16	Auger γ limits.	53
3.17	Auger ν limits.	54
4.1	T3 configurations	60
4.2	VEM histograms	62
4.3	T5 configurations	64
4.4	Time fit	66
4.5	LDF fit example	69

4.6	Angular resolutions	70
4.7	S_{1000} uncertainties	71
4.8	Constant Intensity Cut	72
4.9	Energy Calibration: S_{38} v.s. E_{FD}	74
5.1	Bad PMT example event	79
5.2	SD Station trigger probability	80
5.3	4T5 example events	81
5.4	Station Age	83
5.5	A/P aging	84
5.6	Decay Time distribution	85
5.7	Decay Time deconvolved traces	86
5.8	DT rate	88
5.9	Angular Resolution	93
5.10	Uncertainty on S_{1000}	94
5.11	Herald v5r2-v4r8 differences	94
5.12	Herald v5r2 - offline v10r0 differences	95
5.13	Long term stability	96
5.14	Energy, Zenith angle and multiplicity distribution for vertical and horizontal events	97
5.15	Dataset skymap	98
6.1	Dataset skymap Galactic	99
6.2	VCV update	101
6.3	Autocorrelation	104
6.4	Cross-correlation 2MRS	107
6.5	Cross-correlation Swift-BAT	108
6.6	Cross-correlation RadioGalaxies	109
6.7	Cross-correlation of events with the AGNs in the Swift catalogue as a function of D and \mathcal{L}_{\min} (top-left panel) and detail of the scan in Ψ and E_{th} for the minimum found (top-right panel). The bottom plots are similar, but showing the p-value computed with the analytical method.	110
6.8	Correlation with bright Radiogalaxies	111
6.9	Correlation with CenA	112
6.10	Deflections in the JF2012 GMF model.	115
6.11	Deflections in the PT2011 GMF model.	116
6.12	Magnetic field deflections with regular, striated and random JF2012 GMF-models	116
6.13	Cross correlation with neutrinos	117
7.1	Simulations energy and zenith distributions	122
7.2	Mass-related observables in shower	124
7.3	Curvature distributions	125
7.4	Rise time azimuthal asymmetry	125
7.5	Rise time azimuthal asymmetry distributions	126
7.6	Lateral Trigger Probability	127
7.7	Number of stations distributions	128

7.8	Beta distributions	129
7.9	S_b distributions	130
7.10	Muonic and electromagnetic traces	130
7.11	S_{1500}^μ distributions	131
7.12	Δ distributions	132
7.13	Example of separation quantification	134
7.14	observables discriminating power	135
7.15	Correlation between observables	136
7.16	Decision Tree	137
7.17	Multilayer perceptron	138
7.18	Support Vector Machine	139
7.19	MVA overtraining	140
7.20	MVA output examples	141
7.21	MVA performance	142
8.1	Map of the most proton-like events	148
8.2	Comparison with Universality and X_{SD}	149
8.3	Swift correlation with proton-enriched sample	151
8.4	CenA correlation with proton-enriched sample	152

List of Tables

3.1	Spectrum fit parameters	42
5.1	Daily PMT cuts	77
5.2	List of the new PMTs parameters introduced by this work. They are checked on an event-by-event basis.	78
5.3	Data loss after DT transformation from 1 st January 2004 to 18 th June 2014 (* 2014 is thus only partial). 6T5 events with $\theta < 60^\circ$ are considered.	87
5.4	Dataset Statistics	92
6.1	Joint Auger-TA-IceCube stacking analysis results	117
6.2	Summary of cross-correlation analyses	119
7.1	Simulations	133
8.1	Candidate protons	147
A.1	Parameters chosen for BDT method.	157
A.2	Parameters chosen for MLP method.	157
A.3	Parameters chosen for SVM method.	158

Acknowledgements

First of all I'd like to thank my exceptional tutors, Piera and Julien, for following me during all these years and teaching me how to be a good scientists. This work could not have been possible without them.

I would like also to thank Viviana, who sent me in this wonderful city in the first place and who taught me all the basis about Auger. I thank Antoine for encouraging me to have this PhD and for helping me in these years spent in his group. I thank also Pierre who helped me understanding the physics of cosmic rays and with whom I developed part of this thesis work. A special thank goes in particular to Mariangela for helping me in an uncountable number of situation, dealing with the Offline , with simulations, with root and with everything i had a problem with.

I thank Laura for helping me with the smoothing algorithm and Claudio for helping me with the energy calibration. I also apologize with Isabelle for bringing up much more problems with the SD that I could help solving. I also thank Ioana, Imen, Valerio, Romain, Miguel, Moritz, Olivier, Hugo and all the members of the Collaboration I've been in touch with for constructive discussions. I particularly thank Esteban, Silvia and Diego for welcoming me in Bariloche and Carola for supporting me during my first conference in Palermo.

I wholeheartedly thank my girlfriend Sara for helping me going through the toughest part of this PhD and her father Jonathan for reading this boring manuscript checking my english.

Last but not least, thank to all my family for supporting me in this french adventure that had not ended yet!