THE $\langle\Delta\rangle$ METHOD: AN ESTIMATOR FOR THE MASS COMPOSITION OF ULTRA-HIGH-ENERGY COSMIC RAYS

Patricia Sánchez Lucas

Universidad de Granada

November 2016

Advisor:
Prof. Antonio Bueno Villar

Departamento de Física Teórica y del Cosmos & CAFPE

Programa de Doctorado en Física y Ciencias del Espacio

Editor: Universidad de Granada. Tesis Doctorales
Autora: Patricia Sánchez Lucas
ISBN: 978-84-9163-088-3
URI: http://hdl.handle.net/10481/44643

El doctorando/*The doctoral candidate* **Patricia Sánchez Lucas** y el director de la tesis/and the thesis supervisor **Antonio Bueno Villar**:

Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección del director de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisor/s and, as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when their results or publications have been used) have been respected.

En Granada, a 21 de Noviembre de 2016.

Director de la tesis | Doctorando

Fdo: Antonio Bueno Villar | Patricia Sánchez Lucas

D. Antonio Bueno Villar, Catedrático de Universidad,

CERTIFICA: que la presente tesis doctoral, THE $\langle\Delta\rangle$ METHOD: AN ESTIMATOR FOR THE MASS COMPOSITION OF ULTRA-HIGH-ENERGY COSMIC RAYS, ha sido realizada por D^{a}. Patricia Sínchez Lucas bajo su dirección en el Dpto. de Física Teórica y del Cosmos de la Universidad de Granada, así como que ésta ha realizado una estancia en el extranjero por un periodo superior a tres meses en la *Universidad de Wuppertal* (Alemania).

En Granada a 21 de Noviembre de 2016

Fdo: Antonio Bueno Villar

Contents

Introduction

Cosmic rays were discovered more than 100 years ago, in the early twentieth century, and since then they have been one of most intriguing questions that physicists have had to face. One of the main characteristics of cosmic rays is their vast energy spectrum, extended from few GeV up to tens of EeV (covering more than 13 orders of magnitude). The ultra-high-energy cosmic rays are only those whose energies are larger than 1 EeV. Cosmic rays are continuously bombarding our atmosphere, but their flux is a steeply decreasing function of the energy, so that for the highest energies only one particle per km^2 and per century reaches the Earth. With this extremely low rate at the highest energies the only feasible way to detect a significant number of particles is by deploying detectors covering enormous areas on the ground. Discarding the option of making measurements above the atmosphere.

When cosmic rays with these extreme energies arrive to the Earth, they collide with the atmospheric nuclei giving rise to huge showers of billions of secondary particles which propagate through the atmosphere until they are absorbed or reach the ground. The study of ultra-high-energy cosmic rays is done exclusively through these secondary particles, customarily known as Extensive Air Showers (EAS), with no possibility of studying directly the primary particle.

Our understanding of the properties of the highest-energy cosmic rays has grown enormously over the last 12 years with the advent of data collected with giant arrays of detectors, as the one of the Pierre Auger Observatory. The Pierre Auger Observatory, located in the province of Mendoza, Argentina, is the largest and most sensitive apparatus ever built to record and study EASs. Covering an area of 3000 km^2, this observatory was devised to reveal the nature of ultra-high-energy cosmic rays thanks to a hybrid design which allows the combination of two detection techniques: the detection of fluoresce light and the sampling of the particles that reach the ground.

However, despite the years of work and effort in this field, even these days there is a large number of unanswered questions related to the nature and origin of ultra-high-energy cosmic rays. One of these puzzles is the determination of the mass composition at the highest energies. The study of the mass spectrum of ultra-high-energy cosmic rays is especially difficult due to two reasons. On the one hand, the most adequate observable, X_{max}, is based on fluorescence measurements. This means that the observations are restricted to clear moonless nights, with the subsequent reduction of statistics at energies larger than $10^{19.5}$ eV. On the other hand, to interpret the data, one must use the predictions of hadronic interaction models at centre-of-mass energy around $\sqrt{s}$ $\sim$300 TeV, well beyond what is accessible in particle accelerators as the LHC (14TeV). This fact is particularly problematic taking into account that recent observations of the Pierre Auger Observatory suggest that these predictions are inadequate to describe the had-

ronic component of the EASs.

One of the possible solutions to increase the statistics at the highest energies is the use of alternative observables to X_{max}, based on data collected with arrays of surface detectors, where the duty cycle is nearly 100%. Nevertheless, most of these observables can not be used to make inferences about mass because they are related to the hadronic component of the air showers and thus the comparison with models result in unreliable predictions.

All these obstacles make obvious the necessity of new methods for mass composition studies which allow facing the problem from a new perspective. These new methods should be based on measurements of surface detectors to increase the statistics at the highest energies. It would be also desirable that they are not related to the hadronic component of the EASs, to allow a more reliable comparison with models.

The subject of this thesis follows exactly this approach. Using an observable obtained from the surface detectors of the Pierre Auger Observatory, the risetime, we develop a method to infer the mass composition of ultra-high-energy cosmic rays that fulfills the previous requirements.

This thesis is organized as follows. Chapter 1 gives an overview of cosmic rays, including some of the most relevant experimental results, with special focus on the unclear situation regarding mass composition. In chapter 2 the main features of the Pierre Auger Observatory are described in detail, including the reconstruction of the most relevant observables. Chapter 3 is dedicated exclusively to the experimental determination of the risetime in the Pierre Auger Observatory. In chapter 4 we introduce the specific method used in this thesis to study mass composition: the $\langle\Delta\rangle$ method. In this chapter we cover the main aspects that guarantee a proper application of the method. In chapters 5 and 6 we apply the $\langle\Delta\rangle$ method to data collected with the Pierre Auger Observatory. In chapter 5 we use the data collected with the array of surface detectors whose separation is 1500 m while in chapter 6 we use a smaller array with a separation between detectors amounting to 750 m. The difference between both data sets comes from the different energy ranges probed. The 1500 m array is fully efficient for events with energies above 3 EeV while the 750 m array provides data at lower energies. The chapter 7 shows the combination of the mass composition results obtained in chapters 5 and 6. Finally, in chapter 8 we conclude this thesis using the observable $\langle\Delta\rangle$ for an additional purpose: to assess the level of concordance between data and the predictions provided by hadronic interactions models tuned to reproduced LHC data.

Introducción

Los rayos cósmicos fueron descubiertos hace más de 100 años, a principios del siglo XX, y desde entonces han sido uno de los problemas más intrigantes que los físicos han tenido que afrontar. Una de las principales características de los rayos cósmicos es su amplio espectro de energía, que se extiende desde unos pocos de GeV hasta decenas de EeV (cubriendo más de 13 órdenes de magnitud). Los rayos cósmicos de ultra alta energía son sólo aquellos cuyas energías son superiores a 1 EeV. Los rayos cósmicos están continuamente bombardeado nuestra atmósfera, pero su flujo es una función que decrece abruptamente con la energía, así que con las energías más altas sólo nos llega a la Tierra una partícula por km^2 y por siglo. Con esta tasa tan extremadamente baja para las más altas energías, la única forma factible de detectar un número significativo de estas partículas es desplegando detectores de forma que cubran inmensas áreas en el suelo. Descartando cualquier posibilidad de hacer este tipo de medidas por encima de la atmósfera.

Cuando rayos cósmicos con estas energías tan extremas llegan a la Tierra colisionan con los núcleos atmosféricos dando lugar a una cascada de millones de partículas secundarias que se propagan a través de la atmósfera hasta que son absorbidas o llegan al suelo. El estudio de los rayos cósmicos de ultra alta energía tiene que hacerse exclusivamente a través de estas cascadas de partículas secundarias, comúnmente conocidas como EASs, por su acrónimo del inglés (*Extensive Air Showers*), sin posibilidad de estudiar de forma directa la partícula primaria.

Nuestro conocimiento sobre los rayos cósmicos de ultra alta energía ha crecido enormemente en los últimos 12 años con los datos recogidos por las gigantes extensiones de detectores, como la que tiene el Observatorio Pierre Auger. El Observatorio Pierre Auger, localizado en la provincia de Mendoza, Argentina, es el aparato más sensible y más grande jamás construido para registrar y estudiar EASs. Cubriendo un área de 3000 km^2, este observatorio fue ideado para desvelar la naturaleza de los rayos cósmicos de ultra alta energía gracias a un diseño híbrido que permite la combinación de dos técnicas de detección: la detección de luz de fluorescencia y el muestreo de partículas que llegan al suelo.

Sin embargo, a pesar de todos los años de trabajo y esfuerzo dedicados a este campo, todavía en estos días hay un gran número de preguntas sin resolver relacionadas tanto con la naturaleza como con el origen de los rayos cósmicos más energéticos. Uno de estos misterios es la determinación de la composición química a las energías más altas. El estudio del espectro de masas de los rayos cósmicos de ultra alta energía es especialmente difícil por dos razones. Por un lado, el observable más adecuado, X_{max}, se basa en medidas de fluorescencia. Esto significa que las observaciones están restringidas a noches despejadas en las que no haya luna, con la consecuente reducción

de la estadística a energías superiores a $10^{19,5}$ eV. Por otro lado, para interpretar estos datos, uno debe usar las predicciones de los modelos de interacción hadrónica a energías centro de masas del orden $\sqrt{s} \sim 300$ TeV, mucho más allá de lo que es accesible en aceleradores de partículas como el LHC (14 TeV). Esto hecho es particularmente problemático teniendo en cuenta que recientes observaciones del Observatorio Pierre Auger sugieren que estas predicciones son inadecuadas para describir la componente hadrónica de las EASs.

Una de las posibles soluciones para incrementar la estadística a las más altas energías es el uso de observables alternativos a X_{max}, que estén basados en datos recogidos con los grandes conjuntos de detectores de superficie, donde el ciclo de trabajo es cercano al $100\,\%$. Sin embargo, la mayoría de estos observables no se pueden usar para sacar conclusiones sobre la masa porque están relacionados con la componente hadrónica de las cascadas y por tanto la comparación con los modelos resulta en predicciones poco fiables.

Todos estos obstáculos hacen obvia la necesidad de nuevos métodos para estudiar composición que permitan afrontar el problema desde una nueva perspectiva. Estos nuevos métodos deberían estar basados en medidas de detectores de superficie para incrementar la estadística a las energías más altas. Y además sería deseable que los nuevos métodos no estuviesen relacionados con la componente hadrónica de las EASs, para permitir una comparación más fidedigna con los modelos.

El objetivo de esta tesis sigue exactamente ese enfoque. Usando un observable obtenido con los detectores de superficie del Observatorio Pierre Auger, el risetime, desarrollamos un método para deducir la composición química de los rayos cósmicos de ultra alta energía que cumple con los requerimientos mencionados anteriormente.

Esta tesis se organiza del siguiente modo. El capítulo 1 da una visión general de los rayos cósmicos, incluyendo algunos de los resultados experimentales más relevantes y con especial atención a la incierta situación de la composición química. En el capítulo 2 describimos en detalle las principales características del Observatorio Pierre Auger, incluyendo la reconstrucción de los observables más relevantes. El capítulo 3 está dedicado exclusivamente a la medida experimental del risetime en el Obsevatorio Pierre Auger. En el capítulo 4 introducimos el método específico utilizado en esta tesis para estudiar la composición: el método $\langle\Delta\rangle$. En este capítulo cubrimos los principales aspectos que garantizan la utilización adecuada del método. En los capítulos 5 y 6 aplicamos el método $\langle\Delta\rangle$ a los datos registrados por el Observatorio Pierre Auger. En el capítulo 5 usamos los datos registrados por el conjunto de detectores de superficie que están separados 1500m mientras que en capítulo 6 usamos un conjunto más pequeño cuyos detectores están separados sólo 750 m. La diferencia entre ambos conjuntos de datos procede de los diferentes rangos de energía explorados. El conjunto que tiene una separación de 1500 m es completamente eficiente por encima de 3 EeV mientras que el otro proporciona datos de menor energía. El capítulo 7 muestra la combinación de los resultados sobre composición obtenidos en los capítulos 5 y 6. Finalmente, en el capítulo 8 concluimos esta tesis usando el observable $\langle\Delta\rangle$ para un propósito adicional: evaluar el nivel de concordancia entre los datos y las predicciones proporcionadas por los modelos de interacción hadrónica afinados para reproducir los datos del LHC.

1 Ultra High Energy Cosmic Rays

The Earth's atmosphere is being continuously bombarded by relativistic particles that, in some cases, have a macroscopic kinetic energy equivalent to that of a tennis ball moving at 100 km/h. These singular particles are known as ultra-high-energy cosmic rays (UHECRs) and although they were discovered more than 100 year ago, nowadays they remain one of the biggest mysteries of fundamental physics. The energy of these particles is far beyond the limits reached in any accelerator. The centre of mass energy in LHC is (in excellent conditions) 14×10^{12} eV while for a cosmic ray proton with a momentum of the order of 10^{20} eV hitting an atmospheric nucleus it is around 10^{15} eV. As a consequence, the only theoretical predictions that we can do for cosmic ray studies are extrapolations from much lower-energy data, which introduces large uncertainties in the conclusions that we can obtain from the observations. Because of the rarity of these particles even now we know relatively little about them. In particular we are unable to identify unequivocally their sources of production, their mechanisms of acceleration or their chemical composition.

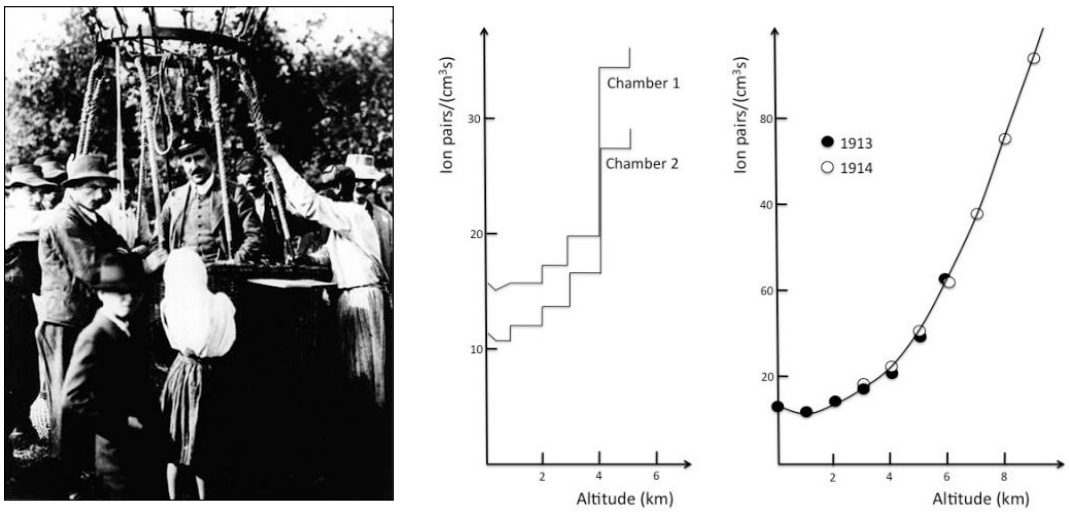
One of the biggest challenges that physicists have to overcome for the study of UHECRs is their extremely low flux. For energies above 1 EeV we expect less than one particle per km^2 and per year. This means that the only feasible way to detect a significant number of these particles is by covering huge areas with detectors having above hundreds of km^2. This is not only a problem from a technological point of view and cost, but also from the point of view of physics, since it makes impossible the access to the first interaction in the top of the atmosphere. All these features make the study of UHECRs one of the most laborious tasks in physics.

1.1 The discovery of cosmic rays

The first evidences of cosmic rays date back to the end of 18th century when the atmospheric ionisation was discovered by Charles Augustin de Coulomb through the observa-

tions of the spontaneous discharge of an electrometer [1]. With the later discovery of natural radioactivity in 1896 by Becquerel [2], a great research interest grew out around this radiation observed in the air. The simplest hypothesis was that the radiation was caused by the radioactive materials on Earth.

At the beginning of 20th century there was a huge controversy about where this radiation came from. To disentangle the problem a lot of measurements of the air ionisation were done in different zones of the Earth's surface without no essential change in the levels of radiation. Theodor Wulf was the first who started measuring changes in radioactivity with height. The idea was simple: if the radioactivity was coming from the Earth's surface, it should decrease with height. It was famous the measurement that he did in the top of the Eiffel tower in 1909 [3]. However the decrease of radiation that he found was too small to discard any hypothesis. There was also a notable importance in the measurements done by Domenico Pacini in 1911 below the sea, in the shore of Liborno [4]. The significant decrease that he measured under the sea was one of the first evidences that the ionizing radiation must come from the atmosphere.



(a) (b)

Figura 1.1: (a) Hess before his balloon flight on 7 August 1912. (b) Increase in ionization with height measured by Hess and Kolhörster [5].

After these evidences several researchers did their measurements on balloon flights. However in 1912 Victor Hess was the first one who achieved enough height (5300 m) to measure that the ionization increased by a factor of 2 compared to the one measured at the surface [6] (see figure 1.1). Besides, given that his experiment coincided with a near-total eclipse of the Sun he was able to deduce that the famous radiation comes from further out in space, ruling out the hypothesis of its terrestrial origin. Kolhörster confirmed Hess's findings in 1914 flying up to a height of 9300 m where he found that the ionization was nine times the value on the ground. For this discovery Hess was awarded the Nobel

Prize in Physics in 1936.

The term *cosmic rays* was coined by Robert Milikan in the 1920s, by similarity with the term *gamma rays*, after a series of experiments to prove the electromagnetic nature of this radiation [7]. However this hypothesis was soon discarded by Kolhörster and Bothe [8] and independently by Bruno Rossi [9] after realizing that some cosmic rays can penetrate very heavy absorbers. In 1930 Bruno Rossi and others organized expeditions at different geomagnetic latitudes to study the interactions of cosmic rays with the geomagnetic field and they proved that the majority of cosmic rays were positive, since more particles come from the West than from the East [10].

Towards the end of the 1930s, Pierre Auger and collaborators made several experiments at high mountain altitudes where they ran in coincidence Geiger-Muller tubes at large distances from each other [11, 12]. They concluded that primary cosmic rays generate showers of particles in the atmosphere, although this fact was suspected even before by Rossi. A shower of particles in the atmosphere was called Extensive Air Shower. The detailed work to deploy these counters in the mountains led Auger to estimate that some of the showers were initiated by cosmic rays of around 10^{15} eV. What was clear at that time is that Auger had detected particles with energy five orders of magnitude greater than anything conceived hitherto. The first claim for a shower with a ultra high energy, above 10^{20} eV, was done by Linsley in 1963 at Volcano Ranch experiment [13]. This measurement opened the field of study of UHECRs.

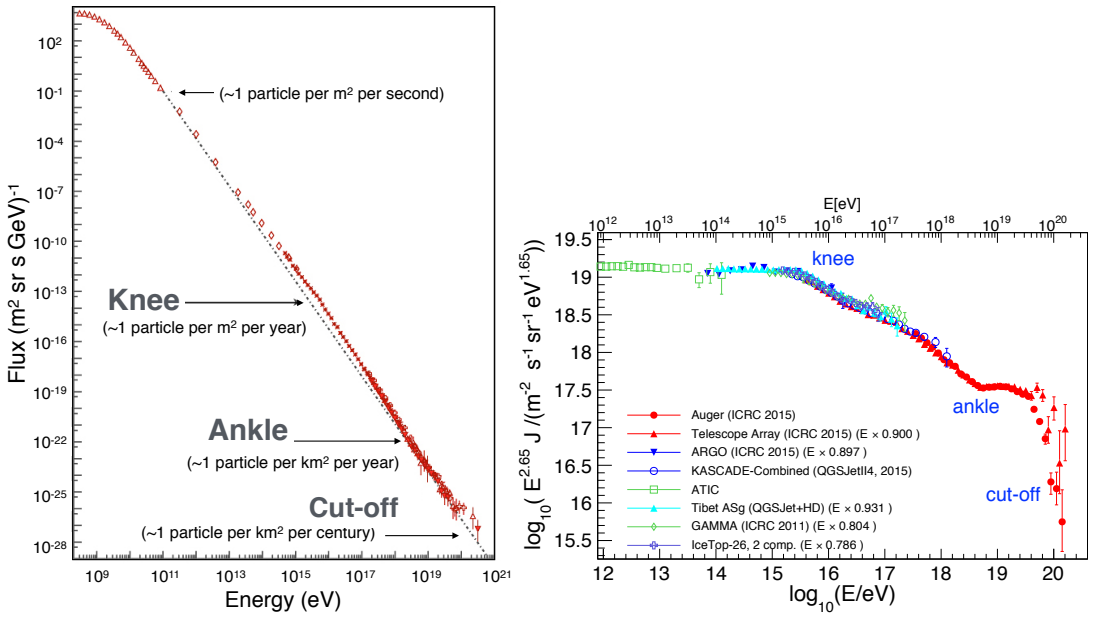
The study of cosmic rays was the basis for the development of Particle Physics that we know nowadays. Discoveries as important as the antimatter, the muons, the mesons, and a whole set of new particles were found doing cosmic rays experiments. In fact, the postulation of the quark model by Gell-Mann and Zweig [14, 15] is based in cosmic ray results, since the kaon, the first strange particle ever seen, was found using a cloud chamber exposed to the cosmic radiation [16].

1.2 The cosmic ray spectrum

The energy spectrum of cosmic rays is the flux of primary particles as a function of the energy, J. It can be described as a rather featureless power law function extending over more than ten orders of magnitude, from energies around 10^9 eV up to more than 10^{20} eV. The exponent of this power law (*the spectral index*), is almost constant and close to 3

$$J \propto \frac{dN}{dE} \propto E^{-\gamma}, \qquad \gamma \approx 3 \tag{1.1}$$

Figure 1.2a shows the cosmic ray flux (number of particles per unit of area, energy, solid angle and time) as a function of energy. Three regions indicated in the figure exhibit a particularly interesting deviation from the average behaviour: *the Knee*, *the Ankle* and *the Cut-off*. The properties of each region are discussed in the next sections. To see clearly the different features of the spectrum at highest energies the flux has been multiplied by $E^{2,65}$ in figure 1.2b. A change in the acceleration mechanism or in the propagation process or in the hadronic interaction cross section with increasing energy can explain the different spectral features.



(a) (b)

Figura 1.2: (a) Cosmic ray flux over 13 decades in energy [17, 18]. (b) Energy spectrum of cosmic rays measured by different experiments above 1TeV. The spectrum is multiplied by $E^{2,65}$ to enhance the features [19].

1.2.1 The Knee

The knee is the point where the spectral index changes from $\sim$ 2.7 to $\sim$ 3.1 and it happens around 10^{15} eV. Several experiments have confirmed the existence of this change in the spectrum: Yakutsk [20], Akeno [21], KASKADE [22] and its extension KASCADE-Grande [23]. Many possible explanations for this spectral feature have been proposed, based mostly on astrophysical reasons, [24]. The standard thinking in the field of cosmic rays is that particles of energy below and around the knee are accelerated at galactic astrophysical objets, mainly at supernova remnants and possible at powerful binary systems [25]. The knee itself is probably a result of reaching the maximum energy of such acceleration process. The current experimental results are not constraining enough to favour one theoretical model over the rest.

1.2.2 The Ankle

The ankle corresponds to a flattening of the spectrum to a spectral index which is again close to $\gamma \sim 2.7$. According to the last measurements of the Pierre Auger Observatory, the ankle is found at E = (4.8 $\pm$ 0.1 $\pm$ 0.8)$\times$ 10^{18} eV, quoting both statistical and systematic uncertainties [26]. This change in the spectrum has been clearly observed by others

experiments as HiRes [27] or TA [28].

The origin of the ankle is frequently understood as the transition from galactic to extragalactic cosmic rays. However the way in which this transition occurs is explained in the different theoretical models with different mass composition of the two components of the flux. In the *ankle model* [29, 30] the extragalactic component is supposed to have a pure proton composition. Likewise, in the *mix composition model* [31], it is favored the possibility of a mixed composition above 10^{19} eV.

Another different explanation is found in the *dip models* [32, 33, 34]. These models predict that the extragalactic component, composed mainly by protons, is extinguished at much lower energies and the ankle is explained by the interaction of protons with the cosmic microwave background: an electron-positron pair is produced reducing the energy of the primary cosmic ray. This causes a suppression of the flux at the largest energies, increasing it at low ones and therefore causing the appearance of the ankle.

1.2.3 The Cut-off

The Pierre Auger Observatory has established unambiguosly the suppression of the cosmic ray flux at E = (4.21 $\pm$ 0.17 $\pm$ 0.76)$\times$ 10^{19} eV, quoting again both statistical and systematic uncertainties [26]. However, the explanation of this cut-off is still a mystery.

Greisen [35] and independently Zatsepin and Kuzmin [36] predicted in 1966, shortly after the discovery of the cosmic microwave background (CMB) by Penzias and Wilson [37], that the cosmic ray energy spectrum should be attenuated at the largest energies if the sources are distributed uniformly in the Universe. The cosmic rays lose part of their energy interacting with the CMB photons and this should produce a drastic flux suppression at the largest energies. This effect is commonly known as the GZK suppression. The reaction taking place is:

$$p + \gamma_{CMB} \longrightarrow \Delta^{+} \longrightarrow p + \pi^{0},\ n + \pi^{+} \tag{1.2}$$

The proton loses about the 20 $\%$ of its energy in the process and this limits the distance from which a high energy particle can travel before its detection to $\sim$ 100 Mpc, see figure 1.3. Analogously, for photons we have the following process:

$$\gamma_{CR} + \gamma_{CMB} \longrightarrow e^{+} + e^{-} \tag{1.3}$$

Therefore, every stable particle apart from neutrinos suffers a significant attenuation at energies above the EeV, meaning that cosmic rays of ultra high energy must have their origin in nearby astrophysical objets.

A key aspect of the GZK effect involves high energy photons and neutrinos. In fact, the pions resulting from equation 1.2 produce photons and neutrinos of very high energy by two possible reactions. The first one would be:

$$p + \gamma_{CMB} \longrightarrow p + \pi^{0} \longrightarrow p + \gamma\gamma \tag{1.4}$$

with two cosmogenic photons in the final state. The second one is:

$$p + \gamma_{CMB} \longrightarrow n + \pi^{+} \longrightarrow n + \mu^{+} + \nu_{\mu} \longrightarrow n + e^{+} + \nu_{e} + \overline{\nu}_{\mu} + \nu_{\mu} \tag{1.5}$$

which produces three neutrinos in the final state.

The discovery of cosmogenic photons and neutrinos would favour a light cosmic ray composition at ultra-high energies. This is because the photo-pion production for heavy primaries has a much higher energy threshold. Hence the importance of ultra-high energy neutrino and photon searches done by the Pierre Auger Collaboration and other experiments.

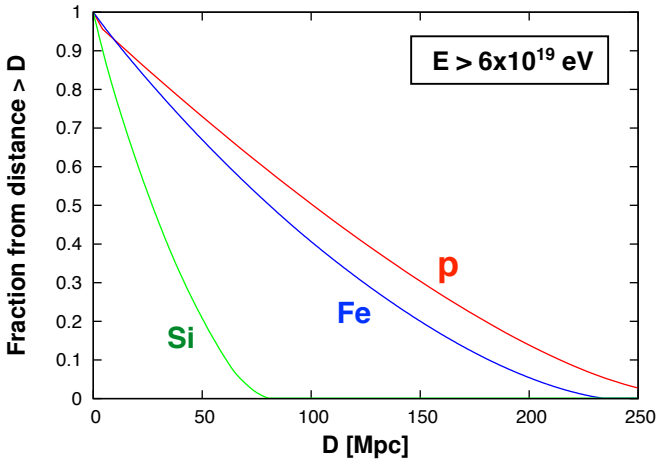


Figura 1.3: GZK horizon: fraction of cosmic rays that arrive at Earth with energy above 6×10^{19} from sources that are farther away than a distance D and that inject protons, silicon and iron nuclei (assuming an uniform distribution of sources with equal intrinsic luminosity and continuous energy loss) [38]. Around 50 $\%$ of the flux of protons and iron nuclei with E >6×10^{19} should come from distances smaller than 100 Mpc.

Nevertheless, the suppression of the cosmic ray flux does not imply that the GZK process exists. The *dissapointing model* [39] attributes this suppression to the maximum energy to which sources can accelerate particles. All in all, the suppression is still one of the most important open questions in the field of cosmic ray physics.

1.3 Origin of cosmic rays

We do not know where UHECRs come from and what astrophysical objects can accelerate those particles in such a vast energy range. Cosmic rays up to some GeV are known to come mostly from the Sun. However, at higher energies they must have a different origin since there is no process taking place in the Sun involving such energies. It would be natural to suppose that the energy in a source capable of accelerating particles up to 10^{20} eV and beyond must be extremely large.

The size of the acceleration region is assumed to be comparable to the Larmor radius of the particle in a magnetic field, which must be sufficiently weak so that the synchrotron losses are not greater that the energy gain. However, this criterion does not specify what acceleration process is involved.

A possible mechanism to accelerate cosmic rays is the diffusive shock acceleration, a subclass of Fermi-acceleration [40]. Charged particles can gain an extremely high energy when they are repeatedly scattered on strong turbulent magnetic fields that arise in shock fronts. This mechanism plays a very important role in many astrophysical models, including jets flared by active galaxy nuclei (AGN) and supernova remnants. Under the assumption of the diffusive shock acceleration, the maximum energy attainable, E_{max}, for a particle of charge Ze when it is accelerated by a magnetic field B in an object of size R is given by [41, 42]:

$$E_{max} = k\beta Ze \left(\frac{B}{\mu G}\right)\left(\frac{R}{kpc}\right) \text{ EeV} \tag{1.6}$$

where β is the shock speed in units of c and k the efficiency of the process (k <1). The different astrophysical objects that satisfy this condition are represented in the so-called Hillas plot, see figure 1.4. This plot only leaves as candidates for UHECRs acceleration Gamma Ray Burst (GRBs), AGN, neutron stars and radio galaxies.

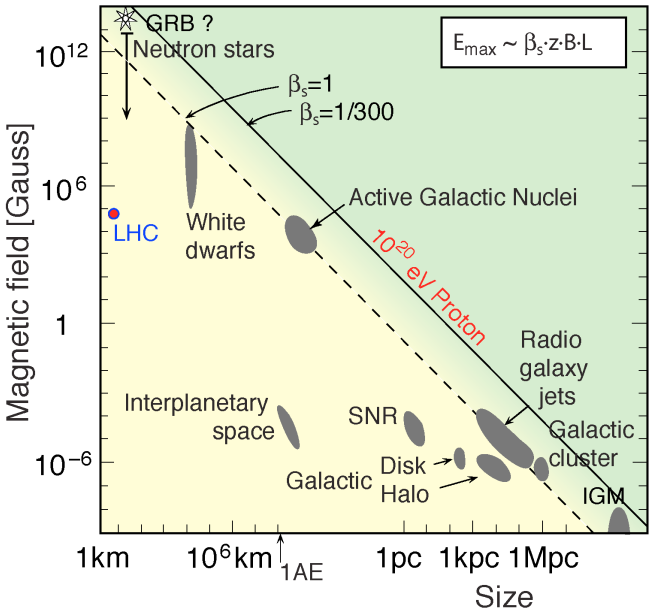


Figura 1.4: Hillas Plot: the magnetic field strength as a function of the size of astrophysical objects that are acceleration candidates for cosmic rays. Objects below the diagonal line cannot accelerate protons to 10^{20} eV.

1.4 Arrival direction and anisotropies in UHECRs

The study of the distributions of arrival directions of cosmic rays gives important hints for our understanding of the transition from a galactic to an extragalactic origin of these particles, and can help us in the location of potential sources.

One of the key pieces in these kind of studies is the distribution of arrival directions on large angular scales. The large-scale anisotropies can be reflective of either a collective motion of cosmic rays or of the global distribution of their sources. A common technique used to study anisotropies is the harmonic expansion (Rayleigh analyses [43]) of the cosmic rays flux in right ascension and in azimuthal angle to obtain the components of a possible dipole along the equatorial plane and along the the Earth's rotation axis, respectively.

With an analysis relying on this technique the Pierre Auger Collaboration has recently observed a dipole structure on a large scale above an energy of E >8 EeV. The largest departure from isotropy appears with an amplitude of the first harmonic in right ascension of $r_{1\alpha}$ = (4.4$\pm$1.0)$\times 10^{-2}$, having a chance probability of P($\geq r_{1\alpha}$) = 6.4$\times 10^{-5}$ [44]. Assuming that the only significant contribution to the anisotropy stems from the dipole component, the result above 8 EeV points to a dipole amplitude of 0.073$\pm$0.015 in the direction (α,δ) = (95°$\pm$13°, -39°$\pm$13°), see figure 1.5a. The origin of this anisotropy is subject of ongoing discussions.

Distributions of arrival directions are also throughly scrutinised to look for potential deviations from isotropy at small angular scales. However, up to now it has not been possible to establish small-angle correlations of arrival directions of Auger data with possible sources beyond any doubt, even though there are some intriguing indications [45, 46, 47].

In a recent study of the Pierre Auger Collaboration, correlations are searched for with known astrophysical structures, such as the Galactic and super-Galactic planes, the Galactic centre and with different populations of nearby extragalactic objects that are considered plausible candidates for UHECR sources. The most significant excess (4.3 σ pre-trial) after a *blind* search was found at an energy threshold of 54 EeV in a radius of 12° close to the super-Galactic plane and in the direction of Centaurus A, [48], see figure 1.5b.

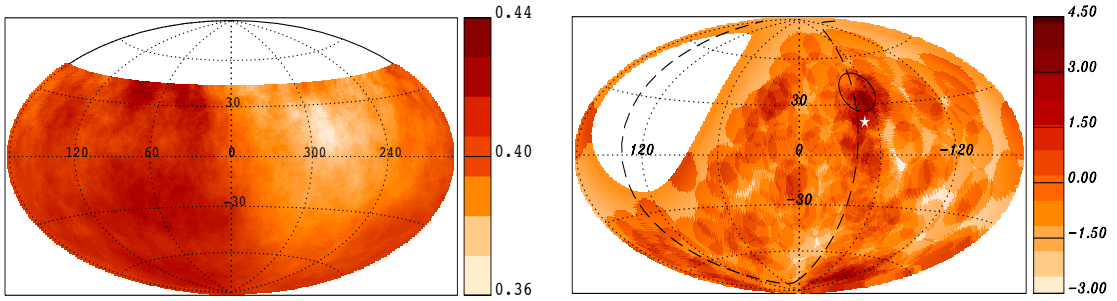


Figura 1.5: Left panel: Sky map in equatorial coordinates of the flux (km^{-2} sr^{-1} yr^{-1}) smoothed in an angular window of 45° for events with E >8 EeV gathered by the Pierre Auger Observatory [44]. Right panel: Map in galactic coordinates of the significance of excesses in 12°-radius window for events with E $\geq$ 54 EeV. The super-Galactic plane (dashed line) and Cen A (white start) are included [48].

1.5 UHE photons and neutrinos

As we pointed out in section 1.2.3, the direct observation of ultra-high energy photons and neutrinos would automatically translate into some restrictions in the mass composition of cosmic rays, favouring a light composition.

The detection of cosmogenic neutrinos and photons of ultra-high energy is even nowadays an experimental challenge for air-shower experiments. The Pierre Auger Collaboration has searched for showers initiated by ultra-high energy neutrinos and photons, but no candidates were found for neutrino-like events and only 4 events survived the photon search, a compatible number with the background expectations [49]. When such a negative result is found, the measured flux of background (in this case high energy cosmic rays), can be used to derive upper bounds on the expected fluxes for cosmogenic neutrinos and photons.

Left panel of figure 1.6 shows the latest limit on the diffuse flux of neutrinos obtained by the Pierre Auger Collaboration [50] together with the current limits from other experiments and the theoretical predictions. The latest results on the flux limit of photons are depicted in the right panel of figure 1.6 [49].

The current limits for the neutrino and photon fluxes highly disfavor top-down models as a possible process of UHECR production. Top-down models is the generic name that refers to every theoretical model in which UHECRs are produced as decay products of some superheavy particle, as for example a dark matter candidate [51]. The current flux limits begin to reach the predicted secondary fluxes from models in which the suppression of cosmic ray flux is originated by the GZK energy loss process for a proton dominated flux. These are the cosmogenic models for neutrinos and photons [52, 53, 54].

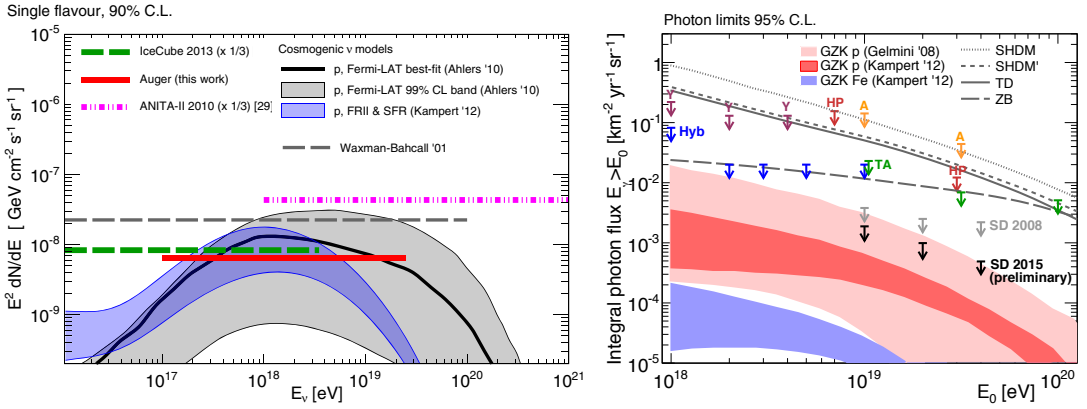
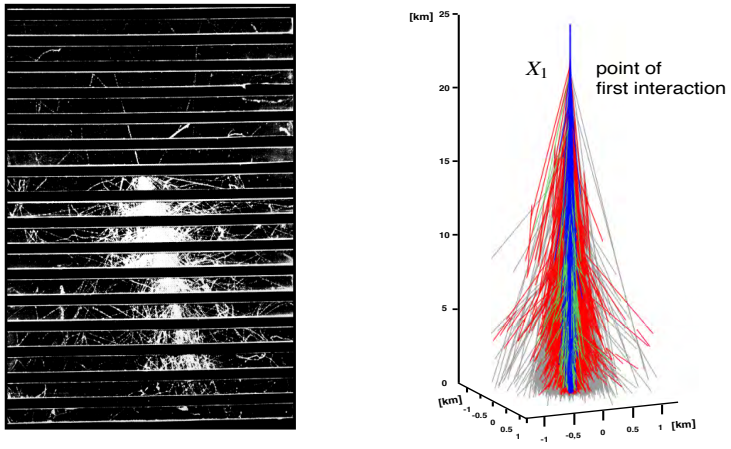


Figura 1.6: Left Panel: Upper limit of the diffuse flux of UHE neutrinos obtained by the Pierre Auger Observatory [50] together the results obtained by other experiments as ANITAII [55] and IceCube [56]. The expected flux for several cosmogenic neutrino models as well as the Waxman-Bahcall bound are also included [52, 57]. Right Panel: Upper limit of the diffuse flux of UHE photons obtained by the Auger Collaboration [49] together with the current limits of other experiments [58, 59, 60, 61, 62] and predictions from top-down models and cosmogenic photons models [53, 54].

1.6 Extensive air showers

When a cosmic ray hits the atmosphere, the interaction with the nuclei results in a highly energetic inelastic collision. The products of such a collision have enough energy to interact again and generate an even a larger set of secondary particles and so on. This process can go on for a few generations, until a vast number of particles (around 10^{10} particles for events of 10^{19} eV) are produced. This mechanism of particle production gives rise to what is known as an EAS. These showers of particles can produce a footprint at ground covering a few kilometers. As shown in the left panel of figure 1.7a, a proton interacts with lead plates and generates a cascade of secondary particles. The tracks observed in the cloud chamber are similar to the tracks produced during the propagation of UHECRs in the atmosphere. An example of an air shower simulated in the atmosphere is shown in the right panel of the same figure.



(a) (b)

Figura 1.7: Left panel: Image of a shower of particles, as seen in a cloud chamber at 3027 m altitude. The primary particle is estimated to be a proton of about 10 GeV [63]. Right panel: Simulation of an extensive air shower. Different colours represent different particles: gray = photons, red = electrons/positrons, green = muons and blue = hadrons.

An EAS can be described as the superposition of different components, see figure 1.8. The most important ones are the hadronic, muonic and electromagnetic components.

After the first interaction, most of the produced secondary particles are pions, photons and electrons (or positrons). Photons, electrons and positrons initiate the electromagnetic shower. Electromagnetic particles are, by a few orders of magnitude, the most abundant component of the shower, and carry the largest fraction of the total energy. Neutral pions decay in two photons with a branching ratio of 98.8 % [64] or to a combination of electrons and photons which contribute to the electromagnetic shower. Charged pions, on the other hand, may interact and produce new secondary particles that will feed the hadronic shower or decay into muons, producing the muonic shower.

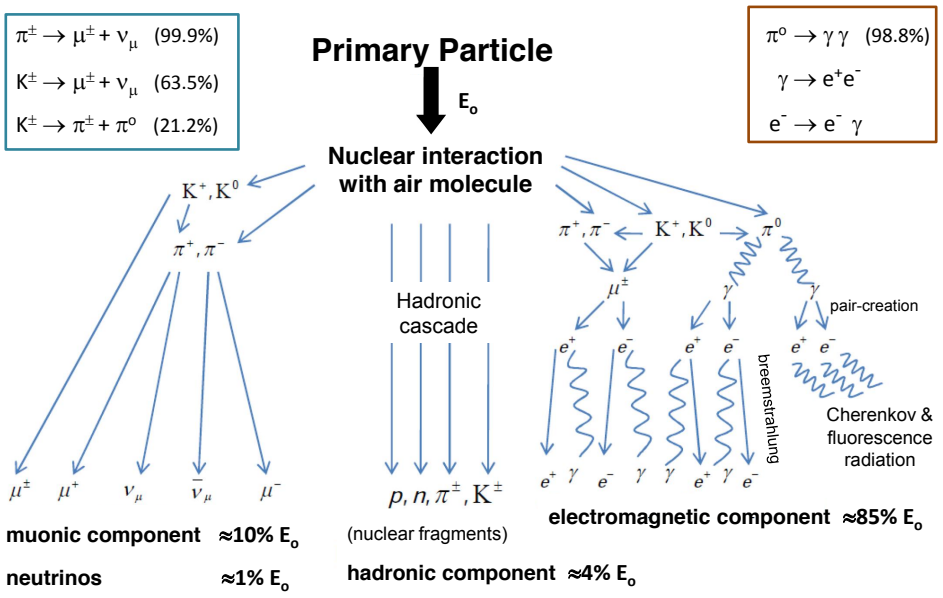


Figura 1.8: Main components of an extensive air shower.

Although the full description of a shower can be fairly complicated, it is dominated by electromagnetic processes [65] and it can be phenomenologically described by a simplified model developed by Heitler [66]. This model was extended by Matthews to include the description of hadronic showers [67].

1.6.1 The Heitler model of electromagnetic showers

In this simplified model at each step all particles interact producing two secondary particles of equal energy. Electrons, positrons and photons interact after traveling an interaction length d = $\lambda_r$ln2, where λ_r is the radiation length of the medium (λ_r = 37 g cm^{-2} in air). After each step electrons divide their energies in half via *bremsstrahlung* emission of a single photon while photons produce a pair e^+e^- of equal energy. After n steps the particle number is N$_n$ = 2^n and their individual energy is E_0/N_n. The process ends when the individual energy drops below the critical value (energy at which the rate of energy loss via *bremsstrahlung* is equal to the one due to ionization), E$_c^\gamma \sim$ 80 MeV in air. At this point the electromagnetic cascade reaches its maximum. Figure 1.9 shows an illustration of the whole process.

Although the model is conceptually very simple, it correctly describes some of the main features of the electromagnetic shower:

- The number of particles at the shower maximum is proportional to the energy of the

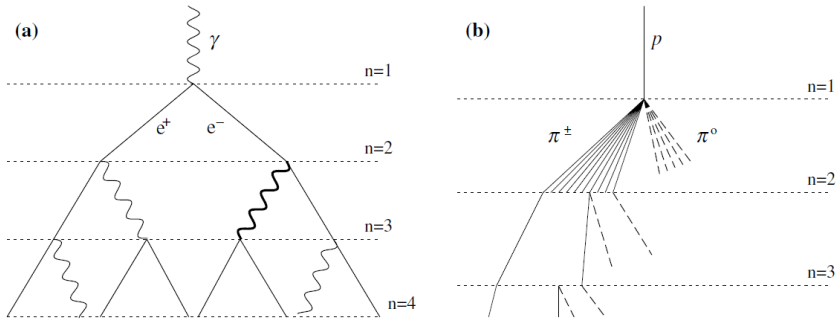


Figura 1.9: Schematic representation of the Heitler model for electromagnetic (a) and hadronic (b) shower development.

primary particle.

$$N_{max} = E_0/E_c^{\gamma} \tag{1.7}$$

- The depth of the maximum of the shower evolves logarithmically with energy. The atmospheric depth of maximum shower development, X_{max}, is determined by the number of interactions needed to reach the critical energy, n_{max}:

$$n_{max} = ln(N_{max})/ln2 = ln(E_0/E_c^{\gamma})/ln2 \tag{1.8}$$

$$X_{max} = X_0 + n_{max}d = X_0 + \lambda_r ln(E_0/E_c\gamma) \tag{1.9}$$

  where X_0 is the depth of the first interaction.

- The rate of evolution of X_{max} with energy is called the **elongation rate** and it is defined as:

$$D_{10} = \frac{dX_{max}}{d\log E_0} = \lambda_r ln10 \sim \text{ 85 g cm}^{-2} \tag{1.10}$$

The results given by this model are broadly confirmed by detailed simulations. However, it presents some discrepancies, since it overestimates the number of particles by a factor 2-3 and the ratio of electrons and positrons to photons by a factor 10-12. This is due to the fact that the absorption of particles below the critical energy is not considered in the model. Besides, more than one photon can be emitted during the *bremsstrahlung* and electrons lose their energy much faster than photons.

1.6.2 Extension of the Heitler model to the hadronic showers

In analogy with the electromagnetic shower, the hadronic component can be described assuming that after each interaction the main products are pions. This was done by Matthews [67] and is thoroughly described in [68]. In this extension, the relevant parameter is the hadronic interaction length λ_I. λ_I is assumed to be constant, a fairly good approximation for interactions in the range 10-1000 GeV. For pions in air, λ_I∼120 g cm^{-2}. After each step of thickness λ_I ln2, 2N$_\pi$ charged pions are produced and N$_\pi$ neutral ones. The

π_0s will decay and fuel the electromagnetic part. At each step roughly one third of the energy is transferred from the hadronic cascade to the electromagnetic one while two thirds continue as hadrons. Therefore the longer it takes for pions to reach the critical energy (20 GeV in air, below which they will decay into muons), the larger will be the electromagnetic component.

Charged pions interactions further carry on the hadronic shower, see figure 1.9. The hadronic cascade continues to grow until charged pions reach an energy where decay is more likely than a new interaction. In this case, the end of the shower is determined by the energy at which pions start to decay into muons.

After n interactions, the total number of pions is $(3N_\pi)^n$, each one of them carrying an energy of $(1/3)^n E_0$. At the critical energy, E_c^π, then:

$$E_c^\pi = \frac{E_0}{(3N_\pi)^{n_c}} \qquad \Longrightarrow \qquad n_c = \frac{\ln(E_0/E_c^\pi)}{\ln(3N_\pi)} \tag{1.11}$$

being n_c the number of steps for the pions to reach E_c^π.

Now, to obtain the number of muons in the shower we assume that all charged pions decay into muons when they reach the critical energy

$$N_\mu = \left(\frac{2}{3}N_\pi\right)^{n_c} \tag{1.12}$$

plugin equation 1.11 into 1.12 we obtain:

$$\ln N_\mu = \beta \ln\left(\frac{E_0}{E_c^\pi}\right) \qquad \text{being} \qquad \beta = \frac{\ln\left(\frac{2}{3}N_\pi\right)}{\ln(3N_\pi)} \tag{1.13}$$

so finally we can write the number of muons in the shower

$$N_\mu = \left(\frac{E_0}{E_c^\pi}\right)^\beta \tag{1.14}$$

This means that the number of muons does not grow linearly with the energy, as does the number of electrons, but it grows as a power-function.

The determination of the position of shower maximum is more complex in the case of the hadronic shower than in the case of a pure electromagnetic one. The larger cross section and the larger multiplicity at each step will reduce the value of X_{max} and will modify the elongation rate. For this reason, to calculate the total elongation rate of the showers a good approximation, made in [67], is to consider the effect of the hadronic cascade only in the first interaction. Therefore, for proton showers

$$D_{10}^p = D_{10}^\gamma + \frac{dX_0}{d\log E_0} \tag{1.15}$$

where D_{10}^γ is the elongation rate for the electromagnetic showers and $X_0 = \lambda_I \ln 2$ the depth of the first interaction. Introducing a realistic parameterization of the dependence of λ_I as a function of energy, such as the one given in [69], the elongation rate is $D_{10}^p \approx 60$ g cm^{-2}. Moreover, since hadronic interaction models predict an approximately logarithmic decrease of λ_I with energy, D_{10}^p is approximately constant too.

An important consequence of equation 1.15 was noted by Linsley [70]: the *Elongation Rate Theorem*. This theorem establishes that regardless of the particular parameterization of λ_I that it is chosen, it will always decrease with increasing energy, and thus, the second term in equation 1.15 will be always negative. Therefore, the elongation rate for electromagnetic showers is always bigger than the one for hadronic showers.

1.6.3 The superposition model

The description explained in the previous section for showers initiated by protons can be extended for showers initiated for different nuclei using a theoretical framework called the *superposition model*. In this model, a primary nucleus of mass A and energy E is described as the superposition of A nucleons of energy $E' = E/A$ (see e.g. [53]).

Showers from heavy nuclei will develop higher, faster and with less shower-to-shower fluctuations than showers initiated by lighter nuclei.

Although the superposition model is a simplification it can describe correctly some of the most important features of showers initiated by different nuclei:

- Nuclei initiated showers will be on average less penetrating than those generated by protons with the same energy

$$X^A_{max}(E_0) = X^p_{max}(E_0/A) = X^p_{max} - \lambda_r A \tag{1.16}$$

- The number of muons is larger for heavier primaries than for light primaries with the same energy

$$N^A_\mu(E_0) = \sum_i^A N^p_\mu(E_0/A) = N^p_\mu(E_0) A^{1-\beta} \tag{1.17}$$

- The elongation rate is the same regardless of the mass of the primary.

$$D^A_{10} = \frac{dX^A_{max}}{d\log E_0} = \frac{d(X^p_{max} - \lambda_r A)}{d\log E_0} = \frac{dX^p_{max}}{d\log E_0} = D^p_{10} \tag{1.18}$$

- The shower to shower fluctuations of X_{max} are smaller for heavy nuclei than for light ones.

Of course in quantitative terms there are differences but all the basic trends regarding the evolution of X_{max} and N_μ with energy and atomic number are reproduced.

The superposition model is a simplification and it cannot fully describe showers initiated by heavy nuclei, since it does not account for nuclear effects such nuclear fragmentation or the re-interaction in the target nucleus. To consider all these processes and others even more complex, more realistic codes are used, such as CORSIKA [71] or AIRES [72], together with hadronic interaction models like EPOS [73], QGSJetII.04 [74] or Sybill [75] (see [76] for a comprehensive review of air shower simulations).

1.7 Mass composition of UHECRs

The acceleration and deflection of a charged cosmic ray in space depends on its atomic number and therefore both are directly related to its mass. If the composition of ultra-high-energy cosmic rays is dominated by a light component we expect small deflections due to the propagation through the magnetic fields. This means that the lightest cosmic rays are observed coming from a direction closer to their source than heavy ones. For this reason it is essential to determine the mass (or the charge) of cosmic rays, to be able to identify the location and the properties of the possible sources.

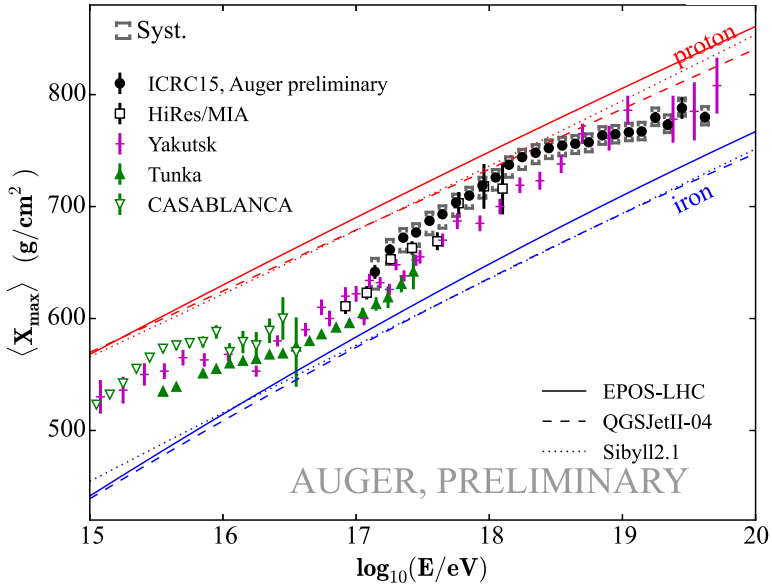


Figura 1.10: Evolution with the energy of the mean of the X_{max} distributions. The experimental results obtained by several experiments are compared to the predictions given by simulations of protons and iron nuclei [77].

UHECRs cannot be measured directly with space-base experiments due to their extremely low fluxes. They have to be studied exclusively through the air showers produced after they collide in the atmosphere. In this situation the only possible way to study their mass composition is by measuring properties or observables of the air showers which depend on the mass. In addition, due to the large fluctuations that can take place during the shower development in the atmosphere, it is extremely difficult to distinguish showers generated by different primaries on an event-by-event basis. Instead, we are forced to follow a statistical approach and to interpret the results through the comparison with the predictions given by simulations.

This comparison with the simulations presents the largest handicap in the study of UHECRs. The discrepancies between the different hadronic models together with the lack of knowledge of key features of hadronic interactions at the highest energies make impossible to establish a firm statement about the primary mass of comic rays.

The main observable used to quantify the mass composition of UHECRs is the atmospheric depth of the maximum of the shower development, X_{max}, since it can be directly measured with fluorescence telescopes. Showers initiated by light nuclei, as protons, develop deeper in the atmosphere and have larger values of X_{max} than showers initiated by heavy primary particles with the same energy (see section 1.6.3). For this reason the study of the evolution of this observable with the energy is one of the most useful techniques to infer the mass composition of cosmic rays. Figure 1.10 shows a compilation of X_{max} measurements done by different experiments together with the predictions given by simulations.

1.7.1 Mass composition measurements in the Pierre Auger Observatory

Since the mass composition is one of most intriguing questions about cosmic rays, it is natural that the Pierre Auger Collaboration had conducted several analyses based on different mass-dependent observables with the purpose of shedding light on this crucial problem.

As we have already mentioned the best-known observable used to quantify the mass composition of UHECRs is X_{max}. It marks the atmospheric depth at which the energy deposited in the atmosphere reaches its maximum. The global X_{max} distribution for a particular energy bin results from the superposition of the distributions $f_i(X_{\text{max}})$ produced by different nuclei of mass A_i:

$$f(X_{\text{max}}) = \sum_i p_i f_i(X_{\text{max}}) \tag{1.19}$$

where the fraction of primary particles of type i is given by p_i. The first two moments of $f(X_{\text{max}})$, i.e., its mean and variance, $\langle X_{\text{max}} \rangle$ and $\sigma(X_{\text{max}})^2$ respectively, allow to study the overall features of primary cosmic ray composition. Measurements of these two moments as a function of energy have been made by the Pierre Auger Collaboration [78], see figure 1.11.

The mean values of the X_{max} distributions as a function of the energy observed in the left panel of figure 1.11 can not be fitted adequately with a straight line, as we expect for a pure composition. In addition, the behavior of $\sigma(X_{\text{max}})$ as a function of energy (right panel of figure 1.11) is not as flat as expected for a pure composition. Then, we can conclude from these plots that the mass composition must be changing as the energy increases. In fact, the conclusion given in [78] is that the cosmic rays flux is composed mainly by light nuclei with a fraction of heavy nuclei increasing with energy.

An extension of this analysis at much lower energies has also been done by the Pierre Auger Collaboration using three telescopes with a high elevation (HEAT) [77]. These are fluorescence telescopes that can operate in an inclination mode, observing showers which develop higher in the atmosphere, and hence with lower energies. Figure 1.12 shows the results obtained with these telescopes at the lowest energies, together with the results given in [78].

If one accepts the hadronic models as the truth, then deductions about the natural logarithm of the atomic mass ($\ln A$) can be made from the values of X_{max}. The trans-

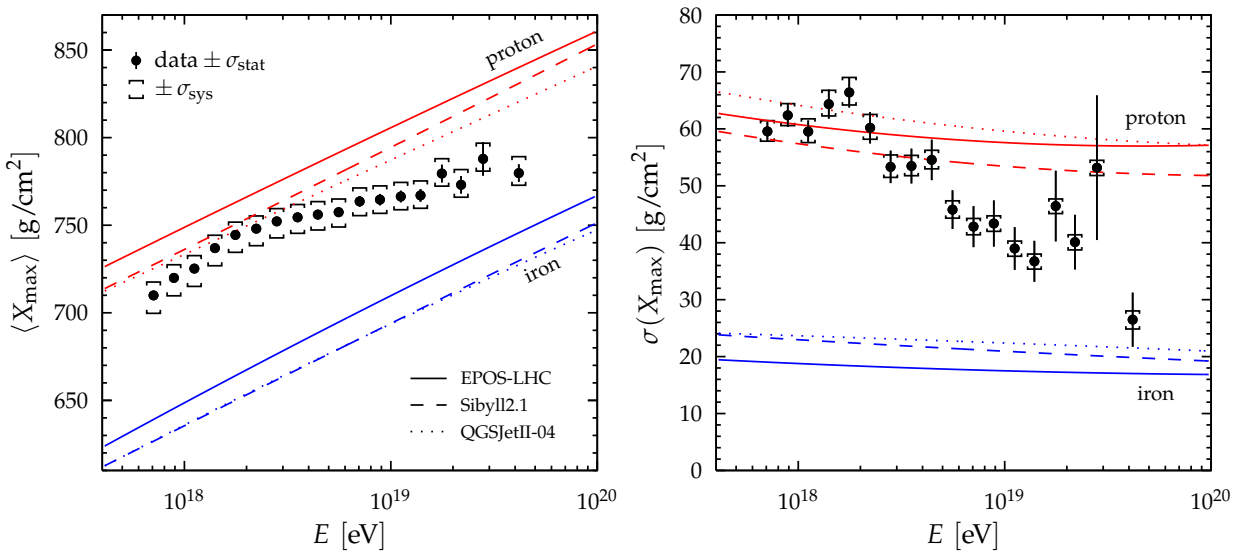


Figura 1.11: Mean (left panel) and standard deviation (right panel) of the X_{max} distributions as a function of energy measured by the Pierre Auger Collaboration [78]. The lines show the prediction given by simulations of protons and iron nuclei for the indicated hadronic models.

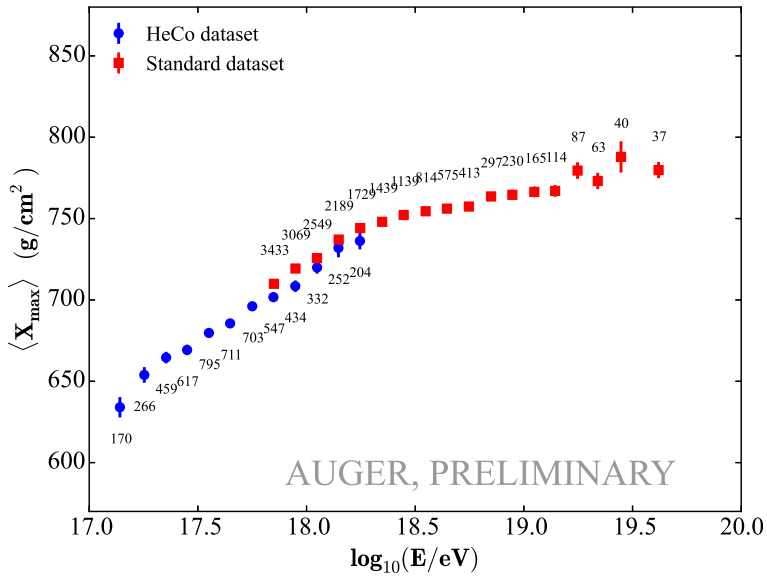


Figura 1.12: Extension of the X_{max} elongation rate obtained by the Pierre Auger Collaboration using the telescopes of high elevation (blue dots) [77]. This result is compared to the one shown in the left panel of figure 1.11 (red squares).

formation of the X_{max} measurements into an average value of $\langle \ln A \rangle$ is done using the following equation:

$$\langle \ln A \rangle = \ln 56 \frac{\langle X_{\mathsf{max}} \rangle_p - \langle X_{\mathsf{max}} \rangle_{data}}{\langle X_{\mathsf{max}} \rangle_p - \langle X_{\mathsf{max}} \rangle_{Fe}} \tag{1.20}$$

This equation is a consequence of the superposition model and it can be generalized for different observables that depend on the mass. It is quite convenient because it provides a straightforward method for the comparison of results coming from different analyses with different selection cuts and systematics uncertainties. Estimates of $\langle \ln A \rangle$, obtained with the X_{max} values shown in figure 1.11 are given in [78] for three different hadronic models: Sybill 2.1 [75], EPOS-LHC [73] and QGSJetII.04 [74] (see figure 1.13).

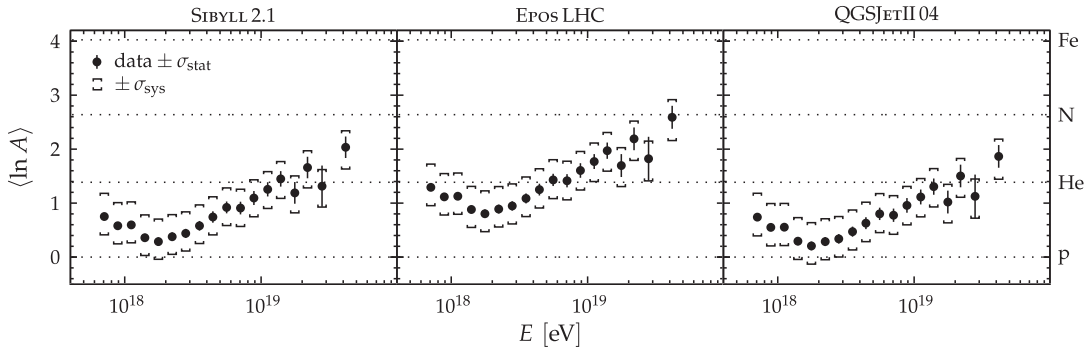


Figura 1.13: $\langle \ln A \rangle$ as a function of the energy as predicted by Sybill 2.1, EPOS-LHC and QGSJetII.04 using the mean values of the X_{max} distributions [78].

Results coming from other analyses can be compared to the results of X_{max} by means of their respective $\langle \ln A \rangle$ values. We discuss here two observables: one is the depth of the maximum of the muonic shower X^{μ}_{max} [80] and the other is (sec θ)$_{max}$ [79]. The last one is related to the distribution of the arrival times of particles at the ground. This time distribution presents an azimuthal asymmetry which strongly depends on the zenith angle, θ. The zenith angle where this asymmetry reaches a maximum, (sec θ)$_{max}$, is an observable sensitive to mass composition [81]. The comparison between the results obtained with these two observables and the results of X_{max} is shown in figure 1.14. The quantity (sec θ)$_{max}$ depends on the radial distance to the shower core, since muons become an increasingly dominant component at the farthest distances. For this reason the results of (sec θ)$_{max}$ are split in two distance ranges in figure 1.14. X_{max} and X^{μ}_{max} do not depend on the radial distance, so the same result is repeated for the two distance ranges.

From figure 1.14 it is evident that the two hadronic models, EPOS-LHC and QGSJetII.04, do not adequately describe the data gathered by the Pierre Auger Observatory: an accurate hadronic model would be expected to give compatible values of $\langle \ln A \rangle$ for different analyses.

The cause of the discrepancies between the predictions given by the different hadronic models has been identified as an underlying problem in the number of the muons

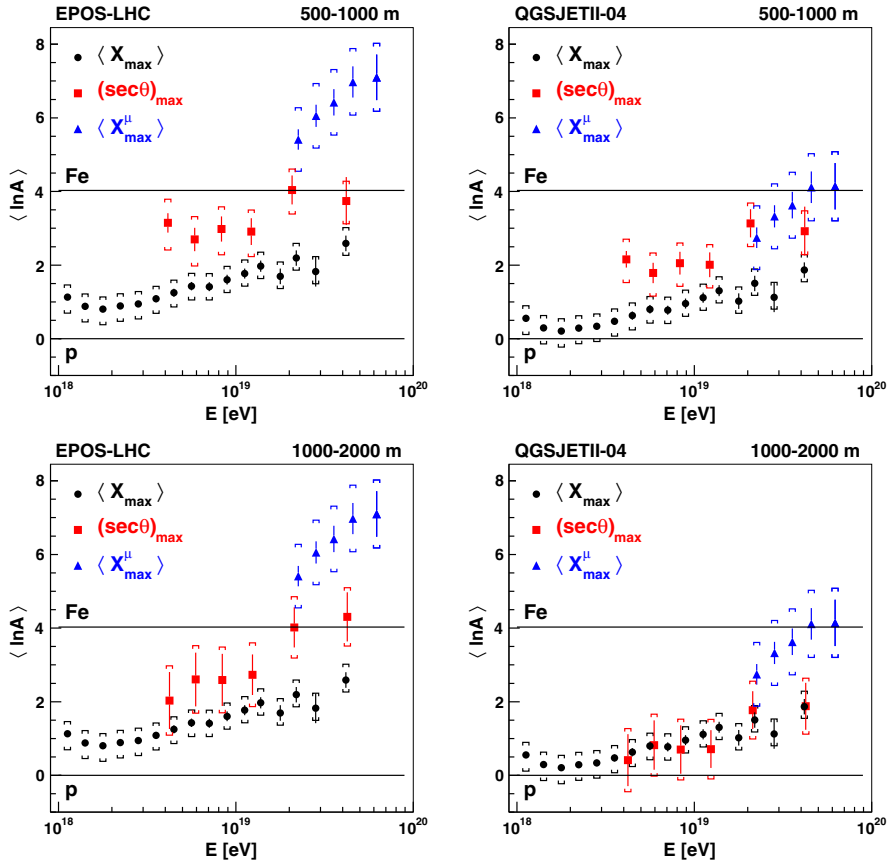


Figura 1.14: $\langle \ln A \rangle$ as a function of the energy as predicted by EPOS-LHC and QGSJetII.04 using the results obtained with (sec $\theta)_{max}$ [79] and with $X^{\mu}_{\max}$ [80]. The results of figure 1.13, obtained with $X_{\max}$ are included for the sake of comparison.

observed in the simulations. The Pierre Auger Observatory has the unique capability of detecting a large number of events with high zenith angles, up to 80°. This gives the opportunity to study showers mainly composed by muons, because for those events the electromagnetic part of the shower is largely suppressed.

In [82] the number of muons in the air showers is measured as a function of the energy for events with zenith angles between 62° and 80°. The muon number for each shower is derived by scaling a simulated reference profile of the lateral muon density distribution at the ground until it fits the data. The observable studied in this analysis is called R_{μ}, and it is the average number of observed muons in a shower divided by the number of muons predicted by proton simulations at 10 EeV with the hadronic interaction model QGSJetII.03 [83]. Figure 1.15 shows the results obtained with this analysis. The discrepancies between the number of muons observed in the data and in the simulations are evident in the two panels of this figure. In both cases the number of muons observed in the data sample is larger than the one observed in simulations. The message that

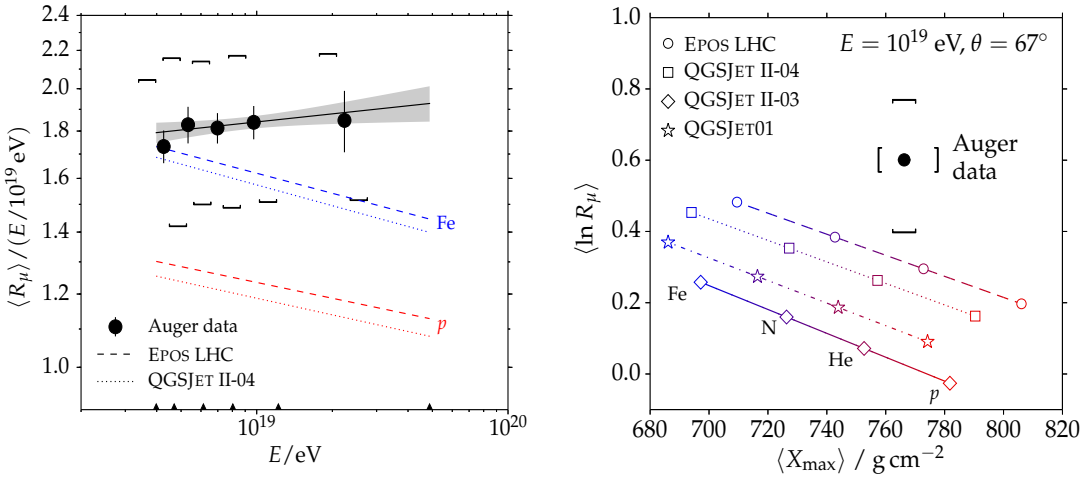


Figura 1.15: Left panel: Evolution of $\langle R_\mu \rangle$ as a function of the energy. Right panel: Evolution of $\langle R_\mu \rangle$ as a function of the $\langle X_{\max} \rangle$. Data gathered at the Pierre Auger observatory are compared to the predictions given by simulations showing the apparent deficit of muons in the simulations [82].

these plots contain is straightforward: there is a muon deficit in the predictions done by the most up-to-date particle-physics models.

An additional analysis which confirms the muon deficit is shown in [84]. In this case the analysis is based on events whose longitudinal development and lateral distribution are simultaneously measured by the Pierre Auger Observatory. The first step of this analysis is to produce simulated showers matching the longitudinal development of each data event until finding the ones having the same reconstructed $X_{\max}$ values as the real events. For those particular showers, the second step is to do a full simulation of the ground signals generated in the surface detectors for comparison with the data.

These signals can be described, as a first approximation, as the sum of two components: the electromagnetic and the hadronic one. Since in simulations these two components are known, the signals can be expressed as the sum of R_E times the electromagnetic signal and R_{had} times the hadronic signal. If the simulations agree with the data the values for R_E and R_{had} should be compatible with one.

The final step of this analysis is to reproduce the ground signals of the real events varying the values of R_E and R_{had} in the simulated events. Figure 1.16 shows the values found for R_E and R_{had} obtained for different hadronic models. The R_E values are compatible with 1, as expected, while the R_{had} values are larger. This means that to reproduce the ground signals of the real events it is necessary more hadronic component than the one provided by the current models. The average hadronic shower is 1.33 $\pm$ 0.16 (1.61 $\pm$ 0.21) times larger than predicted using the leading LHC-tuned models EPOS-LHC (QGSJetII-04), with a corresponding excess of muons. The values of R_{had} needed in the models are comparable to the corresponding muon excess detected in highly inclined air showers [82].

Considering these results the only feasible way to infer the mass composition of UHECRs is through the study of observables whose nature is mainly electromagnetic. To obtain meaningful conclusions about the mass it is necessary to compare the data with models which provide a faithful representation of Nature at the highest energies, and as shown in [84], currently this only seems to be possible for the electromagnetic component of the air showers.

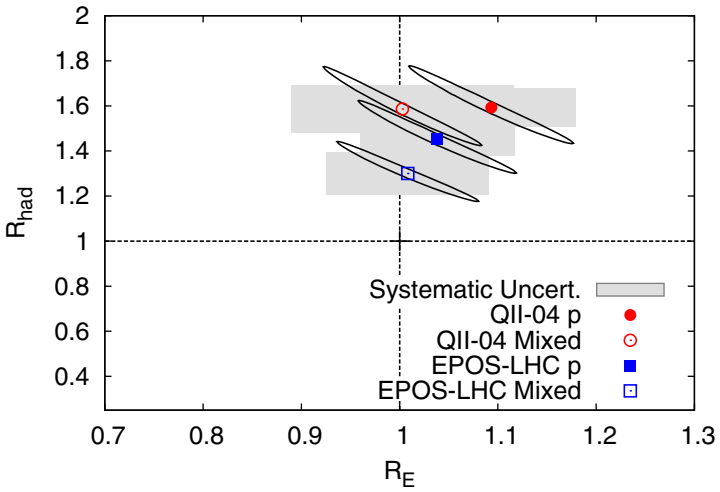


Figura 1.16: Best-fit values of R_E and R_{had} for QGSJetII.04 and EPOS-LHC, for pure proton (solid circle, square) and mixed composition (open circle, square). The ellipses and gray boxes show the 1-σ statistical and systematic uncertainties.

1.8 This thesis in the context of the study of UHECRs

As we have already mentioned, knowing the mass composition of ultra-high-energy cosmic rays is a fundamental question for solving the conundrum of the origin of cosmic rays and of the acceleration mechanisms they are exposed to. For this reason the current experiments for the study of UHECRs are putting an enormous effort to try and give an answer to this problem.

The measurements of X_{max} obtained with fluorescence telescopes are an excellent tool to infer the mass composition of cosmic rays and paramount among them are the results of the Pierre Auger Collaboration due to their accuracy and their good control of the systematic uncertainties (see previous section). However these measurements are limited at the highest energies due to the field of view of the fluorescence detectors and their limited uptime. Fluorescence telescopes only work during the night and, in general, with excellent weather conditions and under very strict requirements of darkness. This limits significantly their duty cycle and as a consequence their statistics. They only provide a precise measurement of the electromagnetic component of the shower.

This problem with the statistics, together with the discrepancies observed in terms of the number of muons between the hadronic models and data, calls for a mass composition analysis independent of simulations and based on data collected by a surface detector array. In this context the Pierre Auger Observatory [85] offers the optimal conditions to develop a method with such characteristics. It operates an array of surface detectors covering a huge area and with a duty cycle close to 100 %, thus increasing the statistics at the highest energies.

These two reasons are the driving principles of the work presented in this thesis. We focus our attention on a particular observable called risetime, which is measured by the surface detectors of the Pierre Auger Observatory. With this observable we develop a data-driven method which allow us not only to infer the mass composition of cosmic rays in a wide range of energies, but also to probe the largest ones with a data sample statistically significant.

2 The Pierre Auger Observatory

2.1 Introduction

The Pierre Auger Project was proposed in 1998 to discover and understand the source or sources of cosmic rays with energies exceeding 10^{19} eV. To accomplish its goals, the Collaboration designed an experiment to achieve high-quality data in a high-statistics study of the most extreme cosmic rays. The Pierre Auger Observatory in the Province of Mendoza, Argentina, has been taking data since 2004, adding detectors as they became active until its completion in 2008. The properties of air showers measured by the Observatory are used to determine the energy, direction and composition of cosmic rays.

The Observatory design features an array of 1600 water Cherenkov surface detectors spread over 3000 km^2 and arranged on a triangular grid, with the sides of the triangles being 1.5 km. This is the Auger Surface Detector (SD). In addition, four fluorescence detector stations, each containing six fixed telescopes designed to detect air-fluorescence light, overlook the surface array. They form the Auger Fluorescence Detector (FD). The layout of the Pierre Auger Observatory is shown in figure 2.1. The surface detectors measure the density distribution of the air shower cascade as it strikes the ground while the fluorescence telescopes measure the light produced by atmospheric nitrogen excited by the cascading shower. This dual approach is called the *hybrid technique*.

The powerful feature of the Auger hybrid design is the capability of observing air showers simultaneously by these two different but complementary techniques. On dark moonless nights, the air fluorescence telescopes record the development of what is essentially the electromagnetic shower that results from the interaction of the primary particle with the upper atmosphere. On the other hand, the surface array measures particle signals as the shower strikes the earth just beyond its maximum development. By recording the light produced by the developing air shower, fluorescence telescopes can make a nearly calorimetric measurement of the energy. This energy calibration can then be transferred to the surface array with its 100 % duty factor and large event-gathering po-

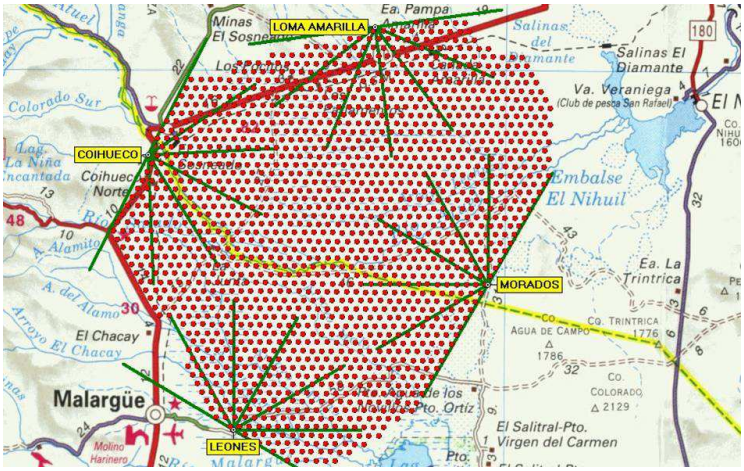


Figura 2.1: The Auger Observatory layout. Each dot corresponds to one of the 1660 surface detector stations. The four fluorescence detector enclosures are shown, each with the field of view of its six telescopes.

wer. Due to this fact, the energy conversion and subsequent determination of the energy spectrum can be done with minimal reliance on numerical simulations or on assumptions about the composition or interaction models.

A water Cherenkov particle detector was chosen for use in the surface array because of robustness, low cost, and sensitivity to showers at high zenith angles. A surface detector station consists of a 12,000 liter polyethylene water tank containing a sealed laminated polyethylene liner with a reflective inner surface. Cherenkov light from the passage of charged particles is collected by three 230 mm photomultiplier tubes (PMTs) that look through windows of clear polyethylene into highly purified water. The surface detector station is self contained. A solar power system provides power for the PMTs and electronics package. The electronics package, consisting of a processor, GPS receiver, radio transceiver and power controller, is mounted on the tank. The surface detectors operate continuously and independently of the FD.

The FD, with a duty cycle of $\sim$15 %, always operates in conjunction with the surface detectors, so that it sometimes is referred as the hybrid detector. Its primary purpose is to measure the longitudinal profile of showers registered by the SD whenever it is dark and clear enough to make reliable measurements of atmospheric fluorescence from air showers. The integral of the longitudinal profile is used to determine the shower energy, and the speed of the shower development is indicative of the mass of the primary particle. The hybrid operation of the FD has better angular resolution than the surface array working alone.

The site of the Observatory is in the Province of Mendoza, at the "pampa amarilla", near the city of Malargüe. It is located at about latitude 35° south with a mean altitude of 1400 m a.s.l. (depth = 870 g cm^{-2}). The site is a relatively flat alluvial plain, sufficiently large to easily encompass the required 3000 km^2 footprint of the array. There are convenient elevated positions on the edge of the array that allow placement of the four fluorescence telescope enclosures slightly above ground level. A campus area in Malargüe includes

Figura 2.2: A fluorescence telescope enclosure, seen on the hill top, and a surface detector station, below.

an office building with a visitor center, a detector assembly area, and a staging area for detectors.

After a period of research and development, the Engineering Array, consisting of 32 prototype surface array detectors and two prototype fluorescence telescopes, was built to validate the design [86]. At the end of 2001, before the end of the scheduled two years, the Engineering Array was able to record and reconstruct air shower events simultaneously with both the surface array and the fluorescence detectors. The Engineering Array was able to demonstrate the validity of the design and the performance of all of the detectors, communications equipment and data systems as well as the deployment methodology. Installation of production detectors was started in 2002. The Observatory started collecting data in January 2004. It has been steadily running in its full configuration since 2008.

Along the years instrumental enhancements have been installed close to one of the fluorescence telescopes enclosure (Coihueco station). These include underground muon detectors, 61 additional water Cherenkov detectors with a smaller spacing (750 m), and three high-elevation fluorescence telescopes for a larger field-of-view (HEAT). Also, research programs are underway to assess the utility of radio and microwave emission from air showers. Details about these enhancements can be found in [85]. The Pierre Auger Observatory has begun a major upgrade of its already impressive capabilities, with an emphasis on improved mass composition determination using the surface detectors. Known as AugerPrime, the upgrade will include new 4 m^2 plastic scintillator detectors on top of all 1660 water Cherenkov detectors, updated and more flexible surface detector electronics, a large array of buried muon detectors, and an extended duty cycle for operations of the fluorescence detectors [87].

In what follows we describe in detail the two main apparatuses that provide the data used in the analyses described in this thesis: the Surface and the Fluorescence Detectors.

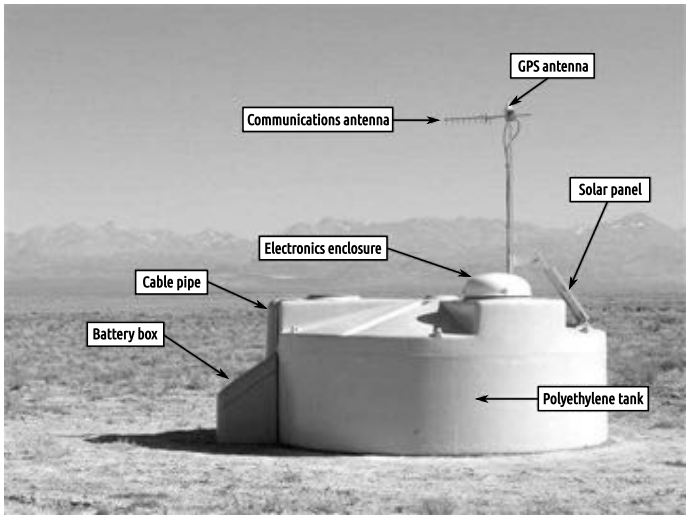


Figura 2.3: A pictorial view of a surface detector station in the field, showing its main components.

2.2 The Surface Detector

Each surface detector station of the Observatory consists of a 3.6 m diameter water tank containing a sealed liner with a reflective inner surface. The liner contains 12,000 l of ultra-high purity water. Three 230 mm diameter photomultiplier tubes are symmetrically distributed at a distance of 1.20 m from the center of the tank and look downwards through windows of clear polyethylene into the water to collect the Cherenkov light produced by the passage of relativistic charged particles through the water. The water height of 1.2 m makes it also sensitive to high energy photons, which convert to electron-positron pairs in the water volume.

The surface detector station is self-contained. A solar power system provides an average of 10 W for the PMTs and the electronics package consisting of a processor, GPS receiver, radio transceiver and power controller. The components of the surface detector station are shown in figure 2.3. The hardware of the surface detector is described extensively in [86, 85].

The tanks are made of high-density polyethylene by the rotomolding process. The exterior is colored beige to minimize the visual impact. The resins are compounded with additives to enhance ultraviolet protection. The interior has added carbon-black to guarantee light-tightness. The tanks have a nominal wall thickness of 1.3 cm and a weight of 530 kg.

Three hatches, located above the PMTs, provide access to the interior of the tank for assembly, water filling and eventual servicing of the interior parts. The hatches are covered with light- and water-tight polyethylene hatch covers. Hatch cover 1 is larger and accommodates the electronics on its top. The electronics is protected by an aluminum dome. The tanks also possess lugs for lifting and four additional lugs to support the solar panel and antenna mast assembly.

Electrical power for the electronics is provided by two 55 Wp solar panels which feed

two 12 V, 105 Ah lead-acid low maintenance batteries. Batteries are charged through a commercial charge controller. The electronics assembly possesses a Tank Power Control Board (TPCB) which also monitors the charging and discharging of batteries and sets the system to hibernation mode if the charge of the batteries falls below a critical level. The batteries are accommodated in a thermally insulated battery box which is installed next to the tank at the shaded southern side.

The solar panels are mounted on aluminum brackets, which also support a mast of a height of 2.15 m. The communications and GPS antennas are mounted at the top of this mast.

The tank liners are right circular cylinders made of a flexible plastic material conforming approximately to the inside surface of the tanks. They enclose the water volume, provide a light-tight environment and diffusively reflect the Cherenkov light produced in the water volume. The liners are produced from a laminate composed of an opaque three-layer co-extruded low-density polyethylene (LDPE) film bonded to a layer of DuPont Tyvek® 1025-BL by a layer of Titanium-dioxide pigmented LDPE. The three-layer co-extruded film consists of a carbon black loaded LDPE formulated to be opaque to single photons, sandwiched between layers of clear LDPE to prevent any carbon black from migrating into the water volume.

The liner has 3 windows through which the PMTs look into the water volume from above. These windows are made of UV-transparent linear low-density polyethylene. The PMTs are optically coupled to the windows with optical silicone and protected with a light-tight plastic cover.

Once deployed in their correct position in the field, the tanks are filled with high purity water produced at a water plant owned by the Auger Project. Water quality exceeds 15 MΩ cm at the output of the water plant and is transported in clean ad-hoc transport tanks. The water is expected to maintain its quality without degradation for the lifetime of the Observatory.

2.2.1 SD electronics

As we have said, to collect the Cherenkov light produced in the water volume of the surface detectors by the air showers, three PMTs look at the water volume from the top. The PMTs (Photonis XP1805/D1) have a 230 mm diameter photocathode and eight dynodes, with the chemical composition of the dynode surfaces optimized by the manufacturer to maximize linearity. Due to their proximity to water they are operated with a positive anode voltage, the photocathode being grounded. The high voltage is provided locally from a module integrated in the PMT base, and is proportional to a DC control voltage provided by the slow control system. The PMTs are operated at a nominal gain of 2×10^5, and are specified for operation at gains up to 10^6. The PMTs are required to be linear within 5 % up to 50 mA anode current. The base, including the high voltage supply, is attached to the tube by soldering to flying leads and is potted in GE silicone RTV-6136 to protect it from the high humidity present in the tank.

Each PMT has two outputs. An AC coupled anode signal is provided. In addition, the signal at the last dynode is amplified and inverted by the PMT base electronics to provide a signal with 32 times the charge gain of the anode. No shaping of the signal is performed

on the PMT base.

Six identical channels of electronics are provided to digitize the anode and amplified dynode signals from each of the PMTs. Each channel consists of a 5-pole Bessel filter with a -3 dB cutoff at 20 MHz and a voltage gain of $-0{,}5$. This filter is implemented using a pair of Analog Devices AD8012 current feedback op-amps. The filtered analog signals are fed to Analog Devices AD9203 10 bit 40 MHz semi-flash ADCs. The ADC negative inputs are biased to -50 mV to bring the input pedestal on scale and allow for amplifier section offsets. The choice of filter cutoff results in 5 % aliasing noise while preserving the time structure of the signals. The use of two 10 bit ADCs with a gain difference of 32 extends the dynamic range of the system to 15 bits with a 3 % precision at the end of the overlap region.

An LED flasher is mounted in a test port of the water tank liner. The LED flasher incorporates two LEDs which can be pulsed independently or simultaneously and with variable amplitude. This allows testing of the linearity of the photomultipliers to be conducted remotely.

Each SD station contains a GPS receiver with its corresponding antenna mounted at the top of the communications mast for event timing and communications synchronization. The receiver is a Motorola (OEM) Oncore UT+. This receiver outputs a timed one-pulse-per-second (1 PPS). The GPS 1 PPS signal is offset from the true GPS second by up to 50 ns, and a correction for this offset is provided periodically by the receiver. Event timing is determined using a custom ASIC which references the timing of shower triggers to the GPS 1 PPS clock. The ASIC implements a 27 bit clock operating at 100 MHz. This clock is latched on the GPS 1 PPS signal at the time of each shower trigger. A counter operating at the 40 MHz ADC clock is also latched on the GPS 1 PPS clock. These data, together with the timing corrections provided by the GPS receiver, are used to calibrate the frequencies of the 40 MHz and 100 MHz clocks and to synchronize the ADC data to GPS time within 10 ns RMS.

The digital data from the ADCs are clocked into a programmable logic device (PLD). In the first half of the deployment, we employed two ALTERA ACEX PLDs (model EP1-K100QI208-2) with 16k $\times$ 36 bits additional external static RAM. In later stations, an Altera Cyclone FPGA replaced the two ACEX devices and external memory. The PLD implements firmware that monitors the ADC outputs for interesting trigger patterns, stores the data in a buffer memory, and informs the station micro-controller when a trigger occurs. There are two local trigger levels (T1 and T2) and a global third level trigger, T3. Details of the local triggers are described in section 2.2.3.

The front end is interfaced to a unified board which implements the station controller, event timing, and slow control functions, together with a serial interface to the communications system. The slow control system consists of DACs and ADCs used to measure temperatures, voltages, and currents relevant to assessment of the operation of the station.

The station controller consists of an IBM PowerPC 403 GCX-80 MHz, with a 32 MB DRAM bank to store data and executable code, and a 2 MB Flash EPROM for the bootstrap and storing of the OS9 operating system. The data acquisition system implemented on the station controller transmits the time stamps of the ~ 20 T2 events collected each second to CDAS (Central Data Acquisition System). CDAS returns T3 requests to

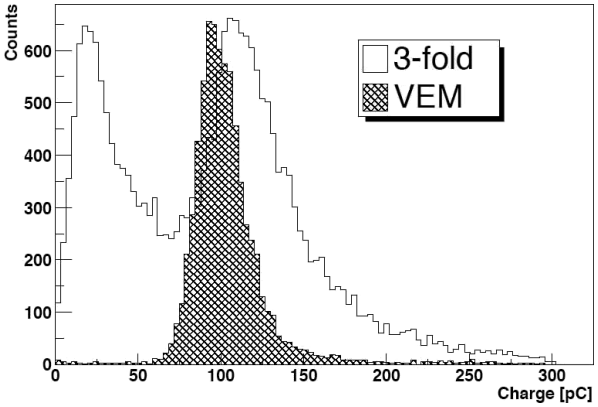


Figura 2.4: Charge spectrum obtained when a surface detector is triggered by a 3-fold coincidence among its photomultipliers (open histogram). The hatched histogram is produced by vertical and central muons. The bin containing the peak of the hatched histogram is defined as a vertical equivalent muon.

the station within ~ 8 seconds of the event (including communications delays due to retransmission). The station controller then selects the T1 and T2 data corresponding to the T3 requests and builds it into an event for transmission to CDAS. Calibration data are included in each transmitted event.

2.2.2 SD calibration

The Cherenkov light recorded by a surface detector is measured in units of the signal produced by a muon traversing the tank on a vertical trajectory (see figure 2.4). This unit is termed the vertical equivalent muon (VEM). The goal of the surface detector calibration is to measure the value of 1 VEM in hardware units (i.e., in integrated Fast Analog to Digital Converter (FADC) channels).

The detector calibration is inferred from background muons. The typical rise time for a muon signal is about 15 ns with a decay time of the order of 60 to 70 ns. The average number of photoelectrons per muon collected by one PMT is 95. By adjusting the trigger rates, the gains of the three PMTs are matched within 6 %. The measurement of the muon charge spectrum allows us to deduce the charge value for the signal produced by a single, central, vertical muon, Q_{VEM}, from which the calibration is inferred for the whole dynamic range. The cross calibration between the anode and dynode output channels is performed by using small shower signals in the overlap region [88].

The decay constant of the muon signal is related to the absorption length of the light produced. This depends of various parameters such as the Tyvek® reflectivity and the purity of the water. The signal decay constant correlates with the so called area-to-peak (A/P) ratio of the signal:

$$\mathrm{A/P} = \frac{Q_{\text{VEM}}}{I_{\text{VEM}}} \tag{2.1}$$

where I_{VEM} is the maximum current of the muon signal. This area-to-peak ratio is a routine monitoring quantity that is directly available from the local station software.

2.2.3 SD local triggers

The front-end electronics implements three types of trigger functions. Shower triggers result in the recording of 768 samples (19.2 μs) of the six ADCs. Muon triggers result in the recording of 24 samples of the three high gain dynode channels for use in calibration. Double buffered scalers are also implemented for use in monitoring rates and for auxiliary physics purposes.

A shower trigger is generated when one of several conditions is satisfied. The first trigger level, called T1, has two independent modes. The first ones is a single bin threshold trigger (TH) requiring the coincidence of all three PMTs being above $1{,}75\, I_{\text{VEM}}$. The rate of this trigger is about 100 Hz, and is used to select large signals that are not necessarily spread in time. It is particularly effective for the detection of very inclined showers. The second T1 mode is a time-over-threshold (ToT) trigger. This trigger requires that single bin threshold trigger be satisfied for at least a minimum number of samples within a sliding time window. A ToT trigger is generated when at least 13 bins within a 3 μs window (120 samples) exceed a threshold of $0{,}2\, I_{\text{VEM}}$ on at least two out of the three PMTs. The ToT trigger is efficient for signals near the core of low-energy showers or signals far from the core of high-energy showers. The rate of the ToT trigger is a few Hz, and depends on the shape of the muon pulse in the tank. A software selection of this trigger with a higher threshold at $3{,}2\, I_{\text{VEM}}$ is also performed.

The second trigger level, called T2, is applied to decrease the global rate of the T1 trigger down to about 23 Hz. While all T1-ToT triggers are promoted to T2-ToT, only T1-TH triggers passing a single threshold of 3 $3{,}2\, I_{\text{VEM}}$ in coincidence for the three PMTs will pass this second level and become T2-TH.

The station controller transmits timestamps for all T2s to the CDAS for global (T3) trigger determination.

Two additional sophisticated triggers were introduced in 2013 to somewhat lower the energy threshold of the array, and improve sensitivity to photon and neutrino initiated showers. See [85] for details.

2.3 The Fluorescence Detector

The 24 telescopes of the FD overlook the SD array from four sites – Los Leones, Los Morados, Loma Amarilla and Coihueco [89]. Six independent telescopes are located at each FD site in a clean climate-controlled building [86], an example of which is seen in figure 2.5. A single telescope has a field of view of $30^\circ \times 30^\circ$ in azimuth and elevation, with a minimum elevation of $1{,}5^\circ$ above the horizon. The telescopes face towards the interior of the array so that the combination of the six telescopes provides 180° coverage in azimuth.

2.3.1 FD telescopes

The details of the fluorescence detector telescope are shown in figure 2.6. The telescope design is based on Schmidt optics because it reduces the coma aberration of large optical systems. Nitrogen fluorescence light, emitted isotropically by an air shower, enters

Figura 2.5: FD building at Los Leones during the day. Behind the building is a communication tower. This photo was taken during daytime when shutters were opened because of maintenance.

through a circular diaphragm of $1{,}1$ m radius covered with a Schott MUG-6 filter glass window. The filter transmission is above 50 % between 310 and 390 nm in the UV range. The filter reduces the background light flux and thus improves the signal-to-noise ratio of the measured air shower signal. It also serves as a window over the aperture and thus keeps the space containing the telescopes and electronics clean and climate controlled. The shutters seen in figure 2.6 are closed during daylight and also close automatically at night when the wind becomes too high or rain is detected. In addition, a fail safe curtain is mounted behind the diaphragm to prevent daylight from illuminating a camera in case of a malfunction of the shutter or a failure of the Slow Control System.

A simplified annular lens, which corrects spherical aberration and eliminates coma aberration, is mounted in the outer part of the aperture. The segmented corrector ring has inner and outer radii of 850 and 1100 mm, respectively. Six corrector rings were made from Schott BK7 glass and Borofloat was used for the rest. More details about the corrector ring can be found in [86, 90].

The light is focused by a spherical mirror of ~ 3400 mm radius of curvature onto a spherical focal surface with radius of curvature ~ 1700 mm. Due to its large area ($\sim 13\,\mathrm{m}^2$), the primary mirror is segmented to reduce the cost and weight of the optical system. Two alternative segmentation configurations are used: one is a tessellation of 36 rectangular anodized aluminum mirrors of three different sizes; the other is a structure of 60 hexagonal glass mirrors (of four shapes and sizes) with vacuum deposited reflective coatings [86]. The average reflectivity of cleaned mirror segments at a wavelength $\lambda = 370$ nm is more than 90 %. Measurements have shown that dust layer deposits could reduce the mirror reflectivity by about 5 % in the bottom part of the spherical mirror, where the segments are turned slightly upward (see, e.g., figure 2.6). Therefore, careful mirror cleaning is performed as needed.

The camera body is machined from a single aluminum block of 60 mm thickness, with an outer radius of curvature of 1701 mm and an inner curvature radius of 1641 mm. The hexagonal photomultiplier tubes, model XP3062 manufactured by Photonis, are positio-

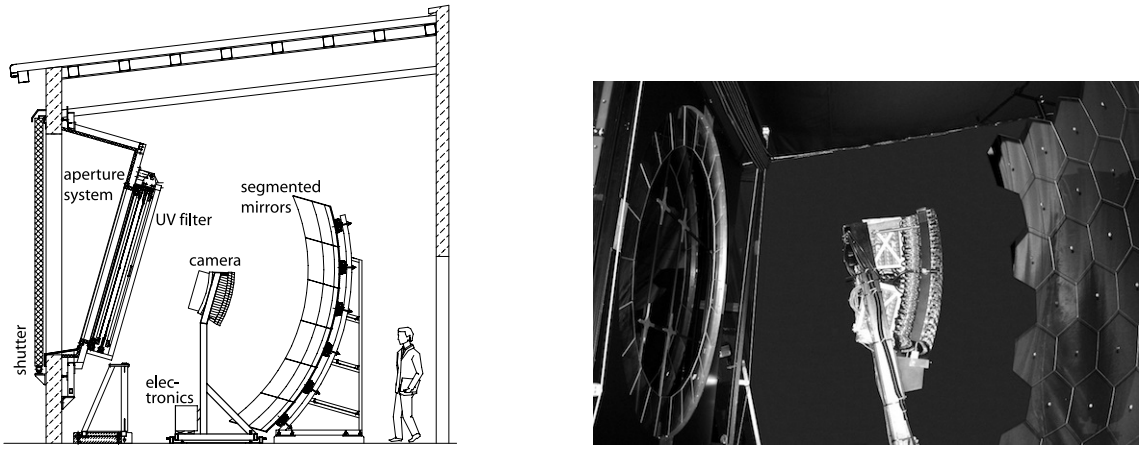


Figura 2.6: Left panel: Schematic view of a fluorescence telescope with a description of its main components. Right panel: Photograph of a fluorescence telescope at Coihueco.

ned inside 40 mm diameter holes drilled through the camera block at the locations of the pixel centers. The pixels are arranged in a matrix of 22 rows by 20 columns.

The PMT boundaries are approximate hexagons with a side to side distance of $45{,}6$ mm. The PMTs are separated by simplified Winston cones secured to the camera body which collect the light to the active cathode of the photomultiplier tube. The light collectors serve to prevent photons from landing in the dead spaces between the PMT cathodes. The upper edge of the light collectors lie on the focal surface of 1743 mm radius. The pixel field of view defined by the upper edges corresponds to an angular size of $1{,}5°$.

All support structures and cables are distributed so as to minimize any obscuration in the light path. The contribution of reflection and scattering inside the optical system of the telescope has been measured in situ and with an airborne remotely controlled platform carrying an isotropic and stabilized UV light source [91]. The measured point spread function of the light distribution in pixels has been implemented in the software used in the air shower reconstruction.

Cleaning and maintenance work has been required during years of detector operation. The cleaning of the UV filter from outside has been performed several times because of deposited dust layers. Less frequently, the inner side of the filter and the corrector ring were washed. Dry and wet methods of mirror cleaning have been adopted over the years and they both improve the reflectivity of mirrors by ≤ 1 % (in the case of mirror segments in the upper rows) up to about 5 % for mirror segments in the bottom rows.

Alignment of individual mirror segments was cross-checked with a laser on site. Moreover, additional methods using data measured by telescopes were used, such as star tracking, Central Laser Facility (CLF) and eXtreme Laser Facility (XLF) shots (section 2.3.4), or a comparison of FD and SD geometry reconstruction. Only in two cases were a realignment of a telescope and a readjustment of camera position needed.

2.3.2 FD electronics

The FD electronics must provide a large dynamic range and strong background rejection, while accepting any physically plausible air shower. Moreover, the electronics is responsible for anti-alias filtering, digitizing, and storing signals from the PMTs.

The XP3062 photomultiplier tube is an 8-stage unit with a bi-alkaline photocathode with quantum efficiency of about $25\,\%$ in the wavelength range 350 to $400\,\mathrm{nm}$. The PMT high voltage is provided by a HV divider chain which forms a single physical unit together with the signal driver circuitry. This head electronics unit is soldered to the flying leads of the PMT [92].

The nominal gain for standard operation of the FD is set to 5×10^4. Stabilization of the HV potential for large pulses, and in the presence of the low but not negligible light intensity of the dark sky background, is realized by employing an active network that uses bipolar transistors in the last three stages of the PMT. The active divider ensures that the gain shift due to the divider chain is less than $1\,\%$ for anode currents up to about $10\,\mathrm{mA}$. The normal dark sky background on moonless nights induces an anode current of about $0{,}8\,\mu\mathrm{A}$ on each PMT.

The head electronics for each PMT is connected to a distribution board located just behind the camera body. Each board serves 44 PMTs, providing high and low voltage and receiving the output signals. The signal is then shaped and digitized in the front-end electronics (FE) unit, where threshold and geometry triggers are also generated. Analog boards in the FE unit are designed to handle the large dynamic range required for air fluorescence measurements; this means a range of $15\,\mathrm{bits}$ and $100\,\mathrm{ns}$ timing.

As the PMT data are processed, they are passed through a flexible three-stage trigger system implemented in firmware and software. The trigger rate of each pixel in a camera (first level trigger) is kept around $100\,\mathrm{Hz}$ by adjusting the pixel threshold level. The algorithm of the second level trigger searches for track segments at least five pixels in length within a camera. The typical trigger rate per camera fluctuates between $0{,}1$ and $10\,\mathrm{Hz}$. The third level trigger is a software algorithm designed to clean the air shower data stream of noise events that survive the low-level hardware triggers. It is optimized for the fast rejection of triggers caused by lightning, triggers caused by cosmic ray muon impacts on the camera and randomly triggered pixels.

The events surviving all trigger levels are sent to the data acquisition computer, which builds an event from the coincident data in all telescopes and generates a hybrid trigger (T3) for the surface array. The event rate is about $0{,}012\,\mathrm{Hz}$ per building for the 24 baseline telescopes.

2.3.3 FD calibration

The reconstruction of air shower profiles and the ability to determine the total energy of a reconstructed shower depend on the conversion of ADC counts to light flux at the telescope aperture for each channel that receives a portion of the signal from a shower. To obtain this important relation, it is necessary to evaluate the response of each pixel to a given flux of incident photons from the solid angle covered by that pixel, including the effects of aperture projection, optical filter transmittance, reflection at optical surfaces, mi-

rror reflectivity, pixel light collection efficiency and area, cathode quantum efficiency, PMT gain, pre-amp and amplifier gains, and digital conversion. This response is measured in a single end-to-end calibration.

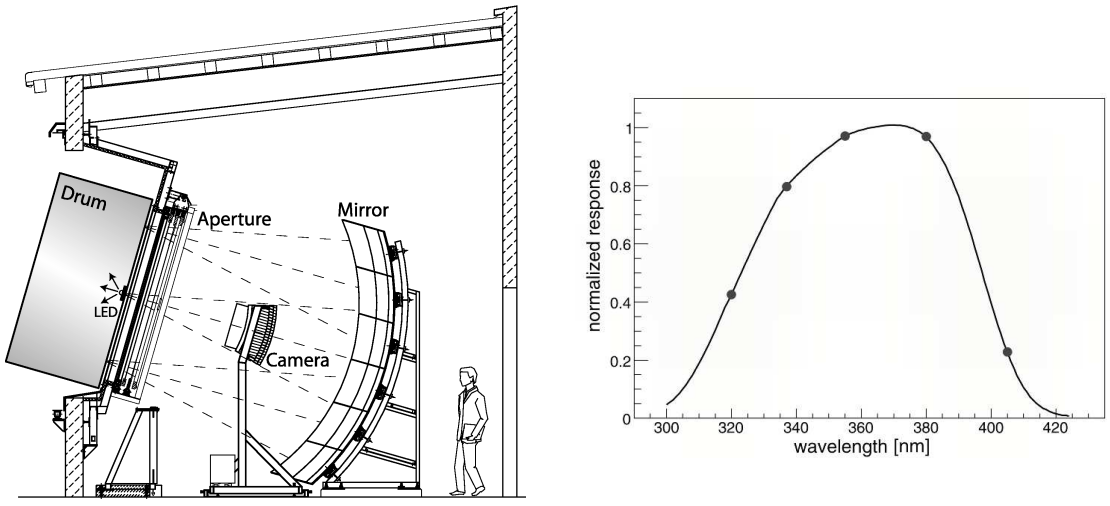
The absolute calibration of the fluorescence detectors uses a portable drum shaped calibrated light source at the telescope aperture, providing uniform illumination to each pixel. The technique [93] is based on a 2.5 m diameter, 1.4 m deep, drum-shaped light source which mounts on the exterior of the FD apertures (see figure 2.7a). The source provides a pulsed photon flux of known intensity and uniformity across the aperture, and simultaneously triggers all the pixels in the camera. In the lab, light source uniformity is studied using CCD images and the intensity is measured relative to NIST calibrated photodiodes. Use of the drum for gain adjustment and calibration provides a known, uniform response for each pixel in a detector.

For calibration at wavelengths spanning the FD acceptance, a xenon flasher is mounted at the back of the drum, with a filter wheel containing 5 notch filters for selection of wavelengths. The xenon flasher [94] provides 0.4 mJ optical output per pulse covering a broad UV spectrum, in a time period of a few hundred nanoseconds. Relative drum intensity measurements at wavelengths of 320, 337, 355, 380 and 405 nm have been made with the same reference PMT used in the absolute measurements. The signals detected at the various wavelengths combine with the lab work to form a curve of relative camera response shown in figure 2.7b. A new detailed measurement procedure was developed that utilized a monochromator and UV light source to measure the FD efficiency in 5 nm steps and found efficiencies consistent with the curve in figure 2.7b [95].

Three additional calibration tools are used at Auger. First, before and after each night of data taking a relative calibration of the PMTs is performed [89]. This relative calibration is used to track both short and long term changes in detector response. Secondly, the relative FD response has been measured at wavelengths of 320, 337, 355, 380 and 405 nm, defining a spectral response curve that has been normalized to the absolute calibration. Thirdly, an independent check of the calibration in some phototubes is performed using vertical shots from a portable laser in the field.

2.3.4 Atmospheric monitoring

The exploitation of the calorimetric measurement of the fluorescence signal in the atmosphere depends essentially on the efficiency of fluorescence light production and subsequent transmission to an FD telescope. In particular, the aerosol content of the atmosphere, in the form of clouds, dust, smoke and other pollutants, needs to be well characterized. The aerosol content of the atmosphere can be variable on short time-scales necessitating the routine monitoring of light transmission conditions in the atmospheric volume above the Pierre Auger Observatory. To account for possible horizontal non-uniformities in the aerosols the area enclosed by the observatory is divided into 5 sub-regions within which only the vertical characteristics of the aerosols are described. Within each region the aerosols are characterized in vertical slices of 200 m thickness, up to a height of 10 km. The aerosol parameters that are important for air shower reconstruction are the VAOD(h), the vertical aerosol optical depth as a function of height, $\alpha(h)$, the aerosol scattering coefficient as a function of height and $\mathrm{d}\sigma/\mathrm{d}\Omega$, the aerosol differential cross section.



(a) Schematic view.

(b) multi-wavelength measurements, normalized at 375 nm.

Figura 2.7: Detector calibration with the 'drum'.

The wavelength dependence of these parameters in the 300 to 400 nm sensitivity range of the FDs is also measured. Aerosol parameters are updated hourly during the periods of FD operation.

These measurements are accomplished using a complex set of instruments including backscatter LIDARs, two laser facilities (CLF and XLF) near the middle of the array, horizontal attenuation monitors, Aerosol Phase Function monitors, star monitors and cloud cameras. The location of these components is shown in figure 2.8 and are described in more detail in [85].

2.4 Data processing and $\overline{\mathrm{Off}\underline{\mathrm{line}}}$ Software

The Pierre Auger Observatory $\overline{\mathrm{Off}\underline{\mathrm{line}}}$ software provides an infrastructure to support development of hybrid event simulation and reconstruction. The software has been designed to accommodate contributions from a large number of physicists developing C++ applications over a long experimental run. The essential features include a "plug-in" mechanism for physics algorithms together with machinery which assists users in retrieving event and detector conditions data from various data sources. A detailed description of the $\overline{\mathrm{Off}\underline{\mathrm{line}}}$ software design, including some example applications, is available in [96]; additional information is also given in [85].

The overall organization of the $\overline{\mathrm{Off}\underline{\mathrm{line}}}$ framework is depicted in figure 2.9. A collection of processing *modules* are be assembled and sequenced through instructions contained in an XML file [97]. An *event* data model allows modules to relay data to one another, accumulates all simulation and reconstruction information, and converts between various formats used to store data on file. Finally, a *detector description* provides a gateway to detector conditions data, including calibration constants and atmospheric properties as a

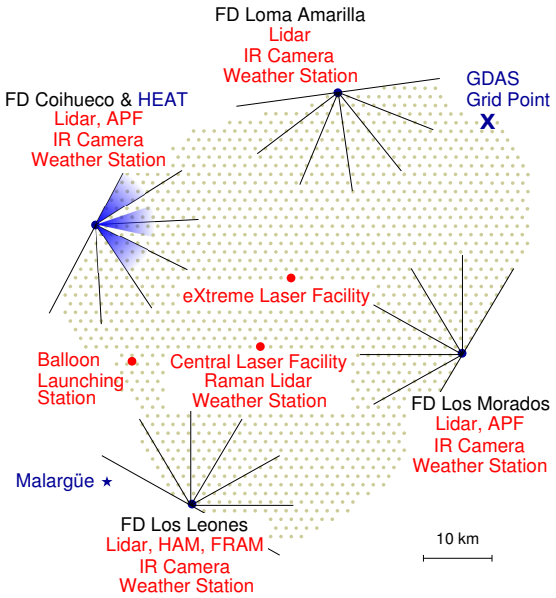


Figura 2.8: Schematic overview of the atmospheric monitoring devices installed at the Pierre Auger Observatory. At each FD site, there is a lidar station, a ground-based weather station, and an infra-red camera for cloud cover detection. In addition, there are devices for measuring the Aerosol Phase Function (APF) at FD Coihueco and Los Morados, a Horizontal Attenuation Monitor at FD Los Leones, and a ph(F)otometric Robotic Atmospheric Monitor also at Los Leones. A steerable backscatter elastic lidar system is installed at each of the 4 FD sites to measure aerosols and the positions of clouds near each site. At central positions within the surface detector array, two laser facilities are installed (CLF and XLF) to measure the vertical aerosol optical depth profile, $\tau_{\rm aer}(h)$, in the line of sight of each FD telescope 4 times per hour. In 2013 the CLF was upgraded with a Raman lidar. At the western boundary of the array, the Balloon Launching Site has been assembled together with a weather station. From this station, the weather balloons were launched so that they were typically carried across the entire array by westerly winds.

function of time.

Simulation and reconstruction tasks are factorized into sequences of processing steps which can simply be pipelined. Physicists prepare processing algorithms in modules, which they register with the $\overline{\rm Off}\underline{\rm line}$ framework. This modular design allows collaborators to exchange code, compare algorithms and build up a variety of applications by combining modules in various sequences. Run-time control over module sequences is obtained through a *run controller*, which invokes the various processing steps within the modules according to a set of user-provided instructions written in XML.

The $\overline{\rm Off}\underline{\rm line}$ framework includes two parallel hierarchies for accessing data: the detector description for retrieving conditions data, including detector geometry, calibration constants, and atmospheric conditions; and an event data model for reading and writing information that changes for each event.

The *detector description* provides a unified interface from which module authors can

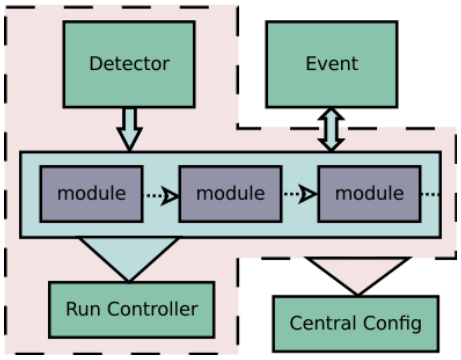


Figura 2.9: General structure of the $\overline{\text{Off}}\underline{\text{line}}$ framework. Simulation and reconstruction tasks are encased in modules. Each module is used to read information from the detector description and/or the event, process the information, and write the results back into the event under command of a *Run Controller*. A *Central Config* object is responsible for handing modules and framework components their configuration data and for tracking provenance.

retrieve conditions data. Data requests are passed by this interface to a back end comprising a registry of so-called managers, each of which is capable of extracting a particular sort of information from a given data source.

The transient (in memory) and persistent (on disk) event models are decoupled. When a request is made to write event contents to file, the data are transferred from the transient event through a so-called *file interface* to the persistent event, which is instrumented with serialization machinery, currently using ROOT [98]. Various file formats are interpreted using the file interface, including raw event and monitoring formats as well as the different formats employed by the AIRES [72], CORSIKA [71], CONEX [99] and SENECA [100] air shower simulation packages.

The $\overline{\text{Off}}\underline{\text{line}}$ framework includes a system to organize and track data used to configure of the software for different applications as well as parameters used in the physics modules. A *central configurator* points modules and framework components to the location of their configuration data, and creates Xerces-based [101] XML parsers to assist in reading information from these locations.

The central configurator keeps track of all configuration data accessed during a run and stores them in an XML log file, which can subsequently be used by the central configurator to reproduce a run with an identical configuration. The logging mechanism is also used to record the versions of modules and external libraries which are used for each run. Syntax and content checking of the configuration files is afforded through W3C XML Schema [102] standard validation. The configuration machinery can also verify configuration file contents against a set of default files by employing MD5 digests [103].

The $\overline{\text{Off}}\underline{\text{line}}$ framework is complemented by a collection of utilities, including an XML parser, an error logger and various mathematics and physics services. We have also developed a novel geometry package which allows the manipulation of abstract geometrical objects independent of coordinate system choice.

Low-level components of the framework are verified with a small test program, known

as a unit test, while full applications are vetted with more detailed acceptance tests. We employ a BuildBot system [104] to automatically compile the $\overline{\text{Off}}\underline{\text{line}}$ software, run the unit and acceptance tests, and inform developers of any problems each time the code is modified.

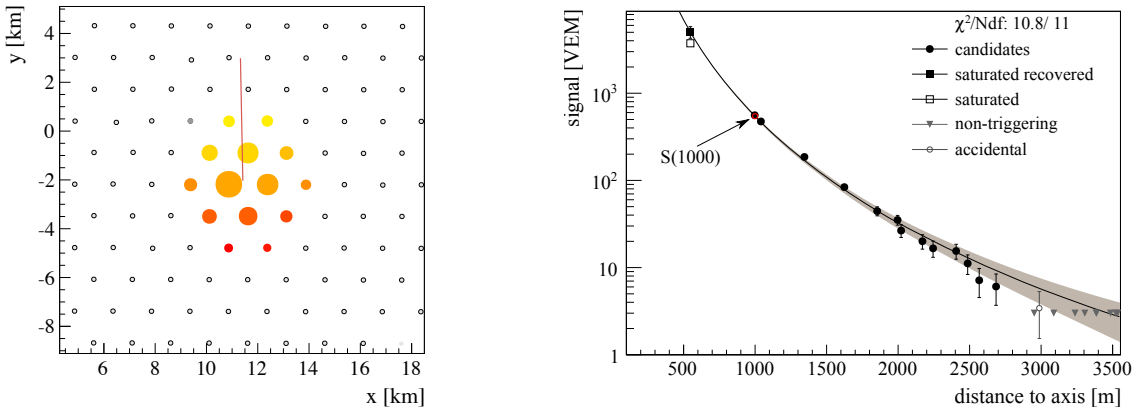
2.5 Event Reconstruction and Aperture

Extensive air showers in the Auger energy range are such dramatic and large scale events that there is essentially no background to both SD and FD measurements. Triggers are easily set up to exclude virtually any possibility of chance coincidences of triggers of the individual SD stations and/or FD pixels that would mimic a real cosmic-ray shower. Therefore, the performance of the detector and its ability to produce high-quality data depend solely on the accuracy of the cosmic-ray shower reconstruction and of the computation of the acceptance of the detector.

2.5.1 SD event reconstruction

Several experiments have proved successful in measuring extensive air shower parameters by use of a surface array. The quantities that can be measured directly are the geometry of the shower axis and the lateral distribution function (LDF), or the particle signal as a function of distance from the shower core. The primary energy can be inferred from the LDF, or, more specifically, from $S(1000)$, the detector signal at 1000 m from the shower core. At smaller distances, close to the core, fluctuations due to the nature of the first interactions of the primary with the atmosphere are dominating, while at larger distances statistical fluctuations become important. The relation between $S(1000)$ and the primary energy established by using shower simulations is therefore model dependent. To avoid the dependence on the models the Observatory takes the advantage coming from the hybrid detection: the air-showers that have triggered independently the FD and SD are used to relate the shower size from SD, $S(1000)$ to the almost-calorimetric measurement of the shower energy, the FD energy. The calibration of the energy is developed with a zenith independent measurement of $S(1000)$, S_{38}, and with the FD energy. The estimator of the primary energy in the case of the 750 m array is the reconstructed signal at 450 m from the shower core, denoted by $S(450)$. The zenith independent measurement of $S(450)$ used for the energy calibration is S_{35}. The SD-only reconstruction takes place in three steps: event and station selection, determination of the shower geometry and measurement of the shower lateral distribution function (LDF).

Event selection. To ensure good data quality for physics analysis there are two additional off-line triggers. The physics trigger, T4, is needed to select real showers from the set of stored T3 data that also contain background signals from low energy air showers. This trigger is mainly based on a coincidence between adjacent detector stations within the propagation time of the shower front. In selected events, random stations are identified by their time incompatibility with the estimated shower front. Time cuts were determined such that 99 % of the stations containing a physical signal from the shower are kept. An



(a) Footprint of the shower.

(b) Lateral distribution. Dependence of the signal size on distance from the shower core.

Figura 2.10: Footprint and lateral distribution of a reconstructed SD event ($E = 104$ EeV, $\theta = 25{,}1°$). In the footprint plot, colors represent the arrival time of the shower front from early (yellow) to late (red) and the size of the markers is proportional to the logarithm of the signal. The line represents the shower arrival direction.

algorithm for the signal search in the time traces is used to reject signals produced by random muons by searching for time-compatible peaks.

To guarantee the selection of well-contained events, a fiducial cut (called the 6T5 trigger) is applied so that only events in which the station with the highest signal is surrounded by all 6 operating neighbors (i.e., a working hexagon) are accepted. This condition assures an accurate reconstruction of the impact point on the ground, and at the same time allowing for a simple geometrical calculation of the aperture/exposure [105], important for, e.g., the spectrum analysis [106]. For arrival-direction studies a less strict cut can be used (5T5 or even 4T5).

Geometry and Energy measurement. An approximate shower geometry solution can be obtained from the simplified linear model assuming that all stations lie within some plane, i.e. here the tangential plane on the reference ellipsoid that contains the signal-weighted barycenter is chosen. In such a case one can expect $z_i \ll x_i,\, y_i$ for the station position (x_i, y_i, z_i). The z-component is neglected and the linear χ^2 is obtained,

$$\chi^2 = \sum_i \frac{[ct_i - ct_0 + ux_i + vy_i]^2}{\sigma_i^2}, \tag{2.2}$$

where t_i is the signal start time in tank i and t_0 is the time when the shower passes the barycenter. Equation 2.2 can be expressed as a set of linear equations and is analytically solved. The approximate solution serves as starting point to more elaborate 3D-fitting attempts taking into account the varying altitude of the stations and a more accurate core location from the LDF fit.

An example of the footprint on the array of an event produced by a cosmic ray with an energy of (104 ± 11) EeV and a zenith angle of $(25{,}1 \pm 0{,}1)°$ is shown in figure 2.10.

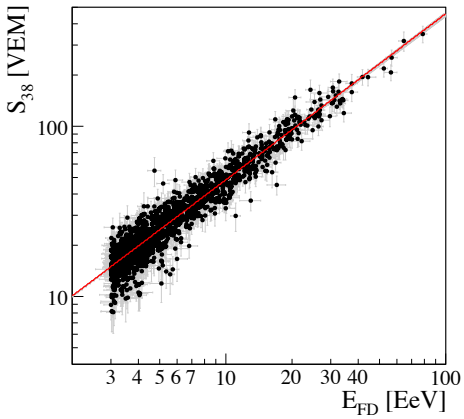


Figura 2.11: Correlation between S_{38} and E_{FD} [110, 111].

The lateral distribution of the signals is depicted in figure 2.10b. The function employed to describe the lateral distribution of the signals on the ground is a modified Nishimura-Kamata-Greisen function [107, 108],

$$S(r) = S(r_{\mathrm{opt}}) \left(\frac{r}{r_{\mathrm{opt}}}\right)^{\beta} \left(\frac{r + r_1}{r_{\mathrm{opt}} + r_1}\right)^{\beta+\gamma} \tag{2.3}$$

where r_{opt} is the optimum distance, r_1= 700 m and $S(r_{\mathrm{opt}})$ is an estimator of the shower size used in an energy assignment. For the SD array with station spacing of 1.5 km the optimum distance [109] is $r_{\mathrm{opt}} = 1000\,\mathrm{m}$ and the shower size is thus $S(1000)$. The parameter β depends on the zenith angle and shower size. Events up to zenith angle 60° are observed at an earlier shower age than more inclined ones, thus having a steeper LDF due to the different contributions from the muonic and the electromagnetic components at the ground. For events with only 3 stations, the reconstruction of the air showers can be obtained only by fixing the two parameters, β and γ to a parameterization obtained using events with a number of stations larger than 4.

The primary particle energy is determined from $S(1000)$ and the shower zenith angle θ. For a given energy, the value of $S(1000)$ decreases with θ due to the attenuation of the shower particles and geometrical effects. Assuming an isotropic flux of primary cosmic rays at the top of the atmosphere, we extract the shape of the attenuation behavior from the data using the Constant Intensity Cut (CIC) method [112]. An attenuation curve $f_{\mathrm{CIC}}(\theta)$ has been fitted with a third degree polynomial in $x = \cos^2\theta - \cos^2\bar{\theta}$, i.e., $f_{\mathrm{CIC}}(\theta) = 1 + a\,x + b\,x^2 + c\,x^3$, where $a = 0{,}980 \pm 0{,}004$, $b = -1{,}68 \pm 0{,}01$, and $c = -1{,}30 \pm 0{,}45$ [110].

The median angle, $\bar{\theta} = 38^\circ$, is taken as a reference point to convert $S(1000)$ to $S_{38} \equiv S(1000)/f_{\mathrm{CIC}}(\theta)$. S_{38} may be regarded as the signal a particular shower with size $S(1000)$ would have produced had it arrived at $\theta = 38^\circ$. For the 750 m array the median angle is $\bar{\theta} = 35^\circ$.

High quality hybrid events, events seen by both the SD and FD, are used to calibrate S_{38} with the near-calorimetric measurement of the primary energy by the FD, E_{FD}. The 1475 high quality hybrid events recorded between Jan 2004 and Dec 2012 which have

an energy above the SD full efficiency trigger threshold [105] are used in the calibration. The correlation between the two variables is obtained from a maximum likelihood method [111, 113] which takes into account the evolution of uncertainties with energy, as well as event migrations due to the finite energy resolution of the SD. The relation between S_{38} and E_{FD} is well described by a single power-law function,

$$E_{\mathsf{FD}} = A\,(S_{38}/\mathsf{VEM})^{B} \tag{2.4}$$

where the resulting parameters from the data fit are $A = (1{,}90 \pm 0{,}05)\times 10^{17}$ eV and $B = 1{,}025 \pm 0{,}007$ [110, 114]. As can be seen in figure 2.11, the most energetic event used in this analysis has an energy of 79 EeV.

The resolution of the final SD energy estimator,

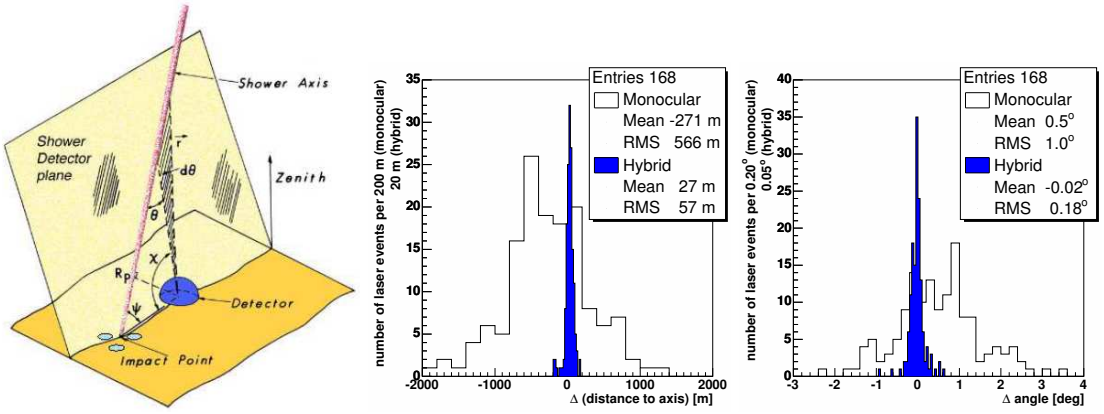
$$E_{\mathsf{SD}} = A(S(1000)/f_{\mathsf{CIC}}(\theta)/\mathsf{VEM})^{B}, \tag{2.5}$$

can be inferred from the distribution of the ratio $E_{\mathsf{SD}}/E_{\mathsf{FD}}$. Using the FD energy resolution of 7.6 %, the resulting SD energy resolution with its statistical uncertainty is $\sigma_{E_{\mathsf{SD}}}/E_{\mathsf{SD}} = (16 \pm 1)$ % at the lower energy edge in figure 2.11 and (12 ± 1) % at the highest energies. Due to the large number of events accumulated until December 2012, the systematic uncertainty on the SD energy due to the calibration is better than 2 % over the whole energy range. The systematic uncertainties are dominated by the FD energy scale uncertainty of 14 % [114]. The main contributions to this uncertainty are related to the knowledge of the fluorescence yield (3.6 %), the atmospheric conditions (3.4 to 6.2 %), the absolute calibration of the telescopes (9.9 %), the shower profile reconstruction(6.5 to 5.6 %) and the invisible energy (3 to 1.5 %).

2.5.2 Hybrid event reconstruction

Geometry Reconstruction. A hybrid detector achieves the best geometrical accuracy by using timing information from all the detector components, both FD pixels and SD stations. Each element records a pulse of light from which one can determine the central time of the pulse and its uncertainty. Each trial geometry for the shower axis yields a prediction for the times at each detector component. Differences between actual and predicted times are weighted using their corresponding uncertainties, squared, and summed to construct a χ^2 value. The hypothesis with the minimum value of χ^2 is the reconstructed shower axis. In the FD, cosmic ray showers are detected as a sequence of triggered pixels in the camera. The first step in the analysis is the determination of the shower-detector plane (SDP) that is the plane that includes the location of the eye and the line of the shower axis (figure 2.12a). Experimentally, it is determined by minimizing the signal weighted sum of scalar product of its normal and the pixel pointing directions.

Next, the timing information of the pixels is used for reconstructing the shower axis within the SDP. As illustrated in figure 2.12a, the shower axis can be characterized by two parameters: the perpendicular distance R_{p} from the eye to the track and the angle ψ that the track makes with the horizontal line in the SDP. Each pixel which observes the track has a pointing direction which makes an angle χ_i with the horizontal line. If t_0 is the time



(a) Hybrid geometry variables. (b) Core resolution. (c) Angular resolution.

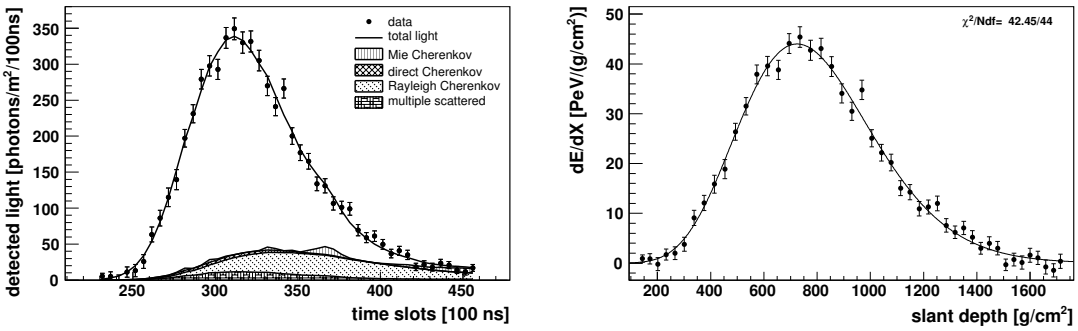
Figura 2.12: (a) Illustration of the geometrical shower reconstruction from the observables of the fluorescence detector. (b) Core location for monocular and hybrid reconstruction. (c) Same as b for the angular resolution.

when the shower front on the axis passes the point of closest approach R_p to the eye, then the light arrives at the ith pixel at the time

$$t_i = t_0 + \frac{R_\mathrm{p}}{c} \cot\left[(\psi + \chi_i)/2\right]. \tag{2.6}$$

The shower parameters are then determined by fitting the data points to this functional form. The accuracy of the monocular (FD-only) reconstruction is limited when the measured angular speed $\mathrm{d}\chi/\mathrm{d}t$ does not change much over the observed track length. For such showers, degeneracy in the fitting parameters can be broken by combining the timing information from the SD stations with that of the FD telescopes. This is called the *hybrid* reconstruction. Example results are shown in figure 2.12b and figure 2.12c for reconstruction of a vertical laser beam at the CLF where some laser light is also injected into a neighboring SD station. There we compare the mono and hybrid reconstructions of the distance to the laser and the zenith angle. With the monocular reconstruction, the location of the CLF can be determined with a resolution of $\sim$500 m. After including the timing information of the single SD station, the resolution improves by one order of magnitude with no systematic shift.

Profile Reconstruction and Energy Determination. Once the geometry of the shower is known, the light collected at the aperture as a function of time can be converted to energy deposit, $\mathrm{d}E/\mathrm{d}X$, at the shower as a function of slant depth. For this purpose, the light attenuation from the shower to the detector needs to be accounted for and all contributing light sources need to be disentangled: fluorescence light [115, 116, 117], direct and scattered Cherenkov light [118, 119] as well as multiply scattered light [120, 121, 122]. Since the Cherenkov and fluorescence light produced by an air shower are connected to the energy deposit by a linear set of equations, the shower profile is obtained by an analytic linear least square minimization [123]. Due to the lateral extent of air showers,



(a) Light at aperture. (b) Energy deposit.

Figura 2.13: Example of a reconstructed shower profile.

a small fraction of shower light is not contained within the optimal light collection area. This is corrected for by taking into account the universal lateral fluorescence [124] and Cherenkov light distributions [125]. The calorimetric energy, E_{cal}, of a shower is given by the integral over the longitudinal energy deposit profile,

$$E_{\mathrm{cal}} = \int_0^\infty \mathrm{d}E/\mathrm{d}X(X)\,\mathrm{d}X. \tag{2.7}$$

Since usually the full profile cannot be observed within the field of view of the FD, this integral is evaluated from a Gaisser-Hillas function [126] that is fitted to the reconstructed energy deposit. In addition, this fit yields an estimate of X_{max}, the mass sensitive position of the shower maximum. An example of the measured light at aperture and the reconstructed light contributions and energy deposit profile is shown in figures 2.13a and 2.13b. The total energy of the shower is obtained from E_{cal} by correcting for the 'invisible energy' carried away by neutrinos and high energy muons [127].

The resolution of the profile measurement can be determined by reconstructing simulated showers, that have passed a full detector simulation [128]. Moreover, at the high energies it can be determined from the data itself by comparing independent measurements of the same shower by different eyes (stereo events). Both studies show that the energy of a shower can be determined with a precision of 8 % above 10 EeV. For the shower maximum, the resolution is 20 g/cm^2 [129].

2.5.3 SD aperture

An important feature of the SD is that it allows for a straightforward control of the shower detection volume. With the requirement that the shower core of the events be reconstructed within the limits of the region covered by the SD stations, the SD aperture can be obtained from a simple geometric calculation of the actual size of the active array on the ground, at any given time. This sets the effective detection surface on the ground, to be weighted by the energy-dependent detection efficiency of cosmic ray showers. This efficiency can be measured directly from the hybrid data: restricting oneself to conditions

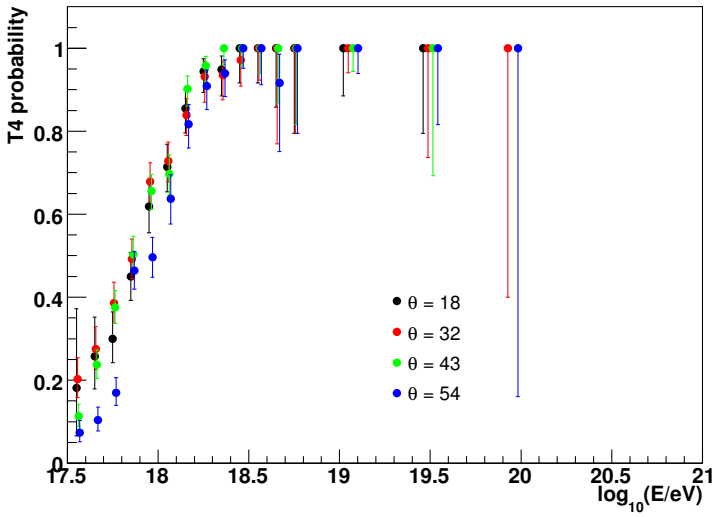


Figura 2.14: Detection efficiency of the SD array, as a function of energy, as measured from the data using the hybrid data set, for different zenith angles.

where the showers are known to be detectable with 100 % efficiency in the 1-tank hybrid mode. One derives the SD detection efficiency at a given energy as the fraction of the corresponding showers that do trigger the SD, at the "physics trigger" level (T4) described above. This is shown in figure 2.14, where the detection efficiency is seen to reach 100 % at the *saturation energy* $E_{\rm sat} \simeq 3{\times}10^{18}$ eV, in very good agreement with simulations and measurements based on the SD trigger probability and signal fluctuations as a function of distance to the shower axis.

Above $E_{\rm sat}$, the instantaneous aperture of the SD is derived from the total surface covered by the array. In order to avoid border effects and a potentially degraded energy reconstruction for showers hitting the ground close to an edge of the array or in a region where some tanks are momentarily inactive, an additional cut is applied to the data (referred to as the *quality trigger* 6T5 or simply T5) to ensure nominal reconstruction accuracy. As already explained, this 6T5 trigger requires that the station recording the highest signal in a given event be surrounded by at least 6 active stations (an active hexagon). In the case when only 5 neighboring stations are active (5T5), the shower core must also be reconstructed inside an elementary triangle of stations that were active at that time. Figure 2.15 illustrates this requirement, showing the core positions allowed for vertical showers arriving well inside the hexagonal cell (left) or near a missing station (right). The total detection area associated with the central station is then seen to be $D^2\sqrt{3}/2 \simeq 1{,}95\,\mathrm{km}^2$ in the former case, and 2/3 of this in the latter. A final integration over solid angle for showers with zenith angles between 0 and 60° gives the nominal aperture per active station: $A_0 \simeq 4{,}59\,\mathrm{km}^2$ sr.

The computation of the total SD aperture at any given time is then obtained by multiplying the elementary aperture by the number of active stations (with the required number of active neighbors), obtained from a simple census using the SD monitoring data,

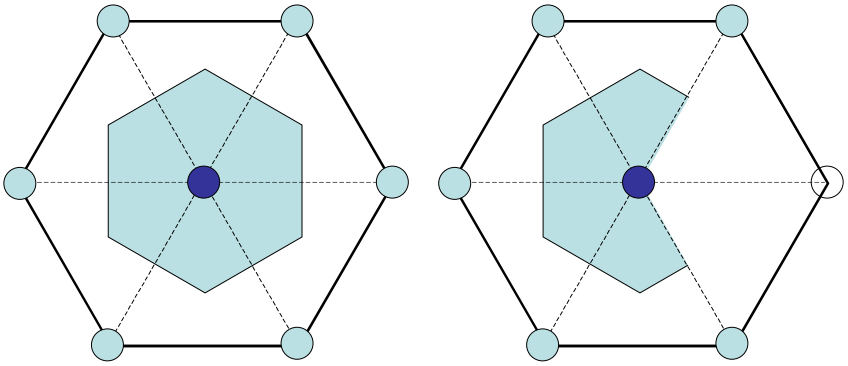


Figura 2.15: Schematic view of the area (shaded region) where the core of a vertical shower must be located inside an elementary hexagonal cell of the SD array to pass the quality trigger. Left: for a complete hexagon with 6 active neighbors. Right: for a hexagon with 5 active neighbors.

which give the list of active stations on a second-by-second basis. Finally, the integrated SD exposure in *linsleys* ($1\,\mathrm{L} = 1\,\mathrm{km}^2\,\mathrm{sr\,yr}$) is obtained by the time integration of the instantaneous aperture of the SD array, taking into account any changes in the array configuration, using the same monitoring data. In this way, the growth of the array during the SD deployment period could be automatically included in the exposure calculation, whatever the shape and duration of the intermediate configurations.

Overall, the above-mentioned technique provides a very accurate determination of the SD acceptance, with an uncertainty of $\sim 3\,\%$, which can be considered as negligible with respect to the uncertainty on the energy reconstruction.

In Table 2.1 we summarize some of the important parameters that characterize the performance of the Observatory. These parameters include the event rate of the detectors and the resolutions of the different reconstructed observables.

SD	
SD Annual Exposure	$\sim 5500\,\text{km}^2\,\text{sr}\,\text{yr}$
T3 rate	0.1 Hz
T5 events/yr, $E > 3\,\text{EeV}$	$\sim 14,500$
T5 events/yr, $E > 10\,\text{EeV}$	~ 1500
Reconstruction accuracy ($S(1000)$)	22 % (low E) to 12 % (high E)
Angular resolution	$1{,}6^\circ$ (3 stations)
	$0{,}9^\circ$ (>5 stations)
Energy resolution	16 % (low E) to 12 % (high E)
FD	
Duty cycle	~ 15 %
Rate per building	0.012 Hz
Rate per HEAT	0.026 Hz
Hybrid	
Core resolution	50 m
Angular resolution	$0{,}6^\circ$
Energy resolution (FD)	8 %
X_{max} resolution	$<20\,\text{g}\,\text{cm}^{-2}$

Cuadro 2.1: Key performance parameters for the Auger Observatory.

3

The risetime and its experimental measurement

In the previous chapter we discussed the discrepancies between data taken by the Pierre Auger Observatory and the predictions given by Monte Carlo simulations. These discrepancies, together with the scant FD statistics at the largest energies to measure X_{max}, lead to the necessity of a method for mass composition studies independent of simulations and based on SD measurements, where the duty cycle is $\sim$100 %.

In this chapter we introduce an observable obtained from the surface detectors, the risetime, and we discuss how it is effectively measured at the Pierre Auger Observatory. This observable will be used in the next chapter to develop a method to infer the position of the shower maximum.

3.1 The risetime as a mass sensitive parameter

The spread of the arrival times of particles at ground carries information about the longitudinal development of a UHECR shower [130, 131]. This can be explained if we consider the geometry of the particles in the shower. Suppose one particle is created at the beginning of a shower and it follows a straight line to arrive at one detector at the ground. Compare this to another particle created later in the shower, close to the shower maximum, which also follows a straight line to arrive at the same detector. Assuming the particles travel at the same speed, they will arrive at the surface detectors at different times, arriving first the first created particle. Therefore, a spread in the arrival times of particles will be observed at the surface detectors. Following this thought we can conclude that the spread in the arrival times of particles is larger for showers which penetrate deeper into the atmosphere. This is demonstrated schematically in figure 3.1.

For the same energy, lighter cosmic rays induce showers with maxima deeper in the atmosphere than showers induced by heavier primary particles. Thus, lighter primary

cosmic rays will have a larger spread of the arrival times at the surface detectors. In this way, it is possible to establish a relationship between the spread of the arrival times and the mass of the primary cosmic rays.

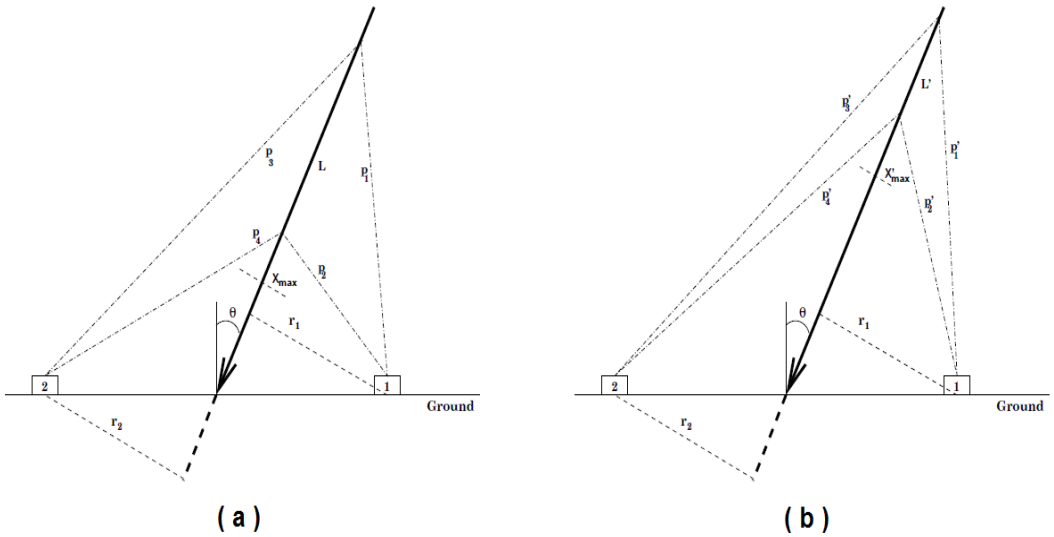
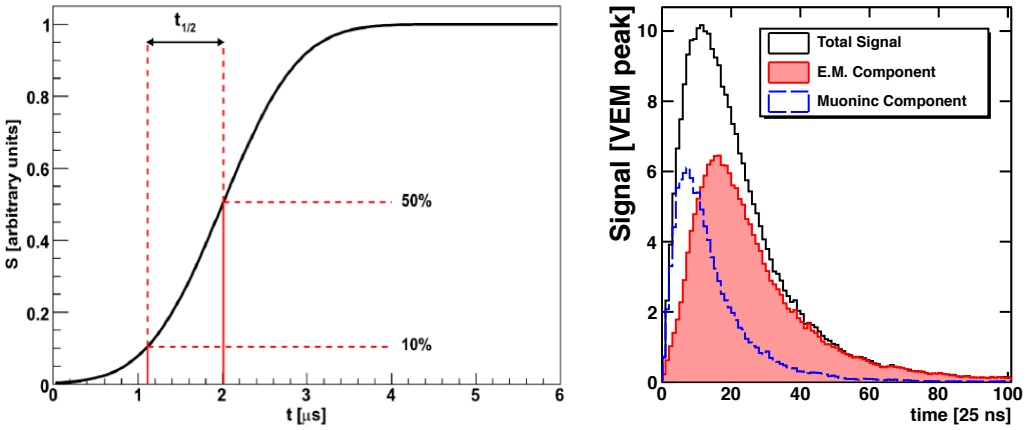


Figura 3.1: Schematic diagrams showing the difference in the spread of the arrival times at the ground for showers with different longitudinal developments. Diagram (a) represents a deeply penetrating shower, whereas diagram (b) represents a shower which develops higher in the atmosphere.

The information related to the arrival times of particles is registered in the FADC traces of the surface detectors. However the full time structure of the pulse is not used in the investigation of the spread of arrival times. This is because early and late parts of the pulse contain features whose use affects mass composition studies. On the one hand, the late part carries little information about the development of the shower because it is dominated by electromagnetic particles. On the other hand, the early part, although it is rich in information, is measured with difficulty using our surface detectors. As a consequence, it is necessary to define a parameter which only incorporates the optimum part of the time pulse.

The risetime, $t_{1/2}$, is the parameter which has been devised for mass composition studies since the pioneering works [131]. This is defined as the time taken by the integrated signal to rise from 10 % to 50 % of its total value, see figure 3.2a. This definition makes it sensitive to the muon-to-electron ratio of the total signal, as muons mostly arrive as a pack in the first time bins, while electrons and photons arrive later due to the attenuation and the scattering that they suffer in the atmosphere, see figure 3.2b.

Before the evaluation of the risetime, some detector effects have to be taken into account to obtain the correct information from the FADC traces. In the next sections we summarise how these effects may distort the risetime values and how they can be overcome.



(a) (b)

Figura 3.2: (a) Schematic view of the risetime definition. (b) Different components to the total signal as a function of the time. Average total signal at 1000 m (log(E/eV) = 19.1 and sec θ = 1.15).

3.2 The negative signal problem

In late 2009, after one of the updates of the $\overline{\text{Off}}\underline{\text{line}}$ software, a large number of unusually small risetime values were observed in the data registered by surface detectors [132]. The cause of this anomaly, found by Ronald Bruijn (University of Leeds), indicated an underlying problem with the calculation of the baselines and the stop-times of FADC traces [133]. This problem was called *the negative signal problem* due to the numerous bins with negative signal values found in a sizable sample of FADC traces.

The negative signal problem particularly affects surface detectors where the low-gain channel is used to record the FADC traces due to the saturation occurring in the high-gain channel. These detectors are called HG saturated detectors. The traces affected by the this problem have smaller risetime values due to two effects: the overestimation of the baseline and an erroneously late stop time.

To grasp why the negative signal problem appears for HG saturated detectors in the new $\overline{\text{Off}}\underline{\text{line}}$ versions we will first discuss the way in which the baselines and the start (stop)-times are calculated. The section of a FADC trace containing the shower signal is determined by a systematic scanning of the bins that searches for *segments* and *gaps* [134] . A shower segment is defined as a set of N_{seg} consecutive bins with signal greater than a threshold value or tolerance, S_{th}, separated by a gap (flat section of trace) of N_{gap} consecutive bins below S_{th}. The largest segment found determines the trace section attributed to the air shower. The signal found in the gaps is not related to the air shower and it is used to calculate the baseline. Figure 3.3 shows a portion of a FADC trace where the segments and the gaps (in green boxes) have been identified.

The baseline is the part of a FADC trace not directly attributable to the air shower

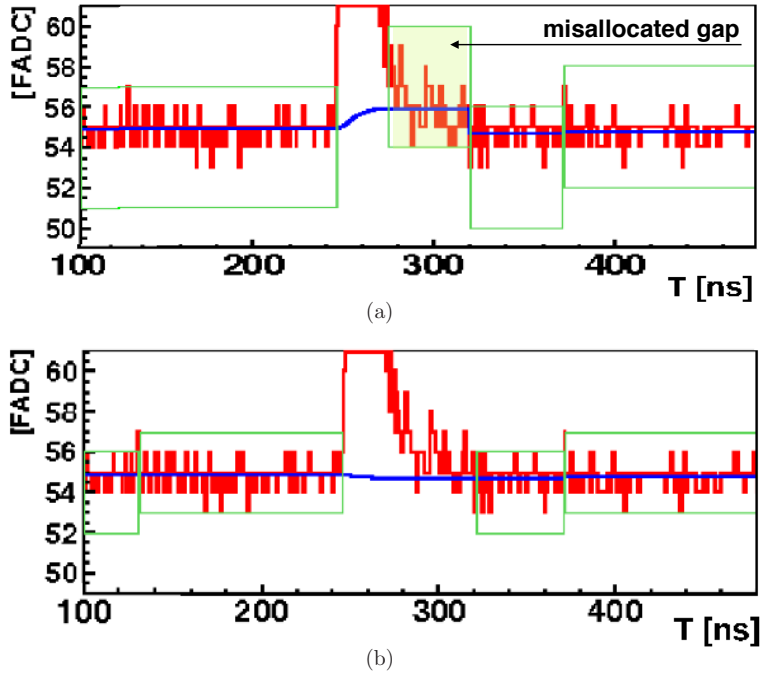


Figura 3.3: Example showing a portion of a FADC trace before the baseline subtraction. The FADC signals (red line), the baseline (blue line) and the gaps used to calculate the baseline (green boxes) are shown. The upper and lower panels show the baseline calculated using the old and new stop-times respectively.

signal or coincident muons. The baseline is a combination of an artificial constant off-set from zero, the undershoot and the fluctuations due to the electronic noise.

In gaps, the baseline is simply the average of the recorded signal. While in the segments, the corresponding baseline is an interpolation of the baseline between surrounding gaps with a time-dynamic correction for the undershoot. As a consequence of this definition, if the baseline is incorrectly calculated for one gap then the baseline interpolation for adjacent segments will also be incorrect.

In the FADC traces with *the negative signal problem*, the tail of the signal has been mistakenly included as part of the next gap rather than as a part of the segment, increasing in that way the average signal in the gap. See the shaded green box in the upper panel of figure 3.3. The level of baseline to be subtracted in this segment is therefore overestimated, creating negative signal values in the time bins toward the end of the trace. This leads to an underestimation of the total signal and the risetime.

The method to search for segments and gaps in the $\overline{\text{Off}}\underline{\text{line}}$ is the same for FADC traces originated from the low-gain and high-gain channels, which means that the problem of identifying gaps and segments happens for both channels. However, the effects in risetimes and in total signals are significant when the low-gain channel is used. This

happens because the ratio dynode/anode is $\sim$30 for surface detectors. This means that an ADC count in the low-gain channel has 30 times more signal in VEM units compared to a count in the high-gain channel.

The late stop-times exacerbate the negative signal problem, although the main contribution comes from the overestimation of the baseline. Now the problem comes from the fact that the start (stop)-times are always calculated using the high-gain channel independently of the level of saturation of the FADC traces. To find the start (stop)-times, the same procedure, previously explained to find segment and gaps, is used but only in the high-gain channel. Once the gaps and segments are located, the start (stop)-time of the trace is the position of the first (last) time bin in the largest segment. The late stop-times appear when the saturated high-gain channel is used to determine the stop-times. The value of S_{th} to distinguish gaps and segments is the same for both channels, without taking account the huge difference between the signal size registered in each channel. As a consequence, some small peaks from the high-gain channel are mistakenly included as part of the segment rather than as a part of the next gap, see figure 3.4. This mistake produces a larger segment than it should be and an unnecessary late stop-time.

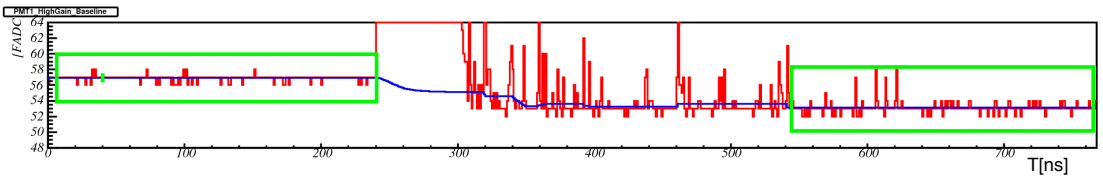


Figura 3.4: Schematic view of the stop-time calculation: the high-gain channel is saturated, nonetheless it is used to calculate the stop-time bin.

3.2.1 Algorithm to solve the negative signal problem

An algorithm was developed by Ronald Bruijn [135] to avoid the negative signal problem. The algorithm improves both the calculation of the baseline and of the stop-time for the low gain channel. In the first step, the algorithm establishes the use of the respective channel for the search of the start (stop)-time instead of using by default the high-gain channel. It also reduces the tolerance, S_{th}, for the low-gain channel. Secondly the procedure to identify segments and gaps is applied as mentioned before, but now, once the end of a gap is reached, the end of the previous segment is checked by working backward through the bins (using a lower signal threshold) to ensure that the end of the signal has not been mistaken as a part of the gap. Where this does occur, the size of the gap is recalculated to ensure that the tail of the signal is included in the signal segment. As a result, the baseline is no longer overestimated, see lower panel in figure 3.3, and the stop-time is well defined. Note that the equations governing the baseline calculation and subtraction have not been altered but the values acting as inputs to these equations have changed. This algorithm is commonly called *Ronald's correction*.

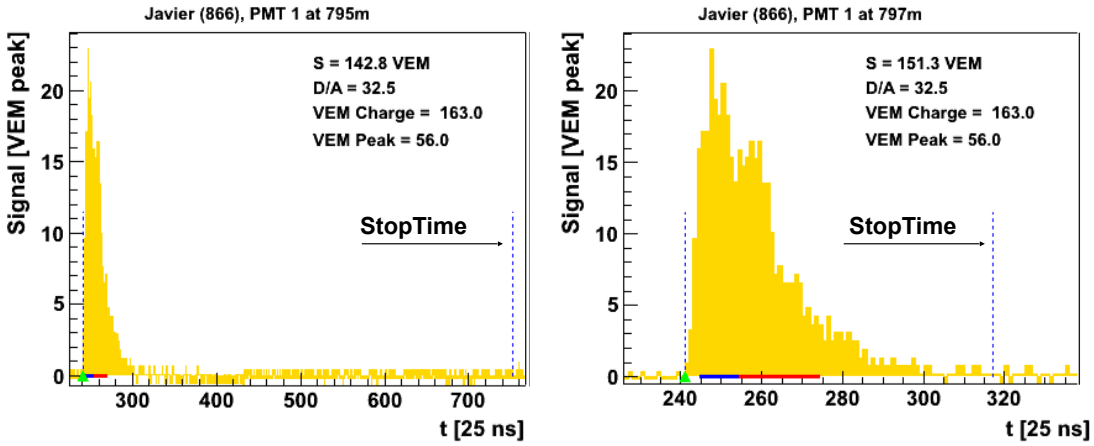


Figura 3.5: Trace of the detector 866 before (left) and after (right) applying the algorithm developed by R. Bruijn (Event 14542888, 05 May 2012). The risetime of the trace shown in the left panel is 225 ns. The risetime of the right panel is 250 ns.

Figure 3.5 shows the effect of the algorithm in a particular trace. With the official $\overline{\text{Off}}\underline{\text{line}}$ reconstruction, the stop-time (indicated by a dash blue line at the end of the trace) is around the time bin 750. Applying the algorithm, the stop-time is reduced up to the time bin 320 and the total signal has increased 8 VEM. Figure 3.6 shows the number of negative signals recorded by surface detectors before and after Ronald's correction. In this figure it is evident that the problem mainly happens when the low-gain channel is used. With this algorithm the negative signals are removed for 80 % of the HG saturated detectors. A residual number of negative signals remains after the correction, but this only affects 9 % of HG saturated detectors.

This algorithm, included in the $\overline{\text{Off}}\underline{\text{line}}$ module *SdCalibrator*, was implemented for the first time in the $\overline{\text{Off}}\underline{\text{line}}$ version v2r9p3-icrc13-final (May 2013) and has been included in subsequent releases. However at this moment the algorithm is off by default in the official release used by the Pierre Auger Collaboration. This choice stems from the fact that final checks about the effects of this algorithm on the event reconstruction are being finished. The main effects on the event reconstruction have been studied as a part of this work and they are included in the next section. No flaws have been identified so far. However a last bunch of cross-checks are being done by members of the collaboration specialized in the event reconstruction.

3.2.2 Effect of Ronald's correction in the event reconstruction

When a new algorithm is proposed to be included in the official release of the $\overline{\text{Off}}\underline{\text{line}}$ package several cross-checks have to be done to guarantee that the modification does not distort the event reconstruction. In the case of the algorithm to solve the negative signal problem these cross-check are especially important because the algorithm modifies both the total signal and the start (stop)-times. These changes only happen for HG satura-

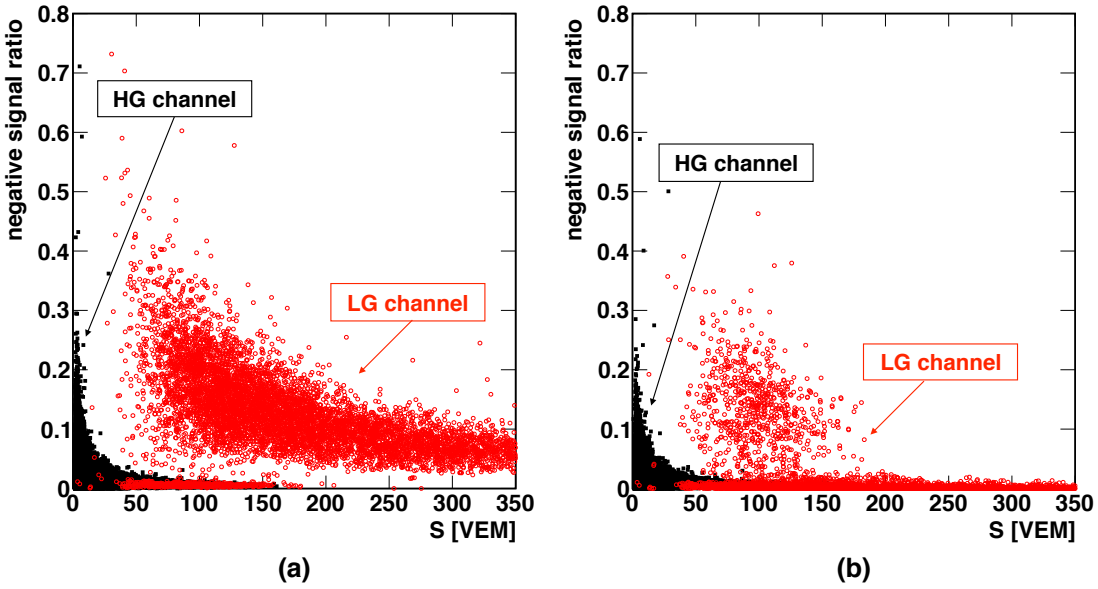


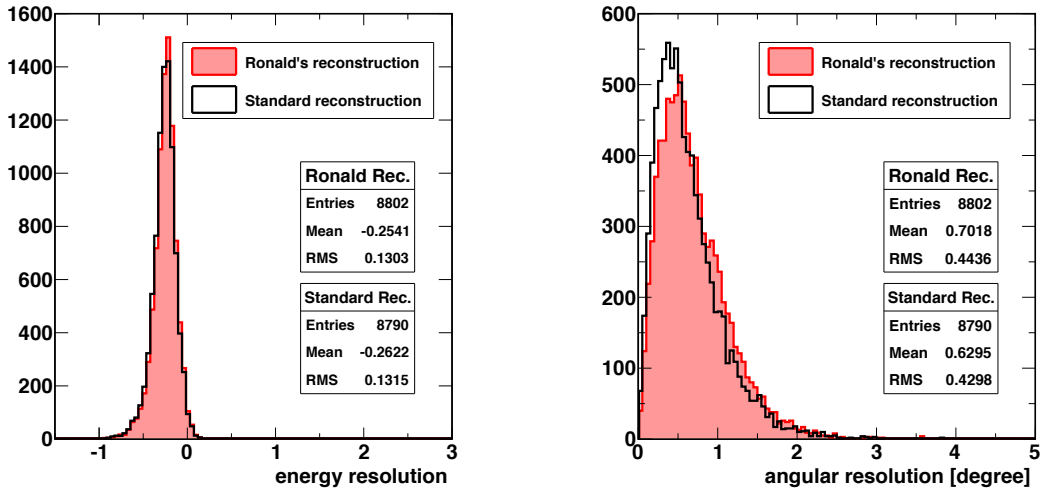
Figura 3.6: Ratio of negative signal over the total signal of the PMTs as a function of the total signal for the high-gain and low-gain channels. Events with energies in the range 10 <E <15 EeV and zenith angle in the range 1.00 <sec θ <1.45 are used. (a) Before applying Ronald's correction. (b) After Ronald's correction.

ted detectors but these detectors are also included in the fits to obtain the reconstructed energy and the axis of the shower.

To study how this modification affects the event reconstruction we use simulations. In particular we use proton nuclei simulated with the hadronic model QGSJetII.04. The goal is to compare the values of the reconstructed parameters to the true ones, which are always known for simulations. This is done for both reconstructions, the official or standard one and the one using Ronald's algorithm. The most important parameters that we have to study are the energy and the angular resolution.

Figure 3.7 shows the values obtained for the energy and the angular resolution. The effect of Ronald's correction in the energy reconstruction is negligible as shown in figure 3.7a. A slight discrepancy is observed for the angular resolution, see figure 3.7b. In this case the histogram found for the angular resolution with the new reconstruction is wider and the mean value is larger. Due to this small discrepancy we have studied the behaviour of the angular resolution as function of the energy, figure 3.8a, and as a function of the sec θ, figure 3.8b. The difference between both reconstructions is constant, decreases with energy and with sec θ and it is never larger than 0.1 degree.

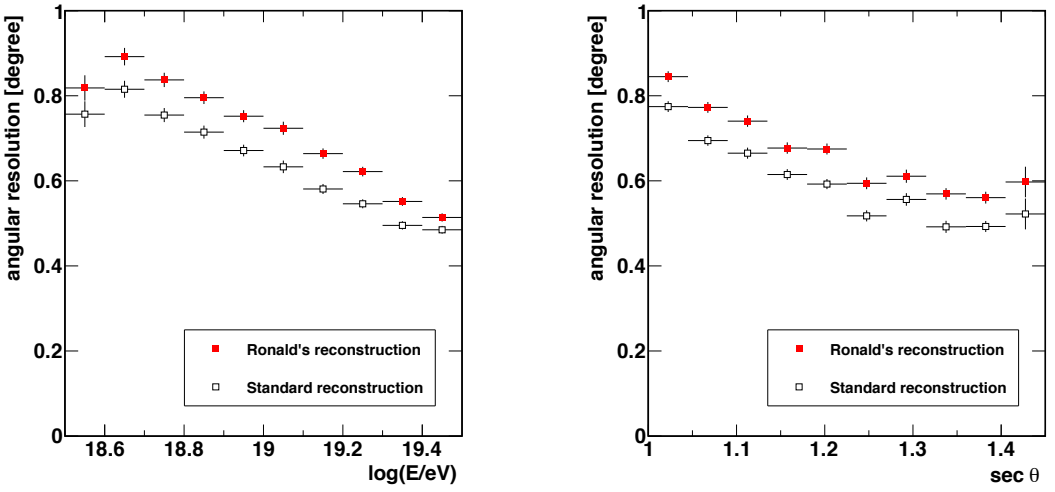
We have also studied the effect of Ronald's correction on the event reconstruction using data. Figure 3.9a shows the difference between both reconstructed energies as a function of $\cos^2\theta$. For θ <45° ($\cos^2\theta$ >0.5) this difference follows a flat trend and it is less than 1.5 %. This implies that it is not necessary to modify the CIC for the reconstruction



(a) Energy resolution histogram

(b) Angular resolution histogram

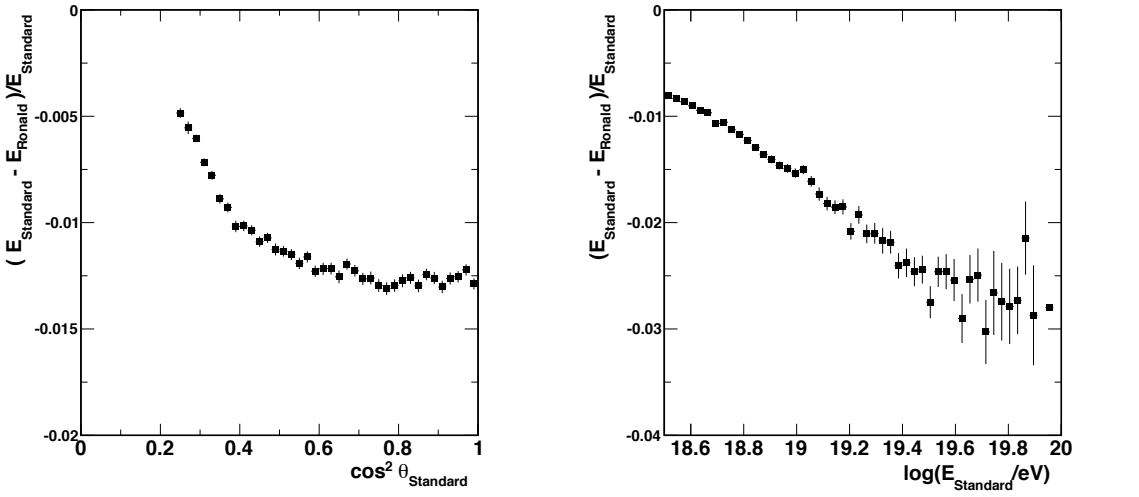
Figura 3.7: Energy and angular resolution obtained for the standard reconstruction and the new one, with the Ronald's correction. To obtain the plots we have used proton nuclei with energies in range 18.5 <log(E/eV) <19.5 and sec θ below 1.45.



(a) Angular resolution as a function of energy.

(b) Angular resolution as a function of sec θ

Figura 3.8: Trend of the angular resolution as a function of the energy and the sec θ for the standard reconstruction and the new one, with Ronald's correction.



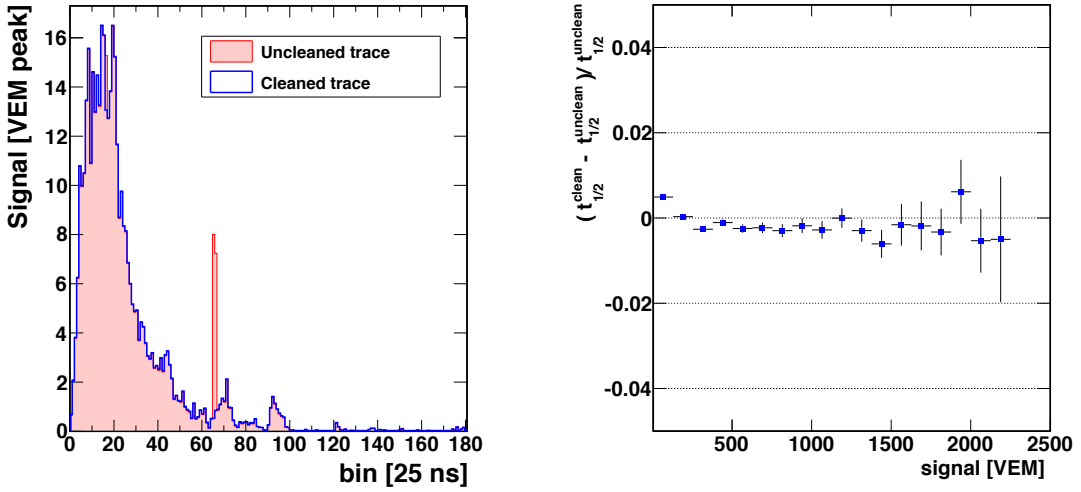
(a) Energy differences as a function of the $\cos^2\theta$ (b) Energy differences as a function of the log(E/eV)

Figura 3.9: Comparison of the standard reconstruction and the new one, with Ronald's correction, using data.

which uses Ronald's correction. Figure 3.9b shows the energy difference as a function of the energy. This is not larger than 3 $\%$ above 50 EeV. Given the results presented in this section, the new reconstruction modifies the angular resolution by less than 0.1 degree. However this small distortion in the angular reconstruction is not significant compared to the improvements it makes in the determination of the baseline, the stop-time and the signal size.

3.3 Direct Light Effect

In a number of FADC traces, sharp peaks not associated with the main signal are observed. They are referred to as *direct light*. What differentiates these peaks from random muons is that these peaks are very narrow, only two or three bins wide, and they are in only one of the three PMTs. An example is presented in figure 3.10a, where one PMT observes a large signal peak at the end of the trace. We assume that the direct light effect is caused by particles entering in the detectors in such a direction that only one PMT preferentially sees the light over the other two. Several processes has been proposed to explain this. The first is that a particle, possibly a low energy electron, is being back-scattered into the detector directly toward one of the PMTs. The second possible mechanism, is photon production within the glass of the PMT. If a particle impacts the glass of a PMT with enough energy to produce Cherenkov photons, a large signal would be produced in this PMT. The third process would be the Cherenkov emission from electrons from the muon decays such that the Cherenkov cone is directly incident on one of



(a) Detector 110 of the event 629572 (PMT1). (b) Risetime difference as a function of the total signal.

Figura 3.10: (a) FADC trace of a PMT affected by the direct light effect, before and after the cleaning done by the correction. (b) Effect of the direct light correction on risetimes as a function of the total signal. Events with energies larger than 3 EeV and sec θ lower than 1.45 are used (S >5 VEM and r <1400 m).

the PMTs. In each scenario a big amount of signal would be seen for only one of the three PMTs.

The $\overline{\text{Off}}\underline{\text{line}}$ software uses an algorithm to remove these peaks of direct light. This algorithm tests each FADC bin individually. In each bin the mean of the three PMTs is calculated along with the RMS spread. If one bin is further than certain tolerance (1.0 σ) away from the mean, this is removed and replaced by the average of the other two bins. The effect of this algorithm before and after its application can be seen in figure 3.10a. In the cleaned trace the large peak around the time bin 62 has disappeared.

In figure 3.10b the effect of the direct light correction on risetimes as a function of the total signal is studied. It can be seen that the correction used to remove this effect changes the mean values of the risetimes in less than 1 $\%$. This change remains constant as a function of the total signal.

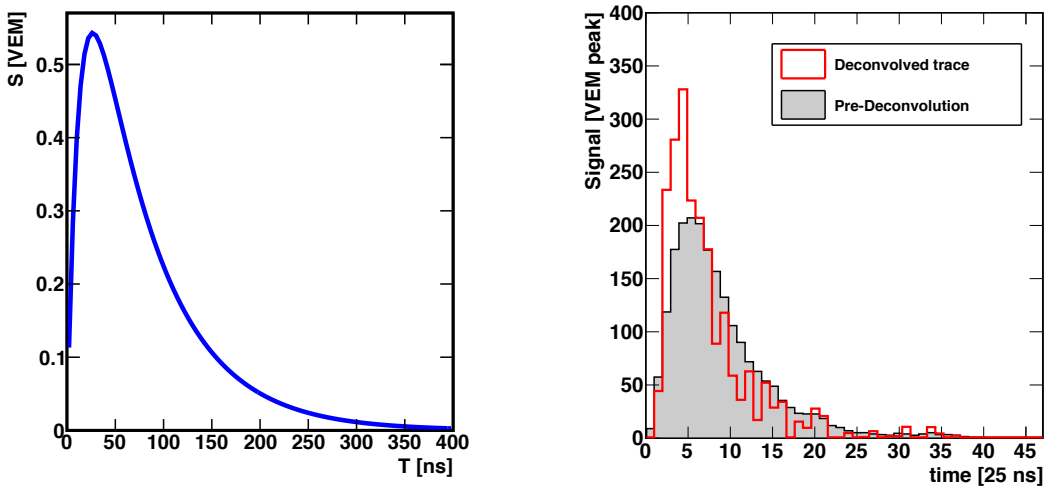
3.4 Deconvolution

The arrival times of particles entering in the surface detectors are artificially lengthened due to detector effects. The main contribution to this lengthening comes from the multiple reflections of Cherenkov photons inside the surface detectors, in particular in the Tyvek liner, although other sources are the physical size of the detectors and the bandwidth limitations of the electronic. The consequence of this lengthening is that the time structure

of the FADC traces do not contain direct information about the air shower, but it is a convolution of the arrival times of particles with the detector response.

This smearing is known as the Single Particle Response (SPR), and it is the detector response to a single ideal vertical particle (described theoretically by a delta function). The average SPR for the surface detectors of the Pierre Auger Observatory has been measured experimentally using background muons and it can be described with equation 3.1, where A is a normalization factor and t the time measured in ns [136] . The values 67 ns and 13 ns correspond to the decay time and the risetime of the muon signal respectively.

$$F(t) = A(\mathrm{e}^{-t/67} - \mathrm{e}^{-t/13}) \tag{3.1}$$



(a) (b)

Figura 3.11: (a) SPR for a single vertical muon. (b) FADC trace from the detector 110 (PMT1) in the event 15576963, before (gray shadow) and after (red line) the deconvolution procedure with the GDA.

Figure 3.11a shows the empirical SPR, where the trace has been normalized to represent a single vertical muon. The actual time taken by a relativistic particle to traverse the tank is of the order of 5 ns, so the comparison with the length of SPR pulse shows that the lengthening is quite significant. The risetime of the SPR pulse is $\sim$40 ns, which causes an artificial lower limit on the risetime measurements.

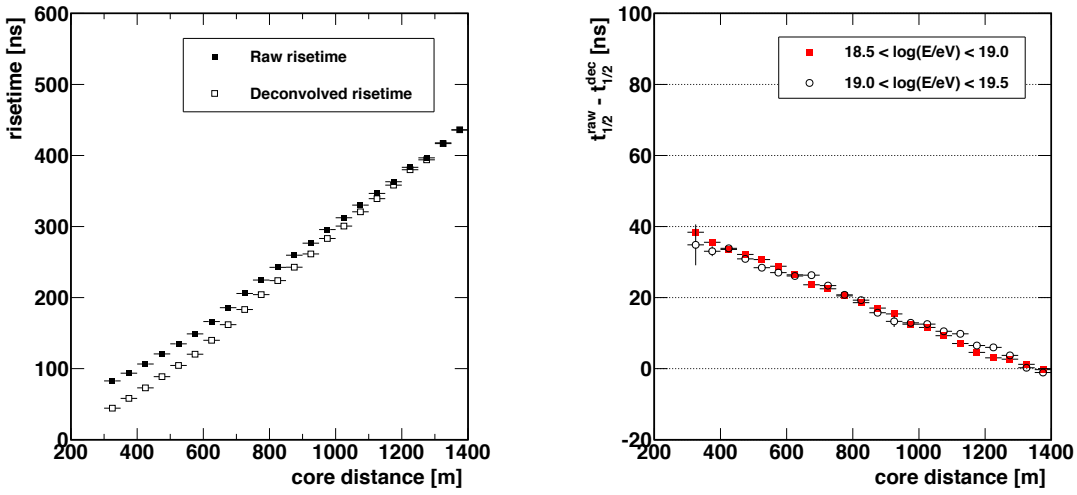
The Gold Deconvolution Algorithm (GDA) implemented in the ROOT package [98] can be used to deconvolve the traces and to obtain a more realistic time distribution of the incoming particles. An example of a trace before and after the deconvolution is shown in figure 3.11b.

Signals close to the shower core are the most affected by the deconvolution, as the

smearing due to the SPR dominates in fast risetimes, see figure 3.12a.

A deconvolution procedure with the GDA was used in previous works based on risetime measurements, see [137] and [138]. However, concerns over the stability of the GDA have lead to its exclusion in this work. The GDA requires that there are no negative signals in any time bins. In traces with several negative signal bins, the algorithm does not work properly since it can convert a value as small as -2 VEM peak in a disproportionate negative value as large as -200 VEM peak, distorting the value of total signal. It was previously shown in section 3.2 that the new $\overline{\text{Off}}\underline{\text{line}}$ versions produce small amounts of negative signal in the FADC traces even after Ronald's correction. This fact makes impractical the application of the deconvolution procedure in this work.

Not using the deconvolution in this work does not affect the analysis negatively. As it will be explained in detail in the next chapter, the Delta Method is based on the comparison of risetimes located at the same core distance, so we will compare risetimes affected equally by the SPR. For this reason, the only aspect that we have to guarantee is that the effect of the SPR as a function of the core distance is the same for different energies. Figure 3.12b shows that the difference between raw and deconvolved risetimes as a function of core distance is the same for the two studied energy bins.



(a) (b)

Figura 3.12: (a) Raw and deconvolved risetimes as a function of the core distance for simulated events with energies in the range 18.5 <log(E/eV) <19.5 and zenith angle lower than 45°. Only detectors with a total signal larger than 15 VEM are used. (b) Difference between raw and deconvolved risetimes as a function of the core distance for two energy bins. Same events as in (a).

3.5 A new approach for the risetime calculation

The risetime is a parameter defined for each individual FADC trace. Therefore, a surface detector with three PMTs working correctly has three different risetime values. The risetime value which characterizes this detector is calculated as the average of the risetimes of each PMT.

In this section we introduce a new procedure to calculate the risetimes of individual PMTs which differs from the procedure implemented in the $\overline{\mathrm{Off}\underline{\mathrm{line}}}$. The new procedure is motivated by the presence of risetimes as low as 15 ns in the early studies done for this thesis, when it is known that FADC traces sample arrival times in bins of 25 ns. To understand the presence of these small values it is necessary to study the usual procedure implemented in the $\overline{\mathrm{Off}\underline{\mathrm{line}}}$ for the risetime calculation.

3.5.1 Usual procedure for the risetime calculation in $\overline{\mathrm{Off}\underline{\mathrm{line}}}$

As it was shown in figure 3.2a the risetime calculation should be quite straightforward taken into account its definition. Nevertheless several rules and assumptions must be established if we take into account that the surface detectors register the signal with a sampling of 40 MHz. This sampling implies that the registered signal is not continuous but binned in intervals of 25 ns.

We will summarize briefly the different steps carried out by the $\overline{\mathrm{Off}\underline{\mathrm{line}}}$ to obtain the risetime for a given FACS trace:

1. The total signal together with 50 % and 10 % of its value are calculated.

2. Find the time bins where 50 % and 10 % of the total signal are located.

3. $\overline{\mathrm{Off}\underline{\mathrm{line}}}$ supposes that signals increase as a linear function inside these two time bins.

4. $\overline{\mathrm{Off}\underline{\mathrm{line}}}$ calculates the linear functions which describe the behaviour of the signal inside each bin.

5. With the linear functions it is possible to relate the 50 % and the 10 % of the total signal to their corresponding times. These times are called t_{50} and t_{10}.

6. Once these times are calculated, the risetime is the difference of t_{50} and t_{10}.

The steps listed before are illustrated in figure 3.13a. With this procedure, t_{50} and t_{10} can take any value inside their corresponding time bins of 25 ns and, thus, the risetime value is continuous. This procedure is correct to calculate risetimes but it is not the most appropriate one, since it is necessary to suppose that the signal increases as a linear function inside the time bins. A method which takes into account our ignorance about the behaviour of the signal inside the time bins would be more appropriate.

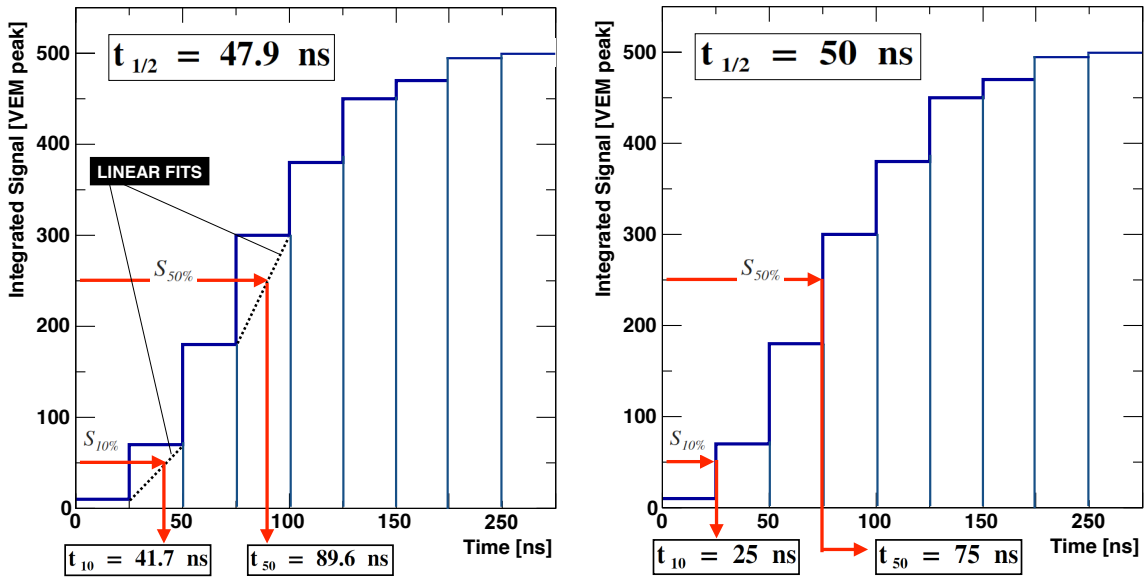
3.5.2 New approach for the risetime calculation

We propose a new approach for the risetime calculation which takes into account the lack of information derived from the sampling of 40 MHz. As in the previous section we will summarize the steps to calculate risetimes:

1. The total signal together with 50 % and 10 % of its value are calculated.
2. Find the time bins where 50 % and 10 % of the total signal are located.
3. **Nothing is supposed about the behavior of the signal inside the bins.**
4. We take for the values of t_{50} and t_{10} the lowest edge of the bins selected in the step 2.
5. Once these times are calculated, the risetime is the difference of t_{50} and t_{10}.

These steps are shown in figure 3.13b. The two first steps are the same in both methods. The only difference between the two approaches is that in the new one, nothing is supposed about the signals inside bins, what makes it more appropriate. The new procedure forces risetime values of each PMT to be a multiple of 25 ns. This increases the risetime uncertainty in a fixed quantity but the advantage is that we have a better control on the uncertainty.

All the risetimes used in this thesis are calculated with this new procedure.



(a) Old Procedure

(b) New Procedure.

Figura 3.13: Schematic view of the risetime calculation in $\overline{\text{Off}}\underline{\text{line}}$ (a) and the approach proposed in this thesis (b).

3.6 Correction for the azimuthal asymmetry of the risetime

In the previous chapter the risetime azimuthal asymmetry was briefly introduced to explain the observable (sec $\theta)_{\text{max}}$. In that chapter it was explained that the zenith angle where the asymmetry reaches a maximum can be used to perform composition studies. In this section we show that for our particular analysis, this asymmetry has to be taken into account and corrected for.

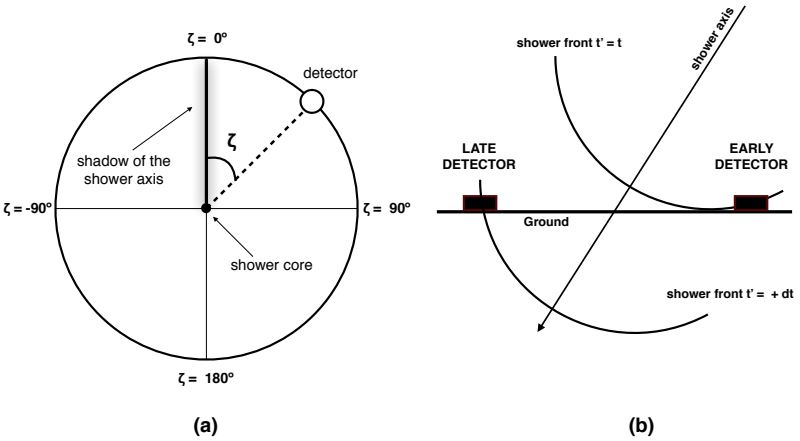


Figura 3.14: Schematic view of the azimuthal asymmetry for inclined showers. Left panel shows the definition of the azimuthal angle, ζ, around the shower core. Right panel shows the concept of early (| ζ | <$\pi/2$) and late (| ζ | >$\pi/2$) detectors.

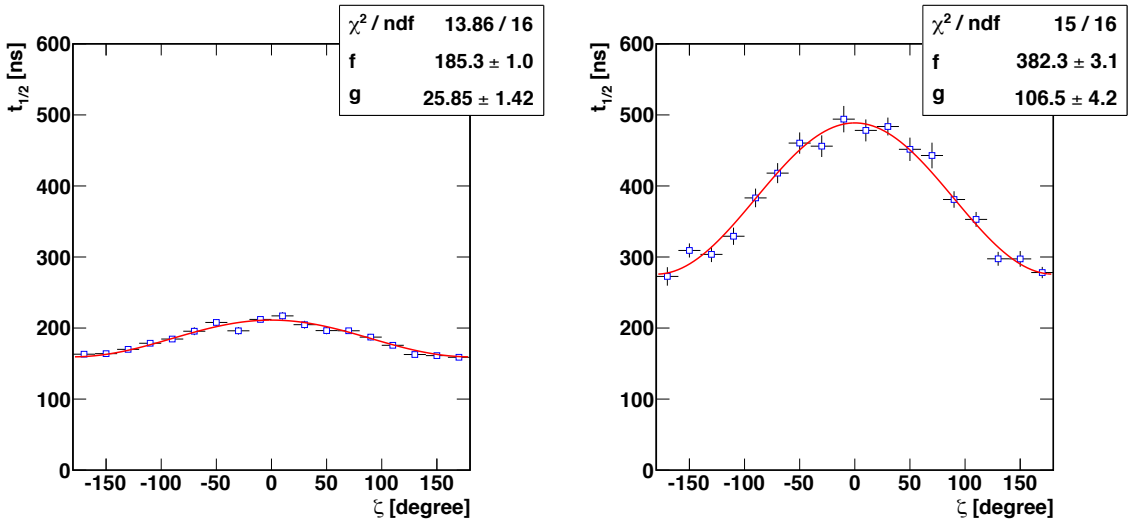
The azimuthal asymmetry in the risetime is the change in the risetime as a function of the azimuthal angle at the ground, ζ. This angle is characteristic of each triggered surface detector in a particular event. It is defined as the angle between the projection of the shower axis at the ground and the direction determined by the line which connects the individual surface detector and the shower core, see figure 3.14a. This asymmetry in risetimes is due to two effects. On the one hand, the particles that reach late detectors traverse longer atmospheric paths. Thus, we expect a bigger attenuation of the electromagnetic component compared to early detectors, (see figure 3.14b). On the other hand, there are contributions coming from geometrical effects. In this case the muonic component has an important role. The angular distribution, at production, of the muons that reach a particular surface detector is different and the late detectors record more muons emitted closer to the shower axis [79]. Geometrical effects only predominate for small zenith angles, for θ >30° the electromagnetic attenuation is the main contribution.

The dependence of the risetime on the azimuthal angle can be described using a cosine function:

$$t_{1/2}(\zeta) = f + g\cos\zeta \tag{3.2}$$

where f and g are free parameters determined from data. The parameter g is the

amplitude of the asymmetry and it depends on zenith angle and core distance. For the same bin of sec θ, the asymmetry is larger for larger core distances, as it is shown in figure 3.15. For the same core distance bin, the dependence of the asymmetry on zenith angle, in particular on ln(sec θ), follows in average a gaussian trend [79].



(a) 600 m <r <800m

(b) 1200 m <r <1400m

Figura 3.15: Average risetime behaviour as a function of azimuthal angle for two core distance bins. Events with energies in the range 18.9 <log(E/eV) <19.1 and zenith angles between 1.20 <sec θ <1.30 are used.

It was mentioned in the previous section that the Delta Method is based on the comparison of risetimes with the same core distance. This fact makes it crucial to keep the circular symmetry around the shower core. For this reason, any effect coming from the asymmetry has to be removed.

To correct the azimuthal asymmetry of a certain detector, whose azimuthal position is ζ, we have to calculate (using an asymmetry parameterization) the risetime that would have been measured by this detector at a particular azimuthal angle that is chosen as the reference value. In this case the chosen reference value is ζ = 90°. An example of how the azimuthal asymmetry correction works is shown in figure 3.16. The correction is implemented using the expression:

$$t_{1/2}^{corrected} = t_{1/2}^{measured} - g(r,\theta\,)\cos\zeta \tag{3.3}$$

Given that we use ζ = 90° as a reference value, only the parameterization of g is required. To apply the asymmetry correction, g must be parameterized as a function of the variables on which it depends. For our case the core distance and the zenith angle.

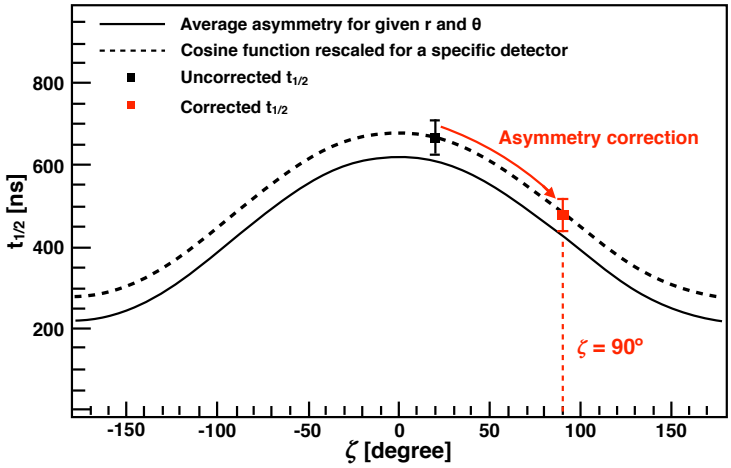


Figura 3.16: A schematic view of how the risetme asymmetry correction works.

3.6.1 Azimuthal asymmetry correction for data of the 1500 m array

The Pierre Auger Collaboration has two parameterizations to describe the parameter g, see [139] and [140]. The main difference between them is the dependence with the core distance. In [139] a quadratic function with two free parameters is used:

$$g(r,\theta\,) = A(\theta) + B(\theta)r^2 \tag{3.4}$$

whilst in [140], although the quadratic function is kept, only a free parameter is used:

$$g(r,\theta\,) = m(\theta)r^2 \tag{3.5}$$

As a part of the work presented in this thesis a comparison of both parameterizations has been made to establish which one works better. Figure 3.17 shows the effect of the correction, for the most vertical events used in this thesis (sec θ <1.10), with the two discussed parameterizations. For this bin of sec θ the correction given in [140] works better. The value of g after the correction is closer to zero for the parameterization given in [140]. This means that on average, the asymmetry has been removed in a larger proportion of events. In the rest of the sec θ bins the effect of both corrections is similar.

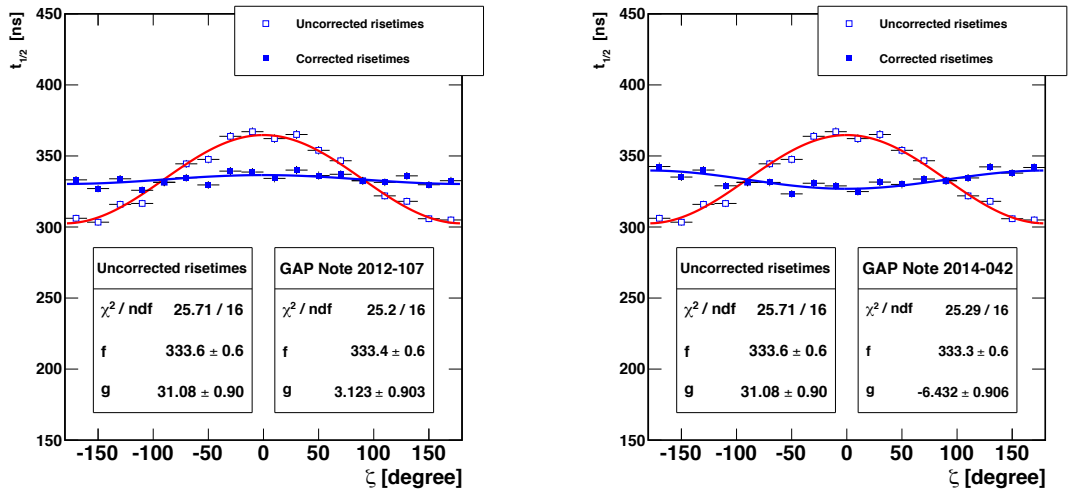
The parameterization used to describe the m parameter in [140] is:

$$m = (a\sec\theta + b\sec^3\theta + c)\sqrt{\sec\theta - 1} \tag{3.6}$$

where:

$$\begin{aligned} a &= (-3{,}9 \pm 2{,}3) \times 10^{-5} \text{ ns m}^{-2} \\ b &= (-1{,}9 \pm 0{,}4) \times 10^{-5} \text{ ns m}^{-2} \\ c &= (2{,}0 \pm 0{,}2) \times 10^{-4} \text{ ns m}^{-2} \end{aligned} \tag{3.7}$$

All risetimes used in this work are asymmetry-corrected. For the case of the 1500 m array, based on the improvement observed in figure 3.17 we have decided to use the



(a) Asymmetry correction given in [140]

(b) Asymmetry correction given in [139]

Figura 3.17: Effect of the risetime asymmetry correction for the two parameterizations discussed in the text. 18.5 <log(E/eV) <19.5 and 1.00 <sec θ <1.10.

parameterization given in [140]. The case of the 750 m array will be discussed in the next subsection.

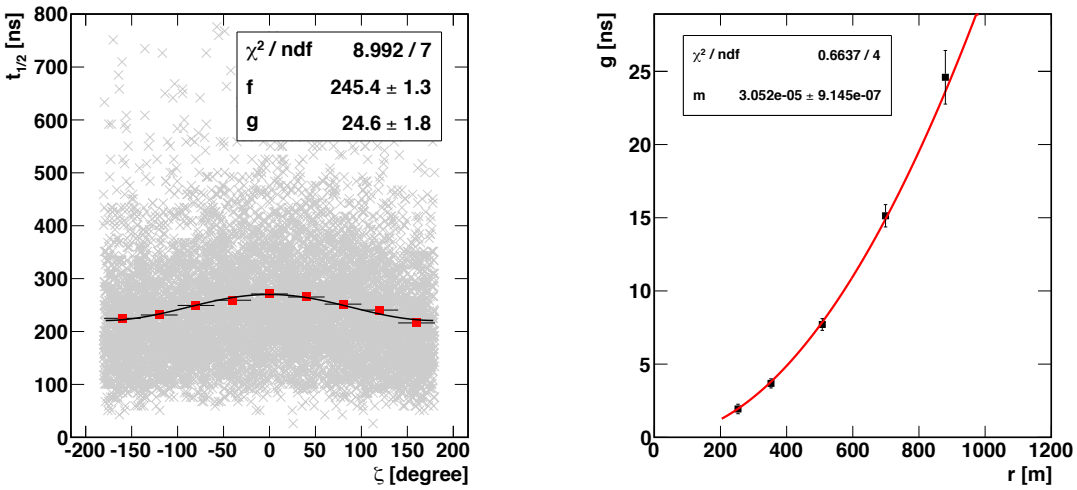
3.6.2 Azimuthal asymmetry correction for data of the 750 m array

For risetimes measured with surface detectors from the 750 m array the two parameterizations mentioned in the previous subsection can not be used. They do not describe properly the risetime dependences for data of the 750 m array. This fact is a consequence of the energy and distance ranges measured by each array. The dependences of risetimes and signals with the core distance and zenith angle are different.

At the time of this work the Pierre Auger Collaboration did not have a parameterization to correct the azimuthal asymmetry in risetimes measured with the 750 m array. Therefore, it was necessary to obtain a new one. The parameterization that we have obtained follows the same structure as the parameterization given in [140]. We use events with energies between 17.5 <log(E/eV) <18.5 and zenith angles in the range (1.00 <sec θ <1.42). Only surface detectors whose total signal is larger than 5 VEM are selected. The surface detectors whose low-gain channel is saturated are discarded.

Data have been divided into six bins of zenith angles and then subdivided into five distance bins from 200 m to 1000 m. For each core distance bin, risetimes are fitted as a function of the azimuthal angle using the equation 3.2 (see figure 3.18a) and we obtain a value for the parameter g. After that, the g values are fitted as a function of the core distance with equation 3.5 (figure 3.18b). In this way we obtain a value of the parameter m for each one of the six zenith angle bins. These values are plotted in figure

3.19a as a function of the sec θ and they are fitted with the equation 3.6. Finally, once we have obtained a parameterization for the parameter m, the azimuthal asymmetry in the risetimes can be corrected using equation 3.3.



(a) risetimes vs. azimuth angle (800 - 1000 m).

(b) g parameter vs. distance.

Figura 3.18: Left panel shows the azimuthal asymmetry in a particular core distance bin for data of the 750 m array (1.07 <sec θ <1.14). The right panel shows the value of g as a function of the core distance for the same sec θ bin.

To parameterize m as a function of sec θ for data of the 750 m array we use equation 3.6. The parameters a, b and c now take the following values:

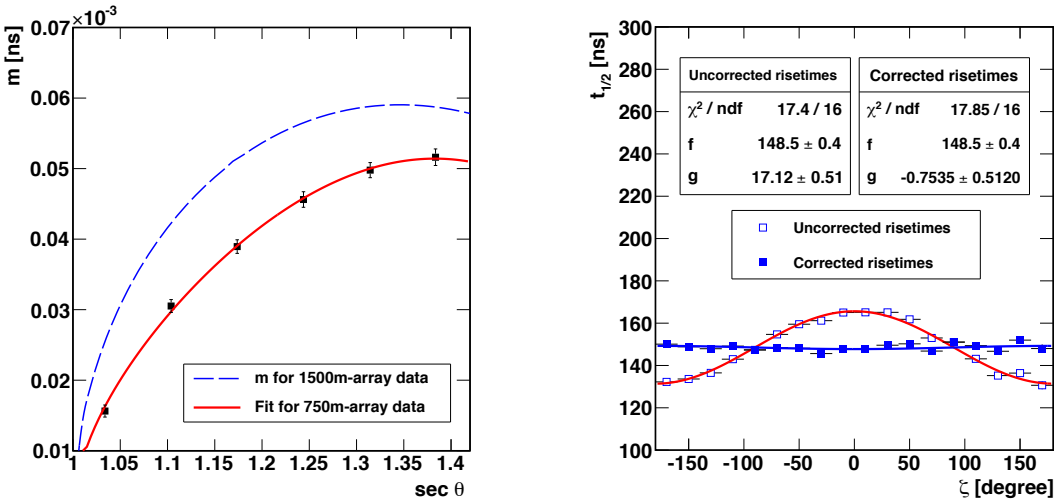
$$\begin{aligned} a &= (2{,}9 \pm 1{,}1) \times 10^{-4} \text{ ns m}^{-2} \\ b &= (-7 \pm 3) \times 10^{-5} \text{ ns m}^{-2} \\ c &= (-13 \pm 9) \times 10^{-5} \text{ ns m}^{-2} \end{aligned} \tag{3.8}$$

After this study we have seen that the azimuthal asymmetry is larger in data measured in the 1500 m array. In figure 3.19a the two parameterizations of the parameter m are compared. The values of m are larger for data of the 1500 m array. This shows that the asymmetry is larger for this data set. This happens because with the 1500 m array the showers are larger so the attenuation across the array is larger than with the 750 m array.

Figure 3.19b shows, for a particular zenith angle bin, that the asymmetry correction presented in this section works correctly for risetimes measured with the 750 m array.

3.7 Risetime uncertainty

The final step for the complete determination of the risetime is the calculation of its experimental uncertainty. The uncertainty in the risetime mainly arises from fluctuations due



(a) m parameter vs. sec θ.

(b) Effect of the asymmetry correction.

Figura 3.19: The left panel shows the parameterization of m as a function of the sec θ for data of the 750 m array (red line). We have also included the parameterization of m used for data of the 1500 m array (blue line). In the right panel the effect of the asymmetry correction for data of 750 m array is shown (1.20 <sec θ <1.30).

to the sampling of the showers. This is a consequence of having surface detectors with a finite size. Other contributions to the uncertainty come from detector effects such as the electronics, the digitalization or the reconstruction procedure.

It is inadvisable to use simulations to estimate the uncertainty in the data, since the extrapolation of models from much lower energies and the technique of thinning, introduce complex uncertainties which cannot be accurately quantified. These disadvantages make necessary the use of an empirical approach to determine the risetime uncertainty.

The standard method to produce an empirical determination of the measurement uncertainty is to make repeated measurements of the observable in a controlled manner. However, the study of air showers does not afford controlled studies. For each event, a different number of surface detectors are involved, and each one of them has different core distances, signals and azimuthal angles. In addition, individual surface detectors from different events cannot be directly compared, even if the energy, the zenith, the core distance, the signal and the azimuth are the same. This is due to the shower to shower fluctuations. The development of each air shower is unique, which would result in fluctuations in the risetimes. Therefore, only risetimes from the same event can be directly compared.

To allow the comparison of risetimes from the same air shower two approaches have been proposed. One is based in the use of *twin detectors* and the other one in the use of *pair of detectors*. We will present in this section a method which combines the use of twins and pairs to obtain a parameterization of the risetime uncertainty. The parameterization

given here will be applied for all the risetimes used in this thesis.

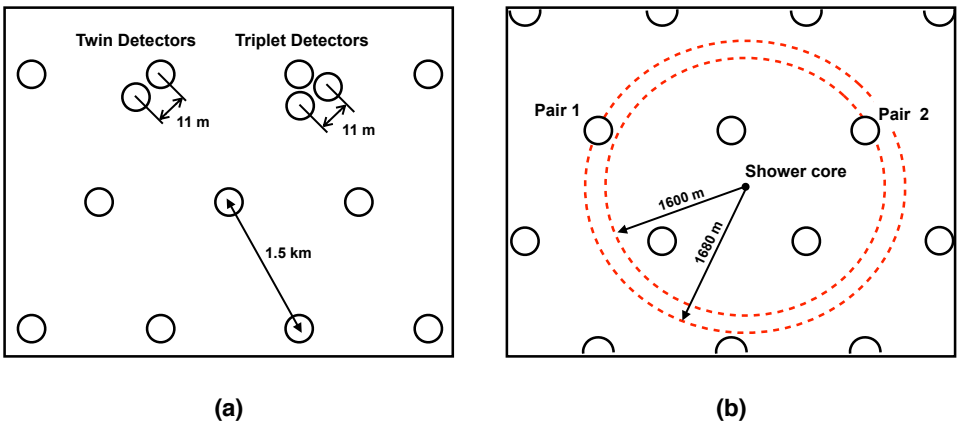


Figura 3.20: (a) Diagram of a couple of twin detectors and a triplet in the 1500 m array. (b) Diagram of a pair of detectors in the 1500 m array.

3.7.1 Uncertainty with Twins

A twin is a set of two surface detectors which are deployed together on the same hexagonal grid point, with a separation of only 11 m, see figure 3.20a. With such a small separation we can consider that both detectors measure the same spot at the ground and therefore, the same amount of electromagnetic and muonic component. Besides, as both detectors are in the same event, shower-to-shower fluctuations can be safely disregarded. At this moment the Surface Detector of the Pierre Auger Observatory has 14 sets of twin detectors and 7 sets of triplets. Triplets consists of a set of three detectors forming a triangular grid and separated each one 11 m. Hereafter twins will refer to both, the triplet and twin detectors. Most of the twin detectors are located in the area of the 750 m array, in positions of the hexagonal grid of the 1500 m array.

For a precise determination of the risetime uncertainty an enormous amount of twins is necessary. However with the small number of twin detectors in the array this is difficult. As twin detectors are located in the 750 m array, two approaches would be possible. The first one would be to calculate the uncertainty using data reconstructed with the 1500 m array and the second one would be with data reconstructed with the 750 m array. The best of these two options will be the one providing a larger statistics. Table 3.1 shows the number of twins found after applying basic quality cuts in both arrays. The SD is not fully efficient below 3 EeV, however here we have used a lower energy cut of 1 EeV to increase the number of twins of our analysis. A trigger efficiency of $\sim 75\,\%$ is achieved for events with energies greater than 1 EeV. After applying the cuts that we require to calculate a parameterization of the uncertainty we find 29 832 twins in the 1500 m array and 28 466 twins in the 750 m array (see table 3.2).

1500 m array	**750 m array**
01/2004 - 12/2012	01/2008 - 12/2013
E >1 EeV	E >0.3 EeV
1.00 < sec θ < 2.00	1.00 < sec θ < 2.00
Events 6T5	Events 6T5
No Bad Periods	No Bad Periods
No LG saturated detectors	No LG saturated detectors
41 934 twins	**41 100 twins**

Cuadro 3.1: Quality cuts applied to select twin detectors in both arrays.

	1500 m array		**750 m array**	
Cuts	Number of twins	Efficiency	Number of twins	Efficiency
Pre-Twin Selection	41 934	1.00	41 100	1.00
5 VEM <S <800 VEM	35 704	0.85	34 461	0.84
r <2000 m	35 620	0.84	34 459	0.83
$\lvert S_i$ - $S_{mean}\rvert$ <0.25S_{mean}	**29 832**	0.71	**28 466**	0.69

Cuadro 3.2: Cuts to obtain the final data set of twins. The last one requires that the signal of each twin must differ by no more than $\pm 25\,\%$ from the average of the two members. It is used to remove outliers.

The statistics found is quite similar for the two data sets and it is larger than the one found in other works [138]. However this is not enough to calculate a precise parameterization of the risetime uncertainty. Since we have to use all the variables that the uncertainty depends on (r,S,θ) and it requires a huge statistics to split data in a large number of bins.

To avoid this problem we have proposed a new approach: Merge the two data sets shown in table 3.2. This means to build a larger data set putting together twins reconstructed with the 1500 m array and with the 750 m array. This is possible because the surface detectors used in the 1500 m array and the 750 m array are the same. The only thing that we have to check is if both reconstructions are compatible. That is events seen in both arrays should have the same reconstructed parameters. We have shown that this only happens if we require in the 750 m array 6T5 events. Without this cut border effects introduce mis-reconstructed events in the data set of the 750 m array. After merging the two data sets we have a final sample with 58 298 twins to obtain a risetime uncertainty parameterization.

Once we have a large enough set of twins, the risetime uncertainty can be obtained from the mean of a half-normal distribution, resulting in the expression [141]:

$$\sigma_{1/2} = \frac{\sqrt{\pi}}{2} < |t^1_{1/2} - t^2_{1/2}| > \tag{3.9}$$

where $\sigma_{1/2}$ is the measurement risetime uncertainty and $t^1_{1/2}$ and $t^2_{1/2}$ are the risetimes

measured for each one of the members of a twin. Note that as the members of a twin are located in the same position the azimuthal asymmetry correction need not be applied.

The selected data have been divided into equidistant bins of sec θ (0.1 wide) and then subdivided into six bins of distance. Finally a further subdivision into ten bins of signal up to 100 VEM, plus a final bin with all signals larger than 100 VEM has been made. We have required also for the parameterization that in any bin we must have at least 10 twins to avoid large statistical fluctuations. The left panel of figure 3.21 shows the average uncertainties as a function of the total signal for the most vertical events (1.00 < sec θ < 1.10).

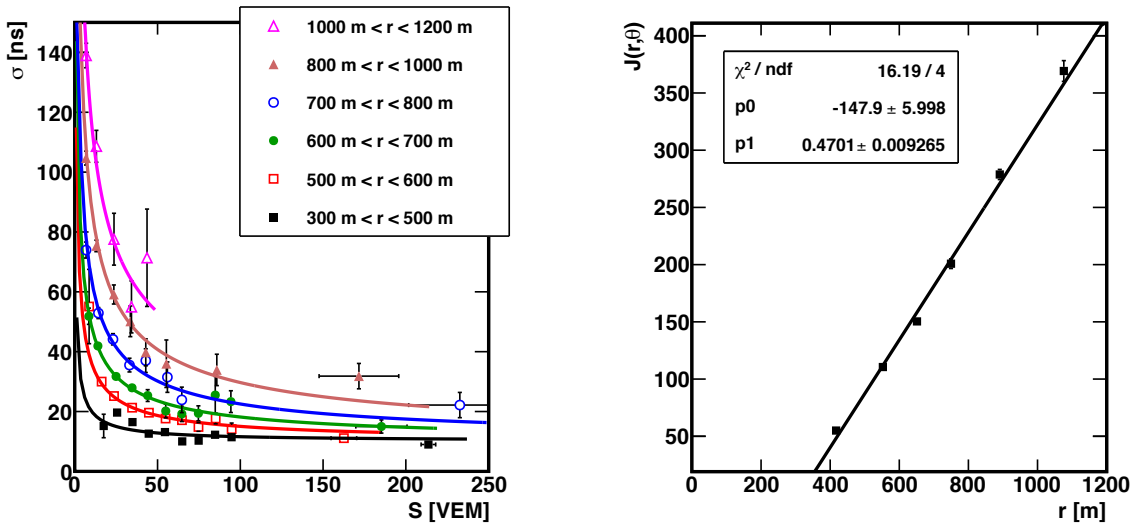


Figura 3.21: Left panel shows the uncertainty obtained with the twins as a function of the total signal for the most vertical events (1.00 <sec θ <1.10). Each line represents the fit obtained for each one of the six core distance bins used. Each point is the average of at least 10 twins. Right panel shows the behavior of J as a function of the distance for the same sec θ bin.

The uncertainty behaves as $1/\sqrt{S}$ for all zenith angles bins and distance ranges and it reflects the increase of Poissonian fluctuations due to the decreasing number of particles in the surface detectors. To describe the behaviour of the uncertainty as a function of the square root of the signal we use the equation:

$$\sigma_{1/2} = \sqrt{\left(\frac{J(r,\theta)}{\sqrt{S}}\right)^2 + \left(\sqrt{2}\frac{25}{\sqrt{12}}\right)^2} \tag{3.10}$$

This equation has two independent components. One empirical part due to risetime fluctuations between different detectors (first term of the square root) and one component due to the binning of the signals in time bins of 25 ns (second term of the square root).

The parameter J behaves linearly as a function of the core distance for each zenith angle bin, see the right panel of figure 3.21. The trend of the parameter J as a function of the distance is fitted using a polynomial function:

$$J(r,\theta\,) = p_0(\theta) + p_1(\theta)r \tag{3.11}$$

The parameters p_0 and p_1 also behave linearly as a function of the sec θ. They are shown in figure 3.22. The parameterization used to describe these two parameters is:

$$\begin{aligned} p_0 &= (-340 \pm 30) + (186 \pm 21)\sec\theta \\ p_1 &= (0{,}94 \pm 0{,}03) + (-0{,}44 \pm 0{,}01)\sec\theta \end{aligned} \tag{3.12}$$

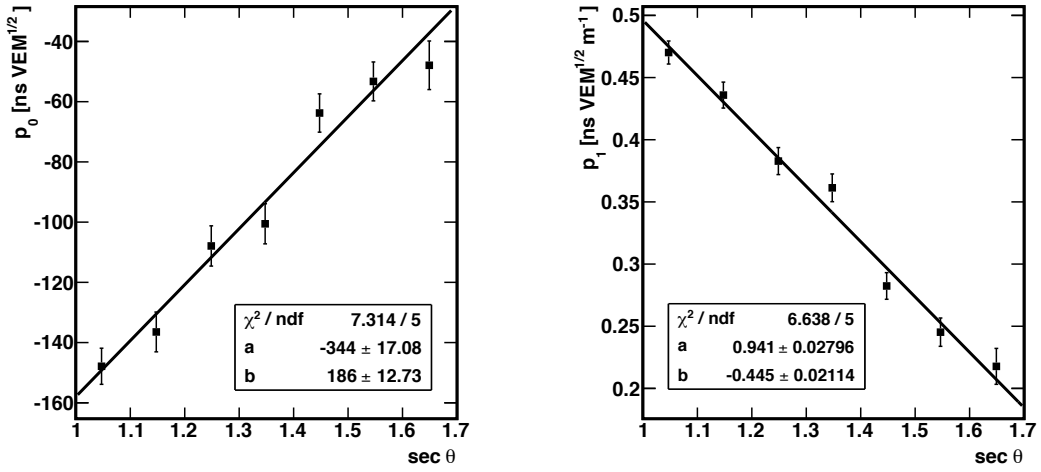


Figura 3.22: Behaviour of the parameters p_0 and p_1 as a function of the secθ for twin detectors.

3.7.2 Uncertainty with Pairs

A pair of detectors is defined as any two detectors in the same shower whose difference in distance to the shower core ($|r_2 - r_1|$) is less than 100 m, irrespective of azimuthal angle, see figure 3.20b. Working with pairs has one advantage. They increase significantly the statistics of duplicated measurements. Nevertheless they have three disadvantages. The first one is that each one of the members of the pair has a different azimuthal angle, what implies that risetimes have to be corrected of the azimuthal asymmetry before being compared. The second one is that there are not pairs below 600 m due to geometrical reasons. And the last and most important one is that a new correction has to be done to compensate the difference in distance between the two members of the pair if we want to compare the two risetimes values.

The correction to compensate the difference in distance is applied as follows. For each event where one of our selected pairs is located we fit the risetimes as a function of the core distance using all the detectors which form this event. To do the fit we use the function given in [137] to describe the average behaviour of the risetime as a function of the core distance:

$$t_{1/2} = 40 + \sqrt{A^2 + Br^2} - A \tag{3.13}$$

The risetime value of the pair located closer to the core is fixed and the risetime value of the detector located farther from the core is corrected. With the fit mentioned previously (unique for each event) we can calculate the difference in risetime predicted by this function when two detectors are separated by a distance of $r_2 - r_1$. Once this difference is obtained we correct the second risetime subtracting this value. Before applying this correction the average difference between the risetimes of the two members of a pair was (14.750 $\pm$ 0.002) ns. After the correction the average of this difference is (0.140 $\pm$ 0.002) ns, closer to zero as expected.

To select pairs of detectors only data of the 1500 m array are used. The basic quality cuts used are the same as the ones used to select twins with the 1500 m array, see table 3.3. The only difference is that in the case of the study with pairs, due to the huge statistics, it is not necessary to lower the cut on the energy up to 1 EeV. After applying the final cuts over the detectors we find 77 023 pairs (table 3.4). The statistics have been increased in more than 30 $\%$ compared to the study with twins.

1500 m - array
01/2004 - 12/2012
E >3 EeV
1.00 < sec θ < 2.00
Events 6T5
No Bad Periods
No LG saturated detectors
117 001 pairs

Cuadro 3.3: Quality cuts applied to select pair detectors.

Cuts	Number of pairs	Efficiency
Pre-Twin Selection	117 001	1.00
5 VEM <S <800 VEM	98 040	0.83
r <2000 m	94 197	0.81
$\lvert S_i$ - $S_{mean}\rvert$ <0.25S_{mean}	**77 023**	0.66

Cuadro 3.4: Cuts on detectors to obtain the final data set of pairs.

The procedure applied to obtain the parameterization and the equations used for the fits are the same than the ones explained in the previous subsection for twins. The

behaviour of the average uncertainties as a function of the total signal for the most vertical events (1.0 <sec θ <1.10) is shown in the left panel of figure 3.23. This figure is similar to the one shown in the left panel of figure 3.21. The behaviour of the parameter J for the mentioned bin of sec θ is shown in the right panel of figure 3.23. The trend of the parameter J as a function of the distance is fitted using the polynomial function shown in equation 3.11 for each bin of the sec θ. The parameters p_0 and p_1, as in the case of the twins, behave linearly as a function of the sec θ. They are shown in figure 3.24.

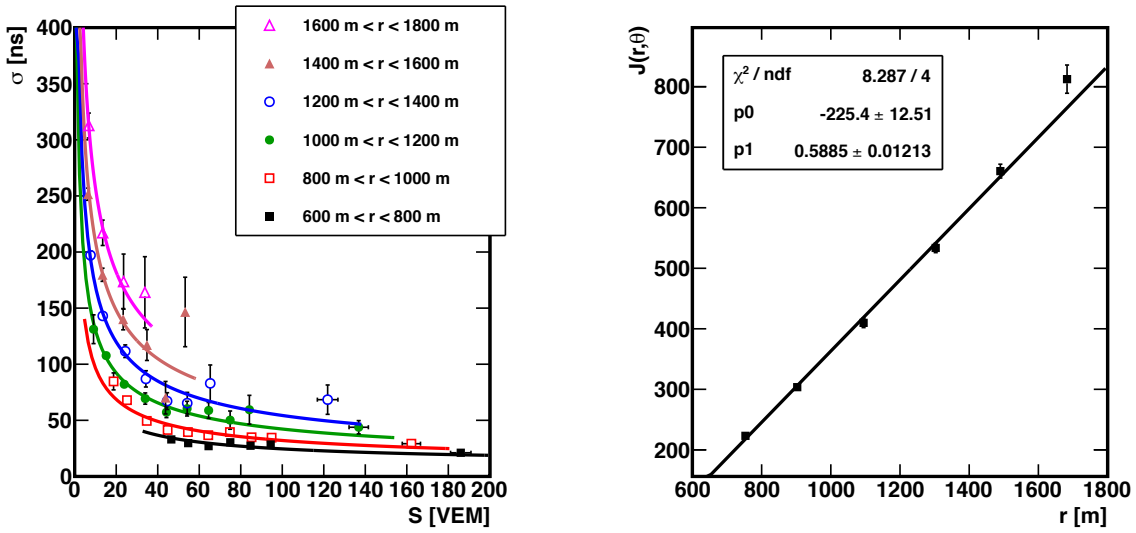


Figura 3.23: Left panel shows the uncertainty obtained with pairs as a function of the total signal for the most vertical events (1.0 <sec θ <1.10). Each line represents the fit obtained for each one of the six core distance bins used. Each point is the average of at least 10 twins. Right panel shows the behavior of J as a function of the distance for the same sec θ bin.

The parameterization used is:

$$\begin{aligned} p_0 &= (-447 \pm 30) + (224 \pm 22)\sec\theta \\ p_1 &= (1{,}12 \pm 0{,}03) + (-0{,}51 \pm 0{,}02)\sec\theta \end{aligned} \tag{3.14}$$

The values of p_0 and p_1 obtained in this case are different from the ones obtained for twin detectors due to the different distance ranges used in each case.

3.7.3 Combination of both results

In the previous two sections we have obtained two different parameterizations for the risetime uncertainty. Due to geometrical reasons the parameterization obtained for twins is only known for core distances between 300 m and 1000 m. Outside this range we do not have enough information to calculate the uncertainty and we should use an extrapolation

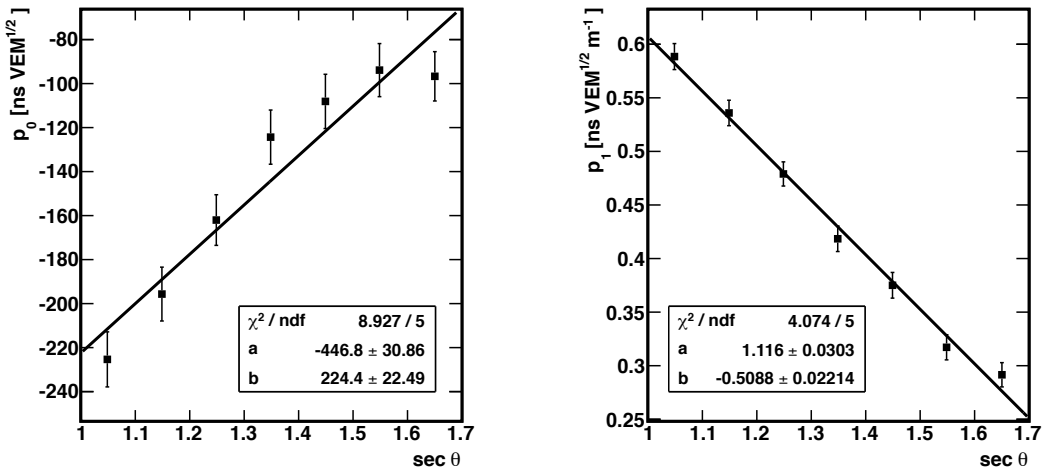


Figura 3.24: Behaviour of the parameters p_0 and p_1 as a function of the sec θ for pair detectors.

of our parameterization. The same happens for the parameterization obtained for pairs. In this case the parameterization is only known for distances between 600 m and 1800 m, and outside this range there is not information related to pairs of detectors.

In the left panel of figure 3.25 the two uncertainty parameterizations are shown, using the extrapolation where it is necessary. In the right panel the difference between both parameterizations divided by the mean value is drawn as a function of the core distance for an easier comparison. Close to the core both parameterizations are similar. This is due to the lower limit of the uncertainty imposed for the time binning of the signals (second term in equation 3.10) which is the same for both parameterizations. Farther from the core, above 1200 m, the difference between both results becomes more important, being larger than 10 ns. This suggests that the extrapolation of the twin parameterization is not good enough. The same happens close to the core. As it is shown in the right panel of figure 3.25, below 600 m the difference reaches negative values. This indicates that the extrapolation of the pair parameterization overestimates the uncertainty values.

The ranges of each parameterization are different. While one covers distances close to the core, the other describe the farther ones. This fact complicates the choice of which should be used in our analysis. For this reason, to avoid the extrapolation of any parameterization (which is always undesirable) and to obtain a parameterization covering the wider possible range of distances we have proposed a new parameterization which combines both. For lower distances we will use the parameterization given by twins and for larger distances the parameterization given by pairs. We have chosen 650 m as a point to divide the core distance ranges because this is the point where both parameterizations are equal, see right panel of figure 3.25.

After taking into account all those issues the final expression for the risetime uncer-

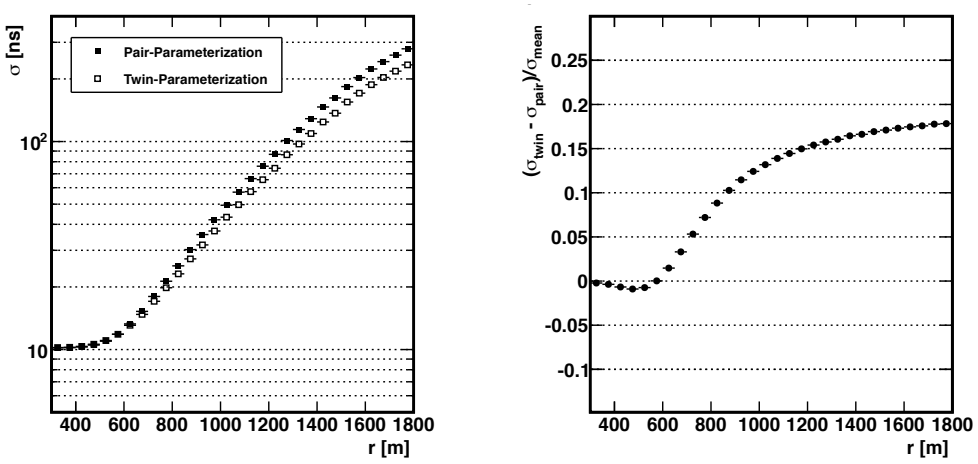


Figura 3.25: Left panel: Comparison of the risetime uncertainty parameterization as a function of the core distance. Events with energies in the range 19.0 <log(E/eV) <19.1 and zenith angle between 1.00 <sec θ<1.42 are used. Right panel: Average difference of the two uncertainty parameterizations as a function of the core distance. In both graphs we use the same sample of events.

tainty is:

$$\sigma_{1/2} = \sqrt{\left(\frac{J(r,\theta\,)}{\sqrt{S}}\right)^2 + \left(\sqrt{2}\frac{25}{\sqrt{12}}\right)^2} \tag{3.15}$$

where J has the expression:

$$J(r,\theta\,) = p_0(\theta) + p_1(\theta)r \tag{3.16}$$

and the parameters p_0 and p_1 are given by:

$$p_0(\theta) = \begin{cases} (-340 \pm 30) + (186 \pm 21)\sec\theta & \text{if r} \leq \text{650 m} \\ (-447 \pm 30) + (224 \pm 22)\sec\theta, & \text{if r >650 m} \end{cases} \tag{3.17}$$

$$p_1(\theta) = \begin{cases} (0{,}94 \pm 0{,}03) + (-0{,}44 \pm 0{,}01)\sec\theta & \text{if r} \leq \text{650 m} \\ p_1 = (1{,}12 \pm 0{,}03) + (-0{,}51 \pm 0{,}02)\sec\theta & \text{if r >650 m} \end{cases} \tag{3.18}$$

This is the risetime uncertainty parameterization which will be use for all the risetimes which appear in this thesis.

3.8 Outliers

In this section we will show briefly several surface detectors with problems in the electronics which have been found in our data set. They are not listed as *Bad PMTs* in the

standard list of Pierre Auger Collaboration and due to that they are not automatically removed by the $\overline{\text{Off}}\underline{\text{line}}$. The events which are reconstructed using one of these surface detectors have to be removed due to a possible mis-reconstruction.

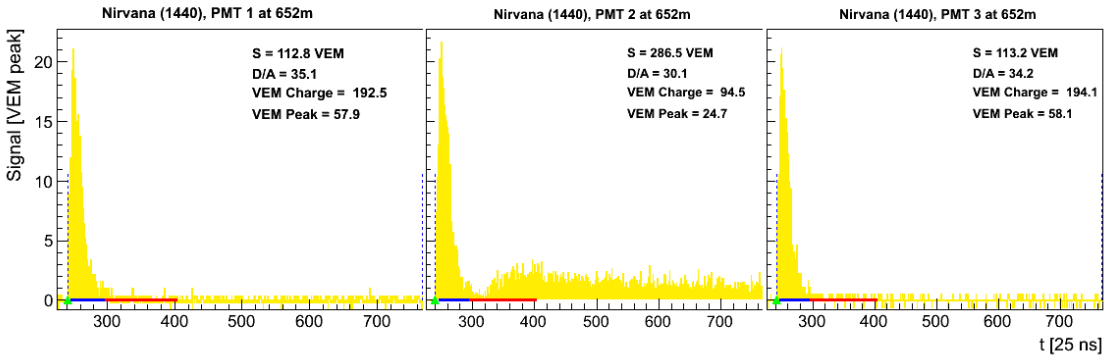


Figura 3.26: Example of one these problematic surface detectors. In this case the PMT2 registers an anomalous signal after the signal attributed to the air shower. Detector 1440 in the event 4336572.

The problem which we have found can be seen in figure 3.26. The second PMT registers an anomalous signal after the signal attributed to the air shower. Due to this large amount of signal in the last part of the trace, the second PMT has a total signal $\sim$170 VEM larger than the other two PMTs. The effect on the risetime is larger. The PMT1 and PMT3 have risetime values of 200 ns while the PMT2 has a risetime of 2600 ns. We have found in our data set only three surface detectors with this problem: the detectors 1440, 1228 and 1387. The correct operation of each one of these detectors has been studied along the years and the problem only appears for a brief period time. This indicates a faulty operation due to a breakdown until the repair done by the local staff of the Observatory.

Bad PMTs					
PMT2 in detector 1440		**PMT3 in detector 1228**		**PMT3 in detector 1387**	
Event	Date	Event	Date	Event	Date
4138094	01-11-2007	4926033	10-05-2008	10144166	30-08-2010
4202367	13-11-2007	5014966	02-06-2008	10148896	31-08-2010
4212175	15-11-2007	6311703	25-08-2008	10220886	10-09-2010
4336572	16-12-2007	6408298	08-09-2008	-	-
4355664	21-12-2007	-	-	-	-

Cuadro 3.5: List of the events in our data set with one of these problematic PMTs.

The events where we have found PMTs with this problem are listed in table 3.5. All of them are removed from the data set used in the analysis presented in this thesis.

4
The $\langle\Delta\rangle$ Method

The risetime carries information about the development of the UHECR showers, as we showed in section 3.1. However, this observable can not be used for mass composition studies in event-by-event analyses. The risetime is an observable obtained for each one of the triggered detectors of a particular event and therefore it does not characterise the whole event. Individual risetimes can only be used for mass studies in terms of averages, studying the risetime distribution of all the detectors that belong to a particular energy bin.

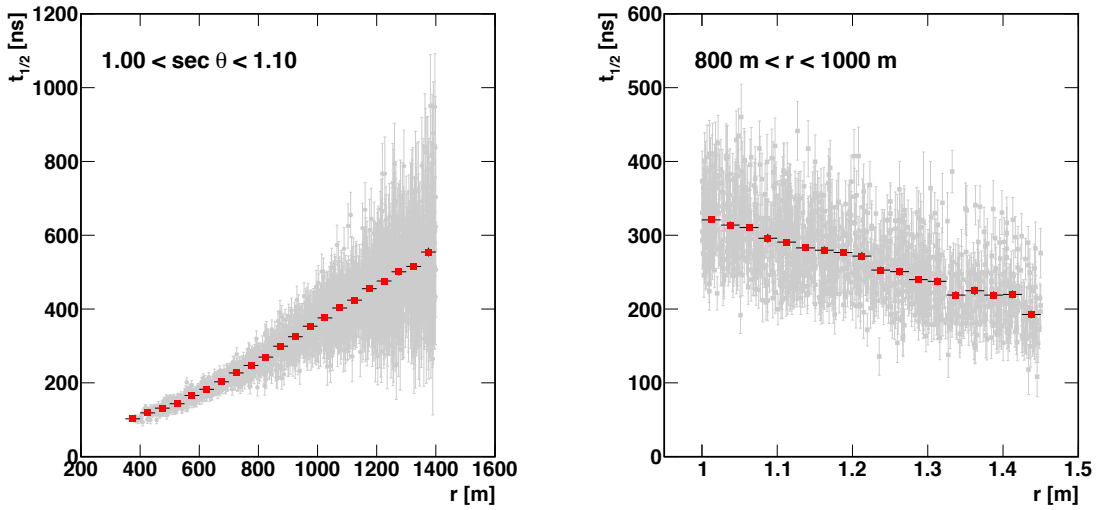
Although different studies based on averages have proven to be a powerful tool for mass composition inferences [79, 142], analyses with an event-by-event base are more convenient because the event-by-event observables can be used a posteriori for other physics goals, as the searches of UHECR sources through correlations with AGNs or the searches of ultra high energy photons [49].

In this chapter we introduce a parameter which combines the risetime information of individual surface detectors in a single value which characterises the whole event. This parameter is called *parameter* $\langle\Delta\rangle$ or simply $\langle\Delta\rangle$. As the risetime is sensitive to mass composition, $\langle\Delta\rangle$ will be sensitive too. For this reason we expect a correlation between $\langle\Delta\rangle$ and X_{max}. If this correlation is found, then $\langle\Delta\rangle$ can be calibrated with X_{max} using a small number of Golden Hybrid events and be used as a surrogate for X_{max} in the SD-only events.

4.1 Risetime dependences

In this section we study the dependence of the risetime with the zenith angle, θ, the core distance, r, and the energy, E. A deep understanding of these behaviours is crucial to combine in a proper way all the risetime information into a single parameter.

Figure 4.1a shows the behaviour of the risetime as a function of core distance for a



(a) $t_{1/2}$ vs r.

(b) $t_{1/2}$ vs sec θ.

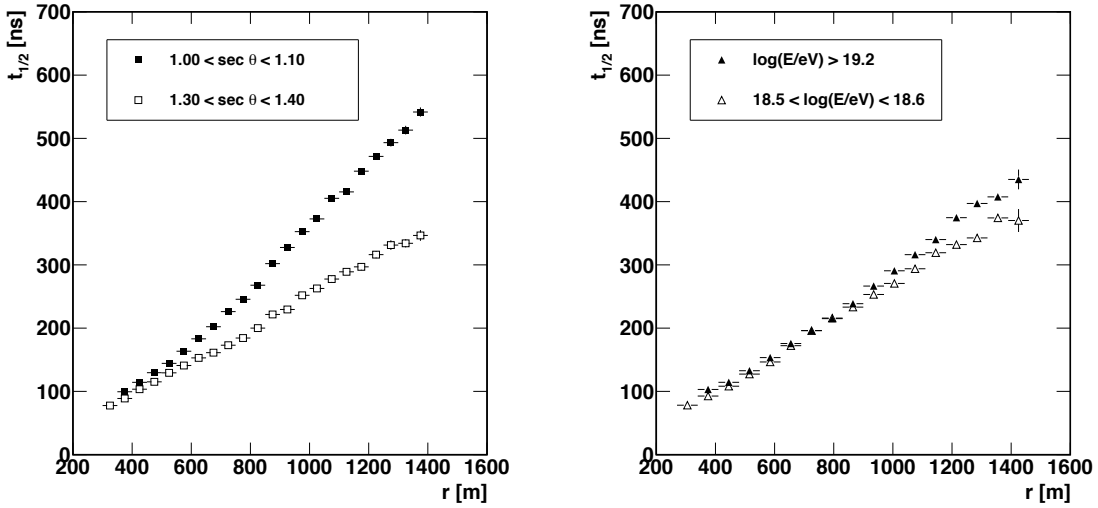
Figura 4.1: Risetime behaviour as a function of core distance (a) and the zenith angle (b). In both panels the scatter distribution and the profile are plotted together. Events with energies in the range 19.1 <log(E/eV) <19.2 are used.

particular sec θ bin. We observe that the risetime increases with the core distance, as expected, since farther from the core the spread of arrival times of particles is larger. The behaviour of the risetime with sec θ is shown in figure 4.1b. In this case, for a fixed core distance, the risetime decreases with the sec θ. This happens because the more inclined the shower is, the larger the attenuation of the electromagnetic component in the atmosphere. Thus, the signals in the surface detectors are composed mainly by muons, which have a smaller spread of arrival times.

Figure 4.2a shows the average risetime as a function of the core distance for two sec θ bins. In this figure we see the trend mentioned before: given the same energy, inclined showers have smaller risetimes than the vertical ones. The risetime behaviour as a function of core distance for two different energy bins can be seen in figure 4.2b. We observe that showers with more energy have on average larger risetimes. This is the expected trend since, provided the composition is the same, events with larger energies develop deeper in the atmosphere and therefore the presence of a larger electromagnetic component gives rise to bigger risetimes.

4.2 The parameter $\langle\Delta\rangle$

$\langle\Delta\rangle$ is the parameter designed to combine the risetime values of individual surface detectors in only one observable which characterises the whole event. The aim of this parameter is to produce a single value which evaluates how fast or slow are the risetimes of



(a) $t_{1/2}$ for two different sec θ bins.

(b) $t_{1/2}$ for two different energy bins.

Figura 4.2: Left panel: Risetime behaviour as a function of core distance for two different sec θ bins. The energy takes any value in the range 19.0 <log(E/eV) <19.2. Right panel: Risetime behaviour as a function of core distance for two different energy ranges. Zenith angles belong to the range 1.20 <sec θ <1.30.

an event in relation to an average behaviour that we take as a reference. This reference will be dubbed hereafter as the benchmark.

The benchmark is a function which describes the average behaviour of the risetimes as a function of core distance and zenith angle for a chosen energy bin. The functional form of the benchmark is obtained using experimental data. The chosen energy bin selected to calculate the benchmark is arbitrary. This means that the choice of a particular energy bin does not change the trend of $\langle\Delta\rangle$ with the energy. This trend is the most important feature for mass composition analyses. Therefore the results stemming from this method do not depend on this choice.

The first step to define $\langle\Delta\rangle$ is to calculate individual Δ_i values for each surface detector which passes certain quality cuts. Δ_i is defined as:

$$\Delta_i = \frac{t_{1/2} - t_{1/2}^{bench}(r,\theta\,)}{\sigma_{1/2}} \tag{4.1}$$

where $t_{1/2}$ is the risetime measured by each detector, $t_{1/2}^{bench}(r,\theta\,)$ is the benchmark parameterization evaluated in the core distance and the zenith angle of each detector and $\sigma_{1/2}$ is the risetime uncertainty.

Once these values are determined, the parameter $\langle\Delta\rangle$ which characterizes an event

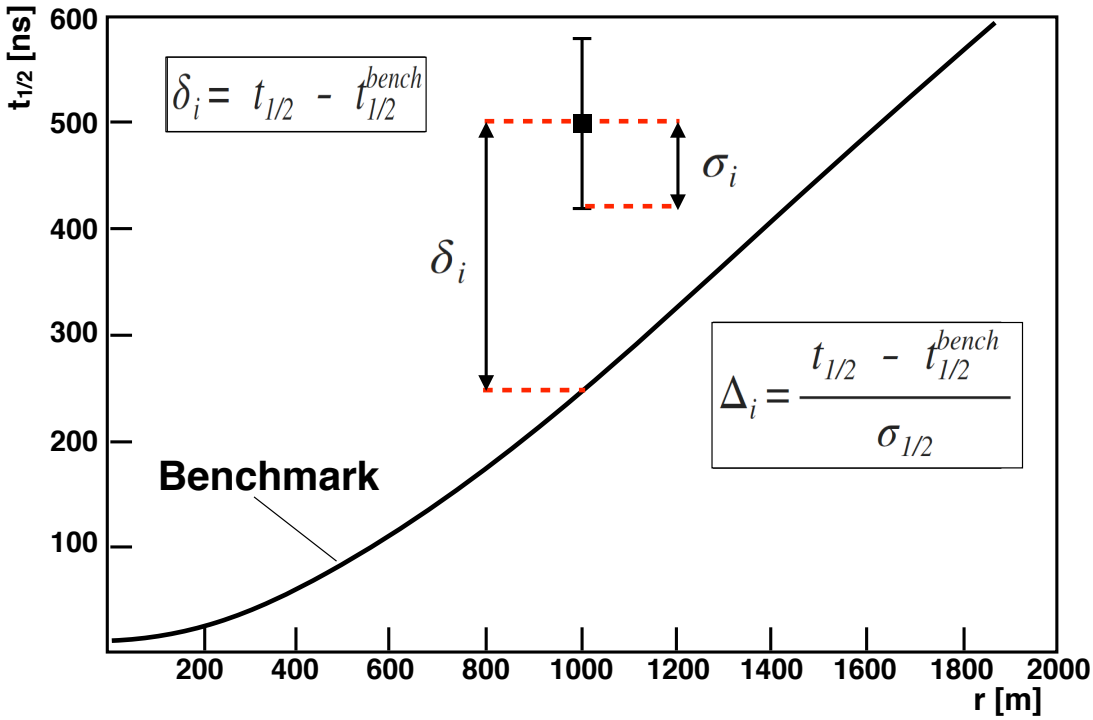


Figura 4.3: Schematic diagram showing the definition of Δ_i.

is obtained as the average of all the Δ_i values belonging to this event:

$$\langle\Delta\rangle = \frac{1}{N}\sum_{i=1}^{N}\Delta_i = \frac{1}{N}\sum_{i=1}^{N}\frac{t_{1/2} - t_{1/2}^{bench}}{\sigma_{1/2}} \tag{4.2}$$

being N the number of surface detectors used in the calculation of $\langle\Delta\rangle$.

Considering this definition, $\langle\Delta\rangle$ is the average deviation from the benchmark of the risetimes of an event, expressed in terms of the risetime uncertainty. This fact implies that $\langle\Delta\rangle$ is a dimensionless parameter. A schematic view of the Δ_i definition is shown in figure 4.3.

The advantage of this parameter against other parameters which also use risetimes, as $t_{1/2}(1000)$ [143] or Δ_{1000} [144], is that $\langle\Delta\rangle$ can be calculated with only one risetime measurement per event since its definition does not require a minimum number of detectors. This feature allow us to apply the $\langle\Delta\rangle$ method in events measured in the 1500 m array with energies as low as 3 EeV. Another important advantage of the use of $\langle\Delta\rangle$ is that it is not necessary to do a fit for each one of the studied events. This fact minimizes the systematic uncertainties in the analysis.

Taking into account the definition of $\langle\Delta\rangle$, if the benchmark describes sucessfully the

average risetime behaviour, the mean value of the $\langle\Delta\rangle$ distribution in the benchmark energy bin must be compatible with zero by definition. Following this logic, events with larger risetimes than the benchmark have a positive value of $\langle\Delta\rangle$, while events with smaller risetimes have a negative value. In terms of mass composition this means that deeper showers than the showers used for the benchmark calculation have positive values of $\langle\Delta\rangle$. Likewise showers which develop higher in the atmosphere have negative values.

The parameter Δ_i is defined so that it is, in statistical terms, a pull and therefore the distributions of Δ_i are pull distributions [145]. Due to that, the mean values of these distributions are centered on zero (only for the benchmark energy bin) and the standard deviations are compatible with 1. According to this statistical approach, the distributions of $\langle\Delta\rangle$ are the sum of several pull distributions taken into account the correlations between the different Δ_is and normalized by N, the number of detectors used to calculate $\langle\Delta\rangle$. The direct implication of this is that the standard deviation of the $\langle\Delta\rangle$ distributions is always lower than one if we build $\langle\Delta\rangle$ with more than one risetime measurement.

The final issue to complete the calculation of the parameter $\langle\Delta\rangle$ is to determine its uncertainty. The uncertainty of each Δ_i has contributions from two different sources: on the one hand the uncertainty of the risetime and on the other hand the uncertainty coming from the parameterizations used in the analysis. We use two different parameterizations, one to describe the benchmark and other to describe the risetime uncertainty. The uncertainty of the risetime contributes to $\sim$95 $\%$ of the Δ_i uncertainty while the uncertainty coming from the parameterizations represents only $\sim$5 $\%$. In view of these values we can assume that the contribution from the risetime uncertainty dominates the uncertainty of each Δ_i. With this assumption and using propagation of errors the expression obtained for the uncertainty of $\langle\Delta\rangle$ is:

$$\sigma_{\langle\Delta\rangle} = \frac{1}{\sqrt{N}} \tag{4.3}$$

where N is the number of surface detectors used in the calculation of $\langle\Delta\rangle$.

4.3 Data Selection

The approach followed by the $\langle\Delta\rangle$ method was thought to be a data-driven analysis. For this reason before carrying out the first step of the analysis, the determination of the benchmark, we need to perform a data selection. Given that we can apply the $\langle\Delta\rangle$ method independently in data registered by the two arrays of the Pierre Auger Observatory, we distinguish between two data sets: data from the 1500 m array and data from the 750 m array.

4.3.1 Data from the 1500 m array

In this case events from January 2004 to December 2014, which pass the 6T5 trigger condition and which are not part of a bad-period have been pre-selected. Of these events, only those whose energy is log(E/eV) $\geq$ 18.5 and sec θ <1.45 have been used in this analysis. The cut on the energy is applied because the 1500 m array is fully efficient

above this value. The cut on sec θ is established to avoid very inclined events where the risetimes are so small that the resolution of our surface detectors is not enough to measure them properly.

Besides these cuts we require surface detectors whose total signal is larger than 5 VEM and whose distance is in a specific range.

To choose the distance range we have to guarantee that we do not select risetimes values with a reconstruction bias. One way to evaluate this possible reconstruction bias is to study the behaviour of the risetime divided by the distance, $t_{1/2}/r$ as a function of core distance for different energy bins, see figure 4.4. This figure shows that the observable $t_{1/2}/r$ behaves linearly up to certain distance, then it presents a plateau extended around $\sim$ 400 m and finally it decreases drastically. This trend has been also observed in simulations, for both protons and iron nuclei, and the distances where the deviations from the linearity occur match with the ones observed in data. This effect happens because we have too few particles in these far surface detectors to probe the full particle arrival time distributions, what results in bias for the risetime values. This reconstruction bias in the risetime values has also been found by other group inside the Collaboration which works with simulated showers with no thinning level [146].

According to the previous explanation, to avoid any bias in our data selection we have to select detectors in a core distance range such that keeps the linearity of the observable $t_{1/2}/r$. We have studied the linearity in each energy bin in addition to the two ones shown in figure 4.4 and we can establish that the linearity is kept if we select the next distance ranges:

$$\begin{aligned} 18.5 \leq \log(E/eV) < 19.6 \quad &\longrightarrow \quad 300 \text{ m} < r < 1400 \text{ m}. \\ 19.6 \geq \log(E/eV) \quad &\longrightarrow \quad 300 \text{ m} < r < 2000 \text{ m}. \end{aligned} \tag{4.4}$$

In addition to the cuts on signal and core distance, a lower limit on risetimes of 40 ns is also imposed to avoid unrealistic risetime values due to the asymmetry correction. Surface detectors whose low-gain channel is saturated are discarded. All these cuts have been carefully selected to avoid any bias in our data set. The version of $\overline{\text{Off}}\underline{\text{line}}$ used for the data reconstructions is the version v3r0p0.

The $\langle\Delta\rangle$ method can be used, in principle, in events where only one detector survives the previous cuts. However in this work we require events with at least 3 detectors passing these cuts. We do that because $\langle\Delta\rangle$ correlates better with X_{max} if we use a minimum of 3 detectors to calculate it. This decision is based on simulations and it will be justified in a forthcoming section.

4.3.2 Data from the 750 m array

To apply the $\langle\Delta\rangle$ Method in data registered by the 750 m array the cuts mentioned above have to be changed due to the different energy ranges that each array can measure. In this case, events from January 2008 to December 2014, which pass the 6T5 trigger condition and which are not part of a bad-period, have been pre-seleted. Of these events, only those with energies in the range 17.5 <log(E/eV) <18.5 and sec θ <1.30 are required for the analysis. The choice of these cuts is motivated for the same reasons as in the 1500

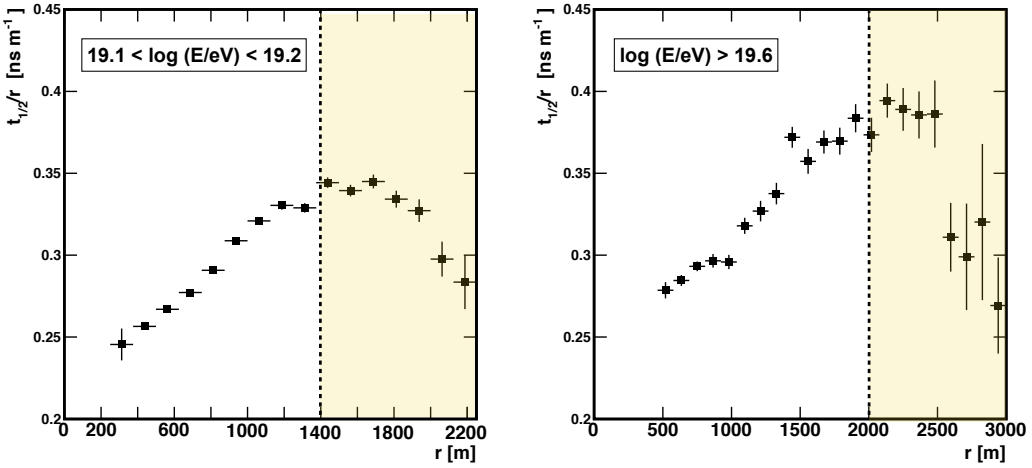


Figura 4.4: $t_{1/2}/r$ as a function of distance, r, for two energy bins. The shaded zone indicates the distance range where the linearity of $t_{1/2}/r$ disappears.

m array: the full efficiency of the array and the lack of resolution for very small risetimes in inclined showers.

Besides these cuts we require surface detectors whose total signal is larger than 3 VEM and whose core distance is between 300 m and 800 m. This core distance range is established taking into account the criterion of linearity explained for the 1500 m array. This distance range ensures the linearity of $t_{1/2}/r$ for the selected energies. A lower limit on risetimes of 40 ns is also used in this case. Surface detectors whose low-gain channel is saturated are discarded. With these cuts we can guarantee that our data set is not biased. As in the case of the 15000 m array, at least 3 detectors are required to calculate $\langle\Delta\rangle$. The $\overline{\mathrm{Off}}\underline{\mathrm{line}}$ version used to reconstruct data is again the version v3r0p0.

4.4 The Benchmark

The benchmark is the function which describes the average risetime behaviour as a function of core distance and zenith angle for an energy bin of our choice. In principle the benchmark calculation should be straightforward because it is a parameterization of the data. However for its calculation we have to deal with a severe obstacle: the risetime measurement for the two readout channels.

On the one hand, in our data set we have HG saturated detectors. These are detectors where the signals are registered by the the low-gain channel because of saturation in the high-gain one. On the other hand, we have detectors with no saturation, called non-saturated detectors, where the signals are registered by the high-gain channel.

The problem that we observe is that risetimes obtained from the different readout channels behave in a different way as a function of core distance, see figure 4.5. There is a gap between the respective behaviours as a function of the core distance. The left

panel of figure 4.5 shows that HG saturated detectors have smaller risetimes than non-saturated detectors in the distance region where they overlap. In the right panel we have plotted the profiles of these same risetimes and we have fitted them with a straight line to quantify the observed difference. The slopes obtained for these fits are compatible taking into account the errors. This means that the risetime trend with the core distance is the same for both channels and the difference observed between them is only a shift. This shift is a function of core distance and zenith angle. Comparing the intercept values of both fits we observe that the difference between risetimes from both channels in this core distance range is $\sim$7 ns. As shown in the next section, this gap of $\sim$7 ns comes from the fact that the FADC traces are not exactly equal for both channels.

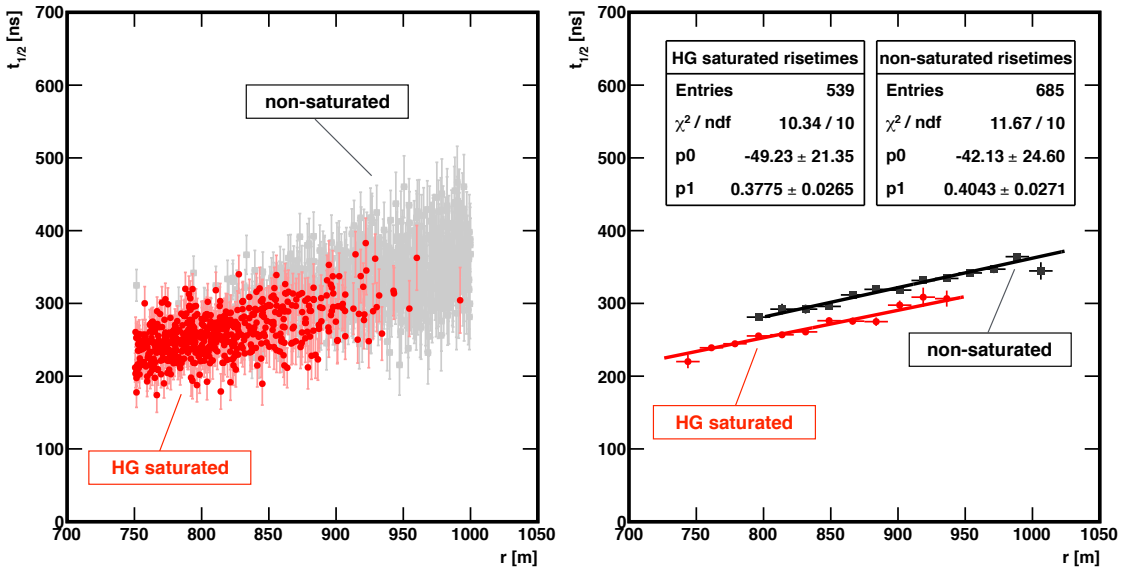


Figura 4.5: Risetimes as a function of core distance in the region where measurements from the two channels overlap (750 m - 1000 m). We have used events with energies in the range 19.0 <log (E/eV) <19.2 and zenith angles between 1.00 <sec θ <1.10. Left panel: scatter distribution. Right panel: bin-by-bin averages of the risetimes.

4.4.1 Risetimes from low-gain and high-gain channel

In this section we explain where the difference between the measurements done by the two channels comes from. We evaluate how large this difference is and we study its trend with the core distance using a small sample of simulated events.

Figure 4.6 shows the FADC traces registered in the two readout channels of one non-saturated PMT. The shape of the traces is similar, as expected, but not exactly the same. In fact, some small differences are observed in the last time bins, where the registered signals are smaller. In spite of the similarity of both traces the risetimes obtained for each one differ in 25 ns.

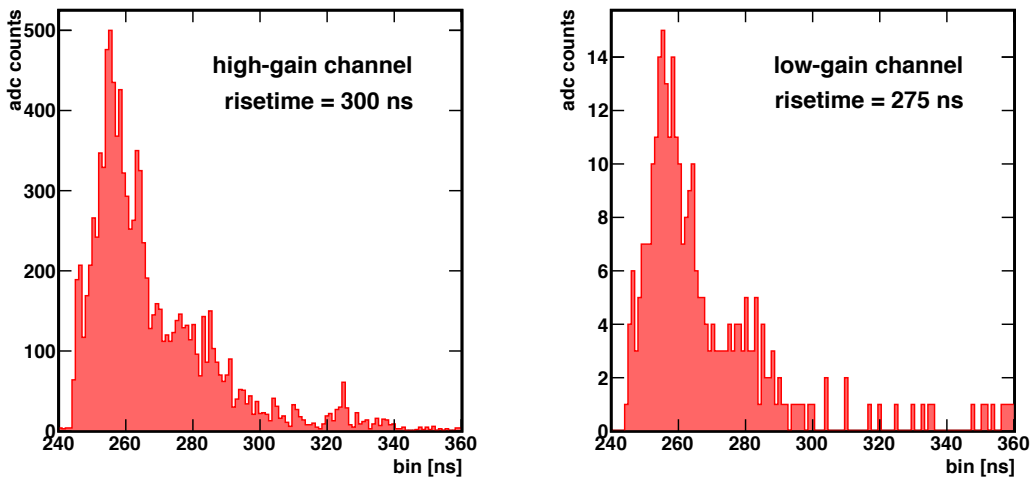


Figura 4.6: FADC traces registered in the high-gain and low-gain channel from a simulated surface detector. The surface detector belongs to an event whose energy is 10 EeV and zenith angle is 24°. The surface detector is located 1015 m away from the shower core and it has a total signal of 54 VEM. The event corresponds to a proton simulated with QGSJetII.04.

The subtle difference in the shape of both traces comes from the fact that the signal at the high channel is amplified to provide a signal with 32 times the charge gain of the low-gain channel. Due to that the low-gain channel has a lower resolution, especially in the end part of the traces where the signals are smaller.This effect acquires a greater importance in detectors whose total signal is very small.

To quantify the difference between the risetimes measured by both channels we use exclusively non-saturated detectors since in this way no channel is saturated. Note that this exercise is different to the one shown in figure 4.5. In this case, for the same detector we have two risetime measurements, one from the high-gain channel and another from the low-gain channel.

Figure 4.7 (left) exhibits the risetime measurements obtained from the two electronic channels. We observe that the average difference between the risetimes of both channels is $\sim$8 ns, with the largest risetimes measured in the high-gain channel. This difference is shown as a function of core distance in figure 4.7 (right). The difference of 8 ns observed in these plots agrees with the difference found in the figure 4.5.

The fact that the two electronic channels do not provide the same risetime measurement forces us to treat separately HG saturated and non-saturated detectors in the whole analysis presented in this thesis. This means that we have to calculate two different benchmarks. One for HG saturated detectors, another for non-saturated detectors. If this issue is not taken into account we would introduce distortions in the behaviour of $\langle\Delta\rangle$. In particular the Δ_i values coming from intermediate distances (the overlapping region) would be flat as a function of energy instead of increasing.

4.4.2 Benchmark for events measured in the 1500 m array

To calculate the benchmark for the 1500 m array data set we have chosen the energy bin 19.1 <log(E/eV) <19.2. This energy bin is the optimal one to parameterize the risetime behaviour of HG saturated and non-saturated detectors separately.

Distinguishing between risetimes from HG saturated and non-saturated detectors limits the choice of the energy bin for the benchmark. If we take a low-energy bin, for example 18.5 <log(E/eV) <18.6, most of the detectors are non-saturated. Thus, we have an extended distance range to fit the behaviour of the non-saturated detectors but an insufficient lever arm to fit the HG saturated detectors (see left panel of figure 4.8). Moreover if we take an energy bin at the largest energies, for example 19.4 <log(E/eV) <19.6, the distance range to fit risetimes from HG saturated detectors is quite broad but it is not sufficient for non-saturated detectors (right panel in figure 4.8). Due to this distinction between detectors it is necessary to have an equilibrium between the core distance ranges of HG saturated and non-saturated detectors. We find this equilibrium for the intermediate energies: in particular we choose the energy bin 19.1 <log(E/eV) <19.2.

To parameterize the benchmark we divide the data of the chosen energy bin in 9 equal sec θ bins in the range 1.00 <sec θ <1.45. After that we separate risetimes from HG saturated and non-saturated detectors and fit them independently as a function of core distance.

First we fit the behaviour of the HG saturated detectors. The function used to fit the

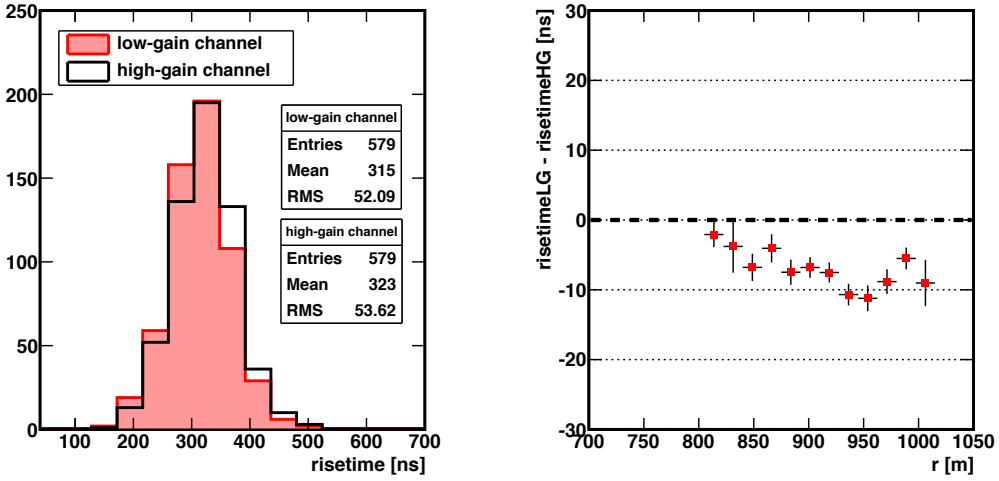
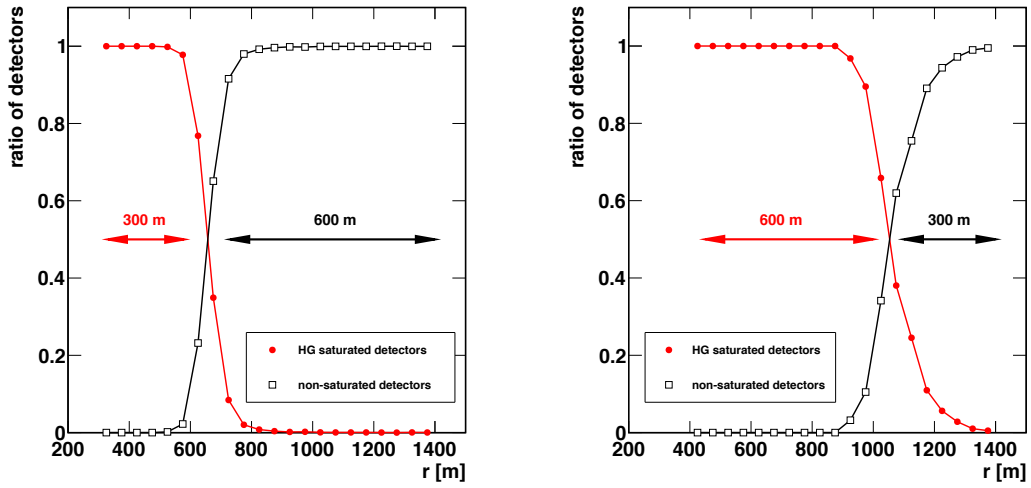


Figura 4.7: Left panel: Distributions of risetimes measured simultaneously in the high-gain and low-gain channel of the non-saturated detectors. Right panel: Difference between risetimes of the two channels as a function of core distance. Simulated events with energies between 19.0 <log(E/eV) <19.2. Only detectors with signals larger than 5 VEM and core distance between 750 m and 1000 m have been used. This difference can be smaller than 25 ns because the risetimes are calculated as the average of the value of the three PMTs.



(a) 18.5 <log(E/eV) <18.6.

(b) 19.4 <log(E/eV) <19.6.

Figura 4.8: Ratio of HG saturated detectors and non-saturated detectors over the total of detectors as a function of core distance. The distances where each kind of detector predominates are indicated.

risetimes from HG saturated detectors is:

$$t_{1/2}^{sat} = 40 + \sqrt{A_{sat}(\theta)^2 + B_{sat}(\theta) r^2} - A_{sat}(\theta) \tag{4.5}$$

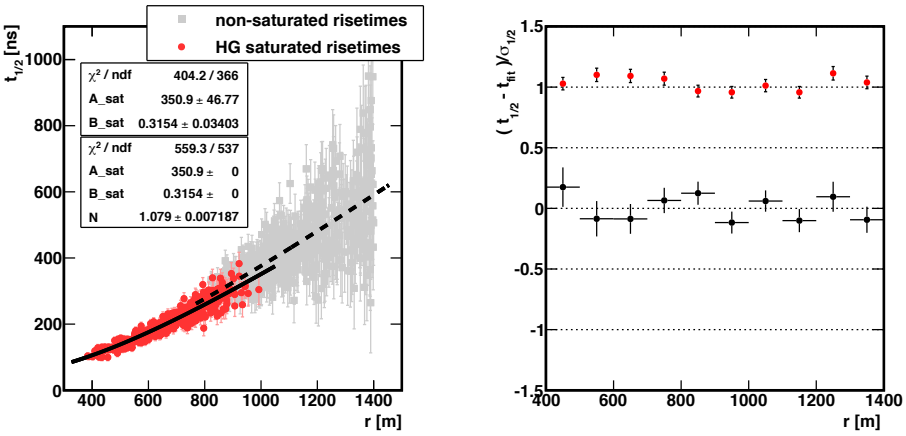
This function for the benchmark parameterization is especially convenient because when r = 0 it takes the value of 40 ns, the limit imposed by the SPR. It was originally introduced in [138] and [137] because it was more suitable to fit the risetime behaviour. The reduced χ^2 values obtained with this function were consistently lower than those obtained with other functions.

Once we have obtained A_{sat} and B_{sat} we use these values to fit the risetimes from non-saturated detectors with the function:

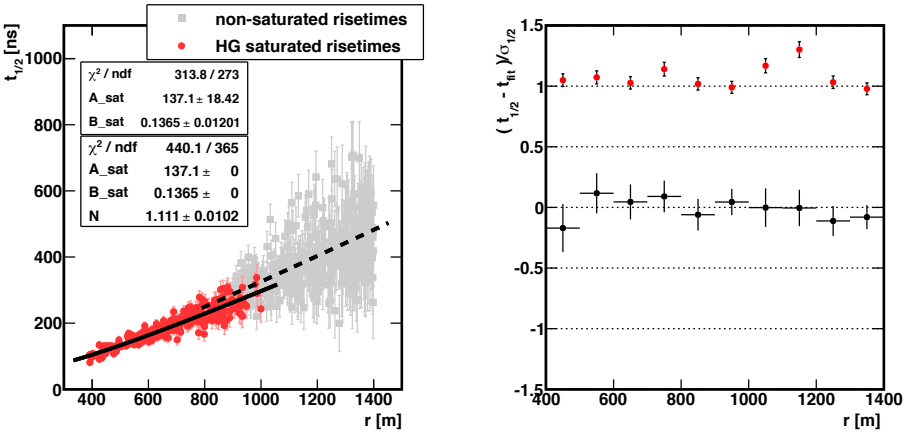
$$t_{1/2}^{non-sat} = 40 + N(\theta)(\sqrt{A_{sat}(\theta)^2 + B_{sat}(\theta) r^2} - A_{sat}(\theta)) \tag{4.6}$$

This function has only one free parameter, $N(\theta)$. The parameters A_{sat} and B_{sat} are fixed and correspond to the values obtained from the HG saturated fit. If we use these two functions for the benchmark parameterization, risetimes from HG saturated and non-saturated detectors have the same dependence with the core distance. The parameter $N(\theta)$ only describes the shift observed between the two risetime behaviours.

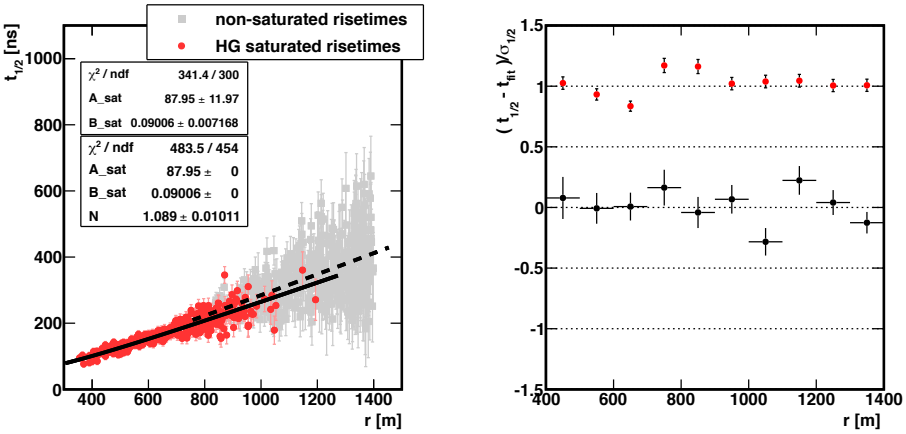
Left panels of figure 4.9 show the risetimes as a function of core distance with their respective fits for three sec θ bins. The risetime residuals have been also evaluated in each case to study the goodness of the fits. The right panels of figure 4.9 show the mean values and the standard deviations of these residuals as a function of core distance.



(a) 1.00 <sec θ <1.05



(b) 1.15 <sec θ <1.20



(c) 1.25 <sec θ <1.30

Figura 4.9: Left panels: Risetimes as a function of core distance with their respective fits for three sec θ bins. Right panels: Mean values and standard deviations of the residuals as a function of core distance. Data of the 1500 m array.

The mean values of the residuals are centered around zero and the standard deviations are compatible with one along the whole distance range. This indicates that the fits reproduce the risetime behaviour in a proper way for each sec θ bin. The uncertainty of the core distance is not taken into account for the fits because it is negligible compared to the risetime uncertainty.

Once we have the risetime fits for each sec θ bin, we plot the parameters A_{sat}, B_{sat} and N as a function of sec θ and fit them with the functions:

$$\begin{aligned} A_{sat}(\theta) &= a_0 + a_1(\sec\theta)^{-4} \\ B_{sat}(\theta) &= b_0 + b_1(\sec\theta)^{-4} \end{aligned} \tag{4.7}$$

$$N(\theta) = n_0 + n_1(\sec\theta)^2 + n_2(\mathrm{e}^{\sec\theta}) \tag{4.8}$$

These functions have an empirical origin and they are chosen because they provide the best description of the data. The functions used to describe A_{sat} and B_{sat} have changed compared to previous works [140]. The new ones provide lower values of the reduced χ^2. The behaviour of A_{sat} and B_{sat} as a function of sec θ with their respective fits are shown in figure 4.10a. The parameter N and its fit can be seen in figure 4.10b.

Finally the parameters which describe $A_{sat}(\theta)$, $B_{sat}(\theta)$ and N(θ) take the values shown in the table 4.1.

Parameter	Value $\pm$ error
a_0	-72 $\pm$ 10
a_1	410 $\pm$ 30
b_0	-0.049 $\pm$ 0.007
b_1	0.36 $\pm$ 0.02
n_0	-0.07 $\pm$ 0.02
n_1	-1.14 $\pm$ 0.02
n_2	0.84 $\pm$ 0.01

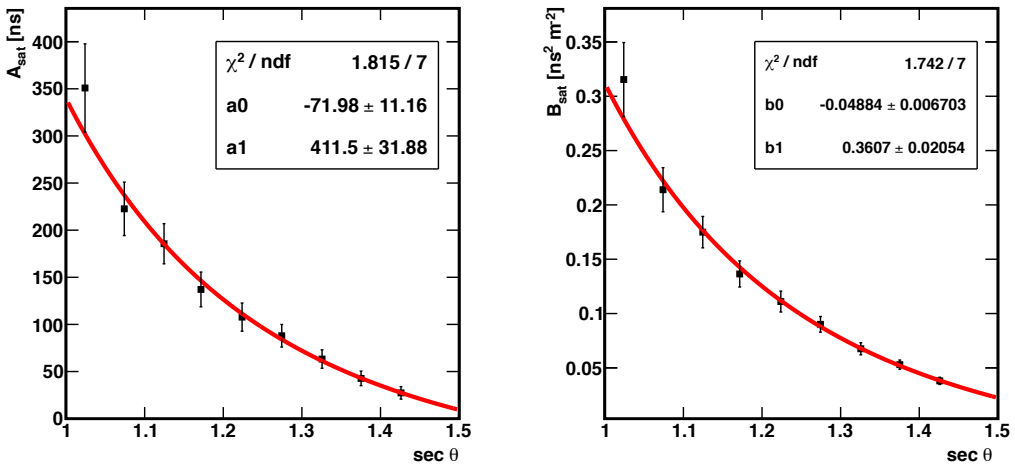
Cuadro 4.1: Values of the parameters which define the benchmark in the 1500 m array.

The parameters $A_{sat}(\theta)$, $B_{sat}(\theta)$ and N(θ) are always positive in the zenith angle range where they are defined. The parameters a_0 and a_1 are expressed in ns, b_0 and b_1 in ns^2 m^{-2} and n_0, n_1 and n_2 are dimensionless.

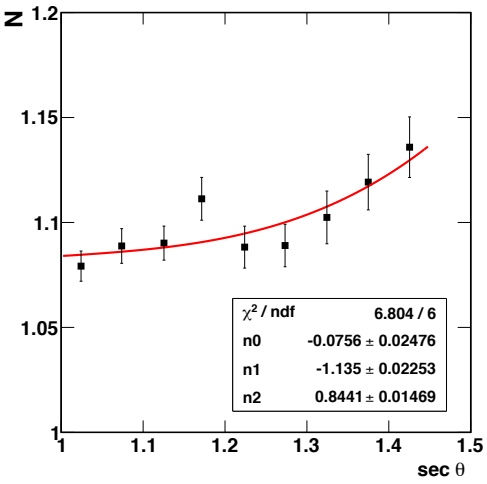
4.4.3 Benchmark for events measured in the 750 m array

To apply the $\langle\Delta\rangle$ method in data registered by the 750 m array, a new benchmark has to be defined. As we mentioned before, this is a consequence of the different energies measured in each array. This fact makes the risetime dependences with the core distance and zenith angle slightly different for both arrays.

For data of the 750 m array the energy bin chosen to calculate the benchmark is: 17.7 <log(E/eV) <17.8. We choose this energy bin because we find here the necessary balance between the core distance ranges of HG saturated and non-saturated detectors.



(a) A_{sat} and B_{sat} as a function of sec θ.



(b) N as a function of sec θ.

Figura 4.10: Parameters for the benchmark definition in the 1500 m array.

The procedure to obtain the benchmark parameterization for data of the 750 m array follows exactly what was explained in the previous section for the 1500 m array. We divide the data of the chosen energy bin in 6 equidistant sec θ bins in the range 1.00 <sec θ <1.30 and after separating risetimes from HG saturated and non-saturated detectors we fit them as a function of the core distance.

The only difference between the procedure explained for the 1500 m array and this one is that now we fit first the non-saturated detectors. We do that because it is more convenient to do the fit with the two free parameters in the set with more risetime measurements. The 750 m array data set have smaller signals than the 1500 m array data set. Hence the number of non-saturated detectors is significantly larger compared to the

number of HG saturated detectors.

The function used to fit the risetimes from the non-saturated detectors is:

$$t_{1/2}^{non-sat} = 40 + \sqrt{A_{non-sat}(\theta)^2 + B_{non-sat}(\theta)r^2} - A_{non-sat}(\theta) \quad (4.9)$$

This function is equal to the function shown in the equation 4.5 but with different parameters. Once this fit is done and we have obtained $A_{non-sat}$ and $B_{non-sat}$, we use these values to fit the risetimes from HG-saturated detectors with the function:

$$t_{1/2}^{sat} = 40 + N(\theta)(\sqrt{A_{non-sat}(\theta)^2 + B_{non-sat}(\theta)r^2} - A_{non-sat}(\theta)) \quad (4.10)$$

This function is equal to equation 4.6 and it has only one free parameter since $A_{non-sat}$ and $B_{non-sat}$ are fixed. Once again, $N(\theta)$ only describes the shift observed between the behaviour of risetimes from non-saturated and HG saturated detectors. Three of the six sec θ bins, where we have fitted the risetime behaviour as a function of core distance, are shown in figure 4.11 (left panels). Right panels of this figure show the mean values and the standard deviations of the risetime residuals as a function of core distance to test the goodness of the fits.

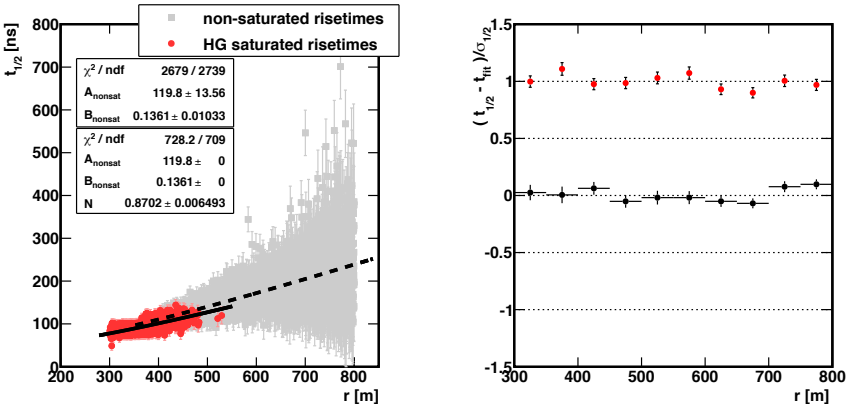
The behaviour of the parameters $A_{non-sat}$, $B_{non-sat}$ and N as a function of the sec θ is fitted using the same equations described for the 1500 m array: equations 4.7 and 4.8. The parameters $A_{non-sat}$ and $B_{non-sat}$ as a function of sec θ with their respective fits are shown in figure 4.12a. For the parameter N and its fit see figure 4.12b.

Finally the parameters which describe $A_{non-sat}(\theta)$, $B_{non-sat}(\theta)$ and N(θ) take the values shown in the table 4.2.

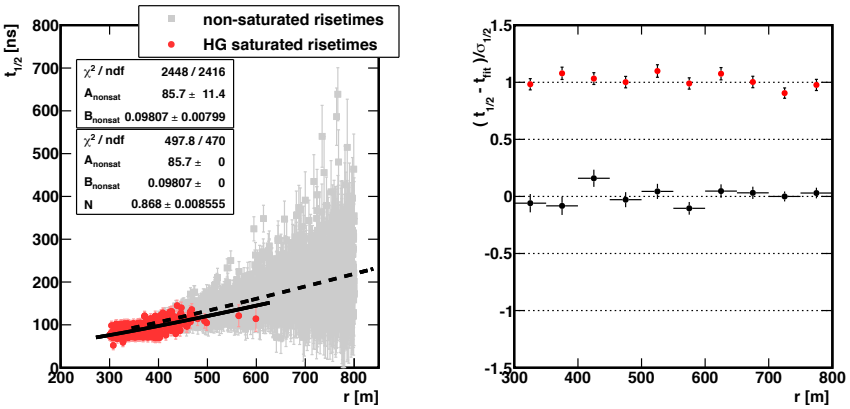
Parameter	Value $\pm$ error
a_0	-12 $\pm$ 13
a_1	150 $\pm$ 20
b_0	-0.005 $\pm$ 0.009
b_1	0.16 $\pm$ 0.02
n_0	8.81 $\pm$ 0.03
n_1	8.08 $\pm$ 0.03
n_2	-5.90 $\pm$ 0.02

Cuadro 4.2: Values of the parameters to define the benchmark in the 750 m array.

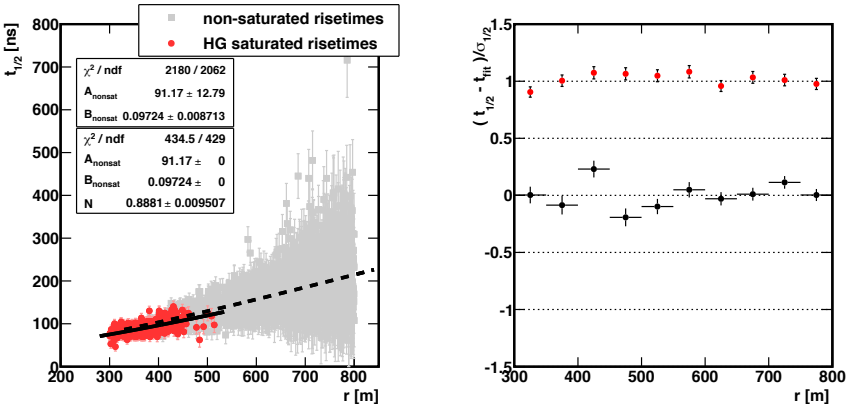
The parameters $A_{non-sat}(\theta)$, $B_{non-sat}(\theta)$ and N(θ) are always positive in the zenith angle range where they are defined. The parameters a_0 and a_1 are expressed in ns, b_0 and b_1 in ns^2 m^{-2} and n_0, n_1 and n_2 are dimensionless. Figure 4.12b reflects that the parameter N of the 750 m array behaves in a way opposite to that in its counterpart for the 1500 m array. Besides in the case of the 750 m array the parameter N is always lower than one whereas in the 1500 m array N is higher than one. This happens because we have inverted the roles of the HG saturated and non-saturated detectors in the benchmark calculation.



(a) 1.00 <sec θ <1.05

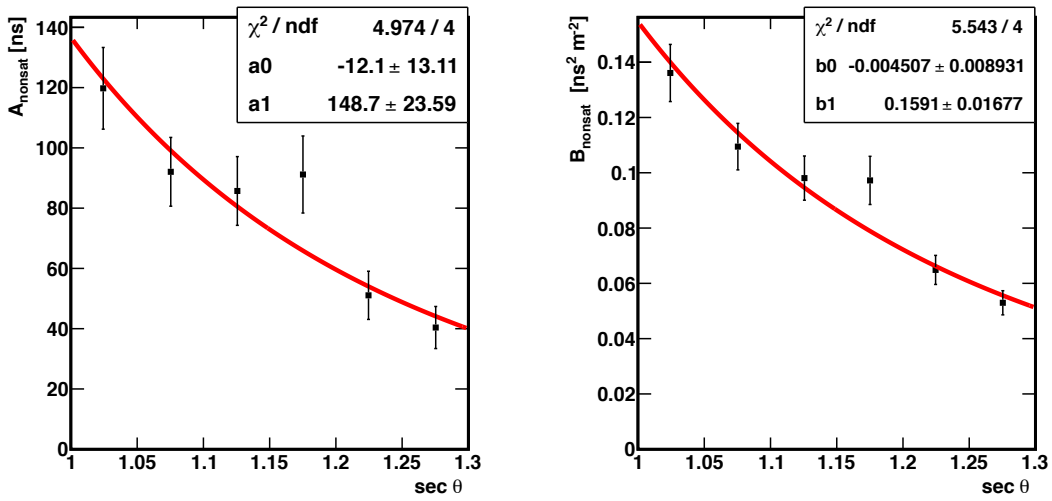


(b) 1.10 <sec θ <1.15

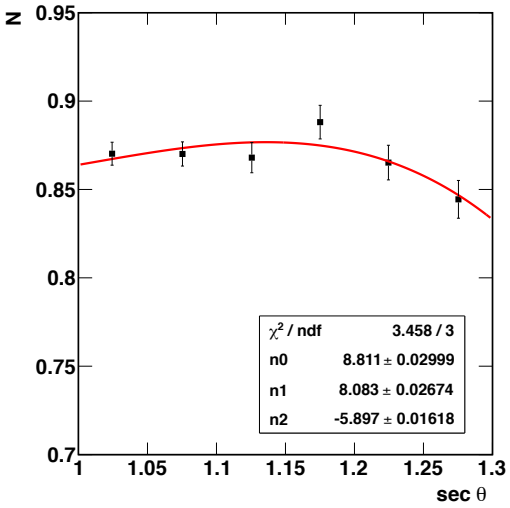


(c) 1.15 <sec θ <1.20

Figura 4.11: Left panels: Risetimes as a function of core distance with their respective fits for three sec θ bins. Right panels: Mean values and standard deviations of the residuals as a function of core distance. Data of the 750 m array.



(a) $A_{non-sat}$ and $B_{non-sat}$ as a function of sec θ.



(b) N as a function of sec θ.

Figura 4.12: Parameters for the benchmark definition in the 750 m array.

4.5 The parameter $\langle\Delta\rangle$ in the benchmark energy bin

We explained in section 4.2 that due to the definition of $\langle\Delta\rangle$ the mean value of its distribution must be compatible with zero for the benchmark energy bin. However this only occurs if the benchmark parameterization describes the risetime average behaviour in a proper way. For this reason the study of $\langle\Delta\rangle$ in the benchmark energy bin is an important cross-check for the goodness of the parameterizations obtained in the previous sections.

Figure 4.13 exhibits the distributions of Δ_i and $\langle\Delta\rangle$ in the benchmark energy bin for data of the 1500 m array. We have used 6371 surface detectors belonging to 1753 events for the determination of the benchmark. The mean values of both distributions are compatible with zero. We obtain 0.005 $\pm$ 0.013 for Δ_i and 0.004 $\pm$ 0.015 for $\langle\Delta\rangle$. Hence we can guarantee that the benchmark parameterization for the 1500 m array describes in a proper way the average risetime behaviour.

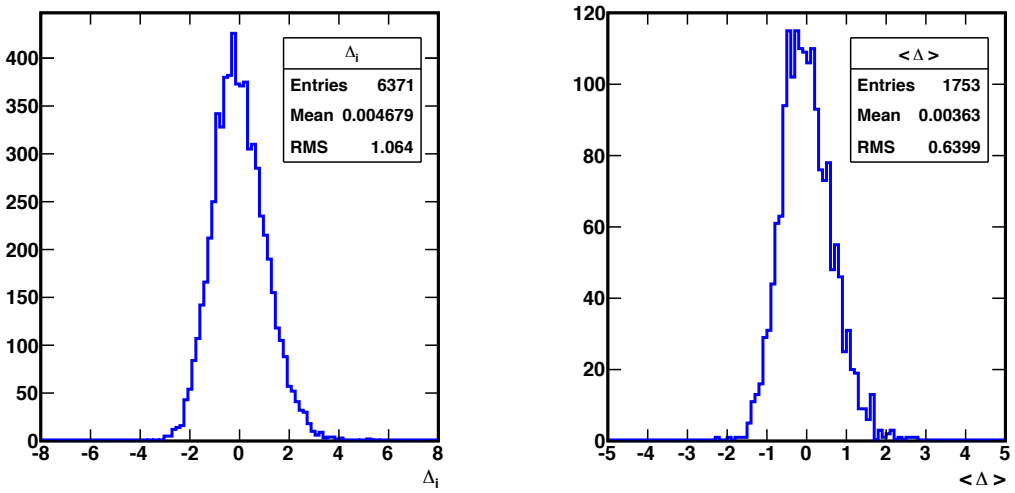


Figura 4.13: Left panel: Δ_i distribution for the benchmark energy bin. Right panel: $\langle\Delta\rangle$ distribution in the same energy bin. Data of the 1500 m array.

We also notice that the RMS value of the distribution of Δ_i is close to one. This is another indication that the values of Δ_i are correctly obtained, since Δ_i really behaves as a pull distribution. Large deviations from one would indicate an underlying problem with the risetime uncertainty parameterization. Note that the RMS value of the distribution of $\langle\Delta\rangle$ is lower than 1. As explained in section 4.2, the distribution of $\langle\Delta\rangle$ can be described as the sum of several pull distributions (the associated with each Δ_i) taken into account the correlations and the normalization coming from the number of used detectors, N. These reasons make the RMS of the $\langle\Delta\rangle$ distribution always smaller than 1.

The same two distributions are shown in figure 4.14 for data of the 750 m array. For this data set we have 16869 surface detectors belonging to 4204 events in the benchmark energy bin. Given that both values are compatible with zero, we obtain 0.006 $\pm$ 0.008 for Δ_i and 0.004 $\pm$ 0.008 for $\langle\Delta\rangle$, we can guarantee that the benchmark parameterization found for data of the 750 m array describes in a proper way the average risetime behaviour. The RMS value of the Δ_i distribution is close to one, so there is no problem with the risetime uncertainty in this case either.

The final cross-check to assure that the benchmark parameterizations describe correctly the risetime dependence with the zenith angle is the study of the $\langle\Delta\rangle$ values as a function of sec θ in the benchmark energy bin. If the functional form used to parameterize the benchmark as a function of sec θ is appropriate we should expect a flat trend of the

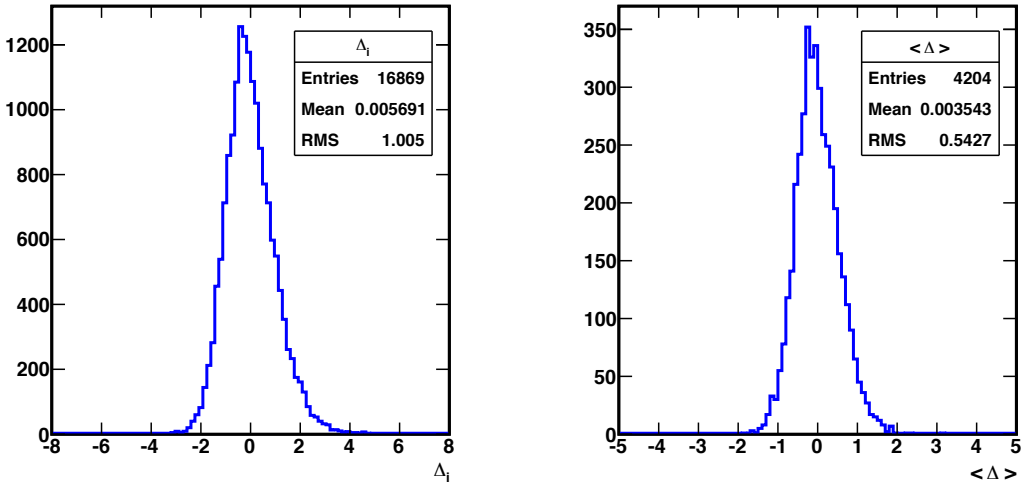
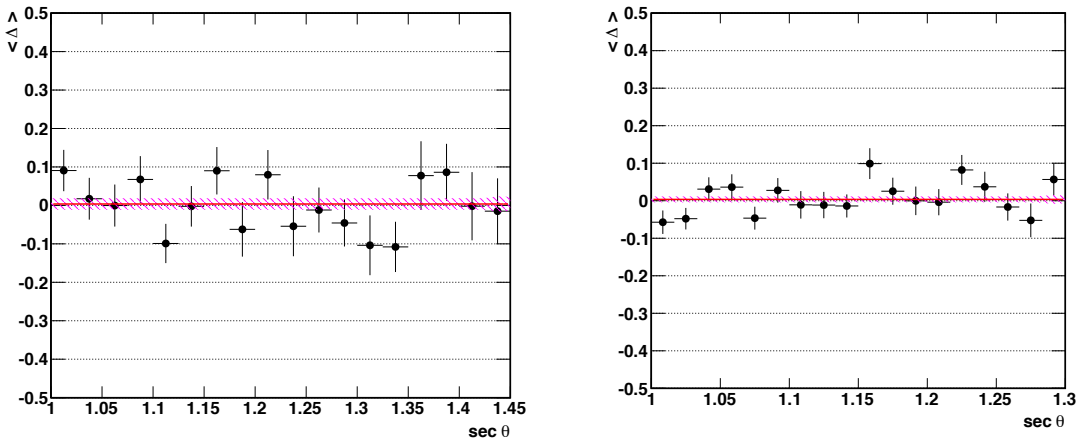


Figura 4.14: Left panel: Δ_i distribution for the benchmark energy bin. Right panel: $\langle\Delta\rangle$ distribution in the same energy bin. Data of the 750 m array.

$\langle\Delta\rangle$ values as a function of sec θ. The benchmark parameterization is calculated so that it contains the whole dependence of risetime with the sec θ. In this way when we calculate the numerator in the Δ_i definition, $t_{1/2}$ - $t_{1/2}^{bench}$, the dependence with the sec θ is canceled. The flat trend of $\langle\Delta\rangle$ with sec θ is observed for both arrays in figure 4.15.



(a) 1500 m array

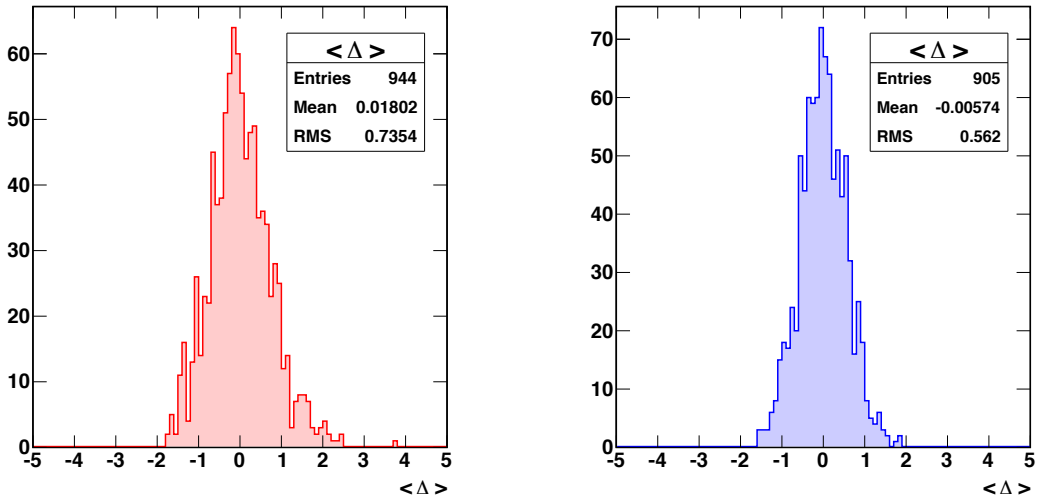
(b) 750 m array

Figura 4.15: $\langle\Delta\rangle$ as a function of sec θ in the benchmark energy bin. The mean values of the data are also indicated with a straight line in both panels.

4.6 Correlation between $\langle\Delta\rangle$ and $X_{\mathbf{max}}$ in simulations

As we mentioned at the beginning of this chapter, the underlying idea behind the $\langle\Delta\rangle$ method is to transform the $\langle\Delta\rangle$ values in values of X_{max} through a calibration procedure to do a mass composition analysis with SD-only events. However to achieve this it is necessary that a certain degree of correlation exists between both observables.

Before applying the $\langle\Delta\rangle$ method to data we would like to test the level of correlation between $\langle\Delta\rangle$ and X_{max} for a small number of events. To understand this issue we use simulations. In particular we use protons and iron nuclei simulated with the hadronic model QGSJetII.04. We use simulations in the 1500 m array with energies in the benchmark energy bin (19.1 <log(E/eV) <19.2). The advantage of using simulations for this test is that the value of X_{max} is known for each event.



(a) Protons (b) Iron Nuclei

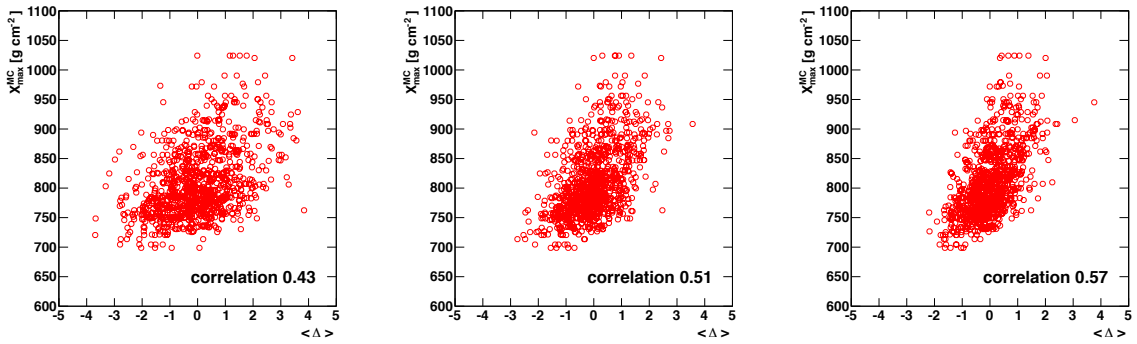
Figura 4.16: $\langle\Delta\rangle$ distributions for QGSJetII.04 simulations in the benchmark energy bin.

To calculate $\langle\Delta\rangle$ in simulations, as in data, the first step is to determine a benchmark for protons and for iron nuclei using exactly the same procedure explained in the section 4.4. We calculate a different benchmark for each data set (one for protons, one for iron nuclei). Using different benchmarks guarantees that protons and iron nuclei are studied independently.

As in data, to guarantee that the benchmark parameterizations obtained for protons and for iron nuclei are correct, we study the distributions of $\langle\Delta\rangle$ in the benchmark energy bin (see figure 4.16) and we observe that both mean values are compatible with zero, 0.018 $\pm$ 0.024 for protons and -0.006 $\pm$ 0.017 for iron nuclei.

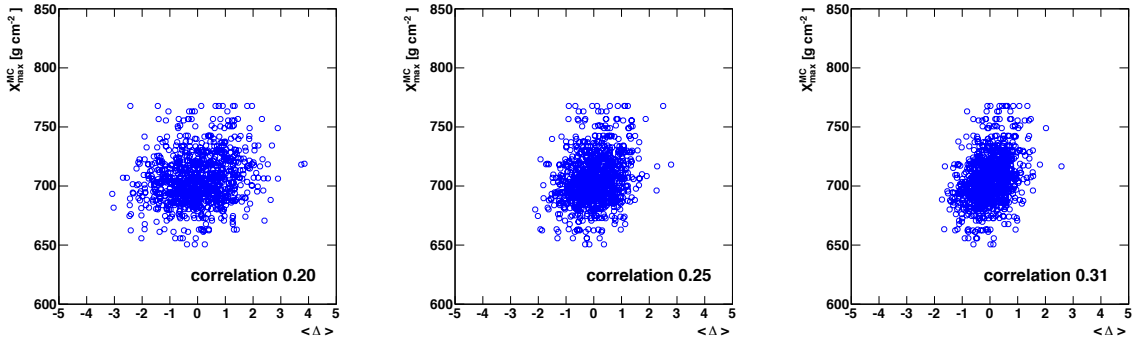
Once the parameter $\langle\Delta\rangle$ is calculated in the simulations we can study the correlation between $\langle\Delta\rangle$ and the values of X_{max}. Figure 4.17 shows the correlation between $\langle\Delta\rangle$ and X_{max} for proton events in the benchmark energy bin. Three different data sets are

studied. In the first one the $\langle\Delta\rangle$ values are calculated using only one detector, the detector with the largest signal. In the second one only two detectors are used, the two with the largest recorded signals. In the third data set the three detectors with the largest signals are used. This figure confirms that the correlation between $\langle\Delta\rangle$ and X_{max} exists and it is larger when we use three detectors to calculate $\langle\Delta\rangle$. The same trend is observed for simulations of iron nuclei in the benchmark energy bin, see figure 4.18.



(a) 1 detector (b) 2 detectors (c) 3 detectors

Figura 4.17: Correlation between $\langle\Delta\rangle$ and X_{max} for proton simulations (QGSJetII.04) in the benchmark energy bin.



(a) 1 detector (b) 2 detectors (c) 3 detectors

Figura 4.18: Correlation between $\langle\Delta\rangle$ and X_{max} for iron nuclei simulations (QGSJetII.04) in the benchmark energy bin.

Two important conclusions are obtained from figures 4.17 and 4.18. The first one is that $\langle\Delta\rangle$ correlates with X_{max} independently of the number of detectors that we use to calculate $\langle\Delta\rangle$. This implies that the $\langle\Delta\rangle$ method effectively can be used for mass composition analyses. The second one is that the correlation increases as we use more detectors to calculate $\langle\Delta\rangle$. This trend is expected because if we use more detectors in the calculation

of $\langle\Delta\rangle$ we are putting into our observable more information related to the development of the shower. However we have to be cautious with the number of detectors that we require to calculate $\langle\Delta\rangle$ since a very strict requirement can introduce a selection bias at the lowest energies. This is the reason to require at least 3 detectors to calculate $\langle\Delta\rangle$. With this number we do not introduce a mass-composition bias at the lowest energies and at the same time we obtain an acceptable level of correlation.

The difference between the correlation factors obtained for protons and iron nuclei deserves a special attention. It is clear that protons present a better correlation with X_{max}. Their correlation factor is almost twice the value found for iron nuclei. This significant difference is expected due to the different proportions of electromagnetic and muonic component present in the air showers. The showers initiated by protons develop deeper in the atmosphere and for this reason they are richer in electromagnetic component. Likewise showers initiated by irons nuclei are richer in muons. Since X_{max} is an observable of a mainly electromagnetic nature, it is expected a larger correlation for protons.

5

Mass composition results with the data of the1500 m array

We have shown that the $\langle\Delta\rangle$ method provides a mass sensitive parameter by studying how it correlates with X_{max} for a sample of simulated events. We also determined the key element of this method, the benchmark. In addition we reviewed all the technicalities that must be taken into account to guarantee an optimal application of the method, such as the specific distance range to avoid the bias in the risetime values, the division between HG saturated and non-saturated detectors or the choice of the new benchmark functions to ensure a flat trend of $\langle\Delta\rangle$ as a function of sec θ.

At this point we have all the tools in hand to apply the $\langle\Delta\rangle$ method to the data from the Pierre Auger Observatory. The final goal of this chapter is to obtain an elongation rate with SD-only events in a energy range that not only overlaps with the data provided by the FD but that is also able to reach even higher energies. As we mentioned in the previous chapter the $\langle\Delta\rangle$ method can be applied independently to data from the two arrays of the Observatory. This chapter is focused on the analysis of the data collected with the 1500 m array.

5.1 Data selection

The cuts used to select data of the 1500 m array were previously introduced, explained and justified in chapter 4 because they were necessary for the benchmark calculation. To avoid unnecessary repetitions, in this section we only list them and show their selection efficiencies, see table 5.1. Data correspond to a period spanning from January 2004 to December 2014.

The most stringent quality cut is the one requiring sec θ <1.45, but it is necessary to discard inclined events with too small risetimes. As we already mentioned, very small risetimes are removed because they cannot be measured with enough precision due to

Quality cuts	**Events**	**Efficiency**
Total events with log(E/eV) >18.5	217 469	100 %
sec θ <1.45	97 981	45 %
6T5 Trigger	67 764	31 %
Reject lightning events	67 764	31 %
Reject bad periods and comms crisis	63 856	29 %
At least 3 detectors to calculate $\langle\Delta\rangle$	**58 583**	27 %

Cuadro 5.1: Quality cuts applied to data and their relative selection efficiency.

the timing resolution of our surface detectors. The rejection of bad periods, communications crisis and lightning events is only required to analyze good events. These cuts have a minimum effect on the resulting statistics.

The last cut in table 5.1 imposes the need for at least 3 detectors to calculate $\langle\Delta\rangle$. An event has information from several surface detectors, but we only use for the calculation of $\langle\Delta\rangle$ those which survive the cuts listed in table 5.2. The selection efficiency of each cut is also indicated in the same table.

Cuts	**Detectors**	**Efficiency**
Detectors before requiring at least 3 det. per event	373 159	100
Reject low-gain saturated	362 196	97.1
S >5 VEM	292 559	78.4
300 m <r <1400 (2000) m	220 394	59.1
$t_{1/2}$ >40 ns	220 180	59.0
After requiring at least 3 det. per event	**210 709**	**56.4**

Cuadro 5.2: Number and percentage of detectors surviving each detector-level cut.

Figure 5.1 shows, for two energy bins, the distribution of the number of detectors per event which survive the detector-level cuts listed in table 5.2. The shaded bars of these histograms indicate the events that will be used in the $\langle\Delta\rangle$ method: only those with at least 3 detectors after detector-level cuts. According to this figure we will use in most cases 3 and 4 detectors to determine $\langle\Delta\rangle$.

It was mentioned in chapter 4 that these cuts are carefully chosen to avoid a selection bias in our data set. To check this issue we have applied exactly these same cuts in different samples of simulated showers of protons and iron nuclei with energies in the range 18.5 $\leq$ log(E/eV) <20.0. We find that the differences between the selection efficiencies for protons and iron nuclei are never larger than 6 %. Below 10 EeV the selection efficiency is always larger for iron nuclei than for protons.

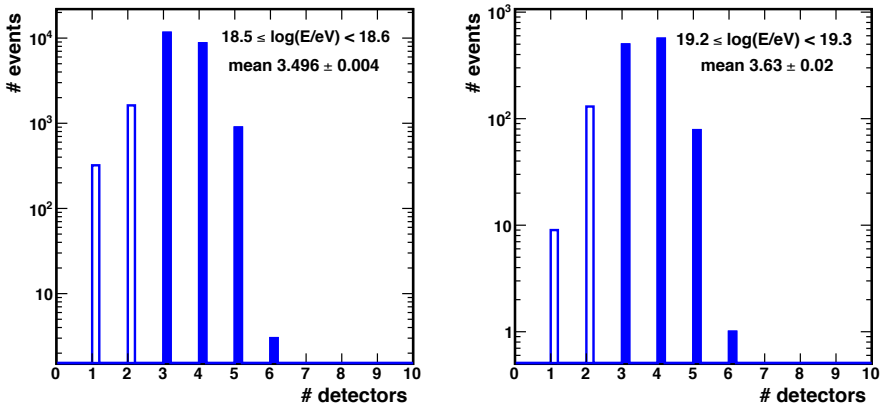


Figura 5.1: Number of detectors per event surviving the cuts given in table 5.2. The colour zone indicates the number of detectors for events selected after requiring at least 3 detectors per event. The mean of the number of detectors after this cut is also indicated.

5.2 Delta as a function of the energy

The parameter $\langle\Delta\rangle$ was already calculated in the previous chapter for events belonging just to the benchmark energy bin ($19.1 \leq \log(E/eV) < 19.2$), but in that case the motivation was only to check if the benchmark parameterization worked properly.

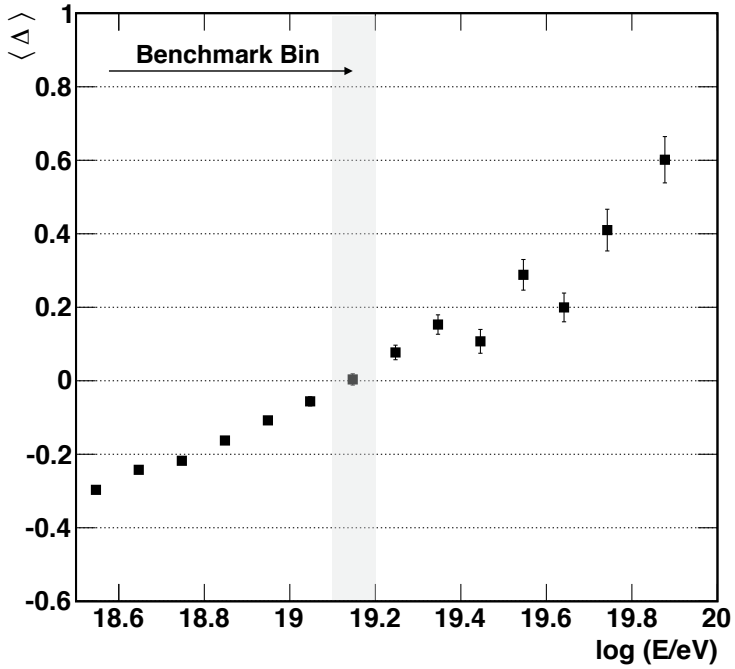


Figura 5.2: $\langle\Delta\rangle$ as a function of energy for data gathered with the 1500 m array.

Now we use the same procedure to calculate $\langle\Delta\rangle$ for all of our selected events, regardless of their measured energy. Just as a reminder, we calculate the $\langle\Delta\rangle$ values using equation 4.2 and taking the benchmark parameterization for the 1500 m array given in section 4.4.2. The mean values of the $\langle\Delta\rangle$ distributions as a function of energy are given in figure 5.2. The uncertainties shown in this figure are of statistical nature only.

The mean value of $\langle\Delta\rangle$ in the benchmark energy bin is compatible with zero as we expected from the discussion of the previous chapter. Furthermore, the mean values of $\langle\Delta\rangle$ go to larger values with increasing energy, as expected, since more energetic events are deeper and as a consequence they have larger risetimes.

The distributions of $\langle\Delta\rangle$ are shown in figure 5.3. In this figure we indicate the number of events that we use in each energy bin to calculate the mean value of each $\langle\Delta\rangle$ distribution. The last two energy bins are regrouped in comparison with the bins shown in figure 5.2. The last energy bin of figure 5.2, log(E/eV) $\geq$ 19.8 contains 51 events. Note that this means that our analysis has 51 events with energies larger than 63 EeV. The mean energy of these 51 events is (76 $\pm$ 2) EeV. The two events with the highest energy exceed 100 EeV. One of them has an energy of (106 $\pm$ 3) EeV (Event 11002850) and the other one of (141 $\pm$ 4) EeV (Event 3036270).

5.2.1 Some cross-checks in the behavior of $\langle\Delta\rangle$

In addition to the cross-checks done in section 4.5 which guarantee that the benchmark parameterization describes properly the risetimes in the benchmark energy bin, we include in this section some additional checks to assess that the behavior of $\langle\Delta\rangle$ is correct in each one of the studied energy bins.

As we explained in section 4.5, the benchmark parameterization, $t_{1/2}^{bench}(r,\theta)$, contains the whole dependence of the risetimes with sec θ and with the core distance. This means that when we calculate the term $(t_{1/2} - t_{1/2}^{bench})$, included in the definition of Δ_i, any dependence on the sec θ and on the distance disappears. Due to this, we expect $\langle\Delta\rangle$ to be independent of sec θ and distance for each energy bin.

Figure 5.4a displays the behaviour of $\langle\Delta\rangle$ as a function of energy for three different sec θ bins. We observe that the trend with energy is exactly the same for the three sec θ bins. This indicates that the functional form used to describe the benchmark is appropriate for all energy bins. The mean values of $\langle\Delta\rangle$ for events with zenith angle in the range 1.30 <sec θ <1.45 are slightly smaller than the values obtained for the rest of events below 19 EeV. This small difference at the lowest energies for the most inclined events is an indication that the resolution of our surface detectors to measure such small risetime values begin to play a role. This issue modifies slightly the trend of the risetimes with sec θ compared to the parameterization obtained in a larger energy bin, and the discrepancy observed stems from there.

This small difference in the trend of the most inclined events will be evaluated in a next section in terms of X_{max} and taken into account as a systematic uncertainty in the final elongation rate.

Figure 5.4b shows the dependence of $\langle\Delta\rangle$ with core distance. There we plot the observables that we have called Δ_{first}, Δ_{second} and Δ_{third}. These observables are Δ_i values for some specific surface detectors. Δ_{first} is the Δ_i value obtained if we use the detector

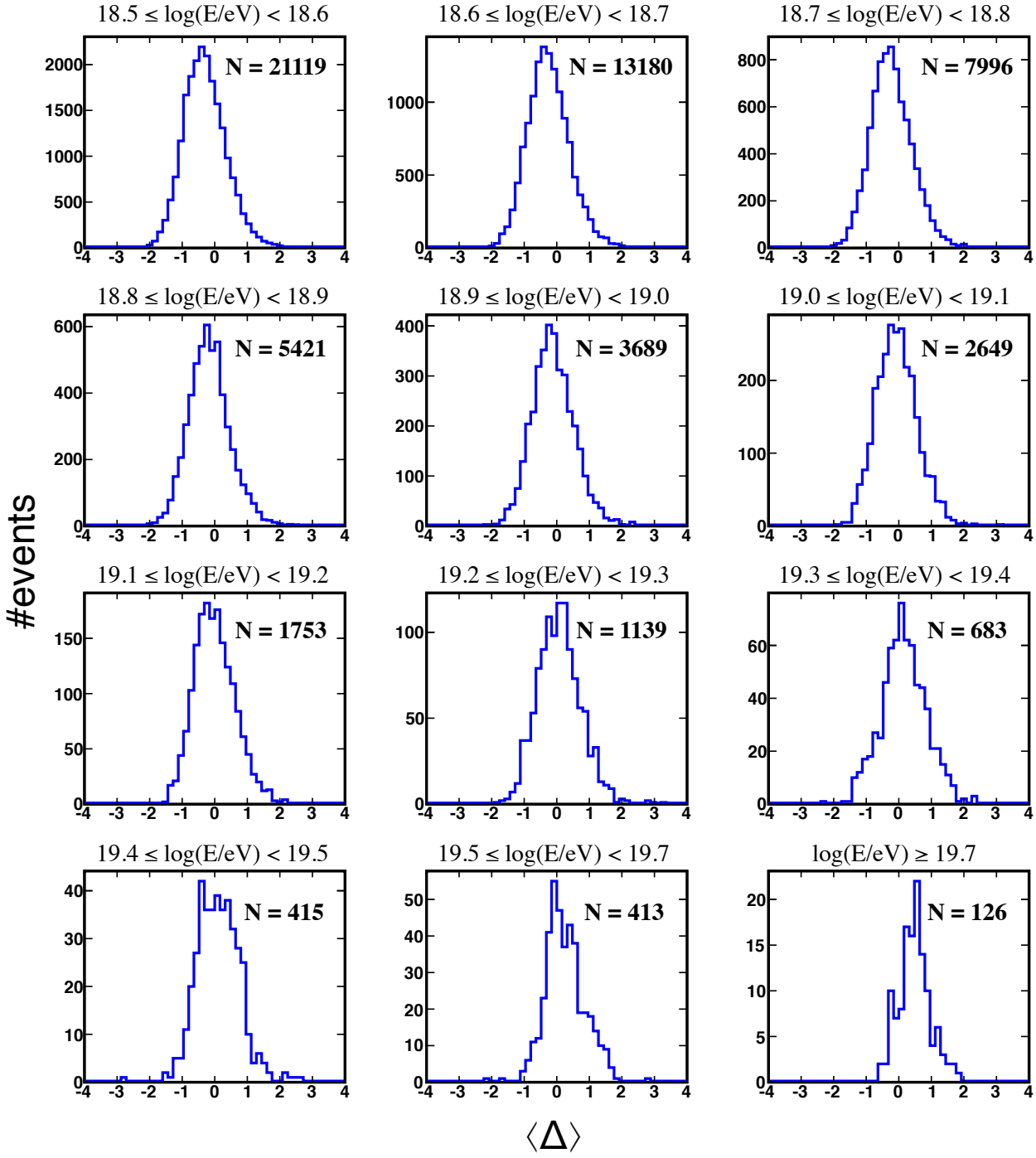
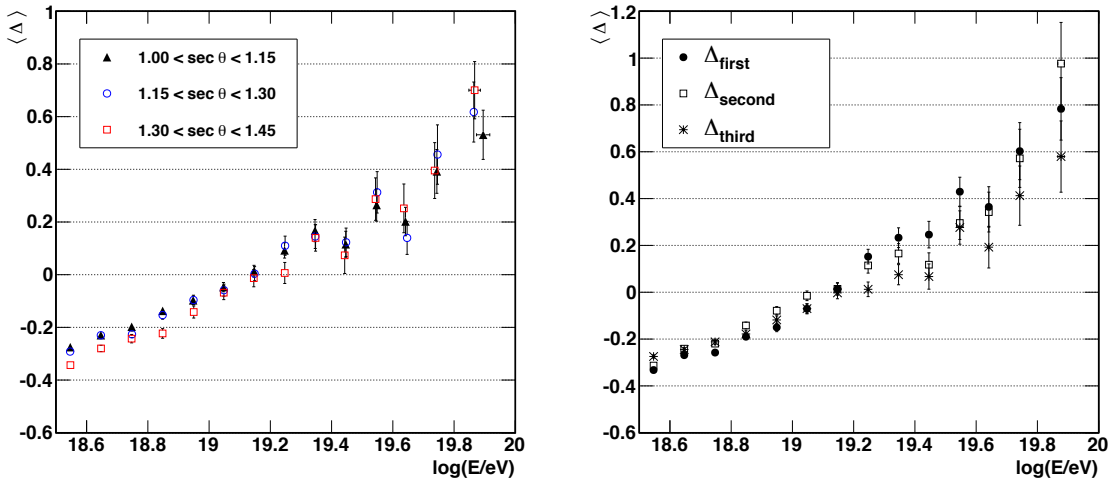


Figura 5.3: $\langle\Delta\rangle$ distributions for the different energy intervals.

with the largest signal of an event, Δ_{second} is the value obtained if we use the detector with the second largest signal and Δ_{third} with the third largest signal. In this way, with Δ_{first} we study the contributions to $\langle\Delta\rangle$ coming from distances close to the core, with Δ_{second} the intermediate distances and with Δ_{third} the farthest distances. We select only the detectors with the three largest signals because our selected events have as least three detectors surviving the selection cuts.

As it is shown in figure 5.4b, the mean values of Δ_{first}, Δ_{second} and Δ_{third} behave in the same way as a function of energy. This is the expected trend and it confirms that $\langle\Delta\rangle$ does not have any dependence on distance in the whole energy range selected for this analysis.



(a) $\langle\Delta\rangle$ vs E split in three sec θ bins.

(b) Δ_{first}, Δ_{second} and Δ_{third} vs E.

Figura 5.4: Study of the dependences of $\langle\Delta\rangle$ with the sec θ and with core distance.

5.3 Calibration with Golden Hybrid events

Once the $\langle\Delta\rangle$ values are completely determined, the next step in our mass composition analysis is the calibration. The goal of the calibration is to find the function which can be used to convert the values of $\langle\Delta\rangle$ into values of X_{max}. The only feasible way to do that without simulations is using the Golden Hybrid events measured by the Pierre Auger Observatory. These events are of high quality, since they are reconstructed with the two independent detectors of the Observatory, the FD and SD. This means that these events can be characterized simultaneously and in a independent way with the two key observables of this thesis, X_{max} and $\langle\Delta\rangle$, and therefore they provide the way to obtain a relationship between them.

5.3.1 Selection of Golden Hybrid events

For the selection of Golden Hybrid events we take the data sample used in [78] to build an elongation rate exclusively with X_{max} measurements obtained with fluorescence telescopes. Doing that we select Golden Hybrid events from a data set which has been previouly cut with the fiducial cuts that guarantee a constant efficiency over a wide range of X_{max} and with the necessary quality cuts to assure a good X_{max} resolution. The selection from such a strict data set restricts our statistics for the calibration procedure, but in this way we ensure that we are not using a biased sample.

The number of Golden Hybrid events included in the data set used in [78] together with the cuts that we have to apply to assure we obtain a meaningful value of $\langle\Delta\rangle$ are shown in table 5.3. After applying these cuts we have 885 Golden Hybrid events to study the correlation between X_{max} and $\langle\Delta\rangle$. Note that we have to apply two cuts in energy, one in terms of the calorimetric energy measured by the FD, E_{FD}, and a second one in terms of the energy obtained by the SD, E_{SD}. This happens because both energies are slightly different and if we do not require the cut on E_{SD} we could find Golden Hybrid events with E_{SD} values as small as log(E_{SD}/eV) = 18.0. This is to be avoided because our selected data set required energies in excess of 18.5 (see section 5.1). The selected Golden Hybrid events contained in each energy bin are listed in table 5.4.

Quality Cuts	**Events**	**Efficiency**
FD events used in [78]	19 759	100 %
Golden Hybrid events	12 825	65 %
sec θ <1.45	9625	49 %
6T5 Trigger	7361	37 %
At least 3 detectors	4025	20 %
log(E_{FD}/eV) $\geq$ 18.5	978	5 %
log(E_{SD}/eV) $\geq$ 18.5	**885**	4.5 %

Cuadro 5.3: Quality cuts and efficiencies for the Golden Hybrid events used in this chapter.

To check that we do not introduce a bias in terms of X_{max} with the cuts listed in table 5.3 we compare the mean vales of X_{max} of our selected Golden Hybrid events with the values given in [78]. This comparison is shown, before applying the cut on E_{SD}, in the left panel of figure 5.5. The right panel of the same figure shows the comparison after the cut on the energy measured by the SD. The mean values of X_{max} for our 885 Golden Hybrid are compatible with the values given in [78] above log(E_{FD}) >18.7. Below this energy the values of X_{max} for the Golden Hybrid events are below the values measured by FD. This is a consequence of the migrations between bins, since we have a cut on E_{SD} and the elongation rates shown in the two panels of figure 5.5 are built in terms of E_{FD}. In fact, note that the elongation rate of the Golden Hybrid events before the cut on E_{SD} (left panel of figure 5.5) it is not biased at all. This effect does not have a direct consequence in our calibration procedure because we do the calibration on a event-by-event basis using the

whole energy range. The effect of the migrations is only relevant when we have to plot the elongation rates, since making the bins in terms of E_{FD} or in terms of E_{SD} introduces a slight difference. This difference will be evaluated in a forthcoming section.

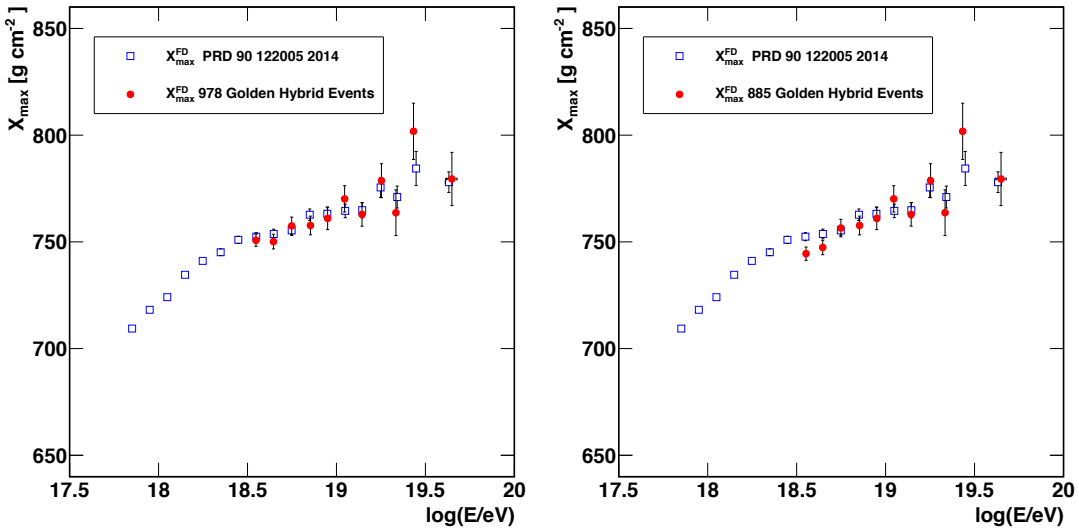


Figura 5.5: Mean values of the $X_{\max}$ distributions obtained for our selection of Golden Hybrid events (red dots). These values are compared to the ones obtained for the whole FD sample used in [78] (blue squares). Left (right) panel: Before (after) the cut $\log(E_{SD}/eV) \geq 18.5$.

Energy bin	Events	Energy bin	Events
$18.5 \leq \log(E_{FD}/eV) < 18.6$	214	$19.1 \leq \log(E_{FD}/eV) < 19.2$	43
$18.6 \leq \log(E_{FD}/eV) < 18.7$	205	$19.2 \leq \log(E_{FD}/eV) < 19.3$	31
$18.7 \leq \log(E_{FD}/eV) < 18.8$	129	$19.3 \leq \log(E_{FD}/eV) < 19.4$	14
$18.8 \leq \log(E_{FD}/eV) < 18.9$	103	$19.4 \leq \log(E_{FD}/eV) < 19.5$	9
$18.9 \leq \log(E_{FD}/eV) < 19.0$	73	$\log(E_{FD}/eV) \geq 19.5$	10
$19.0 \leq \log(E_{FD}/eV) < 19.1$	54		

Cuadro 5.4: Number of the selected Golden Hybrid events listed by energy bin.

5.3.2 Correlation and calibration

To make a clear distinction between the observable $X_{\max}$ measured directly by the fluoresce detectors and the estimator that we obtain in this section we introduce a new name. $X_{\max}$ or $X_{\max}^{FD}$ is the measurement obtained with the fluorescence telescopes whereas $X_{\max}^{\text{Delta}}$ is the estimator obtained after the calibration and which will be used as a surrogate of $X_{\max}$ to study mass composition with SD-only events.

Once we have selected the Golden Hybrid events we can study the level of correlation between X_{max} and $\langle\Delta\rangle$. In figure 5.6 we show the values of X_{max} as a function of $\langle\Delta\rangle$. The value that we have found for the correlation coefficient is 0.46. This value is approximately the mean value of the correlation coefficients that we obtained for simulation of protons (0.57) and iron nuclei (0.31) in chapter 4 with 3 detectors. This fact is an indication that the correlation coefficient of these two observables could be also a good estimator for mass composition analyses. However this study falls outside of the goals of this thesis.

For the calibration of X_{max} in terms of $\langle\Delta\rangle$ we assume that the relationship between both observables can be described with the equation:

$$X_{\mathrm{max}} = a + b\langle\Delta\rangle + c\log(E_{\mathrm{SD}}/\mathrm{eV}) \tag{5.1}$$

where a, b and c are the calibration parameters. The term which dominates the calibration procedure is that which contains $\langle\Delta\rangle$ while the term with the energy is introduced only to accommodate the energy dependence of both observables. We have seen that this term has a minor influence in the calibration and it only smooths the conversion of $\langle\Delta\rangle$ into $X_{\mathrm{max}}^{\mathrm{Delta}}$.

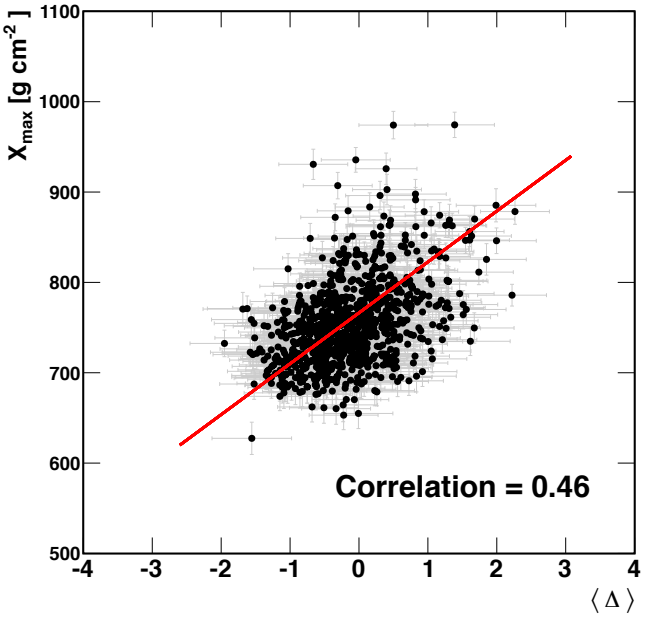


Figura 5.6: X_{max} vs. $\langle\Delta\rangle$ for the 885 Golden Hybrid events that we have selected. The red line is a projection in 2D of the calibration curve obtained with the maximum likelihood method.

The uncertainty used for the X_{max} values is the parameterization given in [78] for the resolution of X_{max}. This parameterization, given as a function of energy, takes into account all the uncertainties affecting the hybrid reconstruction. This includes statistical

and systematic uncertainties as the atmospheric model, the effect of aerosols or the effect of the FD calibration.

The energy measured by the SD is used in the calibration as the shower energy. It is well known that the energy measured by the FD is more accurate than the one measured by the SD, but our goal is to obtain an estimator for X_{max} suitable for SD-only events. This means that at the end of our analysis the only available information comes from the SD. Thus it is more sensible to do the calibration using only SD observables.

The parameters a, b and c are obtained fitting the observables X_{max}, $\langle\Delta\rangle$ and E_{SD} according to equation 5.1. To perform the estimation of these parameters we do the fit using a *maximum likelihood method* [147]. This method is only applicable if the form of the theoretical distribution of the sample is known. In our case we can apply this statistical method because it is known that the X_{max} values follow approximately a Gumbel distribution. The pdf which describes the Gumbel distribution is:

$$f(z) = \frac{ze^{-z}}{\beta} \qquad \text{with} \qquad z = e^{-\frac{x-\mu}{\beta}} \tag{5.2}$$

To find the parameters a, b and c we have to maximize the function:

$$\mathcal{L}(\phi|x) = f(x_1|\phi)f(x_2|\phi)f(3_2|\phi)....f(x_n|\phi) \tag{5.3}$$

where f is the theoretical distribution, x_i is each one of our independent observations (e.g. each Golden Hybrid event) and ϕ represents the set of parameters that we want to estimate.

We use the maximum likelihood method instead of the common least squares method because the first one is more reliable when the errors of the variables are large, especially in the dependent variables, as in the case of $\langle\Delta\rangle$. Besides, the maximum likelihood method allows us to fix the treatment of the errors in the whole maximization procedure.

The values that we obtain for the parameters a, b and c after the maximization procedure are shown in table 5.5 with their respective errors. Since the calibration procedure is done with three variables, the display in 3D of our events together with the surface defined by the values of the parameters a, b and c is not very illustrative. For this reason we show in figure 5.6 a projection in 2D of the calibration curve (red line).

Calibration Parameters	**Values** $\pm$ **Errors**
a	699 $\pm$ 12 g cm^{-2}
b	56 $\pm$ 3 g cm^{-2}
c	3.6 $\pm$ 0.7 g cm^{-2}

Cuadro 5.5: Calibration parameters to obtain $X_{\mathrm{max}}^{\mathrm{Delta}}$ from the values of $\langle\Delta\rangle$ and E_{SD}.

The errors of the calibration parameters have been obtained using a bootstrap method. With this method we essentially produce different samples coming from the original one using sampling with replacement. The new samples have the same number of events as the original one. In this way we can repeat the calibration procedure a huge number of

times and study the distributions of the parameters a, b and c . The RMS values of these distributions are taken as the errors of the calibration parameters.

The goodness of the fit obtained with the maximum likelihood method is checked in the left panel of figure 5.7. This figure shows the distribution obtained for the differences between the estimated and the measured values of X_{max}. In other words, the difference between $X_{\mathrm{max}}^{\mathrm{Delta}}$ and X_{max}. The mean value of this distribution is (4 $\pm$ 2) g cm^{-2}, nearly compatible with zero, and with a RMS value of 44 g cm^{-2}. In the right panel of figure 5.7 we plot the difference between $X_{\mathrm{max}}^{\mathrm{Delta}}$ and X_{max} as a function of sec θ. The flat trend of this dependence indicates that calibration procedure does not introduce any bias in terms of sec θ.

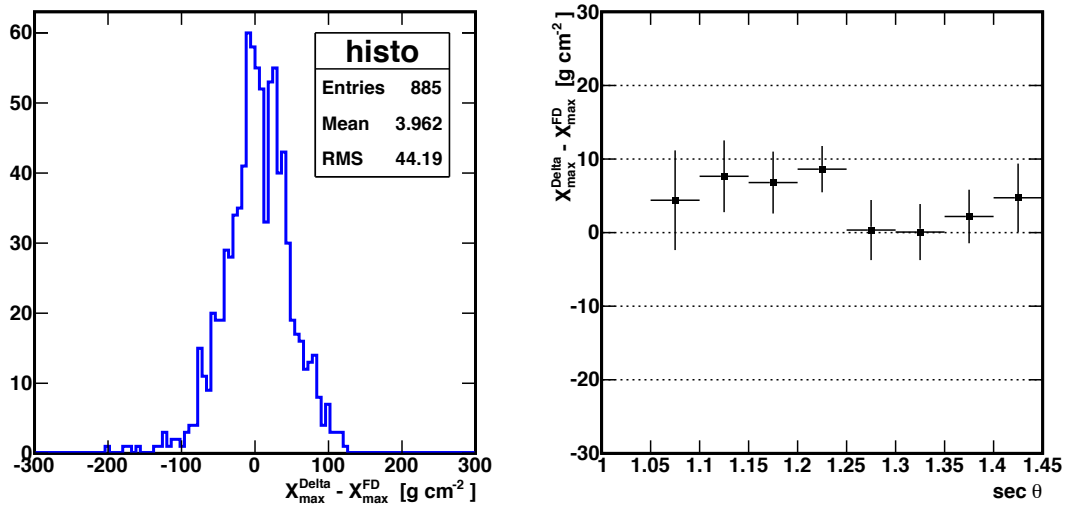


Figura 5.7: Left panel: Differences between $X_{\mathrm{max}}^{\mathrm{Delta}}$ and X_{max} for the 885 Golden Hybrid events used in the calibration procedure. Right panel: Same differences as a function of sec θ.

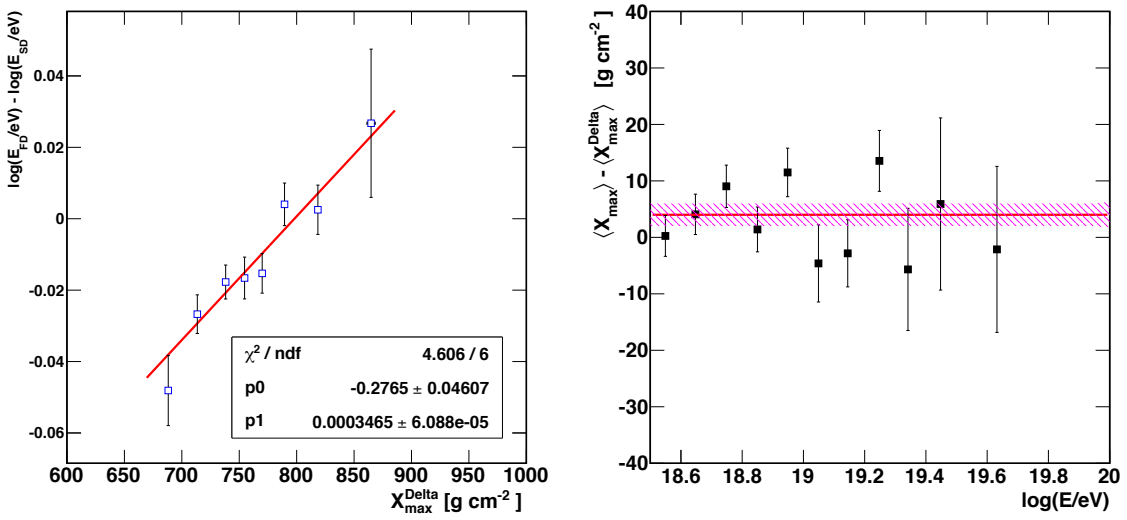
5.4 Energy correction

It was already mentioned that for a given Golden Hybrid event the energy measured by the FD and the SD is in general different. This effect acquires a significant importance in our analysis because one of our goals is to obtain an elongation rate with SD-only events. This makes necessary a common energy to do the comparison. Otherwise the event migrations between bins could introduce distortions in the elongation rate and convey the impression that the results do not match perfectly.

To obtain a common energy to compare the different elongation rates it is necessary to correct the values of the energy measured by the SD. Taking into account the way in which the SD energy is measured in the Pierre Auger Observatory through the value of S(1000) we expect that the difference between both energies will depend on the depth of the

shower, i.e $X_{\rm max}$. Obviously the difference between both energies has to be studied again with Golden Hybrid events, since these are the only events which have a measurement for the two energies.

In figure 5.8a we show for our sample of 885 Golden Hybrid events the difference between the two energies as a function of $X_{\rm max}^{\rm Delta}$. We had also the possibility of doing this study in terms of $X_{\rm max}$, but we considered more convenient the use of the estimator. We look for a correction for the SD energy which will be applied in SD-only events and these events are not characterized by $X_{\rm max}$, but by $X_{\rm max}^{\rm Delta}$.



(a) (b)

Figura 5.8: Left panel: Difference between log(E_{FD}/eV) and log(E_{SD}/eV) as a function of $X_{\rm max}^{\rm Delta}$, fitted with a linear function (red line). Right panel: Differences between the mean values of $X_{\rm max}$ and $X_{\rm max}^{\rm Delta}$ after the energy correction as a function of energy. The mean value of these differences is indicated as a red line.

Figure 5.8a shows also the linear fit that we have used to parameterize the difference between both energies as a function of $X_{\rm max}^{\rm Delta}$. Using this linear fit the final expression to obtain a corrected value for the SD energy is

$$\log(E_{\rm corrected}/{\rm eV}) = \log(E_{\rm SD}/{\rm eV}) + p_0 + p_1 X_{\rm max}^{\rm Delta} \tag{5.4}$$

being p_0 = (-0.28 $\pm$ 0.05) and p_1 = (0.00035 $\pm$ 0.00006) g^{-1}cm^2. This correction for the SD energy will be used hereafter to obtain the elongation rate with data of the 1500 m array.

In figure 5.8b we show the differences between the mean values of $X_{\rm max}$ and $X_{\rm max}^{\rm Delta}$ after the energy correction as a function of energy. The average difference along the whole energy range is (4 $\pm$ 2) g cm^{-2}. This value is compatible with the one observed in case of using the energy measured by the FD, (3 $\pm$ 2) g cm^{-2}.

5.5 Systematic uncertainties

We now turn our attention to identify and evaluate the contributions of the different sources of systematic uncertainties that influence the measurement of $X_{\text{max}}^{\text{Delta}}$. In what follows we discuss the most relevant contributions:

- **Uncertainty on the calibration.** The uncertainties on the calibration parameters introduce a systematic uncertainty in our final results. To calculate this uncertainty we have to propagate the uncertainties of the calibration parameters shown in table 5.5. This propagation has to take into account the correlation of the parameters a, b and c since they are not independent. The correlation coefficients found are ρ_{ab} = -0.20, $\rho_{ac} = -0{,}97$ and $\rho_{bc} = 0{,}34$. After the propagation of these uncertainties we find that the maximum difference in $X_{\text{max}}^{\text{Delta}}$ goes to 3 g cm^{-2} at the lowest energies up to 5 g cm^{-2} at the highest ones. To keep the most conservative scenario we consider a systematic uncertainty of 5 g cm^{-2} for the whole energy range.

- **Seasonal effect.** Data show that $X_{\text{max}}^{\text{Delta}}$ has a dependence of unknown origin with seasons. This effect is shown in figure 5.9a where it is evident that $X_{\text{max}}^{\text{Delta}}$ is deeper in winter. This figure corresponds to the years 2012 and 2013. Data are grouped in four seasons and the behaviour is repeated three times in the plot to have a clearer sight of the effect and its periodicity. The same effect is observed in the rest of the years studied, being always the events deeper in winter. The amplitudes found for the rest of the years are similar to the one shown in figure 5.9a, this one being the largest. We have not identified the source of this behaviour although the effect is relatively small. Therefore we include the amplitude of this effect, 2 g cm^{-2}, as a systematic uncertainty.

- **UTC dependence.** We have also investigated the contribution to the systematic uncertainty due to the UTC time. However we find that this effect is quite small since the difference between the $X_{\text{max}}^{\text{Delta}}$ mean values during the night and day is lower than 1 g cm^{-2}, see figure 5.9b.

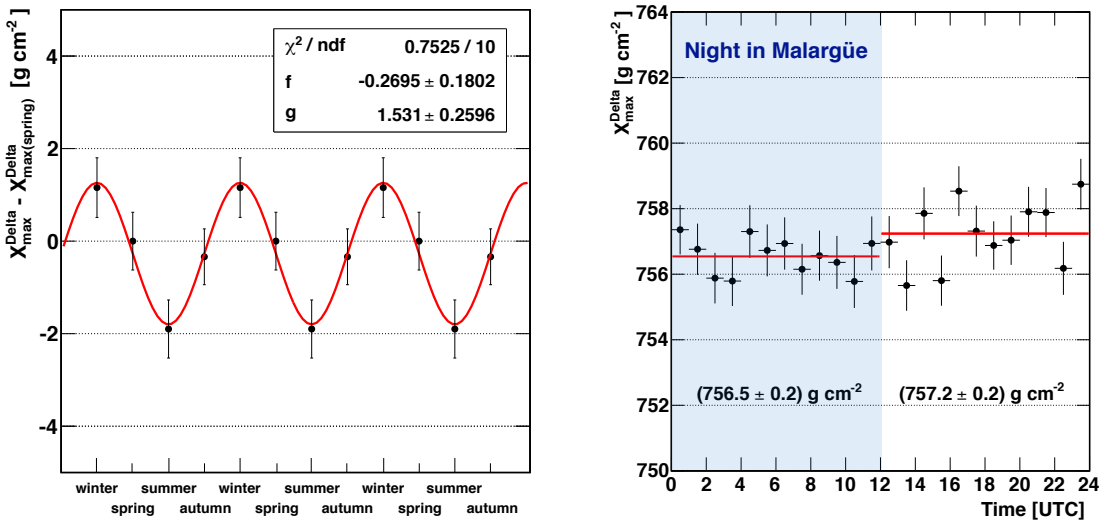
- **Ageing.** $X_{\text{max}}^{\text{Delta}}$ exhibits a dependence with the years, see left panel of figure 5.10. This indicates an ageing effect of the surface detectors which has already been observed before [148]. We discard that this effect is a consequence of the calibration procedure or of the calculation of $\langle\Delta\rangle$ because the same trend is observed when we study the mean values of the risetimes as a function of the years. To estimate this systematic uncertainty we fit separately with a constant data from 2005 to 2010 and from 2012 to 2014 (right panel of figure 5.10). We take as a systematic uncertainty the half difference of these values. In this case the systematic uncertainty is 3 g cm^{-2}.

- **Dependence on sec θ.** Figure 5.11 shows $X_{\text{max}}^{\text{Delta}}$ values of the whole data set used in this work as a function of sec θ. The behaviour is almost flat between 1.00 <sec θ <1.30, then there is a slight gap and after, between 1.30 <sec θ <1.45, the trend is flat again. This decrease was pointed out in section 5.2.1 when we studied the behaviour of $\langle\Delta\rangle$ as a function of energy for different sec θ bins. The half difference

between the mean values of $X_{\mathrm{max}}^{\mathrm{Delta}}$ in the two indicated ranges is taken as the contribution to the systematic uncertainty. It amounts to 1.5 g cm^{-2}.

- **FD systematic uncertainty.** The systematic uncertainty of X_{max} propagates directly into the values of $X_{\mathrm{max}}^{\mathrm{Delta}}$. This systematic uncertainty is given in [78] as a function of the energy. In our energy range (18.5 $\leq$ log(E/eV) <20.0) it goes from 7 to 10 g cm^{-2}. For this reason we take as a systematic uncertainty 8.5 g cm^{-2}, the mean of these two values. This systematic uncertainty contains the contribution coming from the energy scale, for this reason this effect is not studied separately in our list of systematic uncertainties.

Table 5.6 summarizes all the contributions to the systematic uncertainty affecting our $X_{\mathrm{max}}^{\mathrm{Delta}}$ measurement. The overall value amounts to 11.0 g cm^{-2}. The dominant contribution to the systematic uncertainty comes from the FD.



(a) Seasonal effect (2012-2013)

(b) $X_{\mathrm{max}}^{\mathrm{Delta}}$ as a function of sec θ

Figura 5.9: Left panel: Effect of the seasonal variation on the values of $X_{\mathrm{max}}^{\mathrm{Delta}}$. Right panel: $X_{\mathrm{max}}^{\mathrm{Delta}}$ as a function of the UTC time. The shadow zone indicates the duration of the night in Malargüe. The lines indicate the fits with a constant for the different periods of the day.

5.6 The Elongation Rate

At this point of our analysis we have all the necessary information to study the evolution of $X_{\mathrm{max}}^{\mathrm{Delta}}$ with energy. This is shown in figure 5.12. Events are grouped according to their log(E/eV) value in bins of width 0.1, excepting the last energy bin (starting at log(E/eV) = 19.8) which is an accumulative bin containing all events above this energy. The shadow zone represents the systematic uncertainties.

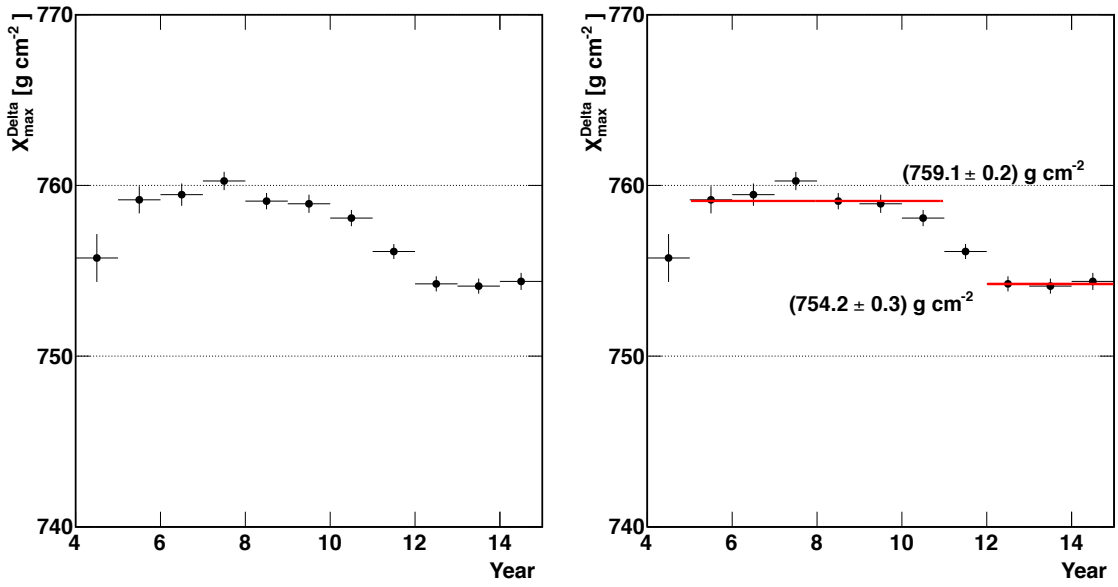


Figura 5.10: Left panel: Aging effect reflected in the $X_{\text{max}}^{\text{Delta}}$ values. Right panel: Same as the left panel including the fits with a constant.

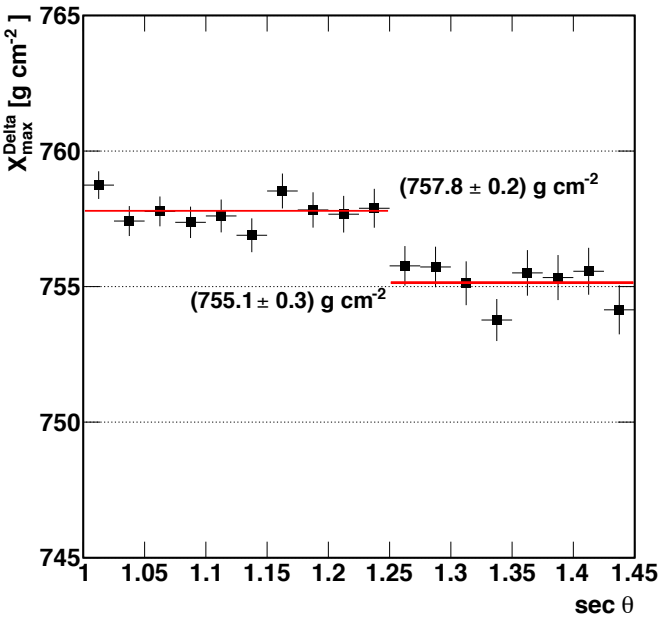


Figura 5.11: $X_{\text{max}}^{\text{Delta}}$ as a function of sec θ for the whole data set used in this work. The lines indicate the fits with a constant for the two behaviours observed in the data sample.

Source	**Sys. uncertainty** [g cm^{-2}]
Uncertainty on calibration	5.0
Seasonal effect	2.0
UTC dependence	1.0
Ageing	3.0
Dependence on sec θ	1.5
FD systematic uncertainty	8.5
Total	11.0

Cuadro 5.6: Contributions to the systematic uncertainty of the $X_{\text{max}}^{\text{Delta}}$. They have been added in quadrature.

This result is compared to the measurements of X_{max} obtained by the FD, shown and explained in chapter 1 [78]. The measurements obtained with the $\langle\Delta\rangle$ method are compatible with the FD results in the energy range where they overlap. Furthermore, we add three more bins at the highest energies because our measurements are done with the SD, where the statistics is larger. This means that we have 517 events above log(E/eV) = 19.5 whereas the FD has only 37 events above this energy. A factor 14 bigger.

Our measurements are also compared in figure 5.12 to the predictions given by the latest version of the packages QGSJetII.04 and EPOS-LHC, which have been tuned to reproduce the most recent results published by the experiments running at the LHC. Our data suggest the flux of cosmic rays is composed mainly of light nuclei with a fraction of heavy nuclei increasing with energy up to $10^{19,6}$ eV. This is the same conclusion found from the FD analysis in [78]. However, above $10^{19,6}$ eV our last energy bins point to an increase of the light component again.

The $X_{\text{max}}^{\text{Delta}}$ values as a function of energy are split into three bins of sec θ in the top panel of figure 5.13. No dependence in sec θ is observed. The number of events used in each energy bin is shown in the bottom panel of this figure. The numbers do not match with the numbers indicated in the distributions of figure 5.3 because the use of the energy correction explained in section 5.4 introduces some migrations between the energy bins. This fact modifies slightly the number of events in each energy bin.

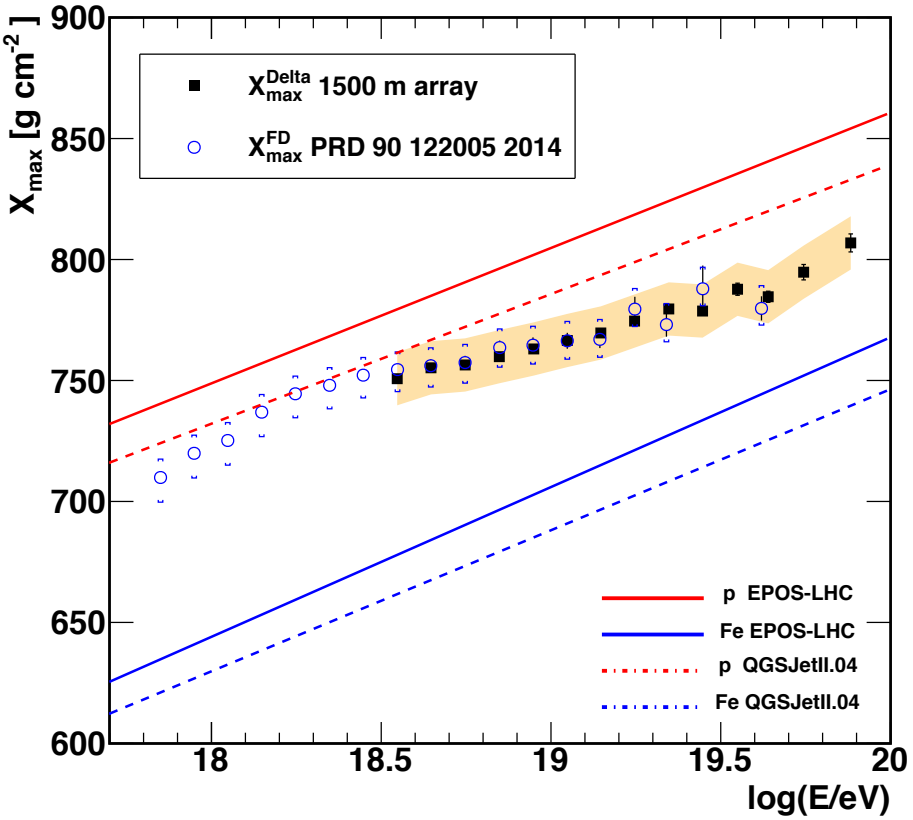
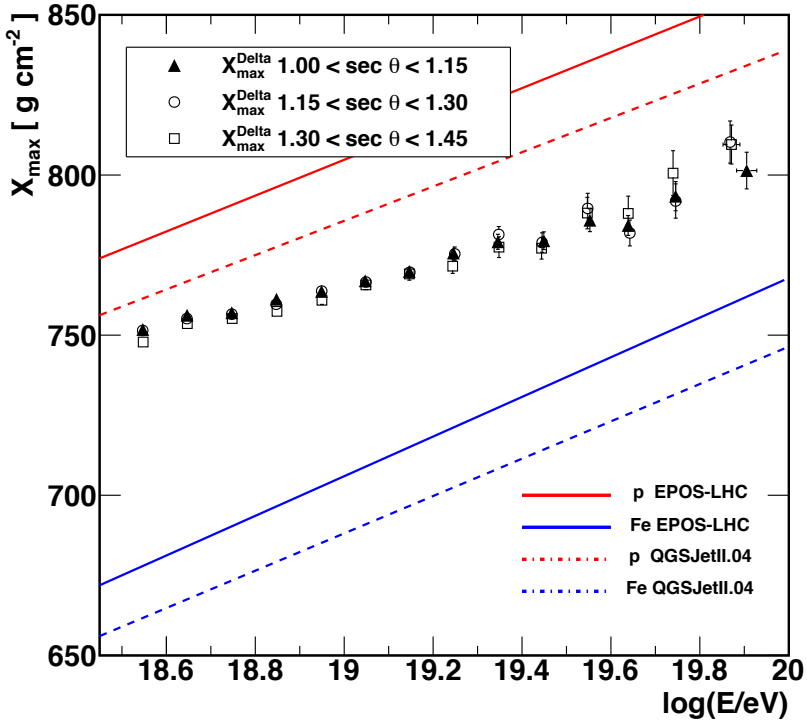
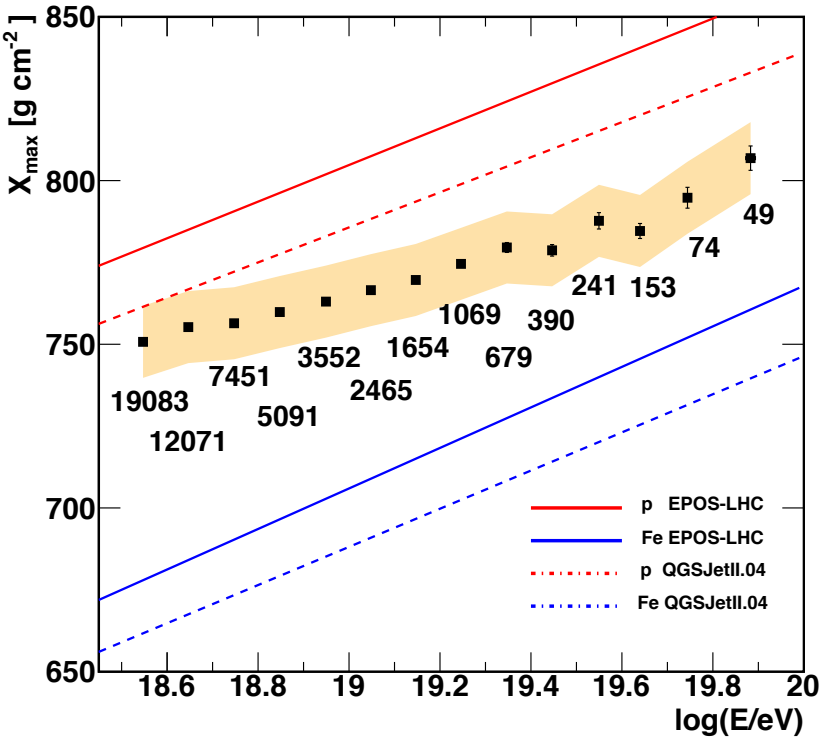


Figura 5.12: Mean values of the $X_{\text{max}}^{\text{Delta}}$ distributions as a function of the energy obtained with data measured by the 1500 m array. The shaded area indicates the systematic uncertainties. Data are compared to the predictions from simulations of protons and iron nuclei and to the measurements of X_{max} done with the FD.



(a)



(b)

Figura 5.13: Top panel: $X_{\mathrm{max}}^{\mathrm{Delta}}$ elongation rate split in three bins of sec θ. Bottom panel: Zoom of the $X_{\mathrm{max}}^{\mathrm{Delta}}$ elongation rate with the number of events used in each energy bin.

6

Mass composition results with the data of the 750 m array

This chapter is focused on the study of the data collected with the 750 m array using the $\langle\Delta\rangle$ method. The procedure followed in this case to apply the $\langle\Delta\rangle$ method in the new data set is exactly the same as the one explained in the previous chapter.

The goal now is to extend to lower energies the elongation rate presented in the previous chapter. Certainly, the most interesting region to study the mass composition of cosmic rays, in the sense of probing an extended energy regime where the FD cannot reach, was covered in the previous chapter. However, the extension to lower energies using the same method in a fully independent data set has also an enormous importance. The agreement between both results would give additional support to the method and prove the robustness of the analysis over nearly three orders of magnitude in energy.

In addition, the importance of this chapter lies in the fact that this is the first mass composition analysis done with the data of the 750 m array.

6.1 Data selection

The cuts used to select data of the 750 m array were previously introduced in chapter 4 because they were required to obtain the data sample used for the benchmark calculation. For this reason in this section we only list them and quote their selection efficiencies, see table 6.1. These data correspond to a period from January 2008 to December 2014. This period of data taking is different from the one used in the 1500 m array because the 750 m array was built four years later.

In table 6.1 we find two very stringent quality cuts. The first one is the cut which requires sec θ <1.30. As we mentioned in the previous chapter this cut is necessary to avoid inclined showers with too small risetimes. In this case, given that we study lower energies the risetimes are proportionally smaller than the ones measured in the 1500 m

Quality cuts	**Events**	**Efficiency**
Total events with 17.5 $\leq$ log(E/eV) <18.5	159 795	100 %
sec θ <1.30	72 907	45.6 %
6T5 Trigger	29 848	18.7 %
Reject lightning events	29 848	18.7 %
Reject bad periods and comms crisis	28 773	18.0 %
At least 3 detectors to calculate $\langle\Delta\rangle$	**28 570**	**17.9** %

Cuadro 6.1: Quality cuts applied to data of the 750 m array with their selection efficiency.

array. Due to this, the limitation imposed by the timing resolution of the surface detectors starts to be important for smaller zenith angles and we are forced to use a stricter cut. The second stringent cut is the one which requires a 6T5 level of trigger. The goal of this trigger is to reject events too close to the border of the array and, indeed, it is very strict for an array with a total area of only 23.5 km^2. But if we do not apply this cut we introduce in our data set events whose footprint is not totally contained in the surface of the array and this leads to a faulty event reconstruction. The rejection of lighting events, bad periods and communtication crisis, as in the 1500 m array, has a negligible effect.

Again, as in the case of the 1500 m array, for the calculation of $\langle\Delta\rangle$ we use at least 3 detectors surviving the cuts listed in table 6.2.

Cuts	**Detectors**	**Efficiency**
Detectors before requiring at least 3 det. per event	216 856	100 %
Reject low-gain saturated	211 701	97.6 %
S >3 VEM	201 566	92.9 %
300 m <r <800 m	114 100	52.6 %
$t_{1/2}$ >40 ns	114 075	52.6 %
After requiring at least 3 det. per event	**113 661**	**52.4** %

Cuadro 6.2: Number of detectors surviving the detector-level cuts for the 750 m array.

Note that in this case we have un upper limit for the energy in table 6.1. The first motivation for using this limit is that the elongation rate obtained for data of the 1500 m array starts at log(E/eV) = 18.5 and this new analysis intends to be an extension of the former one at the lowest energies. Besides, only 2 % of events gathered by the 750 m array are above that energy. The second motivation is that in this energy range we guarantee the criterion of linearity explained in section 4.3.1 with only one upper limit for the core distances of the surface detectors required for this analysis (800 m). Note that in the 1500 m array for energies larger than $10^{19,6}$ eV it is necessary to modify the upper limit for the core distances to avoid introducing a bias. In any case, the remaining 2 % of events with energies larger than $10^{18,5}$ eV could be studied with the $\langle\Delta\rangle$ method if the criterion of linearity is satisfied (see section 7.2).

Figure 6.1 shows, for two energy bins, the distribution of the number of detectors per

event which survive the detector-level cuts listed in table 6.2. The shaded bars of these histograms indicate the events that will be used in the $\langle\Delta\rangle$ method. Comparing this figure with its counterpart from the previous chapter, it is evident that for the 750 m array we use more detectors per event for the calculation of $\langle\Delta\rangle$ (in spite of studying events with lower energies). This is a consequence of having a denser array of surface detectors.

These cuts have also been tested in simulations of protons and iron nuclei to check if they introduce a selection bias in our data set. The differences between the selection efficiencies for protons and iron nuclei are never larger than 1 $\%$.

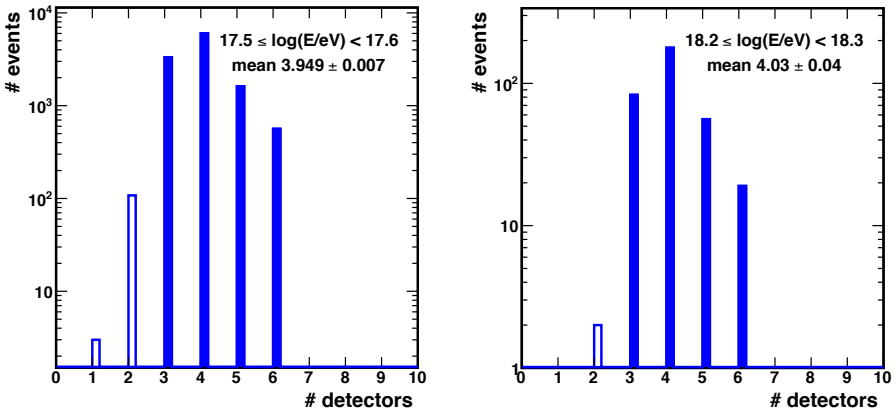


Figura 6.1: Number of detectors per event surviving the cuts given in table 6.2. The color bars indicate the selected events after requiring at least 3 detectors per event.The mean of the number of detectors after this cut is also indicated.

6.2 Delta as a function of the energy

The first $\langle\Delta\rangle$ values for the 750 m array were obtained in chapter 4 for events belonging to the benchmark energy bin (17.7 <log(E/eV) <17.8) to check the goodness of the benchmark parameterization. Now, following the same procedure, we calculate $\langle\Delta\rangle$ for each one of the events that we have selected in the 750 m array. Just as a reminder, we calculate the $\langle\Delta\rangle$ values as we did in the previous chapter, but taking the benchmark parameterization for the 750 m array given in section 4.4.3. The mean values of the $\langle\Delta\rangle$ distributions as a function of the energy are given in figure 6.2, where the uncertainties shown are of statistical nature.

As we anticipated in chapter 4 the mean value of $\langle\Delta\rangle$ in the benchmark energy bin is compatible with zero. The behaviour of $\langle\Delta\rangle$ with energy shows again an increasing trend, as expected, since more energetic events are deeper and therefore they have larger risetimes. Although it is worth noting that the trend is steeper than in the case of the 1500 m array.

The distributions of $\langle\Delta\rangle$ are shown in figure 6.3 together with the number of events that we use in each energy bin.

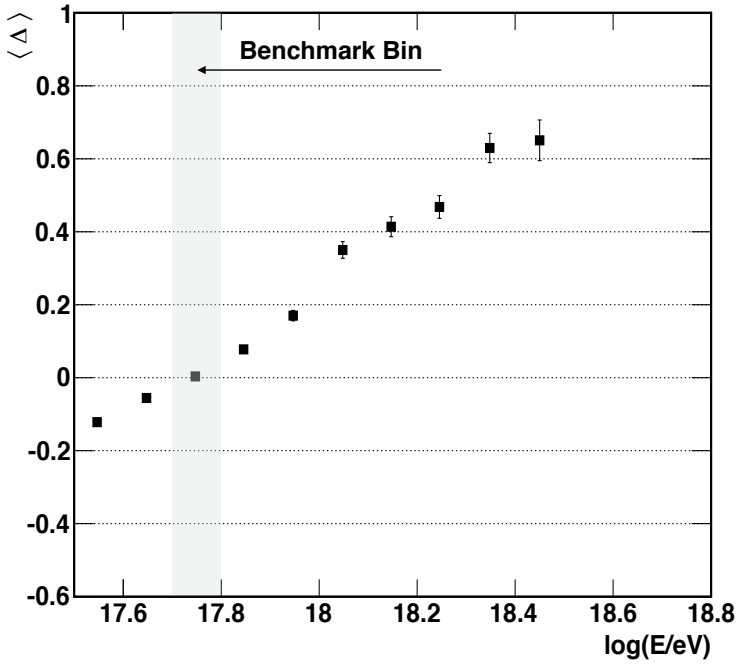


Figura 6.2: $\langle\Delta\rangle$ as a function of energy for data gathered with the 750 m array.

6.2.1 Some cross-checks in the behavior of $\langle\Delta\rangle$

As we discussed in the previous chapter, the benchmark parameterization $t_{1/2}^{bench}(r,\theta)$ contains the whole dependence of the risetimes with sec θ and with core distance. For this reason, no dependence on sec θ or on distance is expected for the behaviour of $\langle\Delta\rangle$ as a function of energy for the data of the 750 m array.

Figure 6.4a shows the behaviour of $\langle\Delta\rangle$ as a function of energy for two bins of sec θ. The trend with the energy is the same for the two bins of sec θ, confirming the above statement. The dependence with the distance is studied in figure 6.4b. The observables Δ_{first}, Δ_{second} and Δ_{third} were introduced in the previous chapter. The fact that the three of them exhibit the same increasing trend with the energy indicates that $\langle\Delta\rangle$ does not have any dependence on distance in the whole energy range selected for this study.

6.3 Calibration with Golden Hybrid events

Following the procedure explained in the previous chapter, the next step in the analysis is the calibration. The Golden Hybrid events used previously were selected from the data sample used in [78] to study the evolution of X_{max} with energy. In this case it is not possible to use the same data sample although it contains events with energies up to $10^{17,8}$ eV. The reason is that the events used in [78] were measured by the 24 fluorescence detectors which oversee the 3000 km^2 of 1500 m array, so the footprints of these events

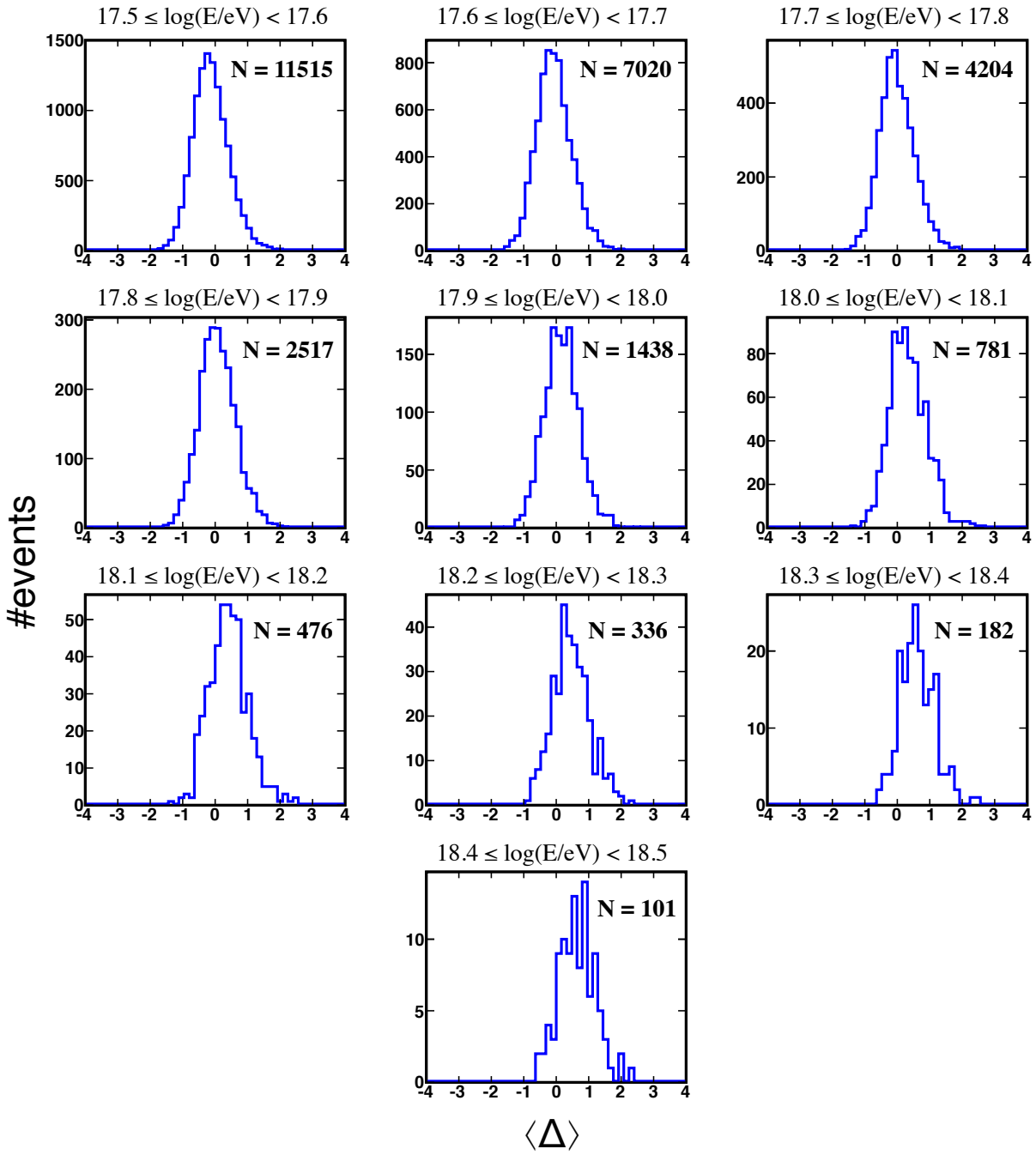
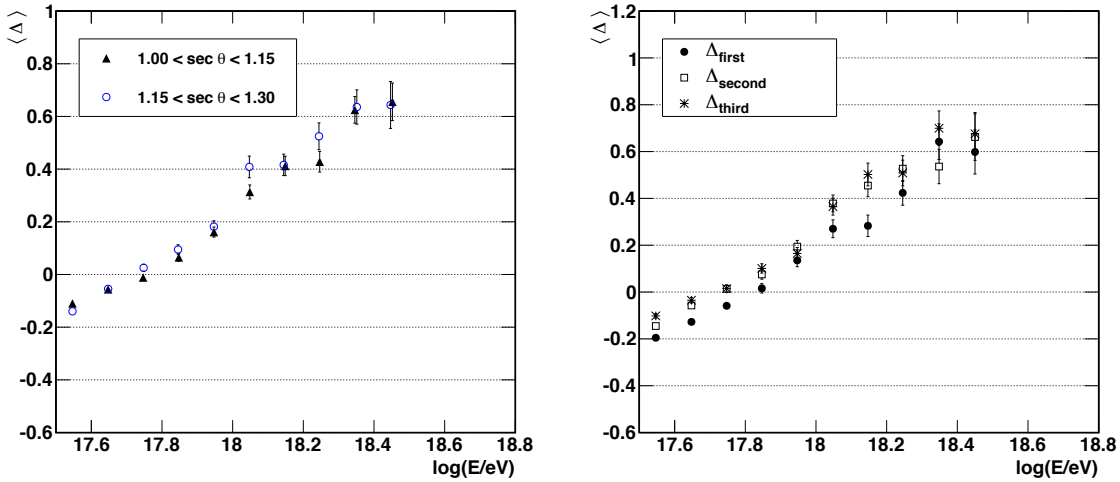


Figura 6.3: $\langle \Delta \rangle$ distributions for the different energy intervals.



(a) $\langle\Delta\rangle$ vs E split in two sec θ bins.

(b) Δ_{first}, Δ_{second} and Δ_{third} vs E.

Figura 6.4: Study of the dependences of $\langle\Delta\rangle$ with the sec θ and with the core distance.

are distributed over this huge area. This implies that the probability of finding showers falling exactly over the 23.5 km^2 of the 750 m array is very limited and only a few events would pass our cuts.

Due to the small size of the 750 m array, to increase the probability of having an event recorded simultaneously by the surface array and by a fluorescence detector it is necessary that the fluorescence detector overlook this small area. This is achieved thanks to the operation of the HEAT telescopes. They are three fluorescence telescopes, identical to the ones which form the FD, that can operate in a inclination mode, observing showers which develop higher in the atmosphere, and hence with lower energies. These telescopes are located next to one of the standard fluorescence detector sites (Coihueco). HEAT and Coihueco telescopes overlook the 750 m array and their combination covers a field of view of 60°. This fact allows measuring showers in an energy range similar to the one measured with the 750 m array. The data collected simultaneously by HEAT and Coihueco telescopes are called HECO measurements.

6.3.1 Selection of Golden Hybrid events

To select Golden Hybrid events we take the data sample used to build an elongation rate exclusively with X_{max} measurements done by HECO. This elongation rate belongs to an analysis which is under internal review [149]. The result of this analysis is the extension of the elongation rate given in [78].

As in the case of the 1500 m array we do the selection of Golden Hybrid events from a data set which has been previously trimmed to guarantee a good reconstruction of X_{max}. The reconstruction biases in the energy and X_{max} have been also corrected. Nevertheless, in this case we have to do the selection before applying the fiducial field

of view (FidFOV) cut, the last one in the selection chain followed in [149]. This is again a consequence of the small size of the 750 m array, since this is the only way to obtain a number of events statistically significant to perform the calibration.

Although we select HECO events without the FidFOV cut, we end up with a correct calibration. The FidFOV cut is introduced to avoid a bias in the X_{max} distributions, but the individual values of X_{max} are correctly determined once applied the necessary quality cuts. A bias in the X_{max} distributions of HECO does not affect our calibration because it is done on an event-by-event basis.

The number of Golden Hybrid events included in the HECO data sample before applying the FidFOV cut, together with the cuts that we have to apply to assure we obtain a meaningful value of $\langle\Delta\rangle$ are listed in table 6.3. After applying these cuts we have only 252 Golden Hybrid events to study the correlation between X_{max} and $\langle\Delta\rangle$ ($\sim$600 events less than in the 1500 m array). The selected Golden Hybrid events included in each energy bin are shown in table 6.4.

Quality Cuts	**Events**	**Efficiency**
HECO data no FidFOV	12003	100 %
Golden Hybrid events	2461	20.5 %
sec θ <1.30	2007	16.7 %
6T5 Trigger	714	5.9 %
At least 3 detectors	660	5.5 %
log(E_{FD}/eV) $\geq$ 17.5	297	2.5 %
log(E_{SD}/eV) $\geq$ 17.5	**252**	**2.1** %

Cuadro 6.3: Quality cuts and efficiencies for the Golden Hybrid events used in this chapter.

Energy bin	**Events**	**Energy bin**	**Events**
17.5 $\leq$ log(E_{FD}/eV) <17.6	69	17.9 $\leq$ log(E_{FD}/eV) <18.0	14
17.6 $\leq$ log(E_{FD}/eV) <17.7	74	18.0 $\leq$ log(E_{FD}/eV) <18.1	10
17.7 $\leq$ log(E_{FD}/eV) <17.8	52	18.1 $\leq$ log(E_{FD}/eV) <18.2	5
17.8 $\leq$ log(E_{FD}/eV) <17.9	25	18.2 $\leq$ log(E_{FD}/eV) <18.3	3

Cuadro 6.4: Number of the selected Golden Hybrid events listed by energy bin.

The cuts listed in table 6.3 do not introduce a bias in terms of X_{max}. This is observed in both panels of figure 6.5, before (left panel) and after (right panel) applying the cut on E_{SD}. In this figure we compare the mean values of X_{max} of our selected Golden Hybrid events with the mean values obtained for the whole HECO data set (including the FidFOV cuts and hence without any bias). In both panels the Golden Hybrid elongation rates are compatible with the HECO values, even using a selection without the FidFOV cut. This happens because the FidFOV cut has a minor effect in the mean values of the X_{max} distributions for energies above $10^{17,7}$ eV.

The small difference observed between the left and the right panels of figure 6.5 indicates that the migrations between energy bins due to the differences between E_{FD} and E_{SD} have a smaller effect in the 750 m array.

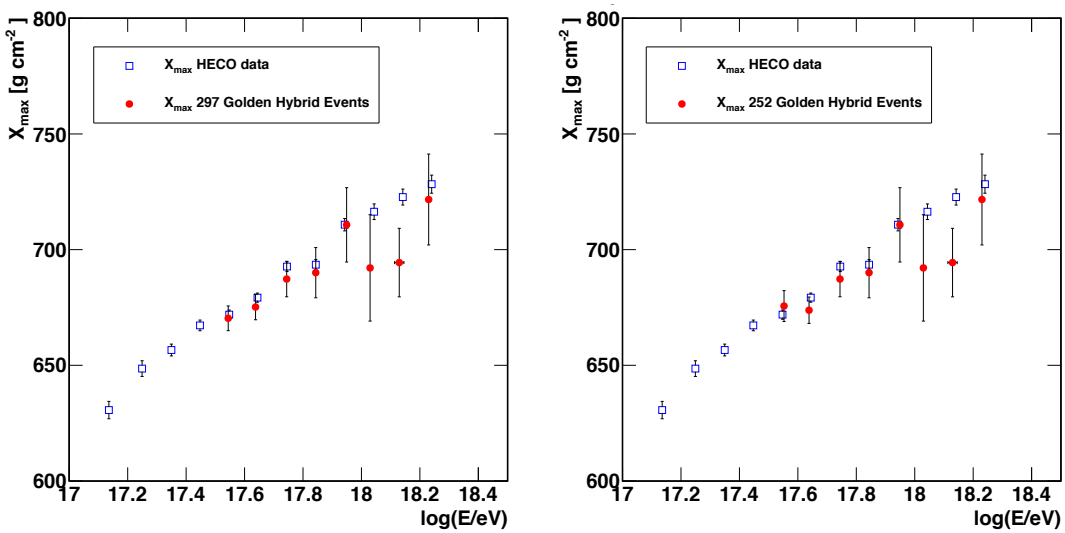


Figura 6.5: Mean values of the X_{max} distributions obtained for our selection of Golden Hybrid events (red dots). These values are compared to the elongation rate provided by HECO (blue squares). Left (right) panel: Before (after) the cut $\log(E_{SD}/\mathrm{eV}) \geq 17.5$.

6.3.2 Correlation and calibration

Once we have selected the Golden Hybrid events we can study again the level of correlation between X_{max} and $\langle\Delta\rangle$. In figure 6.6 we show the values of X_{max} as a function of $\langle\Delta\rangle$. The correlation coefficient found in this case is 0.39. This value is smaller than the one obtained for the data of the 1500 m array (0.46). This difference could be an effect of the scant statistics used in this case, but it could also be a consequence of that these showers, with lower energies, develop higher in the atmosphere and thus we are seeing more muons at the ground. Another possible explanation for this small correlation factor could be that the composition of these showers is different from those ones studied in the previous chapter. As we discussed in chapter 4, the correlation between X_{max} and $\langle\Delta\rangle$ is smaller for iron nuclei, so the decrease of the correlation coefficient could indicate that at the lowest energies the fraction of heavy nuclei increases. As we pointed out in the previous chapter, the correlation coefficient between X_{max} and $\langle\Delta\rangle$ could be a promising observable for mass composition analyses.

For the calibration of X_{max} in terms of $\langle\Delta\rangle$ we assume the same relationship given for the 1500 m array:

$$X_{\mathrm{max}} = a + b\langle\Delta\rangle + c\log(E_{\mathrm{SD}}/\mathrm{eV}) \tag{6.1}$$

using again the energy measured by the SD as the shower energy. In this case the

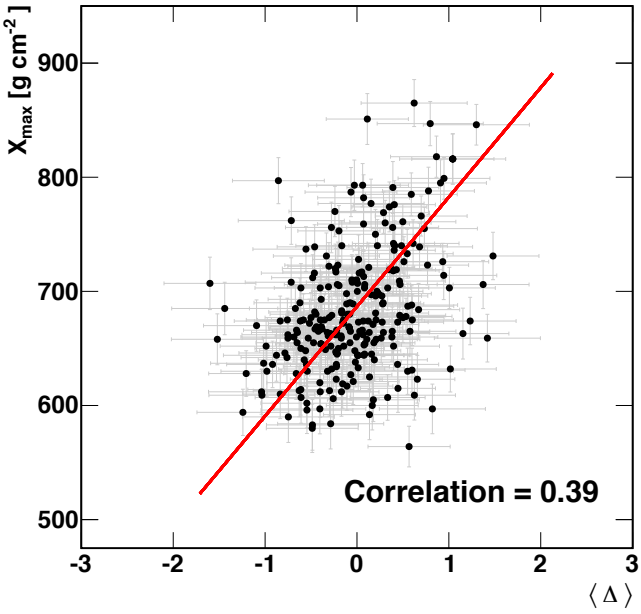


Figura 6.6: X_{max} vs. $\langle\Delta\rangle$ for the 252 Golden Hybrid events. The red line is a projection in 2D of the calibration curve obtained with the maximum likelihood method.

uncertainty used for the X_{max} values is taken from the parameterization given in [150].

To obtain an estimation of the parameters a, b and c we use once more a maximum likelihood method. We follow the same prescriptions explained in the previous chapter, including the same bootstrap method for the calculation of the pertinent errors. The values found for these three parameters after the maximization procedure are given in table 6.5 together with their errors. In this case the errors of these parameters are significantly larger than the ones obtained for their counterparts in the data of the 1500 m array, see table 5.5. This substantial difference comes from the fact that the calibration in the 750 m array is done with far fewer events, only 252. The projection in 2D of the calibration curve (red line) is shown in figure 6.6.

Calibration Parameters	**Values $\pm$ Errors**
a	636 $\pm$ 20 g cm^{-2}
b	96 $\pm$ 10 g cm^{-2}
c	2.9 $\pm$ 1.2 g cm^{-2}

Cuadro 6.5: Calibration parameters to obtain $X_{\text{max}}^{\text{Delta}}$ from the values of $\langle\Delta\rangle$ and E_{SD}.

The goodness of the fit obtained after the maximum likelihood method is checked in the left panel of figure 6.7. This figure shows the difference between $X_{\text{max}}^{\text{Delta}}$ and X_{max} for the 252 Golden Hybrid events used in the calibration. The mean value of this distribution

is (0 $\pm$ 4) g cm^{-2} with a RMS value of 59 g cm^{-2}, fully compatible with zero. In the right panel of this figure we show the difference between $X_{\text{max}}^{\text{Delta}}$ and X_{max} as a function of sec θ. Within the rather large uncertainties, there is no evidence of any trend with the sec θ. Again the calibration procedure does not introduce any bias in terms of the sec θ.

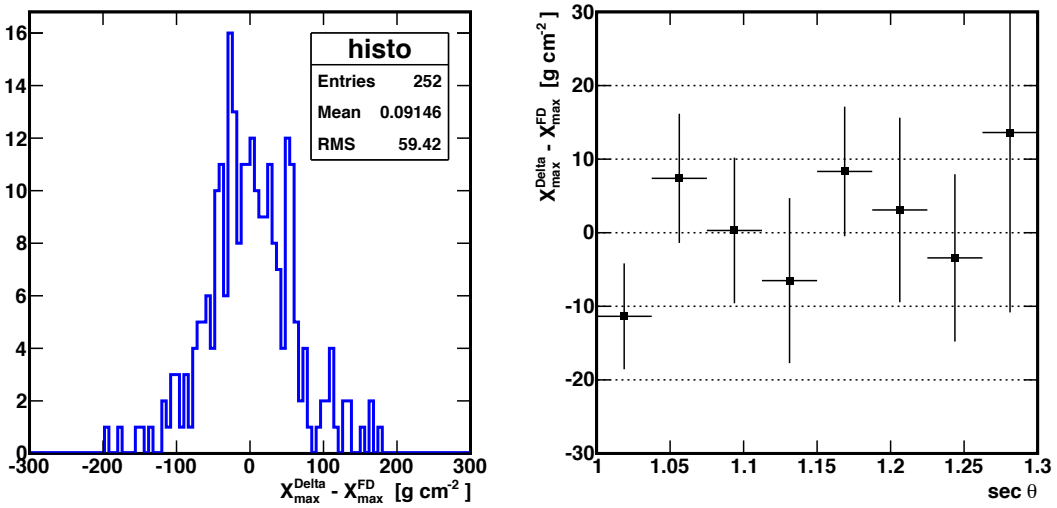


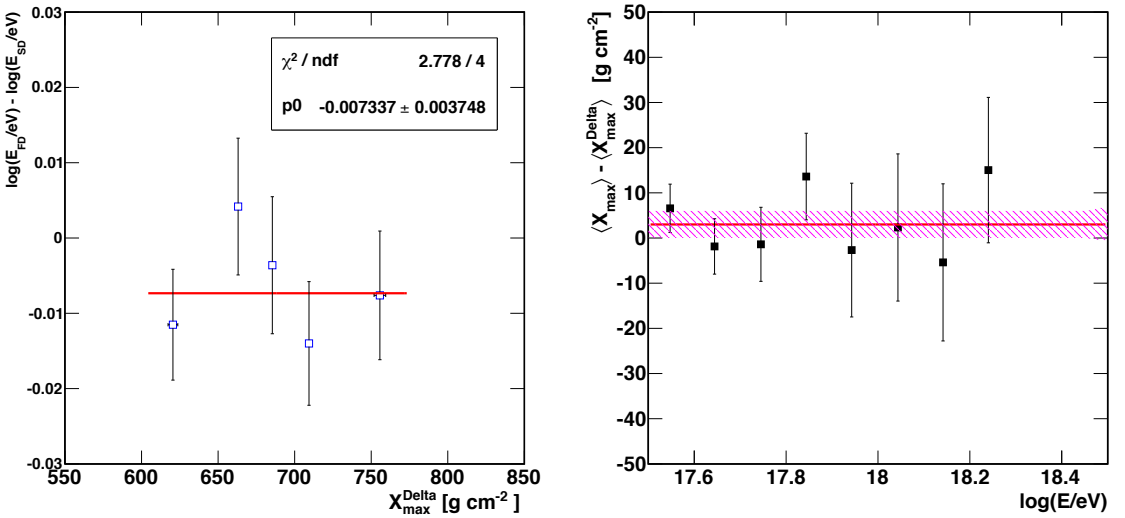
Figura 6.7: Left panel: Differences between $X_{\text{max}}^{\text{Delta}}$ and X_{max} for the 252 Golden Hybrid events used in the calibration procedure. Right panel: Same differences as a function of sec θ.

6.4 Energy correction

We discussed in the previous chapter that to compare an elongation rate obtained with SD-only events to another obtained with FD measurements it is necessary to use a common energy. This is a consequence of the differences between the energies measured by the FD and the SD. The same consideration has to be taken into account now that we use HECO data, since HECO is composed of fluorescence detectors which measure the calorimetric energy as the FD does.

To correct the energy measured by the SD we follow the approach explained in section 5.4. Figure 6.8a shows for our sample of 252 Golden Hybrid events the difference between the two energies as a function of $X_{\text{max}}^{\text{Delta}}$.

In this case the difference between both energies does not show a dependence on the depth of the shower. In fact, the difference is fitted with a constant as a function of $X_{\text{max}}^{\text{Delta}}$ (see figure 6.8a). Possible reasons for observing this flat trend could be the low statistics of Golden Hybrid events or the narrower range of $X_{\text{max}}^{\text{Delta}}$ covered in this case. In figure 6.8a we see that the $X_{\text{max}}^{\text{Delta}}$ values extend over $\sim$140 g cm^{-2} whereas in the case of the 1500 m array (see figure 5.8a) the $X_{\text{max}}^{\text{Delta}}$ values cover a range of $\sim$200 g cm^{-2}.



(a) (b)

Figura 6.8: Left panel: Difference between $\log(E_{FD}/eV)$ and $\log(E_{SD}/eV)$ as a function of X_{max}^{Delta} for the 252 Golden Hybrid events selected in this chapter, fitted with a constant (red line). Right panel: Differences between the mean values of X_{max} and X_{max}^{Delta} after the energy correction as a function of energy. The mean value of these differences is indicated as a red line.

Using the constant fit shown in figure 6.8a, the final expression to correct the SD energy is:

$$\log(E_{corrected}/eV) = \log(E_{SD}/eV) + p_0 \tag{6.2}$$

being p_0 = (-0.007 $\pm$ 0.004). This correction will be used hereafter to obtain the elongation rate with the data of the 750 m array.

In figure 6.8b we show the differences between the mean values of X_{max} and X_{max}^{Delta} after the energy correction as a function of energy. The average difference along the whole energy range is (3 $\pm$ 3) g cm^{-2}. This value is compatible with the one observed in case of using the energy measured by the FD, (3 $\pm$ 3) g cm^{-2}.

6.5 Systematic uncertainties

We now focus our attention on the different sources of systematic uncertainties that influence the measurement of X_{max}^{Delta}.

The main contribution to the systematic uncertainty comes from the uncertainties on the calibration parameters. However, the uncertainties of the calibration parameters found now are significantly larger and therefore the final contribution to the systematic uncertainty increases accordingly. After the propagation of the uncertainties taking into account

the correlations between the parameters, we find that the maximum difference in $X_{\mathrm{max}}^{\mathrm{Delta}}$ is 10 g cm^{-2}. This value is obtained for the largest energies studied in this chapter. To keep the most conservative scenario we use this value as the systematic uncertainty for the whole energy range.

The contributions coming from the seasonal effect, the UTC dependence and the ageing have been evaluated and they are compatible with the ones discussed for the 1500 m array. This was expected because both arrays of surface detectors work in similar circumstances during the different seasons or the different times of the day. For this reason we take as a systematic uncertainty the same values given in the previous chapter.

The dependence on sec θ, discussed for the data of the 1500 m array, is not observed now. Figure 6.9 shows $X_{\mathrm{max}}^{\mathrm{Delta}}$ as a function of sec θ for the whole data set studied in this chapter. The behaviour observed in this figure is compatible with a flat trend and it is not necessary to take into account an additional contribution to the systematic uncertainty.

The systematic uncertainties of the X_{max} values measured by HECO propagate directly into the values of $X_{\mathrm{max}}^{\mathrm{Delta}}$. The systematic uncertainty of HECO is given in [149] and it is very similar to that of the FD [78]. We take as systematic uncertainty the same value considered in the previous chapter, 8.5 g cm^{-2}.

Table 6.6 summarizes all the sources contributing to the systematic uncertainty. The overall value amounts to 14 g cm^{-2}.

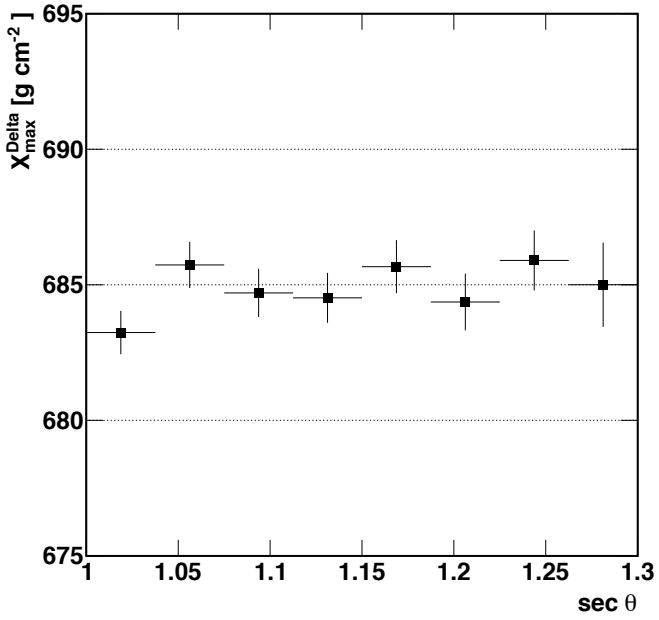


Figura 6.9: $X_{\mathrm{max}}^{\mathrm{Delta}}$ as a function of sec θ for the whole data set used in this chapter.

Source	**Sys. uncertainty** [g cm^{-2}]
Uncertainty on calibration	10.0
Seasonal effect	2.0
UTC dependence	1.0
Ageing	3.0
HECO systematic uncertainty	8.5
Total	14.0

Cuadro 6.6: Contributions to the systematic uncertainty of the $X_{\text{max}}^{\text{Delta}}$.

6.6 The Elongation Rate

The next step is to obtain the evolution of $X_{\text{max}}^{\text{Delta}}$ with the energy for the data of the 750 m array. This is shown in figure 6.10. Events are grouped according to their log(E/eV) value in bins of width 0.1. The shaded area indicates the systematic uncertainty. These results are compared to the measurements of X_{max} obtained with HECO (after the FidFOV cut) [149].

The measurements obtained with the $\langle\Delta\rangle$ method are compatible with HECO results in the energy range where they overlap. Furthermore, thanks to the larger number of events provided by the $\sim$100 $\%$ duty cycle of the surface detectors we can add two more bins above the highest energy bin studied by HECO.

Our results are also compared to the predictions given by the latest version of the packages QGSJetII.04 and EPOS-LHC for the behaviour of protons and iron nuclei (see figure 6.10). Our data suggest that the cosmic ray composition becomes lighter with energy up to around $10^{18,3}$eV and possibly it is dominated by protons at this energy. In fact, our measurement of $X_{\text{max}}^{\text{Delta}}$ at $10^{18,3}$ eV matches with the prediction given by QGSJetII.04 for the X_{max} value of protons with this energy.

The $X_{\text{max}}^{\text{Delta}}$ values as a function of energy are split in two bins of sec θ in figure 6.11a. Both bins of sec θ exhibit the same trend as a function of the energy. The number of events used in each energy bin is shown in figure 6.11b. As in the previous chapter, the numbers do not match with the ones indicated in the distributions of figure 6.3 because of the migrations introduced by the energy correction.

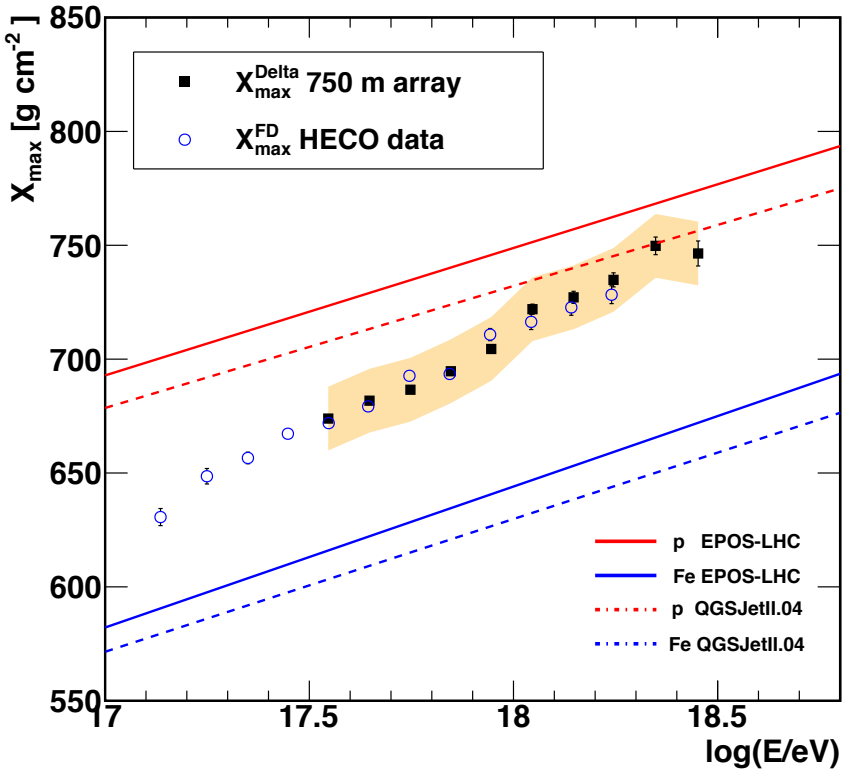
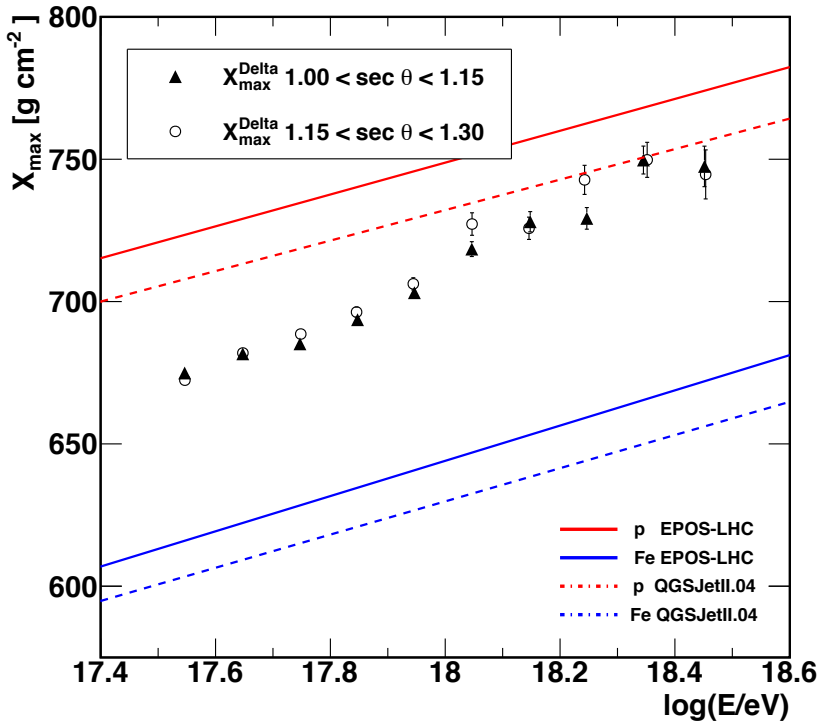
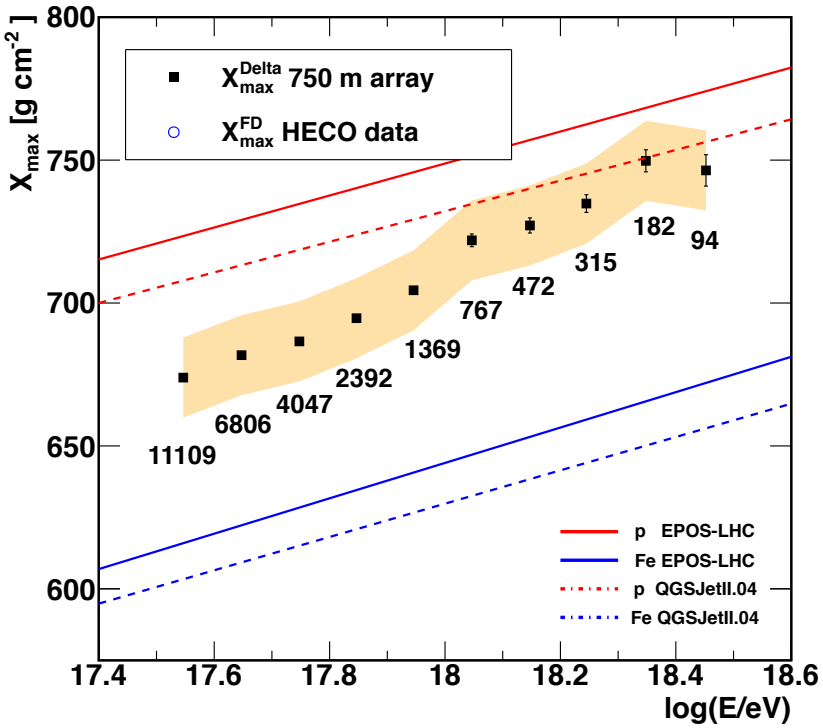


Figura 6.10: Mean values of the $X_{\mathrm{max}}^{\mathrm{Delta}}$ distributions as a function of the energy obtained with data measured by the 750 m array. The shaded area indicates the systematic uncertainty. Data are compared to the predictions for simulations of protons and iron nuclei and to the measurements of X_{max} made by HECO.



(a)



(b)

Figura 6.11: Top panel: $X_{\mathrm{max}}^{\mathrm{Delta}}$ elongation rate split in two bins of sec θ. Bottom panel: Zoom of the $X_{\mathrm{max}}^{\mathrm{Delta}}$ elongation rate with the number of events used in each energy bin.

7

Combination of the results of the $\langle\Delta\rangle$ method

We have applied separately the $\langle\Delta\rangle$ method to the data gathered by the two arrays of surface detectors of the Pierre Auger Observatory. There was a good agreement between our results and the measurements of X_{max} obtained with the two systems of fluoresce telescopes.

Now we assess the agreement between our own results and bring together the two energy ranges where we have measured $X_{\mathrm{max}}^{\mathrm{Delta}}$.

7.1 Combination of results

Figure 7.1 shows jointly the results obtained for $X_{\mathrm{max}}^{\mathrm{Delta}}$ with the data of both arrays of surface detectors. It has been obtained as the combination of figures 5.12 and 6.10.

In the light of this figure several comments are in order. The first one is the smooth transition between the elongation rate obtained with the data of the 750 m array, ending at $10^{18,5}$ eV, and the one obtained with the 1500 m array, starting at that energy. This is a notable achievement taking into account that the data sets used to obtain each elongation rate are completely independent and that the $\langle\Delta\rangle$ method has been applied separately in each data set with different benchmark energy bins and different benchmark parameterizations. Such a high level of agreement between two independent results confirms the robustness of the analysis and the potential of the $\langle\Delta\rangle$ method to study mass composition over extended energy ranges.

The second remarkable aspect is that the results provided by the data of the 750 m array, having shown that they agree with the measurements made with the fluorescence detectors, are the first mass composition results obtained in the Pierre Auger Collaboration using the data collected by this small surface array.

Last but not the least important feature of the result shown in figure 7.1 is that thanks to the low-energy extension allowed by the data of the 750 m array, we have been able

to obtain for the first time an elongation rate with only SD-measurements covering more than 2 orders of magnitude in energy, from $10^{17,5}$ eV up to nearly 10^{20} eV.

Figure 7.2 shows the same results as figure 7.1, but we have included the $X_{\rm max}$ measurements obtained with the fluorescence telescopes (FD and HECO) to underline again the agreement between the results obtained with the $\langle\Delta\rangle$ method and the fluorescence measurements.

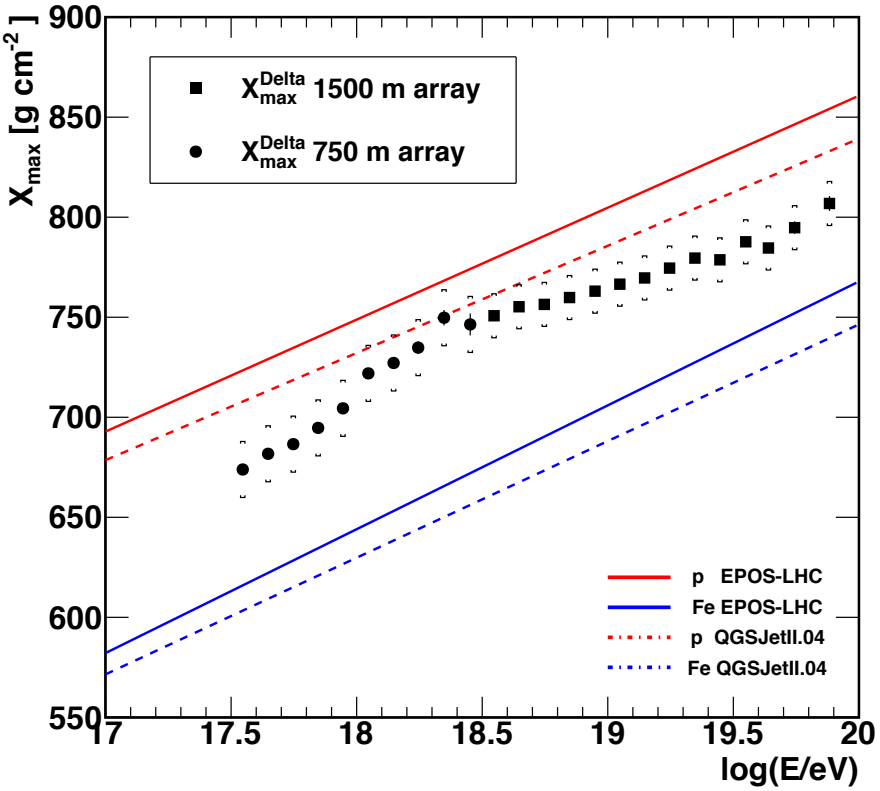


Figura 7.1: Mean values of the $X_{\rm max}^{\rm Delta}$ distributions as a function of the energy obtained with data measured by the 1500 m and 750 m arrays. The brackets represent the systematic uncertainty. Data are compared to simulations of protons and iron nuclei.

7.2 The extension of both elongation rates

An interesting way to check the results presented in the previous section and to verify the agreement between the two independent elongation rates is to study their behaviour in an overlapping energy region. For this purpose it is necessary to extend the results obtained with the data of the 1500 m array below $10^{18,5}$ eV and the results obtained with the data of the 750 m array above this energy. To perform correctly such extensions we have to

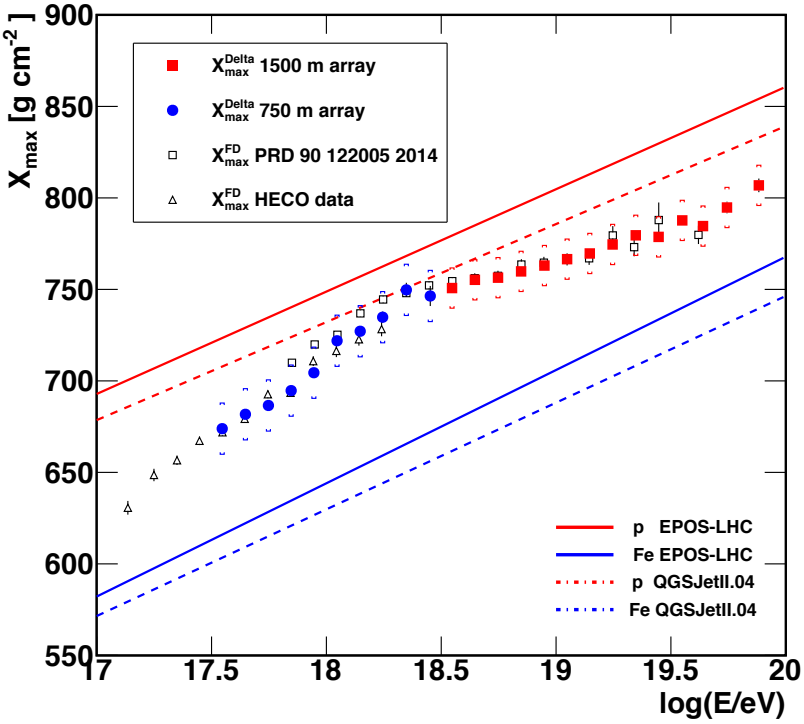


Figura 7.2: Same as figure 7.1 including the X_{max} measurements obtained with the fluorescence telescopes (FD and HECO).

consider three important aspects: the trigger efficiency, the statistics and the linearity of term $t_{1/2}/r$.

In principle it is possible to extend the results of the 1500 m array several energy bins below $10^{18,5}$ eV. Although we have to mention that this is not an optimal energy range for the 1500 m array, since this array is fully efficient above 3 EeV ($\sim 10^{18,5}$ eV). In this case we have decided to extend the measurements up to $10^{18,2}$ eV, three energy bins below our original cut, because in that way we guarantee that the trigger efficiency of the array is always larger than 80 $\%$ [105]. The number of events is not a problem for extending the elongation rate to lower energies because the flux of cosmic rays increases steeply with decreasing energy.

The last aspect to consider in the case of the 1500 m array is the linearity of the term $t_{1/2}/r$. As we explained in chapter 2 to avoid a reconstruction bias in the risetime values it is necessary to guarantee that the term $t_{1/2}/r$ behaves linearly as a function of the core distance. Remember that due to this criterion of linearity we were forced to enlarge the upper limit on distance from 1400 m up to 2000 m for energies larger than $10^{19,6}$ eV. To keep the linearity of the term $t_{1/2}/r$ when we go to lower energies is not a problem because the reduction of the energy together with the cut on the signal size, automatically limit the distance range covered by the showers.

Figure 7.3 shows the extension of the elongation rate obtained with the data of the 1500 m array, three energy bins below our original cut at $10^{18,5}$ eV, together with the $X_{\rm max}$ measurements of the FD. No modification in the $\langle\Delta\rangle$ method has been introduced to obtain this figure. We have applied simply the $\langle\Delta\rangle$ method to a wider data set. The same cuts described in section 5.1 have been applied to select data with energies in the range 18.2 $\leq$ log(E/eV) <18.5. The agreement between the mean values of $X_{\rm max}^{\rm Delta}$ and the measurements of the FD for the three energy bins below $10^{18,5}$ eV indicates that the extension at lower energies has been done properly.

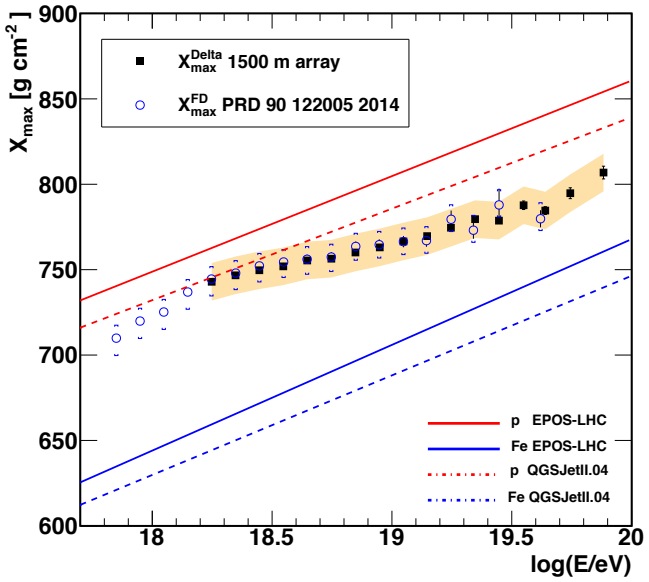


Figura 7.3: Same as figure 5.12 including an extension at lower energies. Three new energy bins have been included below $10^{18,5}$ eV.

We now turn our attention to the extension of the elongation rate obtained with the data of the 750 m array above $10^{18,5}$ eV. In this case the trigger efficiency is not an obstacle because the 750 m array is fully efficiency above 0.3 EeV ($\sim 10^{17,5}$ eV). In this case it is more problematic the low number of events, since it decreases drastically with increasing energy. Nevertheless there are 102 events with energies between 18.5 $\leq$ log(E/eV) <18.7 which can be studied.

In this case, as we do an extension to higher energies, we find a problem with the linearity of $t_{1/2}/r$ similar to the one discussed for the 1500 m array. As shown in section 4.3.1, to avoid any bias in our data sample we have to select detectors in a core distance range that keeps the linearity of the ratio $t_{1/2}/r$. In the energy range 17.5 $\leq$ log(E/eV) <18.5 this condition is satisfied if we select detectors between 300 m and 800 m, but above this energy we are forced again to enlarge the upper limit on distance. In this case from 800 m up to 1400 m (see figure 7.4).

The left panel of figure 7.4 shows that effectively the optimal upper limit for the distance is 800 m in the benchmark energy bin and the right panel of the same figure reflects

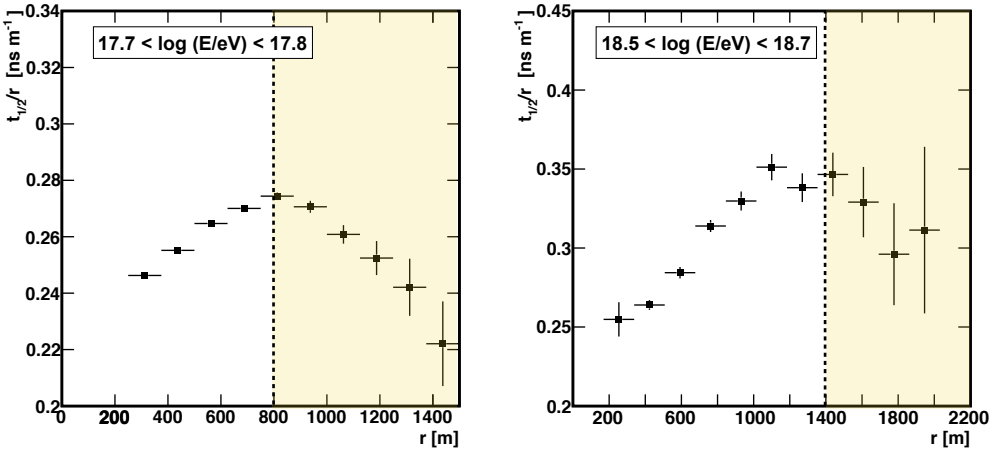


Figura 7.4: $t_{1/2}/r$ as a function of distance, r, for two energy bins. The shaded zone indicates the distance range where the linearity of $t_{1/2}/r$ breaks down.

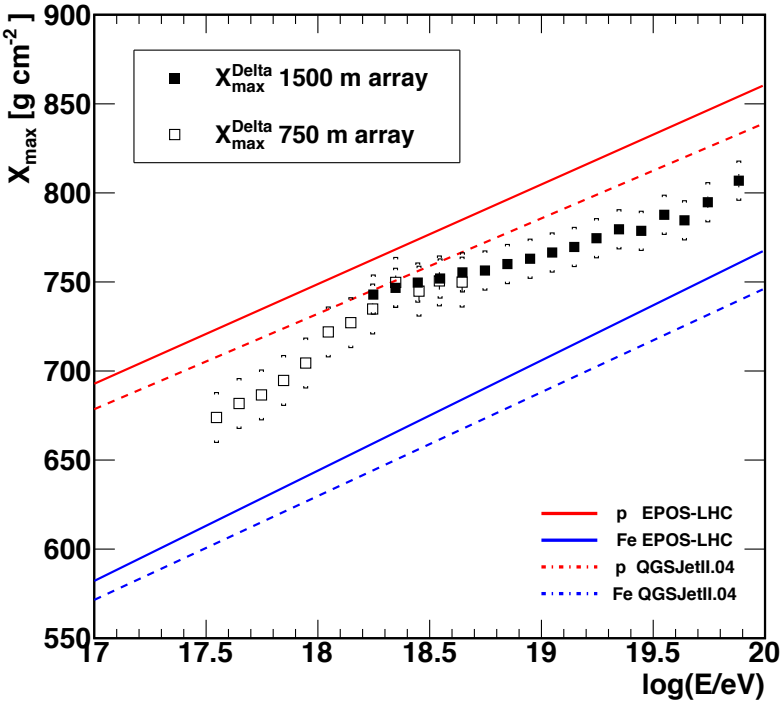


Figura 7.5: Same as figure 7.1 including the extensions above and below $10^{18,5}$ eV. Three new energy bins have been included below $10^{18,5}$ eV with the data gathered by the 1500 m array and two new ones above $10^{18,5}$ eV using the data of the 750 m array.

that it is necessary to open this upper limit up to 1400 m if we want to study energies between 18.5 $\leq$ log(E/eV) <18.7. Note that this upper limit for the distances matches with the upper limit used in the 1500 m array for the same energies.

Figure 7.5 shows the extension of the elongation rate obtained with the data of the 750 m array two energy bins above our original cut at $10^{18,5}$ eV together with the extension obtained for the data of 1500 m array. As in the previous case, no modification in the $\langle\Delta\rangle$ method has been introduced to obtain the mean values of $X_{\text{max}}^{\text{Delta}}$ above $10^{18,5}$ eV. The only variation has been an increase of the upper limit on distance to avoid a bias in the risetimes values.

The agreement between both elongation rates in the overlapping region (18.2 $\leq$ log(E/eV) <18.7) corroborates the accuracy of each one of the elongation rates, since we obtain compatible values for $X_{\text{max}}^{\text{Delta}}$ in the same energy range with two measurements totally independent.

7.3 Interpretation of the results of $X_{\text{max}}^{\text{Delta}}$: $\langle \ln A \rangle$

Inferring the mass composition of UHECRs from the measurements of $X_{\text{max}}^{\text{Delta}}$ is only possible by comparing the results to the predictions given by hadronic models. It was shown in chapter 1 that if one assumes that the hadronic models are a faithful representation of Nature at the highest energies ever probed, then we can infer the value of the natural logarithm of the atomic mass ($\ln A$).

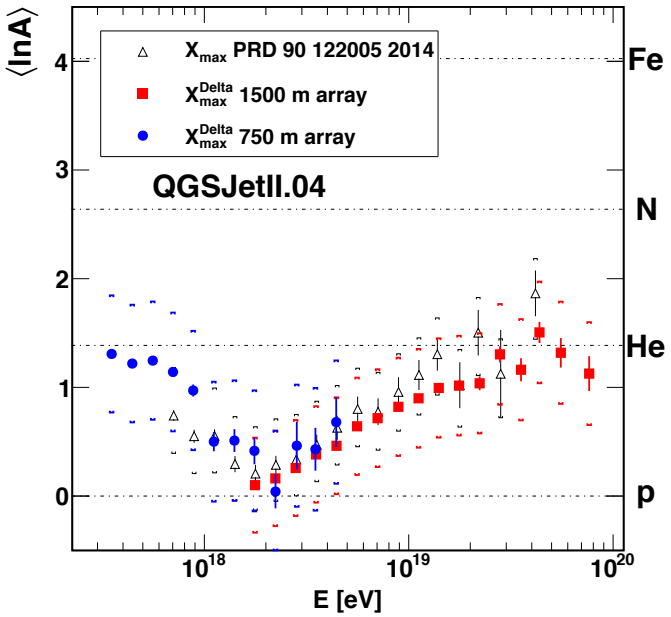
For each interaction model, the values of $\langle \ln A \rangle$ derived from data have been computed using the following relationship:

$$\langle \ln A \rangle = \ln 56 \frac{\langle X_{\text{max}}^{\text{Delta}} \rangle_p - \langle X_{\text{max}}^{\text{Delta}} \rangle_{data}}{\langle X_{\text{max}}^{\text{Delta}} \rangle_p - \langle X_{\text{max}}^{\text{Delta}} \rangle_{Fe}} \tag{7.1}$$

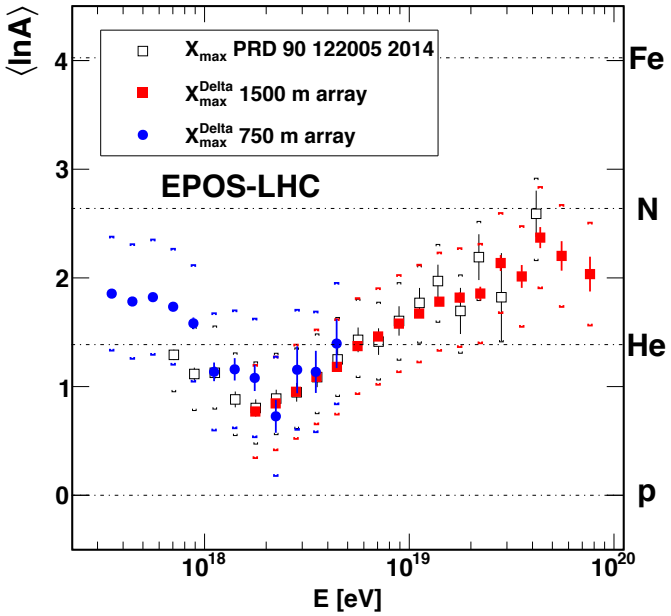
Figure 7.6 shows the results of such conversion for the two hadronic models considered throughout this thesis: QGSJetII.04 (top panel) and EPOS-LHC (bottom panel). As expected from the agreement between the elongation rates obtained with the $\langle\Delta\rangle$ method and the measurements made with the fluorescence detectors, the estimation of $\langle \ln A \rangle$ provided by the $\langle\Delta\rangle$ method agrees with the results given by the FD [78]. For this reason the mass composition inferences arising from the $\langle\Delta\rangle$ method and from the FD are similar in the energy range where they overlap.

Although the trend of $\langle \ln A \rangle$ with energy is the same for both hadronic models, the different features of the hadronic interactions implemented in each model give rise to differences in terms of the absolute values of $\langle \ln A \rangle$.The interpretation with EPOS-LHC leads to a heavier average composition than the one obtained for QGSJetII.04.

Our data suggest that the flux of cosmic rays is composed of predominantly light nuclei at around $10^{18,3}$ eV, possibly protons as it is indicated by QGSJetII.04, and that the fraction of heavy nuclei is increasing up to energies of $10^{19,6}$ eV since the $\langle \ln A \rangle$ values increase up to this energy. At this point we observe a change in the trend of $\langle \ln A \rangle$: our two highest energy bins show that $\langle \ln A \rangle$ decreases again. This fact suggests an increase of the light component in the cosmic ray flux at the highest energies.



(a)



(b)

Figura 7.6: Conversion of $X_{\text{max}}^{\text{Delta}}$ to $\langle \ln A \rangle$ for the independent analysis presented in this thesis, one obtained with the data of the 1500 m array (red squares) and the other with the data of the 750 m array (blue squares). The values of $\langle \ln A \rangle$ coming from the measurements of X_{max} made with the FD are also included for the comparison (empty squares) [78]. On the top (bottom) panel we use QGSJetII.04 (EPOS-LHC) as the reference hadronic model. Brackets correspond to the systematic uncertainties.

8

Comparison of $\langle\Delta\rangle$ with the models

The aim of this thesis has been to find an observable, based on the risetimes of the signals registered by the surface detectors, capable of providing information about the mass of the primary cosmic ray, through a procedure based solely on the use of data. Such observable is $\langle\Delta\rangle$, and the method developed throughout the previous chapters exploits the correlation between $\langle\Delta\rangle$ and X_{max} present in a small number of Golden Hybrid events to obtain an estimator of X_{max} for the entire dataset analyzed. This estimator is what we call $X_{\text{max}}^{\text{Delta}}$. The robustness of this analysis as well as the potential of the $\langle\Delta\rangle$ method to study mass composition has been broadly demonstrated in the previous chapters.

The observable $\langle\Delta\rangle$, and in particular the risetime, is characterized by containing information about both components of extensive air showers: the hadronic and the electromagnetic component (see section 3.1). When we perform the calibration with X_{max} (an observable of a mainly electromagnetic nature) we maximize the information coming from the electromagnetic component in the observable $X_{\text{max}}^{\text{Delta}}$. This is the key feature which makes possible to obtain valuable information about the mass composition of cosmic rays using the $\langle\Delta\rangle$ method.

However, the observable $\langle\Delta\rangle$ itself can also be used to extract precious information on the validity of the hadronic models and add further insight to the problem connected to the deficiency of muons. By studying the evolution of the $\langle\Delta\rangle$ in terms of energy, we can assess the level of agreement between data and simulations. In particular, as $\langle\Delta\rangle$ contains information about the muonic part of the shower, this observable allows us to analyze, from another perspective, the problem disclosed by the data of the Pierre Auger Observatory: the description given by current models is inadequate to describe the development of the hadronic component of an EAS. This is the objective of the analysis described in this chapter.

8.1 $\langle\Delta\rangle$ for different mass compositions

In this section we study the behaviour of $\langle\Delta\rangle$ for proton and iron primaries predicted by the different hadronic models used in this thesis. For this purpose, a library of Monte Carlo events has been produced using the EPOS-LHC and QGSJetII.04 hadronic interaction models for the primary species aforesaid. The showers are produced with continuous distributions of energies and zenith angles. The energies cover the range 18.5 $\leq$ log(E/eV) <20.0 and the zenith angles are lower than 45° (1.00<sec θ <1.42). To reduce statistical fluctuations we have distributed the events according to their value of log (E/eV) in bins of width 0.1, so that we have $\sim$1000 showers per energy bin (for each primary and hadronic model). The thinning level of the showers is 10^{-6}. With this thinning level simulated risetimes show no significant biases or enhanced fluctuations [151]. Events have been fully simulated using the 1500 m array configuration.

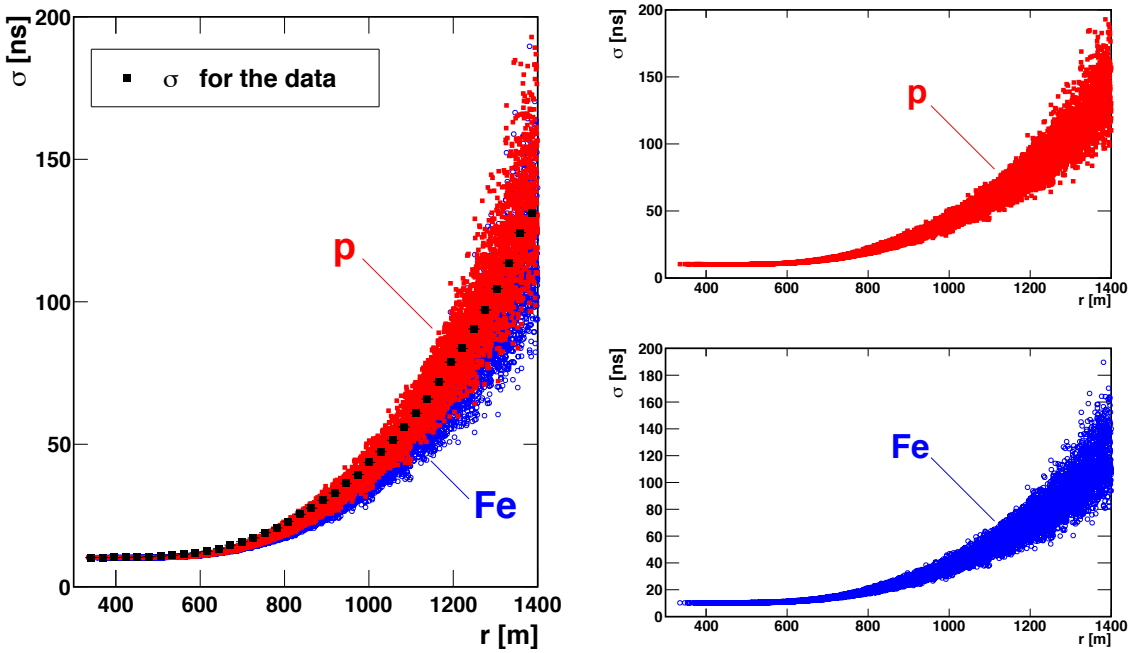
When comparing the behaviour of $\langle\Delta\rangle$ between different sets of events with different mass compositions, for instance between proton and iron primaries, we have to take into account a subtlety: we have to take as a reference for all of them the benchmark of only one set. Otherwise the comparison is meaningless. The choice of a particular benchmark as a reference is arbitrary and the conclusions stemming from the comparison are independent of this decision. The choice of a different reference only implies a shift in the mean values of $\langle\Delta\rangle$, but the shift will be the same for each set and the absolute differences between the $\langle\Delta\rangle$ values will be the same. We have decided to use as a reference for the whole analysis presented in this chapter the benchmark parameterization obtained for the data in section 4.4.2. In that way we do not have to modify the $\langle\Delta\rangle$ values obtained for the data of the 1500 m array shown in chapter 5.

Once we have set the benchmark, we can calculate the $\langle\Delta\rangle$ value for each one of the simulated events which pass the cuts listed in tables 5.1 and 5.2. These are the same cuts used to select the data of the 1500 m array. The same selection has to be applied in data and in simulations to make a proper comparison. The simulated risetimes have also been corrected for their respective azimuthal asymmetries. Likewise the risetime uncertainty has also been evaluated independently for protons and iron nuclei following the approach used with twin detectors in section 3.7.1. Figure 8.1 shows a comparison of the risetime uncertainties estimated for protons, iron nuclei and data.

The mean values of $\langle\Delta\rangle$ distributions have been obtained as a function of the energy and they have been fitted with linear functions. The result of these fits for protons and iron nuclei are shown in figure 8.2 for QGSJetII.04 (left panel) and EPOS-LHC (right panel).

Both panels of figure 8.2 show a marked difference between the values of $\langle\Delta\rangle$ for protons and for iron nuclei as a function of the energy. This means that if the hadronic models were a fair representation of reality and they described correctly both components of the air showers, then it would be possible to study the mass composition of cosmic rays only in terms of $\langle\Delta\rangle$, without the necessity of the calibration with X_{max}.

An additional test to confirm that we could used $\langle\Delta\rangle$ itself for mass composition analyses (only if the models were correct) is to study the behaviour of different mass composition simulations in relation to the trend observe for protons and iron nuclei. In this case we have produced two new samples of simulated showers, with the same characteristics of the ones produced for protons and iron nuclei, but with different compositions. One of



(a) (b)

Figura 8.1: Left panel: Risetime uncertainties estimated for protons (red squares), iron nuclei (empty circle) and data as a function of core distance. The uncertainties of the data have been plotted as a profile for a clearer view. The uncertainties have been evaluated in events with energies in the range (19.0 <log(E/eV) <19.2). Right panels: Same uncertainties shown in the left panel split in two panels for a better comparison.

the productions is a mix 50 %p-50 %Fe which is constant with the energy. The other one is also a mix p-Fe, but whose fractions change gradually with the energy. At the lowest energies it is composed by 100 % iron nuclei and at the largest energies by 100 % protons. Figure 8.3 shows the values of $\langle\Delta\rangle$ obtained for these new mass compositions as a function of energy. The mix 50 %p-50 %Fe is between the predictions given for protons and iron nuclei, as expected, and the composition changing with the energy also follows the expected trend: at the lowest energies agrees with the predictions of the iron nuclei and at the highest energies agrees with the protons. In both cases we have taken as a reference the benchmark parameterization obtained for the data. We have done this illustrative study only for the hadronic model QGSJetII.04, but we expect a similar outcome for EPOS-LHC.

An interesting remark emerging from figure 8.3 is the behaviour of the slope of $\langle\Delta\rangle$ as a function of the energy according to the different simulated mass compositions. The slopes of $\langle\Delta\rangle$ as a function of the energy are listed in table 8.1 for the different compositions shown in figure 8.3. These values point out that the compositions which are pure or constant with the energy have compatible values of the slope of $\langle\Delta\rangle$ whereas the composition

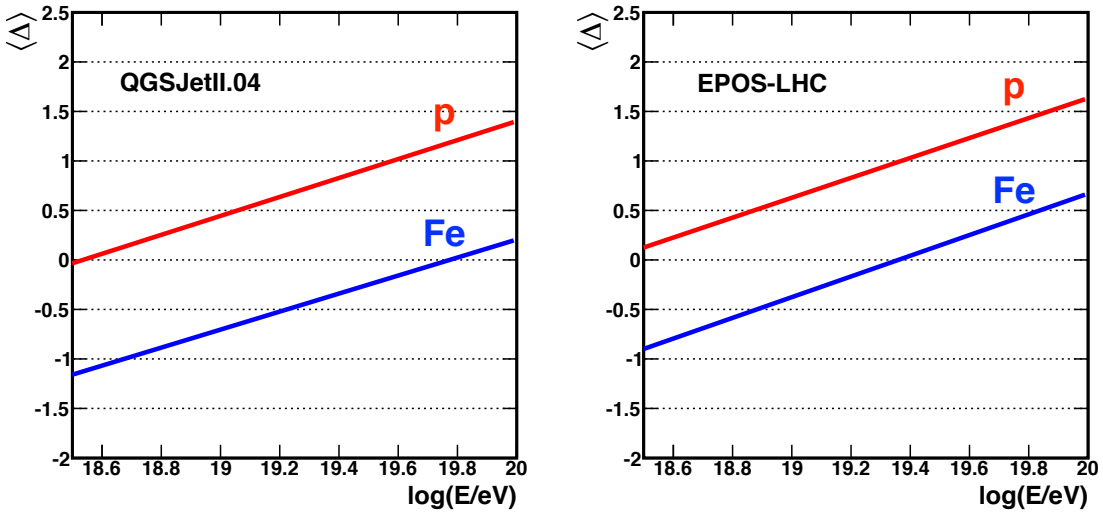


Figura 8.2: Predictions given by the hadronic models QGSJetII.04 (left panel) and EPOS-LHC (right panel) for the behaviour of $\langle\Delta\rangle$ as a function of the energy for two different primary species: proton and iron.

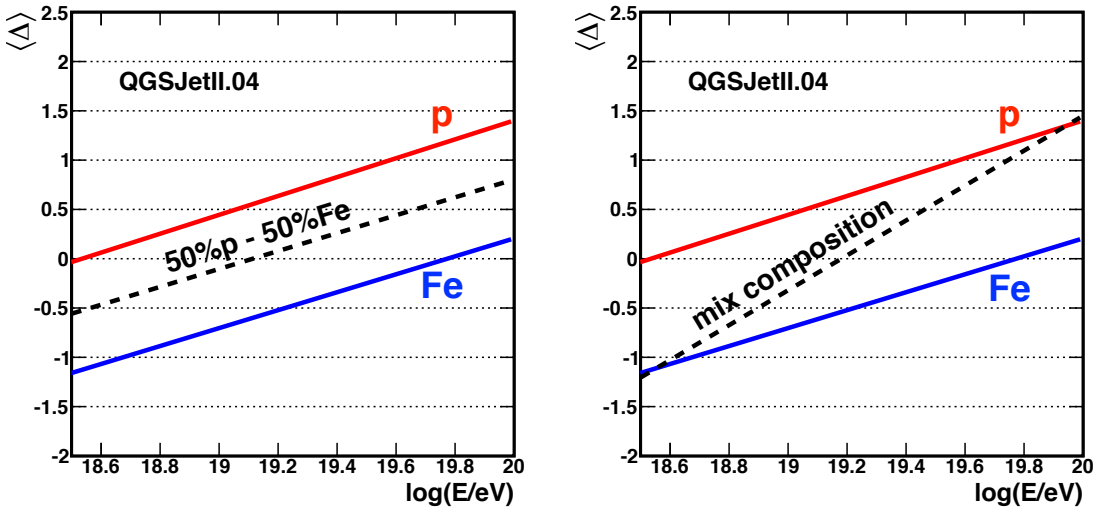


Figura 8.3: Predictions given by the hadronic model QGSJetII.04 for the behaviour of $\langle\Delta\rangle$ in the case of two new simulated compositions. A mix 50 % p - 50 %Fe which is constant with the energy (left panel) and a mix p-Fe whose fractions change gradually with the energy (right panel). These predictions are compared to the behaviour of iron nuclei and protons.

changing with the energy has a completely different value. This implies that we could use this slope to distinguish if the cosmic rays have a constant or a variable composition with the energy.

Mass Composition	Slope
p	0.95 ± 0.03
Fe	0.91 ± 0.02
50 %p- 50 %Fe	0.91 ± 0.02
p-Fe fractions changing with energy	1.78 ± 0.03

Cuadro 8.1: Slopes of $\langle\Delta\rangle$ as a function of the energy for the different compositions shown in figure 8.3.

8.2 $\langle\Delta\rangle$ for data

Now, having the behaviour of $\langle\Delta\rangle$ for different mass compositions, we can compare the values of $\langle\Delta\rangle$ obtained with the data of the 1500 m array to the predicted values for proton and iron primaries, and for both hadronic models. This is shown in figure 8.4 (see appendix B for further details). Note that for the data, the mean value of the $\langle\Delta\rangle$ distribution in the benchmark energy bin (19.1 < log(E/eV) < 19.2) is compatible with zero, as expected, since we use the benchmark parameterization of the data. In fact, the mean values of $\langle\Delta\rangle$ shown in this figure are the same values shown in figure 5.2. The only difference between both figures is that now we have included the systematic uncertainties of $\langle\Delta\rangle$ and the predictions given by the hadronic models for the behaviour of protons and iron nuclei.

The evaluation of the systematic uncertainties is straightforward because the contributing sources are essentially the same as the ones discussed in chapter 5 for $X_{\text{max}}^{\text{Delta}}$. Although, evidently, in this case we do not have to consider the contribution of the uncertainties on the calibration parameters and the systematic uncertainties of X_{max} either, because we work in terms of $\langle\Delta\rangle$. The rest of the contributions, the seasonal effect, the UTC dependence, the ageing and the dependence on sec θ have been evaluated following the same approach explained in chapter 5. All of them behave in same way as the uncertainties of $X_{\text{max}}^{\text{Delta}}$, but in this case we have to express these contributions in units of $\langle\Delta\rangle$ instead of using g cm^{-2}. In addition to these sources, we have evaluated separately the contribution coming from the energy scale. Unlike the study done in chapter 5, now we do not consider the term of the FD uncertainty, which included the contribution of the energy scale, and for this reason the effect is studied separately.

Table 8.2 summarizes all the sources contributing to the systematic uncertainties of our $\langle\Delta\rangle$ measurements. The overall value amounts to 0.11. The systematic uncertainty of the measurements of $\langle\Delta\rangle$ is 10 % of the predicted separation between the $\langle\Delta\rangle$ values of proton and iron primaries for both models.

The evolution of the mean values of $\langle\Delta\rangle$ with the energy indicates a change in the composition as energy increases. It seems that heavier compositions become more do-

Source	**Sys. uncertainty**
Seasonal effect	0.03
UTC dependence	0.01
Ageing	0.04
Dependence on sec θ	0.02
Energy scale	0.1
Total	0.11

Cuadro 8.2: Contributions to the systematic uncertainty of $\langle\Delta\rangle$.

minant till the energy approaches $10^{19,6}$ eV. Beyond this energy our two last energy bins seem to indicate the emergence of a lighter component. However it is evident from figure 8.4 that the mass predictions resulting of this study will depend strongly on the hadronic model adopted.

Figure 8.4 shows that the Auger data are fully bracketed by the proton and iron predictions given by QGSJetII.04. In the case of the hadronic model EPOS-LHC, the data

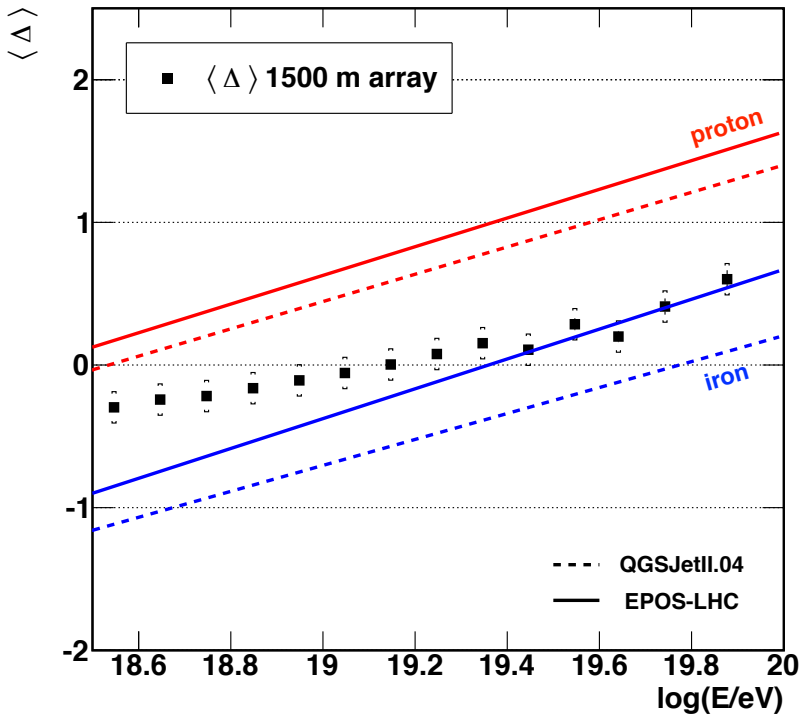


Figura 8.4: Mean values of the $\langle\Delta\rangle$ distributions as a function of the energy obtained with the data measured by the 1500 m array. Brackets correspond to the systematic uncertainties. Data are compared to the predictions of $\langle\Delta\rangle$ obtained from simulations.

below $10^{19,4}$ eV are bracketed by the proton and iron predictions whereas the data above this energy are compatible with the behaviour expected for iron nuclei. There is even a energy bin (19.6 <log(E/eV) <19.7) where the mean value of $\langle \Delta \rangle$ falls bellow the iron prediction. This is a first indication of $\langle \Delta \rangle$ as a valuable observable to identify potential problems in the description that models do of the hadronic component of the EAS.

Considering the behavior of $\langle \Delta \rangle$ shown in figure 8.4 an interesting exercise would be the evaluation of the elongation rate, i.e. the slope of $\langle \Delta \rangle$ as function of the energy, for consecutive groups of 5 energy bins. With this exercise we can evaluate how the elongation rate changes with the energy. Table 8.3 shows the different groups of 5 energy bins that we have used to fit with a straight line the behaviour of $\langle \Delta \rangle$ as a function of energy. The table also shows the slopes of each fit together with their respective uncertainties.

Groups of 5 energy bins	**Slopes from the fit**
18.5 $\leq$ log(E/eV) <19.0	0.45 $\pm$ 0.02
18.6 $\leq$ log(E/eV) <19.1	0.46 $\pm$ 0.03
18.7 $\leq$ log(E/eV) <19.2	0.55 $\pm$ 0.03
18.8 $\leq$ log(E/eV) <19.3	0.57 $\pm$ 0.04
18.9 $\leq$ log(E/eV) <19.4	0.63 $\pm$ 0.05
19.0 $\leq$ log(E/eV) <19.5	0.56 $\pm$ 0.07
19.1 $\leq$ log(E/eV) <19.6	0.58 $\pm$ 0.08
19.2 $\leq$ log(E/eV) <19.7	0.38 $\pm$ 0.09
19.3 $\leq$ log(E/eV) <19.8	0.48 $\pm$ 0.12
log(E/eV) $\geq$ 19.4	0.99 $\pm$ 0.14

Cuadro 8.3: Slopes with their respective uncertainties obtained from fitting with a straight line the behaviour of $\langle \Delta \rangle$ as a function of the energy for different groups of 5 energy bins.

The values shown in table 8.3 indicate that the slope of $\langle \Delta \rangle$ slightly increases as a function of energy up to log(E/eV) = 19.6. When we consider energies between 19.2 $\leq$ log(E/eV) < 19.7 we find a smaller value for the slope (0.38 $\pm$0.09). For the last interval of energies we observe that the slope is much greater than all the previous ones (0.99 $\pm$ 0.14). This might be indicative of a change in mass composition towards lighter nuclei.

8.3 Interpretation of the results in term of $\langle \ln A \rangle$

To study the discrepancies between models and data further, and to make comparisons with other estimates, we have transformed the measurements of $\langle \Delta \rangle$ (and their corresponding uncertainties) into mass units. For each hadronic model, the value of $\langle \ln A \rangle$ derived from data has been computed using the following relationship:

$$\langle \ln A \rangle = \ln 56 \frac{\overline{\langle \Delta \rangle}_p - \overline{\langle \Delta \rangle}_{data}}{\overline{\langle \Delta \rangle}_p - \overline{\langle \Delta \rangle}_{Fe}} \tag{8.1}$$

In this case, due to the definition of $\langle\Delta\rangle$ using brackets, $\langle\rangle$, we have indicated the use of mean values with an overline.

The results of this transformation are shown in figure 8.5 for QGSJetII.04 (top panel) and EPOS-LHC (bottom panel). As expected, neither result coincides with the values of $\langle \ln A\rangle$ obtained with the FD measurements of X_{max} [78]. This happens because $\langle\Delta\rangle$ is an observable with information about the muonic part of the showers, as we explained at the beginning of this chapter. Despite the disagreement between the predictions obtained with $\langle\Delta\rangle$ and with X_{max}, it is interesting to note that for the range of energies where the FD analysis and this one overlap, the evolution of $\langle \ln A\rangle$ as a function of energy follows a very similar increasing trend for both hadronic models.

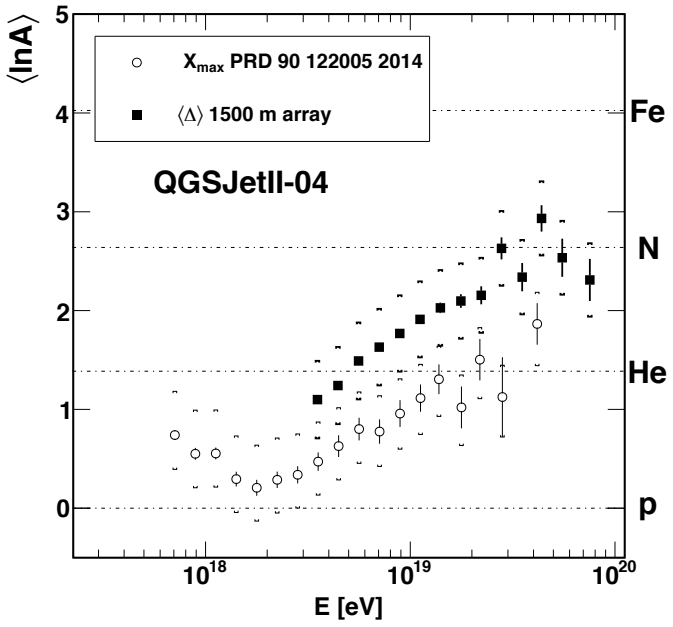
When looking at the predictions for each particular model, we see that the interpretation with EPOS-LHC leads to a heavier average composition than with QGSJetII.04. Regardless of this difference in terms of absolute values, for both hadronic models we observe a change in the increasing trend of $\langle \ln A\rangle$ above $10^{19,6}$ eV. The last two energy bins indicate a turn towards a composition with lighter elements.

We conclude by making a comparison in figure 8.6 of the $\langle \ln A\rangle$ values obtained from the measurements of $\langle\Delta\rangle$ to the ones obtained with two different mass estimators already used by the Pierre Auger Collaboration: (sec $\theta)_{max}$ [79] (top panel) and X^{μ}_{max} [80] (bottom panel). Both estimators and their results were described in chapter 1.

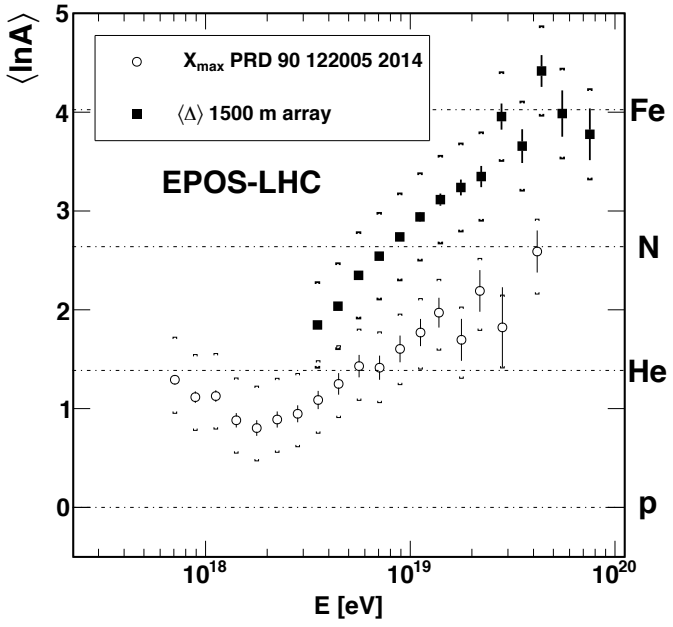
The four mass measurements shown in figure 8.6 are sensitive to different types of hadronic interactions since the importance of the muonic component is different within each one of them. In the direct determination of X_{max} [78], the dominant component is the electromagnetic one and the proportion of muons in the shower is of minor importance. By contrast, X^{μ}_{max} [80] is dominated by the muon component which is the result of a long cascade of lower energy hadronic interactions. This explains why the maximum difference between the values of $\langle \ln A\rangle$ is observed between the predictions of X_{max} and X^{μ}_{max}. The situation for (sec $\theta)_{max}$ [79] is between the previous ones. (sec $\theta)_{max}$ is associated with a complex interplay between these two components and for this reason the predicted $\langle \ln A\rangle$ values are in the middle of the values given by X_{max} and X^{μ}_{max}.

Our observable, $\langle\Delta\rangle$, is in a similar situation to the one of (sec $\theta)_{max}$ (recall that both observables are based on risetime measurements). This means that $\langle\Delta\rangle$ is also related to the two components of an EAS. For this reason we expect that the predicted values for $\langle \ln A\rangle$ using $\langle\Delta\rangle$ to be between the predictions given by X_{max} and X^{μ}_{max}. This is exactly what we see in figure 8.6b. In fact, the values of $\langle \ln A\rangle$ obtained with $\langle\Delta\rangle$ are compatible with the results of (sec $\theta)_{max}$ taking into account the systematic uncertainties and the larger statistical uncertainties in (sec $\theta)_{max}$, see figure 8.6a. The agreement between the results of $\langle\Delta\rangle$ and (sec $\theta)_{max}$ is larger for the predictions given by the hadronic model EPOS-LHC.

These results imply, as in the case of (sec $\theta)_{max}$, that the study of $\langle\Delta\rangle$ itself can be used to probe the validity of hadronic interaction models.

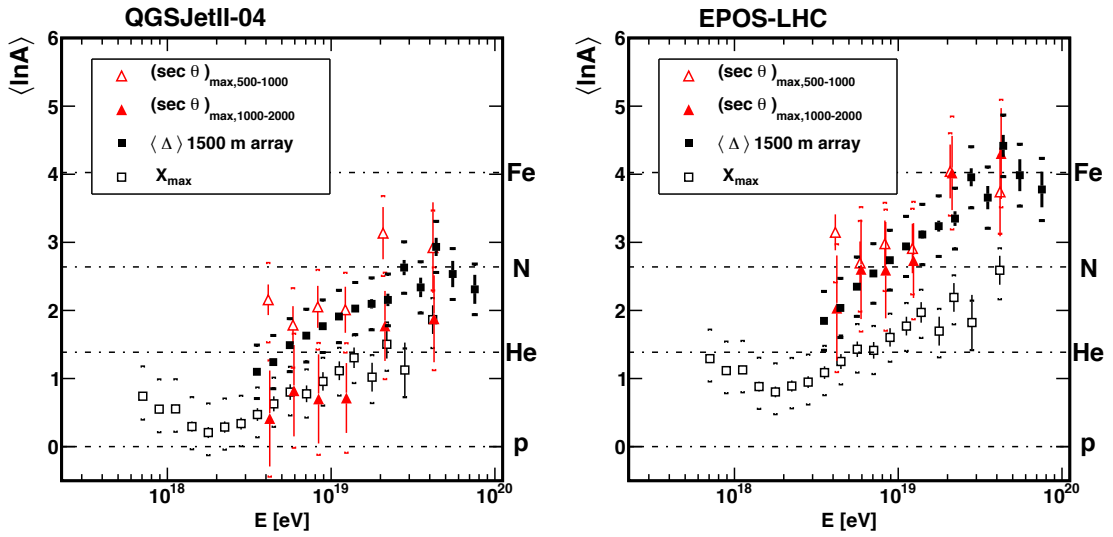


(a)

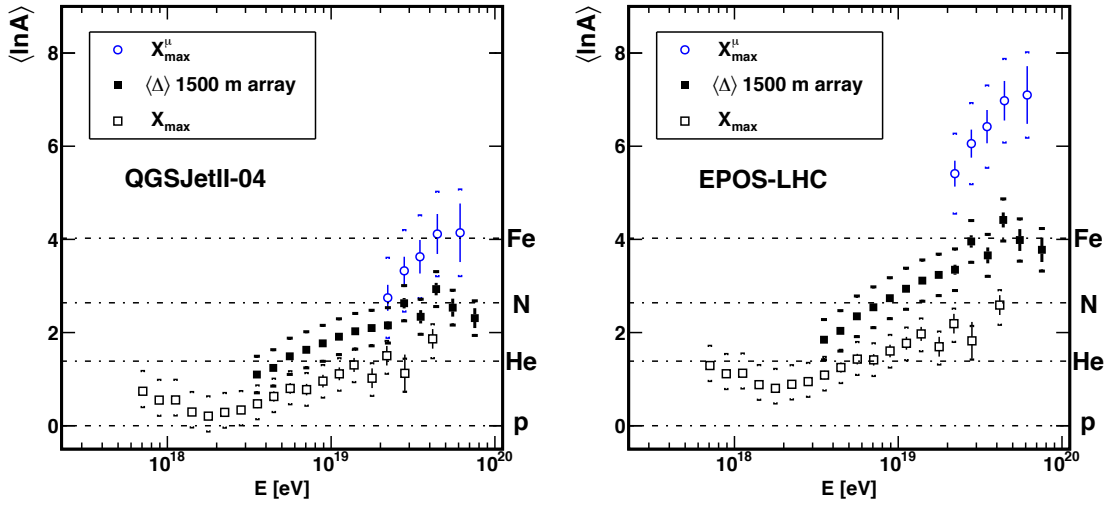


(b)

Figura 8.5: Conversion of $\langle \Delta \rangle$ to $\langle \ln A \rangle$. The values of $\langle \ln A \rangle$ coming from the measurements of X_{max} done with the FD are also included for comparison [78]. On the top (bottom) panel we use QGSJetII.04 (EPOS-LHC) as a reference hadronic model. Brackets correspond to the systematic uncertainties.



(a)



(b)

Figura 8.6: $\langle \ln A \rangle$ as a function of energy as predicted by QGSJetII.04 and EPOS-LHC. The values given in figure 8.5 are compared with those from the analysis of (sec θ)$_{max}$ (top panel) and from the MPD method (bottom panel).Brackets correspond to the systematic uncertainties.

Abridgement of results

Nowadays there are a large number of unanswered questions related to the nature and the origin of cosmic rays. One of these puzzles is the determination of the mass composition of ultra-high-energy cosmic rays, crucial for the understanding of which are the sources of cosmic rays and the mechanisms of acceleration that they are exposed to.

In this thesis we have introduced a method, the $\langle\Delta\rangle$ method, which is able to provide, for SD data, mass composition inferences over a broad energy range and with a statistical sample that is larger than the one used in fluorescence measurements. Thanks to our calibration procedure it is possible to convert the observable $\langle\Delta\rangle$ in an estimator of the depth of air shower maximum, $X_{\mathrm{max}}^{\mathrm{Delta}}$ (mainly depending on the electromagnetic component), making possible the comparison with the models.

The improvements implemented in the $\langle\Delta\rangle$ method as well as the successful results achieved from it are summarized next:

■ Improvements in the experimental determination of the risetime and its uncertainty

The $\langle\Delta\rangle$ method is based on the risetime of the signals measured by the surface detectors of the Pierre Auger Observatory. For this reason, it is crucial to guarantee the correct determination of the risetime and its uncertainty.

- One of the key pieces for the correct determination of the risetime is the use of a specific correction to avoid a problem with the baseline of the signals registered in the low-gain channel. Although the correction was already implemented in the Auger software it has been necessary to check the effect of this software modification in the standard event reconstruction. We have done this study with data and simulated events, showing that the effect is negligible for the energy resolution and it only modifies the angular resolution by less than 0.1 degree.

- We have introduced a new approach for the risetime calculation which takes into account the lack of information derived from the 40 MHz sampling of the signals of the surface detectors. The new approach is more appropriate because it does not assume anything about the behaviour of the signals inside each time bin and allows a better control of the risetime uncertainty.

- The risetimes are affected by an azimuthal asymmetry which depends on the core distance and zenith angle. For the $\langle\Delta\rangle$ method is crucial to keep the circular symmetry around the shower core and this asymmetry has to be corrected. Whereas for the data of the 1500 m array this correction already existed, there was no study

for the data of the 750 m array. With the aim of using the data of the 750 m array for mass composition analyses we have obtained the first asymmetry correction for this data set.

- The last step in the risetime determination is the calculation of its uncertainty. Although in previous works the risetime uncertainty had been already studied using both twin and pair of detectors, this is the first analysis with enough statistics to obtain an accurate parameterization for the uncertainty that is exclusively based on data. Furthermore, our novel idea of combining the results of twins and pairs allows us to obtain a parameterization covering a wider distance range. Thanks to this there is no longer the need of performing extrapolations.

■ Improvements in the technicalities of the $\langle\Delta\rangle$ method

The $\langle\Delta\rangle$ method was already used in previous works with inconclusive results. However in this thesis we have been able to bring to fruition the potentialities of this method.

- The core concept of the $\langle\Delta\rangle$ method is the benchmark. For this reason it is essential that the benchmark represents a faithful description of the risetime behaviour. We have realized that risetimes obtained from different readout electronic channels behave in a different way as a function of the core distance. This fact has forced us to treat separately the risetimes from the high- and low-gain channels in the whole analysis presented in this thesis, including the determination of a different benchmark parameterization for each one. This is the first time that this distinction has been implemented and it has has been crucial for the proper development of the $\langle\Delta\rangle$ method.
- Besides the benchmark the other key aspect of the $\langle\Delta\rangle$ method is the calibration procedure. The correlation observed between the depth of air shower maximum, X_{max}, and $\langle\Delta\rangle$ for a sample of simulated events shows that $\langle\Delta\rangle$ can be used to infer the composition of cosmic rays.

■ Mass composition results at the largest energies

- With the previous considerations we have applied the $\langle\Delta\rangle$ method to data of the 1500 m array and **we have obtained an elongation rate for SD-only events extended from $10^{18,5}$ eV up to nearly 10^{20} eV.** This SD elongation rate is compatible with the one obtained by the Pierre Auger Observatory with the FD in the energy range where they overlap.
- The systematic uncertainties of our measurements are $\sim$11 g cm^{-2}, similar to the values achieved with the FD. The main contribution comes from the calibration procedure, since the uncertainties of X_{max} directly propagate into our measurements.
- Due to the huge statistics provided by the 1500 m array **we have added three more bins at the highest energies measured by the FD**. We have 517 events above $10^{19,5}$ eV whereas the FD only has 37 events above this energy. A factor 14 bigger.

- Our results suggest that the flux of cosmic rays is composed mainly by light nuclei with a fraction of heavy nuclei increasing with energy up to $10^{19,6}$ eV. The same conclusions are derived from the FD. **However, above $10^{19,6}$ eV our results seem to suggest an increase of the light component in the cosmic ray flux**.

■ Extension of the mass composition results at lower energies

- The deep understanding of the $\langle \Delta \rangle$ method has allowed us to apply it to the data of the 750 m array. **These new measurements cover an energy range from $10^{17,5}$ eV up to $10^{18,5}$eV extending to lower energies the results mentioned above**. The extension of the elongation rate to lower energies using the same method in a fully independent data set has an enormous importance because this gives additional support to our method. **This is the first mass composition analysis done with data of the 750 m array.**

- The systematic uncertainty in this case is $\sim$14 g cm^{-2}, larger than the one found for the data of the 1500 m array. This increase of the uncertainty is due to the fact that the calibration procedure is done with a smaller data sample.

- **Thanks to the low-energy extension allowed by the data of the 750 m array, we have obtained for the first time an elongation rate with only SD-measurements covering more than 2 orders of magnitude in energy, from $10^{17,5}$ eV up to nearly 10^{20}eV.**

■ Confirmation of the inadequate predictions provided by hadronic models

In addition to the mass composition results obtained with the $\langle \Delta \rangle$ method, the observable $\langle \Delta \rangle$ itself can also be a powerful tool to extract precious information about the validity of the hadronic models. By studying the behaviour of $\langle \Delta \rangle$ as a function of the energy we can assess the level of agreement between data and models. In particular, since $\langle \Delta \rangle$ contains information about the muonic part of the showers, we can gauge how good the description of muons is.

- The predicted values for $\langle \ln A \rangle$ obtained from this comparison are incompatible with the predictions obtained from observables based on the electromagnetic part of the air showers. **Our result confirms the tension between the hadronic interaction models and the data in a energy range extending from $10^{18,5}$ eV up to nearly 10^{20}eV**.

Compendio de resultados

Todavía hoy existen un gran número de preguntas por resolver sobre la naturaleza y el origen de los rayos cósmicos. Uno de esos misterios es la determinación de la composición química de los rayos cósmicos más energéticos, crucial para la comprensión de cuáles son las fuentes y los mecanismos de aceleración a los que están expuestos.

En esta tesis hemos introducido un método, el método $\langle\Delta\rangle$, que es capaz de proporcionar, sólo para datos del SD, conclusiones sobre la composición en rango de energías muy amplio y con una muestra estadística mayor que la usada en las medidas de fluorescencia. Gracias a nuestro proceso de calibración es posible transformar el observable $\langle\Delta\rangle$ en un estimador de la profundidad del máximo de las cascadas, $X_{\mathrm{max}}^{\mathrm{Delys}}$ (principalmente basado en la componente electromagnética), haciendo posible así la comparación con los modelos.

Las mejoras implementadas en el método $\langle\Delta\rangle$ así como los exitosos resultados logrados con él se resumen a continuación.

■ Mejoras en la determinación experimental del risetime y su error

El método $\langle\Delta\rangle$ está basado en el risetime de las señales medidas por los detectores de superficie del Observatorio Pierre Auger. Por este motivo, es crucial garantizar la correcta determinación del mismo y su error.

- Una de las piezas clave para la correcta determinación de los risetimes es el uso de una corrección que evita un problema con el baseline de las señales registradas en el canal de baja ganancia. Aunque esta corrección ya había sido implementada en el software de Auger ha sido necesario estudiar su efecto sobre la reconstrucción de los sucesos. Hemos hecho este estudio tanto con datos como con simulaciones y hemos visto que el efecto de la correción es despreciable sobre la resolución de la energía y que sólo cambia la resolución angular en menos de 0.1 grados.

- Hemos introducido una nueva forma de calcular los risetimes que tiene en cuenta la falta de conocimiento derivada del muestreo a 40 MHz de las señales de los detectores de superficie. El nuevo enfoque es más apropiado porque no asume nada sobre el comportamiento de las señales dentro de los bines temporales y además permite un mejor control del error.

- Los risetimes están afectados por una asimetría azimutal que depende de la distancia al núcleo de la cascada y del zénit. Para el método $\langle\Delta\rangle$ es crucial conservar la simetría circular alrededor del núcleo y por eso esta asimetría debe corregirse.

Mientras que para los datos recogidos por los detectores separados 1500 m esta corrección ya existía, no había ningún estudio para los datos de los detectores separados 750 m. Con el objetivo de usar los datos de los detectores separados 750 m para estudios de composición hemos obtenido la primera corrección de la asimetría azimuthal de estos datos.

- El último paso en la determinación del risetimes es el cálculo de su error. Aunque en trabajos anteriores el error ya había sido estudiado usando detectores *pares* y *gemelos*, este es el primer análisis con la suficiente estadística como para obtener un precisa parametrización del error que se basa exclusivamente en el uso de datos. Además nuestra innovadora idea de combinar los resultados de detectores pares y gemelos permite obtener una parametrización que cubre un rango de distancias más amplio. Gracias a esto no es necesario realizar extrapolaciones.

■ Mejoras en los detalles técnicos del método $\langle\Delta\rangle$

El método $\langle\Delta\rangle$ ya había sido usado en trabajos anteriores con resultados poco concluyentes. Sin embargo en esta tesis hemos sido capaces de llevar a buen término todo el potencial de este método.

- El concepto central del método $\langle\Delta\rangle$ es el *comportamiento de referencia* o *benchmark*. Por este motivo es esencial que el benchmark sea una fiel descripción del comportamiento de los risetimes. Aquí nos hemos dado cuenta de que los risetimes obtenidos de los diferente canales de la electrónica se comportan de forma diferente en función de la distancia al núcleo. Este hecho nos ha forzado a tratar de forma separada los risetimes de los canales de alta y baja ganancia en todo el análisis presentado en esta tesis, incluyendo la determinación de un benchmark para cada uno. Esta es la primera vez que esta distinción se ha implementado y ha sido crucial para el correcto desarrollo del método $\langle\Delta\rangle$.
- Además del bechmark el otro aspecto clave del método $\langle\Delta\rangle$ es la calibración. Con la correlación observada entre la profundidad del máximo de las cascadas, X_{max}, y $\langle\Delta\rangle$ para una conjunto de sucesos simulados, hemos demostrado que el observable $\langle\Delta\rangle$ puede utilizarse para deducir la composición de los rayos cósmicos.

■ Resultados de composición a las energías más altas

- Con todas las anteriores consideraciones hemos aplicado el método $\langle\Delta\rangle$ a los datos obtenidos con los detectores separados 1500 m y **hemos obtenido una tasa de elongación sólo para sucesos SD que se extiende desde $10^{18,5}$ eV hasta casi 10^{20} eV.** Esta tasa de elongación del SD es compatible con los resultados obtenidos por el Observatorio Pierre Auger con el FD en el rango de energía donde ambos coinciden.
- Los errores sistemáticos de nuestras medidas son $\sim$11 g cm^{-2}, similares a los obtenidos con el FD. La principal contribución a este error procede de la calibración, ya que los errores de X_{max} se propagan directamente en nuestras medidas.

- Debido a la enorme estadística proporcionada por los detectores que están separados 1500 m **hemos añadido 3 puntos más a las energías más altas medidas por el FD**. Nosotros tenemos 517 sucesos por encima de $10^{19,5}$ mientras que el FD ahí sólo tiene 37 sucesos. Un factor 14 veces más grande.

- Nuestros resultados sugieren que el flujo de los rayos cósmicos está compuesto principalmente por núcleos ligeros con una fracción de núcleos pesados que va incrementándose hasta $10^{19,6}$ eV. Las mismas conclusiones obtenidas por el FD. **Sin embargo, por encima de $10^{19,6}$ eV nuestros resultados sugieren un incremento de los componentes ligeros en el flujo de rayos cósmicos.**

■ Extensión de los resultados de composición a energías más bajas

- La profunda comprensión del método $\langle \Delta \rangle$ nos ha permitido aplicarlo también a los datos recogidos por los detectores separados sólo 750 m. **Estas nuevas medidas cubren un rango de energía que va desde $10^{17,5}$ eV hasta $10^{18,5}$, extendiendo a energías más bajas el resultado antes mencionado**. La extensión de la tasa de elongación a energías más bajas usando el mismo método en un conjunto de datos completamente independiente tiene una enorme importancia porque le da un respaldo adicional a nuestro método. **Este es el primer análisis de composición que se ha hecho usando los datos de los detectores separados 750 m.**

- El error sistemático en este caso es de $\sim$14 g cm^{-2}, mayor que el encontrado para los datos de los detectores separados 1500 m. Este incremento del error se debe al hecho de que la calibración se ha realizado con un conjunto de datos más reducido.

- **Gracias a la extensión a las bajas energías permitida por los datos de los detectores separados 750 m, hemos obtenido por primera vez una tasa de elongación con medidas hechas sólo por el SD que cubre más de 2 órdenes de magnitud en energía, desde $10^{17,5}$ eV hasta cerca de 10^{20} eV.**

■ Confirmación de las inadecuadas predicciones hechas por los modelos hadrónicos

Además de los resultados de composición obtenidos con el método $\langle \Delta \rangle$, el observable $\langle \Delta \rangle$ por si solo también puede ser una poderosa herramienta para extraer información sobre la validez de los modelos hadrónicos. Estudiando el comportamiento de $\langle \Delta \rangle$ como función de la energía podemos evaluar el nivel de acuerdo entre los datos y los modelos. En particular, ya que $\langle \Delta \rangle$ contiene información sobre la parte muónica de las cascadas, podemos estimar como de buena es la descripción de los muones.

- Los valores predichos para $\langle \ln A \rangle$ sacados de esta comparación son incompatibles con las predicciones obtenidas con los observables basados en la parte electromagnética de las cascadas. **Nuestro resultado confirma la tensión entre los modelos hadrónicos y los datos en un rango de energía que se extiende desde $10^{18,5}$ eV hasta casi 10^{20} eV.**

A

End to end calibration with simulations

The full analysis chain of the $\langle\Delta\rangle$ method can be validated by applying it to simulated events and comparing the values of $X_{\text{max}}^{\text{Delta}}$ to the X_{max} values at generator level, i.e. the true X_{max} values before the detector simulation.

This test validates the ability of the $\langle\Delta\rangle$ method to produce reliable values of X_{max} for different mass compositions. This is an important feature needed for the comparison of the values of $X_{\text{max}}^{\text{Delta}}$ obtained for the data with the values of X_{max} predicted for different composition hypotheses.

The goal of this study is to validate independently the proper performance of the $\langle\Delta\rangle$ method for different composition hypotheses. For this reason in this case it is necessary to use different benchmark parameterizations for protons and iron nuclei. Each mass composition has to be described using its own benchmark to follow the same prescriptions given in chapter 4 to apply the $\langle\Delta\rangle$ method. The benchmark parameterizations used here are the same ones used in section 4.6 to study the level of correlation between $\langle\Delta\rangle$ and X_{max}. In that section we saw that those benchmark parameterizations describe successfully the behaviour of simulated events since the mean values of the $\langle\Delta\rangle$ distributions were compatible with zero in the benchmark energy bin (19.1 < log(E/eV) < 19.2) (see figure 4.16).

The results from this test are shown in figure A.1. For both compositions, the values of X_{max} (solid lines) can be reproduced well with the values of $X_{\text{max}}^{\text{Delta}}$ resulting of the $\langle\Delta\rangle$ method. The simulated measurements of $X_{\text{max}}^{\text{Delta}}$ agree within $\pm$ 4 g cm^{-2} with the generated values. A slightly larger bias is visible for the first energy bin with an iron composition. This bias could be partially attributed to the anomalies pointed in [152] for the simulated risetimes of the smaller signals.

Given these results we conclude that the analysis chain developed in this thesis performs well and that the calibration procedure to transform a $\langle\Delta\rangle$ value in a estimator of X_{max} does not introduce any bias in our final elongation rate.

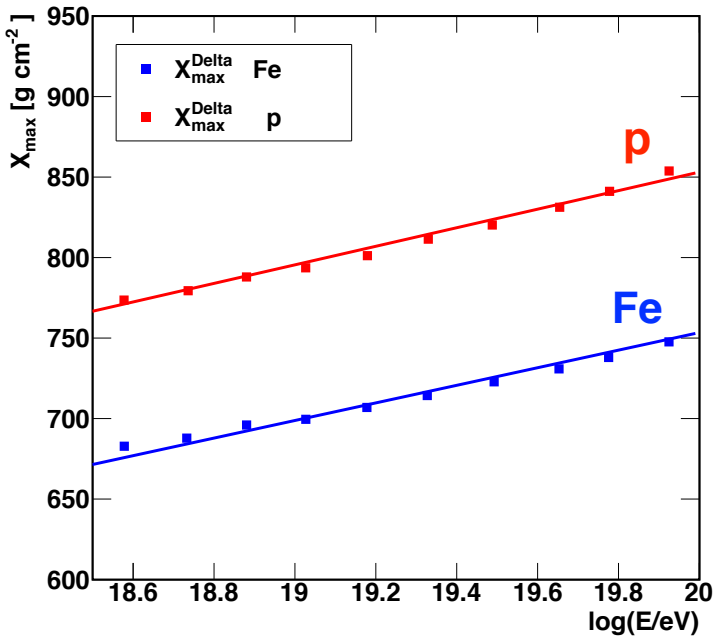


Figure A.1: Values of $X_{\mathrm{max}}^{\mathrm{Delta}}$ obtained after the whole analysis chain is applied to simulated showers of proton and iron primaries using the QGSJetII.04 hadronic model. The values of X_{max} of the generated events before the detector simulation are shown as solid lines.

B

The ten most-energetic events

We list here the main parameters of the ten most-energetic events selected in our analysis. Figure B.1 shows their individual $\langle\Delta\rangle$ values compared to the average values of $\langle\Delta\rangle$ of the highest energy bin. Table B.1 compiles the most relevant features of these events.

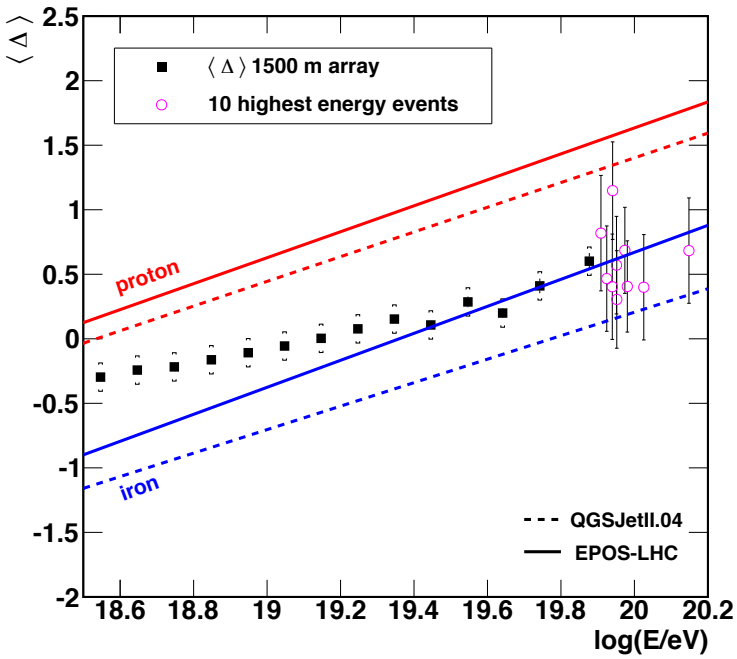


Figure B.1: Same as figure 8.4 including the individual values of $\langle\Delta\rangle$ from the ten most energetic events.

EventId	Date	Time	log(E/eV)	Zenith[°]	Det.	$\langle\Delta\rangle$	$X_{\mathrm{max}}^{\mathrm{Delta}}$ g cm^{-2}
1096757	04/12/04	13:41:20	19.952	44.6	7 (3)	0.305	788
1748196	03/11/05	08:22:50	19.925	14.1	6 (3)	0.466	797
3036270	14/01/07	09:49:46	20.150	14.4	6 (1)	0.679	809
3687808	13/07/07	08:22:26	19.951	17.7	7 (3)	0.571	802
6465006	21/09/08	08:22:47	19.974	44.5	9 (4)	0.685	809
7425978	19/03/09	23:08:25	19.909	27.4	5 (2)	0.818	816
11002850	27/01/11	06:13:37	20.025	24.9	6 (2)	0.399	793
12631868	10/09/11	02:07:16	19.940	24.7	6 (3)	0.403	793
21712764	14/05/13	18:50:00	19.941	45.2	7 (3)	1.148	835
29228282	21/09/14	15:49:51	19.981	36.6	8 (3)	0.406	794

Table B.1: Most relevant features of the 10 events with the most extreme energies observed in our analysis. In the column *Det.* we refer the number of detectors of each event used in this analysis. In brackets we indicate the number of these detectors whose high-gain channel is saturated.

In figures B.2 and B.3 we show the *EventBrowser* view of the two most energetic events listed in table B.1.

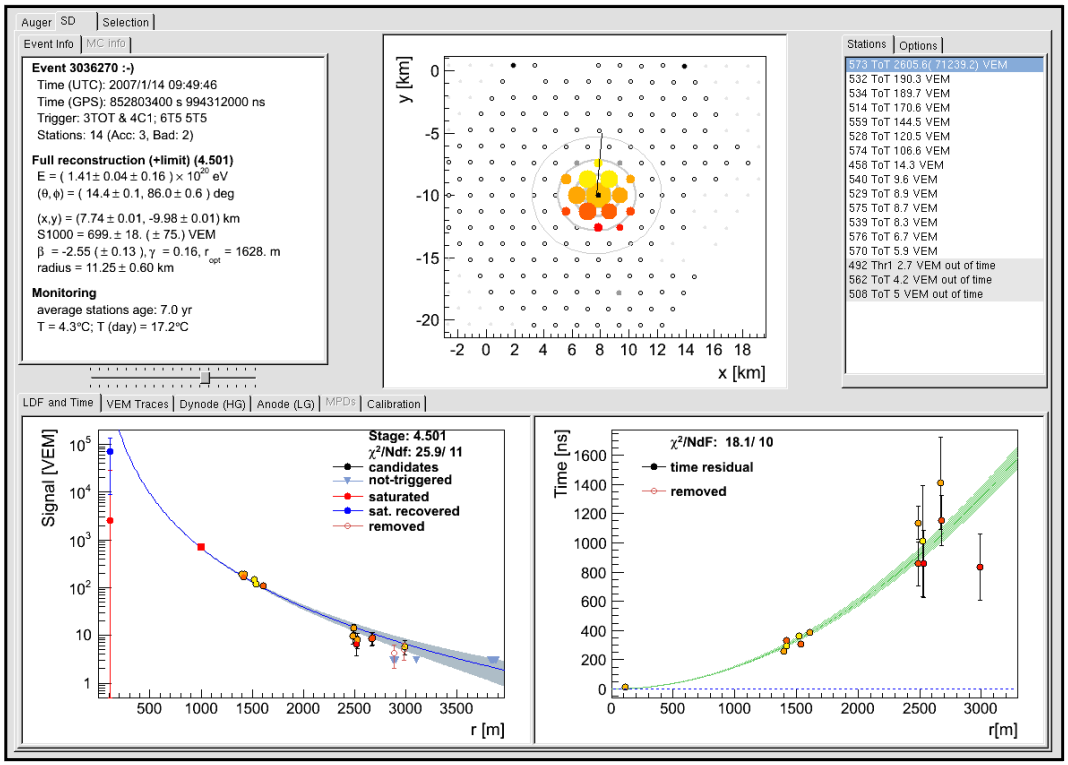


Figure B.2: View in the Auger software of the Event 3036270.

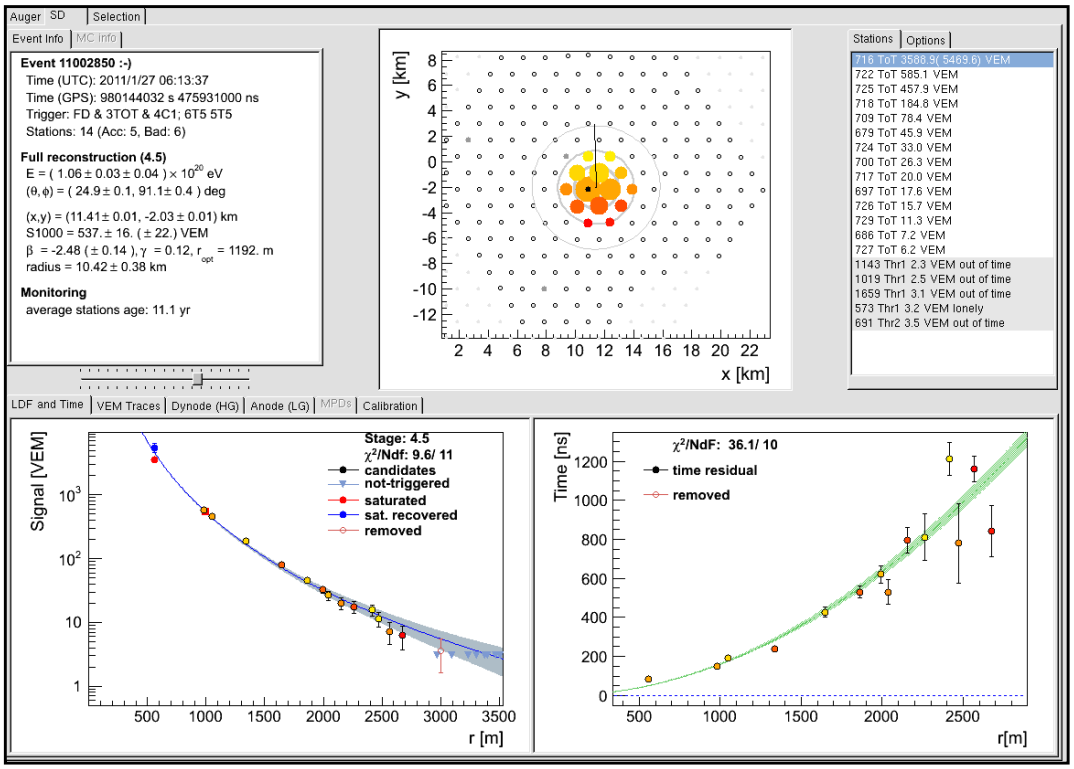


Figure B.3: View in the Auger software of the Event 11002850.

List of Figures

List of Tables

Bibliography

[1] C. Coulomb, *Mdm. de l'Acad. des Sciences*, Académie royale des sciences de Paris , 612 (1875).

[2] H. Becquerel, *Sur les radiations émises par phosphorescence*, Comptes rendus de l'Acad. des Sciences **420** (1896).

[3] T. Wulf, *Beobachtungen über die Strahlung hoher Durchdringungsfähigkeit auf dem Eiffelturm*, Physikalische Zeitschrift , p. 811 (1910).

[4] D. Pacini, *Penetrating Radiation at the Surface of and in Water*, Nuovo Cim. **8**, 93 (1912), arXiv:1002.1810.

[5] Cern Time Line, http://timeline.web.cern.ch/timelines/cosmic-rays.

[6] V. F. Hess, *Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten*, Phys. Z. **13**, 1084 (1912).

[7] R. A. Millikan and G. H. Cameron, *High Frequency Rays of Cosmic Origin III. Measurements in Snow-Fed Lakes at High Altitudes*, Phys. Rev. **28**, 851 (1926).

[8] W. Bothe and W. Kolhörster, *Das Wesen der Höhenstrahlung*, Z. Phys. **56**, 751 (1929).

[9] B. Rossi, *Absorptionsmessungen der durchdringenden Korpuskularstrahlung in einem Meter Blei*, Die Naturwissenschaften **20**, 65 (1932).

[10] B. Rossi, *On the Magnetic Deflection of Cosmic Rays*, Physical Review **36**, 606 (1930).

[11] P. Auger, R. Maze, and T. Grivet-Meyer, *Grandes gerbes cosmiques atmosphériques contenant des corpuscules ultra pénétrantes*, Comptes Rendus **206**, 1721 (1938).

[12] P. Auger, P. Ehrenfest, R. Maze, J. Daudin, and A. F. Robley, *Extensive cosmic ray showers*, Rev. Mod. Phys. **11**, 288 (1939).

[13] J. Linsley, *Evidence for a primary cosmic-ray particle with energy 10^{20} eV*, Phys. Rev. Lett. **10**, 146 (1963).

[14] M. Gell-Mann, *A Schematic Model of Baryons and Mesons*, Phys. Lett. **8**, 214 (1964).

[15] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, Developments in the Quark Theory of Hadrons, Volume 1. Edited by D. Lichtenberg and S. Rosen , 22 (1964).

[16] G. Rochester and C. Butler, *Evidence for the existence of new unstable elementary particles*, Nature **160**, 855 (1947).

[17] T. R. Jaeger, R. L. Mutel, and K. G. Gayley, *Project RESUN, a radio EVLA search for UHE neutrinos*, Astropart. Phys. **34**, 293 (2010), arXiv:0910.5949.

[18] J. W. Cronin, S. P. Swordy, and T. K. Gaisser, *Cosmic rays at the energy frontier*, Sci. Am. **276**, 32 (1997).

[19] Courtesy of Ioana Maris for the GAP Report No. 092, 2015.

[20] A. A. Ivanov, S. P. Knurenko, and I. Y. Sleptsov, *Measuring extensive air showers with Cherenkov light detectors of the Yakutsk array: the energy spectrum of cosmic rays*, New J. Phys. **11**, 065008 (2009).

[21] M. Nagano *et al.*, *Energy spectrum of primary cosmic rays above $10^{17.0}$ eV determined from extensive air shower experiments at Akeno*, J. Phys. G Nucl. Partic. **18**, 423 (1992).

[22] T. Antoni *et al.*, *KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems*, Astropart. Phys. **24**, 1 (2005).

[23] The KASCADE-Grande Collaboration, W. Apel *et al.*, *KASCADE-Grande measurements of energy spectra for elemental groups of cosmic rays*, Astropart. Phys. (2013), arXiv:1306.6283.

[24] J. R. Hörandel, *Models of the knee in the energy spectrum of cosmic rays*, Astropart. Phys. **21**, 241 (2004).

[25] A. Letessier-Selvon and T. Stanev, *Ultrahigh Energy Cosmic Rays*, Rev. Mod. Phys. **83**, 907 (2011), arXiv:1103.0031.

[26] Pierre Auger Collaboration, I. Valino, *The flux of ultra-high energy cosmic rays after ten years of operation of the Pierre Auger Observatory*, PoS **ICRC2015**, 271 (2016).

[27] D. J. Bird *et al.*, *Evidence for correlated changes in the spectrum and composition of cosmic rays at extremely high energies*, Phys. Rev. Lett. **71**, 3401 (1993).

[28] Telescope Array, W. Hanlon and D. Ikeda, *Energy Spectrum and Mass Composition of Ultra-High Energy Cosmic Rays Measured by the hybrid technique in Telescope Array*, PoS **ICRC2015**, 362 (2016).

[29] A. M. Hillas, *Cosmic Rays: Recent Progress and some Current Questions*, (2006), arXiv:astro-ph/0607109.

[30] D. De Marco and T. Stanev, *On the shape of the ultrahigh energy cosmic ray spectrum*, Phys. Rev. D **72**, 081301 (2005).

[31] D. Allard, E. Parizot, and A. Olinto, *On the transition from galactic to extragalactic cosmic-rays: Spectral and composition features from two opposite scenarios*, Astropart. Phys. **27**, 61 (2007).

[32] V. Berezinsky, *Transition from galactic to extragalactic cosmic rays*, (2007), arXiv:0710.2750.

[33] V. Berezinsky, S. Grigorieva, and B. Hnatyk, *Extragalactic UHE proton spectrum and prediction for iron-nuclei flux at* $10^8 - 10^9$ *GeV*, Astropart. Phys. **21**, 617 (2004).

[34] V. S. Berezinsky, S. I. Grigoreva, and B. I. Hnatyk, *Extragalactic UHE proton spectrum and prediction of flux of iron-nuclei at* $10^8 - 10^9$ *GeV*, Nucl. Phys. Proc. Suppl. **151**, 497 (2006), [497(2006)].

[35] K. Greisen, *End to the cosmic ray spectrum?*, Phys. Rev. Lett. **16**, 748 (1966).

[36] G. Zatsepin and V. Kuzmin, *Upper limit of the spectrum of cosmic rays*, JETP Lett. **4**, 78 (1966).

[37] A. A. Penzias and R. W. Wilson, *A Measurement of excess antenna temperature at 4080-Mc/s*, Astrophys.J. **142**, 419 (1965).

[38] D. Harari, *The flux suppression at the highest energies*, Comptes Rendus Physique **15**, 376 (2014), arXiv:1406.1117.

[39] R. Aloisio, V. Berezinsky, and A. Gazizov, *Ultra high energy cosmic rays: The disappointing model*, Astropart. Phys. **34**, 620 (2011).

[40] A. R. Bell, *The acceleration of cosmic rays in shock fronts. I*, Monthly Notices of the Royal Astronomical Society **182**, 147 (1978).

[41] A. M. Hillas, *The Origin of Ultrahigh-Energy Cosmic Rays*, Ann. Rev. Astron. Astrophys. **22**, 425 (1984).

[42] L. O. Drury, *Acceleration of cosmic rays*, Contemporary Physics **35**, 231 (1994).

[43] J. Linsley, *Fluctuation effects on directional data*, Phys. Rev. Lett. **34**, 1530 (1975).

[44] Pierre Auger Collaboration, A. Aab *et al.*, *Large Scale Distribution of Ultra High Energy Cosmic Rays Detected at the Pierre Auger Observatory With Zenith Angles up to 80*$^\circ$, Astrophys. J. **802**, 111 (2015), arXiv:1411.6953.

[45] Pierre Auger Collaboration, J. Abraham *et al.*, *Correlation of the highest energy cosmic rays with nearby extragalactic objects*, Science **318**, 938 (2007), arXiv:0711.2256.

[46] Pierre Auger Collaboration, J. Abraham *et al.*, *Correlation of the highest-energy cosmic rays with the positions of nearby active galactic nuclei*, Astropart. Phys. **29**, 188 (2008), arXiv:0712.2843, [Erratum: Astropart. Phys.30,45(2008)].

[47] Pierre Auger Collaboration, P. Abreu *et al.*, *Update on the correlation of the highest energy cosmic rays with nearby extragalactic matter*, Astropart. Phys. **34**, 314 (2010), arXiv:1009.1855.

[48] Pierre Auger Collaboration, A. Aab *et al.*, *Searches for Anisotropies in the Arrival Directions of the Highest Energy Cosmic Rays Detected by the Pierre Auger Observatory*, Astrophys. J. **804**, 15 (2015), arXiv:1411.6111.

[49] Pierre Auger Collaboration, C. Bleve, *Update of the neutrino and photon limits from the Pierre Auger Observatory*, PoS **ICRC2015**, 271 (2016).

[50] Pierre Auger Collaboration, A. Aab *et al.*, *Improved limit to the diffuse flux of ultrahigh energy neutrinos from the Pierre Auger Observatory*, Phys. Rev. **D91**, 092008 (2015), arXiv:1504.05397.

[51] M. Kachelriess, *The rise and fall of top-down models as main UHECR sources*, (2008), arXiv:0810.3017.

[52] M. Ahlers, L. A. Anchordoqui, M. C. Gonzalez-Garcia, F. Halzen, and S. Sarkar, *GZK Neutrinos after the Fermi-LAT Diffuse Photon Flux Measurement*, Astropart. Phys. **34**, 106 (2010), arXiv:1005.2620.

[53] K.-H. Kampert and M. Unger, *Measurements of the Cosmic Ray Composition with Air Shower Experiments*, Astropart. Phys. **35**, 660 (2012), arXiv:1201.0018.

[54] G. Gelmini, O. E. Kalashev, and D. V. Semikoz, *GZK photons as ultra high energy cosmic rays*, J. Exp. Theor. Phys. **106**, 1061 (2008), arXiv:astro-ph/0506128.

[55] ANITA, P. W. Gorham *et al.*, *Observational Constraints on the Ultra-high Energy Cosmic Neutrino Flux from the Second Flight of the ANITA Experiment*, Phys. Rev. **D82**, 022004 (2010), arXiv:1011.5004, [Erratum: Phys. Rev.D85,049901(2012)].

[56] IceCube, M. G. Aartsen *et al.*, *Probing the origin of cosmic rays with extremely high energy neutrinos using the IceCube Observatory*, Phys. Rev. **D88**, 112008 (2013), arXiv:1310.5477.

[57] J. N. Bahcall and E. Waxman, *High-energy astrophysical neutrinos: The Upper bound is robust*, Phys. Rev. **D64**, 023002 (2001), arXiv:hep-ph/9902383.

[58] Pierre Auger Collaboration, P. Abreu *et al.*, *The Pierre Auger Observatory III: Other Astrophysical Observations*, (2011), arXiv:1107.4805.

[59] Telescope Array, T. Abu-Zayyad *et al.*, *Upper limit on the flux of photons with energies above* 10^{19} *eV using the Telescope Array surface detector*, Phys. Rev. **D88**, 112005 (2013), arXiv:1304.5614.

[60] A. V. Glushkov *et al.*, *Constraints on the flux of primary cosmic-ray photons at energies E > 10*^{18} *eV from Yakutsk muon data*, Phys. Rev. **D82**, 041101 (2010), arXiv:0907.0374.

[61] M. Ave, J. A. Hinton, R. A. Vazquez, A. A. Watson, and E. Zas, *New constraints from Haverah Park data on the photon and iron fluxes of UHE cosmic rays*, Phys. Rev. Lett. **85**, 2244 (2000), arXiv:astro-ph/0007386.

[62] K. Shinozaki *et al.*, *Upper limit on gamma-ray flux above 10^{19} eV estimated by the Akeno Giant Air Shower Array experiment*, Astrophys. J. **571**, L117 (2002).

[63] K.-H. Kampert, A. A. Watson, and A. A. Watson, *Extensive Air Showers and Ultra High-Energy Cosmic Rays: A Historical Review*, Eur. Phys. J. **H37**, 359 (2012), arXiv:1207.4827.

[64] Particle Data Group, C. Patrignani *et al.*, *Review of Particle Physics*, Chin. Phys. **C40**, 100001 (2016).

[65] L. Anchordoqui *et al.*, *High energy physics in the atmosphere: Phenomenology of cosmic ray air showers*, Annals Phys. **314**, 145 (2004), arXiv:hep-ph/0407020.

[66] W. Heitler, *The Quantum Theory of Radiation.* Dover Books on Physics and Chemistry (Dover Publications, 1954).

[67] J. Matthews, *A Heitler model of extensive air showers*, Astropart. Phys. **22**, 387 (2005).

[68] T. Stanev, *High Energy Cosmic Rays* (Springer Berlin Heidelberg, 2010).

[69] J. Alvarez-Muniz, R. Engel, T. Gaisser, J. A. Ortiz, and T. Stanev, *Hybrid simulations of extensive air showers*, Phys. Rev. D **66**, 033011 (2002).

[70] J. Linsley, *Structure of large air showers at depth 834 g/cm^2. III - Applications*, Int. Cosm. Ray Conf. **12**, 89 (1977).

[71] D. Heck, G. Schatz, T. Thouw, J. Knapp, and J. Capdevielle, *CORSIKA: A Monte Carlo code to simulate extensive air showers*, FZKA-6019 (1998).

[72] S. Sciutto, *AIRES: A System for air shower simulations. User's guide and reference manual. Version 2.2.0.*, (1999), arXiv:astro-ph/9911331.

[73] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*, Phys. Rev. **C92**, 034906 (2015), arXiv:1306.0121.

[74] S. Ostapchenko, *Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model*, Phys. Rev. D **83**, 014018 (2011).

[75] E. Ahn, R. Engel, T. K. Gaisser, P. Lipari, and T. Stanev, *Cosmic ray interaction event generator* SIBYLL *2.1.*, Phys. Rev. D **80**, 094003 (2009).

[76] J. Knapp, D. Heck, S. Sciutto, M. Dova, and M. Risse, *Extensive air shower simulations at the highest energies*, Astropart.Phys. **19**, 77 (2003).

[77] A. Porcelli for the Pierre Auger Collaboration, *Measurements of X_{max} above 10^{17} eV with the fluoresce detectors of the Pierre Auger Observatory*, Talk at the 34th International Cosmic Ray Conference, The Hague (2015).

[78] Pierre Auger Collaboration, A. Aab *et al.*, *Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV*, Phys. Rev. **D90**, 122005 (2014), arXiv:1409.4809.

[79] Pierre Auger Collaboration, A. Aab *et al.*, *Azimuthal Asymmetry in the Risetime of the Surface Detector Signals of the Pierre Auger Observatory*, Phys. Rev. **D93**, 072006 (2016), arXiv:1604.00978.

[80] Pierre Auger Collaboration, A. Aab *et al.*, *Muons in air showers at the Pierre Auger Observatory: Measurement of atmospheric production depth*, Phys. Rev. **D90**, 012012 (2014), arXiv:1407.5919, [Erratum: Phys. Rev.D92,no.1,019903(2015)].

[81] M. T. Dova *et al.*, *Time asymmetries in extensive air showers: a novel method to identify UHECR species*, Astropart. Phys. **31**, 312 (2009), arXiv:0903.1755.

[82] Pierre Auger Collaboration, A. Aab *et al.*, *Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events*, Phys. Rev. **D91**, 032003 (2015), arXiv:1408.1421, [Erratum: Phys. Rev.D91,no.5,059901(2015)].

[83] S. Ostapchenko, *QGSJET-II: Towards reliable description of very high energy hadronic interactions*, Nucl. Phys. Proc. Suppl. **151**, 143 (2006), arXiv:hep-ph/0412332.

[84] Pierre Auger Collaboration, A. Aab *et al.*, *Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory*, Phys. Rev. Lett. **117**, 192001 (2016), arXiv:1610.08509.

[85] Pierre Auger Collaboration, A. Aab *et al.*, *The Pierre Auger Cosmic Ray Observatory*, Nucl. Instrum. Meth. **A798**, 172 (2015), arXiv:1502.01323.

[86] Pierre Auger Collaboration, J. Abraham *et al.*, *Properties and performance of the prototype instrument for the Pierre Auger Observatory*, Nucl. Instrum. Meth. **A523**, 50 (2004).

[87] Pierre Auger Collaboration, A. Aab *et al.*, *The Pierre Auger Observatory Upgrade - Preliminary Design Report*, (2016), arXiv:1604.03637.

[88] Pierre Auger Collaboration, X. Bertou *et al.*, *Calibration of the surface array of the Pierre Auger Observatory*, Nucl. Instrum. Meth. **A568**, 839 (2006).

[89] Pierre Auger Collaboration, J. Abraham *et al.*, *The Fluorescence Detector of the Pierre Auger Observatory*, Nucl. Instrum. Meth. **A620**, 227 (2010), arXiv:0907.4282.

[90] M. A. L. de Oliveira, V. de Souza, H. C. Reis, and R. Sato, *Manufacturing the Schmidt corrector lens for the Pierre Auger Observatory*, Nucl. Instrum. Meth. **A522**, 360 (2004).

[91] Pierre Auger Collaboration, J. Baeumi, *Measurement of the Optical Properties of the Auger Fluorescence Telescopes*, (2013), arXiv:1307.5059.

[92] K. H. Becker *et al.*, *Qualification tests of the 11000 photomultipliers for the Pierre Auger Observatory fluorescence detectors*, Nucl. Instrum. Meth. **A576**, 301 (2007).

[93] J. T. Brack *et al.*, *Absolute Calibration of a Large-diameter Light Source*, JINST **8**, P05014 (2013), arXiv:1305.1329.

[94] Pierre Auger Collaboration, P. Bauleo, J. T. Brack, J. L. Harton, R. Knapik, and A. C. Rovero, *Multi-wavelength Calibration Procedure for the Pierre Auger Observatory Fluorescence Detectors*, Astropart. Phys. **31**, 305 (2009), arXiv:0811.1518.

[95] B. Gookin, A. Dorofeev, J. Brack, J. Harton, and A. Rovero, GAP Report No. 091, 2014 (unpublished).

[96] S. Argiro *et al.*, *The Offline Software Framework of the Pierre Auger Observatory*, Nucl. Instrum. Meth. **A580**, 1485 (2007), arXiv:0707.1652.

[97] E. Harold and W. Means, *XML in a Nutshell* (O'Reilly Media, Sebastopol, CA, 2004).

[98] ROOT Data analysis Framework, https://root.cern.ch.

[99] T. Bergmann *et al.*, *One-dimensional Hybrid Approach to Extensive Air Shower Simulation*, Astropart. Phys. **26**, 420 (2007), arXiv:astro-ph/0606564.

[100] H.-J. Drescher and G. R. Farrar, *Air shower simulations in a hybrid approach using cascade equations*, Phys. Rev. **D67**, 116001 (2003), arXiv:astro-ph/0212018.

[101] Apache-Xerces, https://xerces.apache.org/.

[102] XML-SCHEMA, https://www.w3.org/standards/xml/schema.

[103] The MD5 Message-Digest Algorithm, http://tools.ietf.org/html/rfc1321.

[104] Buildbot-The Continuous Integration Framework, https://buildbot.net/.

[105] Pierre Auger Collaboration, J. Abraham *et al.*, *Trigger and aperture of the surface detector array of the Pierre Auger Observatory*, Nucl. Instrum. Meth. **A613**, 29 (2010), arXiv:1111.6764.

[106] Pierre Auger, J. Abraham *et al.*, *Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory*, Phys. Lett. **B685**, 239 (2010), arXiv:1002.1975.

[107] K. Kamata and J. Nishimura, *The Lateral and the Angular Structure Functions of Electron Showers*, Progress of Theoretical Physics Supplement **6**, 93 (1958).

[108] K. Greisen, *The extensive air showers*, Progress in Cosmic Ray Physics **3**, 1 (1956).

[109] D. Newton, J. Knapp, and A. A. Watson, *The Optimum Distance at which to Determine the Size of a Giant Air Shower*, Astropart. Phys. **26**, 414 (2007), arXiv:astro-ph/0608118.

[110] Pierre Auger Collaboration, A. Schulz, *Measurement of the energy spectrum of cosmic rays above 3×10^{17} eV with the Pierre Auger Observatory*, (2013), arXiv:1307.5059.

[111] Pierre Auger Collaboration, R. Pesce, *Energy calibration of data recorded with the surface detectors of the Pierre Auger Observatory: an update*, (2011), arXiv:1107.4809.

[112] J. Hersil, I. Escobar, D. Scott, G. Clark, and S. Olbert, *Observations of Extensive Air Showers near the Maximum of Their Longitudinal Development*, Phys. Rev. Lett. **6**, 22 (1961).

[113] Pierre Auger Collaboration, H. Dembinski, *The Cosmic Ray Spectrum above 4×10^{18} eV as measured with inclined showers recorded at the Pierre Auger Observatory*, (2011), arXiv:1107.4809.

[114] Pierre Auger Collaboration, V. Verzi, *The Energy Scale of the Pierre Auger Observatory*, (2013), arXiv:1307.5059.

[115] F. Arqueros, J. R. Hoerandel, and B. Keilhauer, *Air Fluorescence Relevant for Cosmic-Ray Detection - Summary of the 5th Fluorescence Workshop, El Escorial 2007*, Nucl. Instrum. Meth. **A597**, 1 (2008), arXiv:0807.3760.

[116] AIRFLY, M. Ave *et al.*, *Spectrally resolved pressure dependence measurements of air fluorescence emission with AIRFLY*, Nucl. Instrum. Meth. **A597**, 41 (2008).

[117] AIRFLY, M. Ave *et al.*, *Precise measurement of the absolute fluorescence yield of the 337 nm band in atmospheric gases*, Astropart. Phys. **42**, 90 (2013), arXiv:1210.6734.

[118] M. Giller, G. Wieczorek, A. Kacperczyk, H. Stojek, and W. Tkaczyk, *Energy spectra of electrons in the extensive air showers of ultra-high energy*, J. Phys. **G30**, 97 (2004).

[119] F. Nerling, J. Bluemer, R. Engel, and M. Risse, *Universality of electron distributions in high-energy air showers: Description of Cherenkov light production*, Astropart. Phys. **24**, 421 (2006), arXiv:astro-ph/0506729.

[120] M. D. Roberts, *The role of atmospheric multiple scattering in the transmission of fluorescence light from extensive air showers*, J. Phys. **G31**, 1291 (2005).

[121] J. Pekala, P. Homola, B. Wilczynska, and H. Wilczynski, *Atmospheric multiple scattering of fluorescence and Cherenkov light emitted by extensive air showers*, Nucl. Instrum. Meth. **A605**, 388 (2009), arXiv:0904.3230.

[122] M. Giller and A. Smialkowski, *An analytical approach to the multiply scattered light in the optical images of the extensive air showers of ultra-high energies*, Astropart. Phys. **36**, 166 (2012), arXiv:1201.4052.

[123] M. Unger, B. R. Dawson, R. Engel, F. Schussler, and R. Ulrich, *Reconstruction of Longitudinal Profiles of Ultra-High Energy Cosmic Ray Showers from Fluorescence and Cherenkov Light Measurements*, Nucl. Instrum. Meth. **A588**, 433 (2008), arXiv:0801.4309.

[124] D. Gora *et al.*, *Universal lateral distribution of energy deposit in air showers and its application to shower reconstruction*, Astropart. Phys. **24**, 484 (2006), arXiv:astro-ph/0505371.

[125] M. Giller, B. Dawson, and G. Wieczorek, *Influence of the scattered Cherenkov light on the width of shower images as measured in the EAS fluorescence experiments, in: Proceedings of the 30th International Cosmic Ray Conference, Mexico*, **4**, 401 (2007).

[126] T. Gaisser and A. Hillas, *Reliability of the method of constant intensity cuts for reconstructing the average development of vertical showers, in: Proceedings of the15th International Cosmic Ray Conference: Plovdiv, Bulgaria*, **8**, 353 (1977).

[127] Pierre Auger Collaboration, M. Tueros, *Estimate of the non-calorimetric energy of showers observed with the fluorescence and surface detectors of the Pierre Auger Observatory*, (2013), arXiv:1307.5059.

[128] L. Prado *et al.*, *Simulation of the fluorescence detector of the Pierre Auger Observatory*, Nucl. Instrum. Meth. **A545**, 632 (2005).

[129] Pierre Auger Collaboration, B. R. Dawson, *Hybrid Performance of the Pierre Auger Observatory in: Proceedings of the 30th International Cosmic Ray Conference, Mexico*, **4**, 425 (2007), arXiv:0706.1105.

[130] J. Linsley and L. Scarsi, *Arrival times of air shower particles at large distances from the axis*, Phys. Rev. **128**, 2384 (1962).

[131] A. A. Watson and J. G. Wilson, *Fluctuation studies of large air showers: the composition of primary cosmic ray particles of energy $E_p \sim 10^{18}$ eV*, J. Phys. **A7**, 1199 (1974).

[132] R. Bruijn, Pierre Auger Collaboration Private Oral Communication, October, 2009.

[133] R. Bruijn, Pierre Auger Collaboration Private Oral Communication, March, 2011.

[134] P. Billoir, GAP Report No. 074, 2005 (unpublished).

[135] R. Bruijn, Pierre Auger Collaboration Private Oral Communication, June, 2011.

[136] A. Creusot and D. Vaberic, GAP Report No. 074, 2005 (unpublished).

[137] B. Smith, GAP Report No. 161, 2008 (unpublished).

[138] C. Wileman, GAP Report No. 105, 2008 (unpublished).

[139] C. Jarne *et al.*, GAP Report No. 042, 2014 (unpublished).

[140] H. Cook, GAP Report No. 107, 2012 (unpublished).

[141] E. M. Pugh and G. H. Winslow, *The Analysis of Physical Measurements* (Addison-Wesley, 1966).

[142] L. Molina-Bueno, GAP Report No. 092, 2015 (unpublished).

[143] G. van Aar, GAP Report No. 055, 2016 (unpublished).

[144] K. S. C. Mora, GAP Report No. 037, 2010 (unpublished).

[145] L. Demortier and L. Lyons, Everything you always wanted to know about pulls, cdf/anal/public/5776, 2002.

[146] G. De Mauro, C. Timmermans, and G. van Aar, Pierre Auger Collaboration Private Oral Communication, September, 2016.

[147] W. R. Leo, *Techniques for nuclear and particle physics experiments* (Springer-Verlag, 1987).

[148] P. Billoir, GAP Report No. 038, 2014 (unpublished).

[149] M. Plum, Pierre Auger Collaboration Private Oral Communication, June, 2016.

[150] A. Porcelli, GAP Report No. 058, 2014 (unpublished).

[151] R. Bruijn, J. Knapp, and I. Valino, *Study of statistical thinning with fully-simulated air showers at ultra-high energies, in: Proceedings of the 32nd International Cosmic Ray Conference (ICRC 2011), Beijing, China*, **2**, 39.

[152] A. Bueno, P. Sanchez-Lucas, and A. A. Watson, GAP Report No. 119, 2014 (unpublished).