INTERNATIONAL DOCTORATE ON ASTROPARTICLE PHYSICS

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA
SCUOLA DI DOTTORATO DI RICERCA
CICLO XXV

université PARIS DIDEROT PARIS 7

UNIVERSITÉ PARIS DIDEROT
ÉCOLE DOCTORALE 517:
PARTICULES, NOYAUX ET COSMOS

Measurement of the energy spectrum of cosmic rays above 3×10^{17} eV using the infill array of the Pierre Auger Observatory

THESIS PRESENTED BY
Hugo Marcelo Rivera Bretel

SUPERVISORS
Lino Miramonti — Piera L. Ghia

DIRECTORS OF THE DOCTORAL SCHOOLS
Marco Bersanelli — Philippe Schwemling

RAPPORTEURS
Andrea Chiavassa — Univ. Torino
Isabelle Lhenry-Yvon — IPN CNRS Orsay

DEFENDED ON FEBRUARY 8, 2013 IN FRONT OF THE JURY COMPOSED BY

Pierre Billoir	LPNHE CNRS Paris
Andrea Chiavassa	Univ. Torino
Piera L. Ghia	LPNHE CNRS Paris
Marco Giammarchi	INFN Milano
Isabelle Lhenry-Yvon	IPN CNRS Orsay
Lino Miramonti	Univ. Milano

Abstract

The Pierre Auger Observatory, in Argentina, combines a 3000 km^2 surface array of water Cherenkov detectors with fluorescence telescopes to measure extensive air showers initiated by ultra-high energy cosmic rays. This "hybrid" observatory (in operation since 2004, and completed in 2008) is fully efficient for cosmic rays energies above 10^{18} eV, that is, from just below the "ankle" of the energy spectrum up to the highest energies.

After the completion of the main observatory, the Auger collaboration has started to deploy new instruments to extend the energy range down to about 0.1 EeV. The planned extensions include two infill surface arrays with 750 and 433 m spacing, with muon detection capabilities, and three additional fluorescence telescopes with a more elevated field of view. The 750 m infill array (covering about 24 km^2) and the new telescopes are now operational. Their aim is the measurement of cosmic rays from below the second knee of the spectrum up to the ankle, where data from the extensions overlap those from the main observatory. The study of the evolution of the spectrum through the second knee and the ankle, together with the primary mass composition, are crucial to the understanding of the transition from a galactic cosmic ray origin to an extragalactic one.

This thesis makes use of data from the 750 m infill array: the objective is the measurement of the cosmic ray energy spectrum in the energy region above 3×10^{17} eV, where the array is fully efficient. To get to the energy spectrum, several steps are needed, from the reconstruction of events, through the precise determination of the exposure of the array, up to the determination of the primary energy. The thesis deals with these aspects, before reaching the final result.

The first chapter gives a general introduction to cosmic ray physics and detectors. It also summarizes experimental results above the first knee of the spectrum with particular emphasis on those obtained above 10^{17} eV. The next two chapters describe the Pierre Auger Observatory and the infill array, respectively. In chapter 2, the main Auger results are summarized too, after a schematic description of the different components of the observatory. Chapter 3 sets the stage for the following chapters. It presents a more detailed description of the characteristics of the infill array, in particular the trigger definitions, event selection and reconstruction. In chapter 4 the performance of the reconstruction of the lateral distribution of observed showers is studied in detail. This is particularly important for the energy spectrum, since the signal at a fixed distance from the shower axis is used as the energy estimator of the event. This signal is estimated by means of the measured lateral distribution of the shower. Chapter 5 presents a comparison between the event reconstruction of the infill and main arrays. Using the set of showers detected by both instruments, the derived geometry and energy estimation are compared, showing a good agreement. In chapter 6, the energy threshold of the array, and hence the set of events to be used, is defined. The methods to obtain the exposure of the array are discussed, as well as related systematic uncertainties. Finally, in chapter 7, the technique to derive the primary energy for each detected shower

is presented. The derived energy spectrum is discussed, and the flux is shown to be consistent with that measured by other instruments in the overlapping energy regions.

Résumé

L'observatoire Pierre Auger, situé en Argentine, combine un réseau de surface, étendu sur 3000 km^2 et composé de détecteurs Cherenkov, avec 4 télescopes de fluorescence pour mesurer des gerbes atmosphériques initiées par les rayons cosmiques d'ultra haute énergie. Cet observatoire "hybride" (en fonctionnement depuis 2004, et dont le déploiement a été achevé en 2008) est pleinement efficace pour des énergies de rayons cosmiques primaires supérieures à 10^{18} eV, c'est-à-dire à partir de la cheville du spectre en énergie jusqu'aux plus hautes énergies.

Après l'achèvement de l'observatoire principal, la collaboration Auger a débuté le déploiement de nouveaux instruments afin d'étendre la gamme d'énergie jusqu'à 0.1 EeV. Les extensions prévues comptent deux réseaux de surface resserés, dits "infill", d'espacement de 750 m et 433 m, avec des possibilités de détection de muons, et trois télescopes de fluorescence supplémentaires ayant un champ de vue d'élévation plus grande. Le réseau "infill" de 750 m (couvrant environ 24 km^2) et les nouveaux télescopes sont aujourd'hui en fonctionnement. Leur but est la mesure des rayons cosmiques à partir d'en dessous du deuxième genou du spectre jusqu'à la cheville où les données recouvrent celles de l'observatoire principal. L'étude de l'évolution du spectre du deuxième genou à la cheville, alliée à la composition du rayon cosmique primaire, sont cruciales dans la compréhension de la transition de l'origine galactique à extragalactique des rayons cosmiques.

Cette thèse s'appuie sur les données provenant du réseau "infill" de 750 m : l'objectif est la mesure du spectre en énergie des rayons cosmiques dans la région d'énergie au-dessus de 3×10^{17} eV, où le réseau est pleinement efficace. Pour obtenir le spectre en énergie plusieurs étapes sont nécessaires, à partir de la reconstruction des événements, en passant par la détermination exacte de l'exposition du réseau, jusqu'à la détermination de l'énergie du primaire. La thèse traite de ces aspects, avant d'atteindre le résultat final.

Le premier chapitre donne une introduction générale de la physique des rayons cosmiques et des détecteurs associés. Il résume également les résultats expérimentaux au-dessus du première genou du spectre avec un accent particulier sur ceux obtenus au-dessus de 10^{17} eV. Les deux chapitres suivants décrivent respectivement l'observatoire Pierre Auger et le réseau "infill". Dans le chapitre 2, après une description schématique des différentes composantes de l'observatoire, les principaux résultats obtenus par l'observatoire Auger sont résumés. Le chapitre 3 prépare le terrain pour les chapitres suivants. Il présente une description plus détaillée des caractéristiques du réseau "infill", en particulier la définition du trigger, la sélection et la reconstruction des événements. Dans le chapitre 4, la qualité de la reconstruction de la distribution latérale des gerbes observées est étudiée en détail. Ceci est particulièrement important pour le spectre en énergie, puisque le signal à une distance fixe de l'axe de la gerbe est utilisé comme estimateur de l'énergie de l'événement. Ce signal est estimé au moyen de la distribution latérale de la gerbe mesurée. Le chapitre 5 présente une comparaison entre la reconstruction d'événement du réseau

"infill" et celle du réseau principal. En utilisant les gerbes détectées par les deux instruments, la géométrie et l'estimation de l'énergie obtenues sont comparées, montrant une bonne correspondance. Dans le chapitre 6, le seuil d'énergie du réseau, et donc l'ensemble des événements qui seront utilisés, est défini. Les méthodes pour obtenir l'exposition du réseau et les incertitudes systématiques associées sont discutées. Enfin, dans le chapitre 7, la technique pour obtenir l'énergie du primaire pour chaque gerbe détectée est présentée. Le spectre en énergie obtenu est discuté, et le flux est montré compatible avec celui mesuré par d'autres instruments dans les régions d'énergie communes.

Contents

Chapter 1

High energy cosmic rays

A brief introduction to high energy cosmic rays and Extensive Air Showers (EAS) is given in this chapter. Section 1.1 shortly reminds the history of the discovery of cosmic rays and EAS. Section 1.2 briefly describes the spectrum of cosmic rays and its features. While the answer to the question of the origin of cosmic rays is not yet fully known, the dominant views on their acceleration and propagation are outlined in the same section. As at very high energies cosmic rays can be studied only through EAS, those are discussed in section 1.3 together with the most common techniques employed for EAS detection. Sections 1.4 summarizes the experimental results obtained by different EAS detectors on spectrum, mass and arrival directions of cosmic rays in the energy region around the knee ($\sim 10^{15}$ eV) and in the top end of the spectrum (above 10^{19} eV). Finally section 1.5 focuses on the intermediate energy region, the one that spans between the second knee ($\sim 10^{17}$ eV) and the ankle of the spectrum ($\sim 3 \times 10^{18}$ eV). That is the range where the transition from a galactic to an extragalactic origin of cosmic rays is expected to take place. This is also the energy range that will be studied in this thesis by means of data from the infill array of the Pierre Auger Observatory.

1.1. Short history

A century ago Victor Hess conducted a number of experiments where he launched high accuracy electrometers to measure the ionization rate in the atmosphere. What he discovered was that the ionization rate increased with altitude, contrary to what was expected at the time. For this discovery he was awarded a Nobel prize in physics, in 1936.

It was more than a decade later before Robert Millikan was able to prove that what Hess had observed in the atmosphere was a flux of photons and electrons, that he later termed 'cosmic rays'. He also determined that these particles were not terrestrial in origin, but rather emanated from deep space. Millikan argued that these photons and electrons where produced by the scattering of gamma rays by the Compton effect, but it was Compton himself in the 1930s that correctly surmised that it was not in fact photons and electrons that comprised cosmic rays, but rather

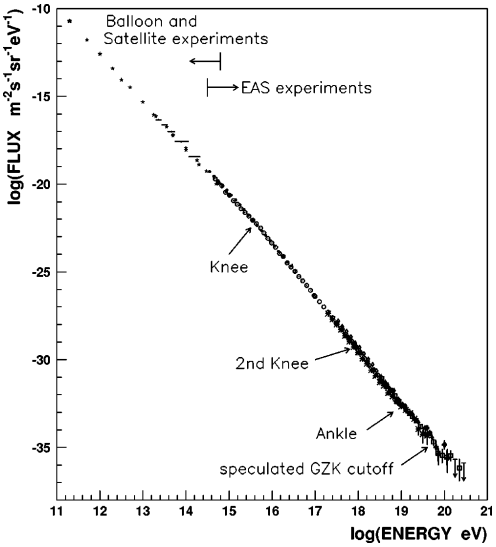


Figure 1.1: The differential cosmic ray flux observed by several experiments. Small structures are noted. Figure taken from [1].

charged particles such as protons. Over the next decade several experiments were carried out to confirm Compton's findings.

In the late 1930s, Pierre Auger and his collaborators found that two Geiger counters placed many meters apart reported the arrival of particles at the same time. They had discovered what is known today as extensive air showers, which presently are the only mean to study cosmic rays of the highest energies. What distinguished the work done by Auger and his group was that they provided an estimate of the primary energy based on the number of particles detected on ground.

The basic experimental methods and techniques used nowadays in cosmic ray research were laid out by the MIT group lead by Bruno Rossi. His group was the first to develop routines capable of estimating the arrival direction of showers as well as the position where the shower signal would be the largest: the shower core. This was accomplished by using the fast timing information coming from the array of detectors they had built on the roof of the physics department at MIT. The group was also the first to report measurements of the differential shower size and took the first steps into determining the energy using models of shower development in the atmosphere.

1.2. Cosmic ray flux, acceleration and propagation

1.2.1. The energy spectrum

The Earth's atmosphere is permanently exposed to a flux of ionized energetic nuclei arriving from all directions, the so called cosmic rays. Their energies range from a few MeV up to at least 10^{20} eV. The differential energy spectrum is depicted in figure 1.1. The steep spectrum reflects the rapid decrease in flux with increasing

energy. The flux decreases from about 1000 particles/m^{2}s at GeV energies, to nearly 1 particle/m^{2}yr at PeV energies and further to less than 1 particle/km^{2}century at 10^{20} eV.

The energy spectrum is well described by an inverse power law

$$\frac{\mathrm{d}N}{\mathrm{d}E} \propto E^{-\gamma} \tag{1.1}$$

over its wide range. It is rather featureless, as can be seen from figure 1.1. Nonetheless, small structures appear as changes in its steepness at different energies. The first of these structures, known as the "knee", is located at $\sim 4 \times 10^{15}$ eV, where the spectral index γ changes from ~ -2.7 to ~ -3.0. A further steepening of the spectrum, the "second knee", occurs at $\sim 4 \times 10^{17}$ eV, where γ decreases to ~ -3.2. The spectrum flattens again at $\sim 3 \times 10^{18}$ eV and γ returns to the value of ~ -2.7. This inflection point is known as the "ankle".

The strong decrease in flux poses a big experimental challenge and knowledge about particles and their origin is more and more limited with increasing energy (and decreasing flux). The chemical composition of cosmic rays at sub-GeV energies is measured with small detectors in outer space, capable of discriminating between individual isotopes. In the TeV regime, individual elements can be resolved with balloon-borne detectors. At energies exceeding 100 TeV, large detection areas are required to collect a suitable number of particles in a reasonable time. At present, such detectors are realized at ground level only and secondary particles generated in the atmosphere (i.e. extensive air showers) are registered. At PeV energies, groups of elements could be resolved, while at the highest energies even a classification into light and heavy particles becomes already an experimental challenge.

1.2.2. Cosmic ray origin and propagation

The dominant view in the scientific community, regarding the origin and propagation of cosmic rays, is that up to 10^{17}–10^{18} eV cosmic rays originate inside our galaxy as products of acceleration mechanisms that take place at astrophysical sites, such as supernova remnants or pulsars. Cosmic rays then travel through the galactic medium eventually reaching the solar system and the Earth's atmosphere. Above 10^{18} eV cosmic rays are believed to originate outside our galaxy.

Acceleration mechanisms

The acceleration of charged particles at astrophysical sites may occur mainly by two different mechanisms. The first of these is a diffusive shock process based on the acceleration model originally proposed by Fermi [2]. Particles are accelerated by the collisions they experience with magnetic clouds in the galactic medium. Statistical equilibrium between cosmic rays and the magnetic clouds implies that the energy gained by the particles could be enormous. The gain in energy is possible because "head-on" collisions with the magnetic fields, which increase the particle's energy, are more likely to occur than "following" collisions, which decrease the energy. In this scenario, the maximum energy for a cosmic ray depends only on the time it

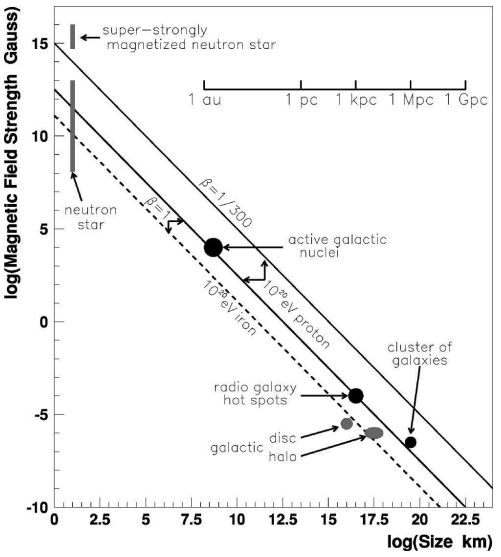


Figure 1.2: The Hillas plot illustrating the acceleration capabilities of some celestial objects. The diagonal lines show the minimum strength a magnetic field must have to accelerate particles to 10^{20} eV. Figure taken from [1].

spends in the acceleration region. It has been shown [3, 4] that if diffuse shock acceleration from supernovae is taken in to account, the acceleration mechanism can be relatively efficient up to energies of the order of 10^{15} eV. A principal advantage of the diffuse shock acceleration model is that it naturally produces a power law spectrum whose spectral index is within the range of experimental measurements. The combination of energy gain per crossing and the escape probability lead to a spectral index γ in the range -2.4– -2.1, depending on the geometry of the shock and on its relativistic properties.

The second acceleration mechanism is that in which direct acceleration takes place by a single shot acceleration process. The mechanism depends on the existence of a strong electromagnetic field that induces a fast acceleration. Pulsars are good candidates as sources of galactic cosmic rays. The inducted electromotive force around these objects is capable of accelerating hydrogen nuclei to $\sim 5 \times 10^{16}$ eV and iron nuclei to $\sim 10^{18}$ eV, according to models [5].

The two mentioned acceleration mechanisms can explain cosmic rays of higher energy too. Indeed, as pointed out by Hillas, the maximum attainable energy is in general given by

$$E_{max} \approx ZeRB\beta \tag{1.2}$$

where Ze is the charge of the particle, R is the size of the accelerating object, B is the strength of the magnetic field and β is the characteristic velocity of the scattering center. If a particle attains an energy E_{max} it escapes the accelerating region and no further acceleration is possible. Using (1.2) Hillas produced the plot in figure 1.2. It presents celestial objects that could act as particle sources as well as acceleration candidates. Only objects above the diagonal line could accelerate protons up to 10^{20} eV: those include, among others, Active Galactic Nuclei (AGN) and cluster of

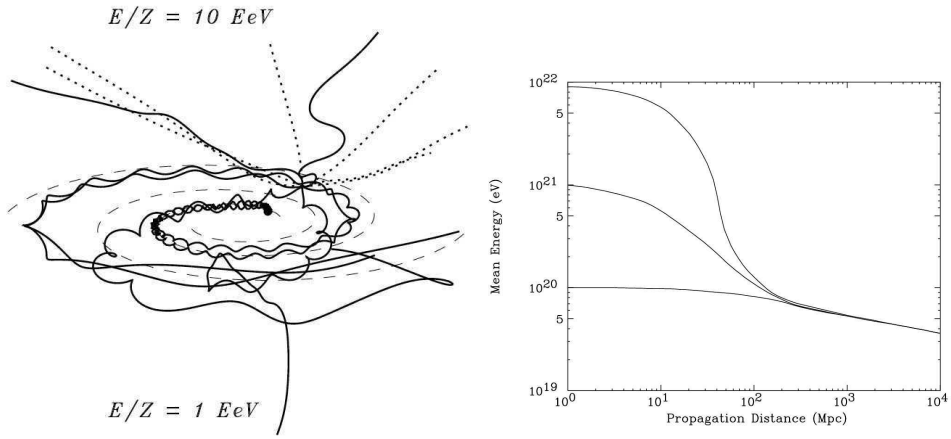


Figure 1.3: Left panel: Examples of trajectories of nuclei in the galactic magnetic fields. Solid lines correspond to $E/Z = 1$ EeV, while dotted lines correspond to $E/Z = 10$ EeV. Right panel: the energy of hydrogen nuclei as a function of propagation distance. Curves represent different initial energies at the source. Figure taken from [9].

galaxies.

Although not acceleration mechanisms per se, top-down models are briefly considered in this section. There is a number of such models that postulate the decay of super-massive particles, or 'X' particles, that lead to the production of cosmic rays of the highest energies. Among these, one is the topological defects model that requires the existence of objects such as cosmic string or magnetic monopoles. The highest energy cosmic rays are originated as products of the decay or annihilation of such defects. Another model proposes super-heavy dark matter particles as a source of cosmic rays of very high energies. For a theoretical overview on these matters see the review by Bhattacharjee and Sigl [6]. Although there is much speculation around these models, some of them make testable predictions on the anisotropy or fluxes of the decay products [7]. The fact that no neutrinos nor gamma rays of ultra high energy have been observed is strong evidence against top-down models [8].

Propagation

Cosmic ray particles are assumed to propagate in a diffusive process through the galaxy, frequently deflected by magnetic fields. Cosmic rays are not confined to the galactic disk and may also propagate in the halo as well. The magnetic field in our galaxy is roughly parallel to the local spiral arms, but with large fluctuations. Galactic cosmic rays are confined within our galaxy, with a mean residence time of the order of 10^7 years.

The cosmic ray spectrum measured on Earth, as well as the distribution of their arrival directions, is significantly affected by processes that take place during the propagation of cosmic rays. From the spectrum point of view, a frequently discussed idea suggests that the knee is indeed a propagation effect. The propagation is accompanied by leakage of particles from the galaxy. With increasing energy it becomes more and more difficult to confine the nuclei to the galaxy. This can be un-

derstood since the Larmor radius of a proton in the galactic magnetic field ($\sim 3\mu$G) becomes, with increasing energy, comparable to and finally exceeds the thickness of the galactic disk. This ultimately leads to a loss of particles with a rigidity dependent cut-off of the flux for individual elements. From the arrival directions point of view, it has to be considered that the Larmor radius of protons with an energy around 1 PeV in the galactic magnetic field is of the order of 0.4 pc. Hence, it is not expected to find any point sources for galactic cosmic rays. It is at around 1 EeV that the Larmor radius of charged particles become comparable to the size of the Galaxy: it is about 300 pc. Above that energy, for containment reasons, cosmic rays are believed to have an extragalactic origin. Also, nuclei with E/Z above 10 EeV are not significantly deflected by the magnetic field. In this case the information on the incoming direction is conserved and the correlation with sources can in principle be established. In figure 1.3 (left), the trajectories of nuclei with $E/Z = 1$ EeV (solid lines) and 10 EeV (dotted lines) in the galactic magnetic field model, are drawn. Dashed lines indicate the spiral arms.

Concerning extragalactic cosmic rays, it must be mentioned that while they propagate in the universe they experience an energy loss due to the interaction with microwave (CMB) and infrared photons (IRB) of the extragalactic background radiation. Above 2×10^{18} eV, CMB photons play an important role in cosmic ray propagation. Pairs of electrons and positrons are produced via a scattering process identical to the pair productions of γ rays in the nuclear field. A dip at 2×10^{18} eV caused by this process would be expected in the cosmic ray spectrum, but it would only be observable if the dominant component at these energies is due to extragalactic protons. At higher energies (above 50–60 EeV), protons undergo photo-pion production in the interaction with CMB photons. Namely, a proton with an energy higher than $\sim 4 \times 10^{19}$ eV produces pions through a resonance

$$\gamma + p \rightarrow \Delta^{+} \rightarrow p + \pi^{0} \quad \text{or} \quad n + \pi^{+}. \tag{1.3}$$

As a consequence, the cosmic ray loses close to $1/6^{\text{th}}$ of its original energy per interaction with the CMB. Given that the interaction length for this process is of the order of 50–100 Mpc, it implies that cosmic rays with energies greater than 4×10^{19} eV must have been produced in a relatively close site. The case for heavier nuclei is similar in the sense of a rapid loss of energy. This energy loss is due to the interactions with the IRB via photo-disintegration. The plot in figure 1.3 (right) shows the energy of a proton as a function of the propagation distance. After 100 Mpc the energy of the particle is virtually independent of its initial energy at the source. If there are no nearby sources capable of accelerating particles above 4×10^{19} eV, a flux suppression in the highest energy region of the cosmic ray spectrum is expected from the interactions of cosmic rays and the CMB. This suppression is usually referred as the GZK cut-off.

1.3. Detection techniques

In the first part of this section, general characteristics of extensive air showers are discussed, as these air showers are the only mean to study cosmic rays with ener-

gies higher than $\sim 10^{15}$ eV, employing ground base detectors. The second part is devoted to a description of some of the instruments utilized to detect air showers.

1.3.1. Extensive air showers

The interactions of cosmic rays with the Earth's atmosphere generate a cascade of secondary particles that may reach ground level. These cascades are called extensive air showers (EAS). A typical EAS can be viewed as the sum of three parts: the hadronic, electromagnetic and muonic components.

Upon entering the atmosphere, the primary cosmic ray interacts with the oxygen and nitrogen molecules and produces a shower primarily composed of pions and kaons; these particles become the hadronic component of the EAS. The cosmic ray loses part of its energy in the first interaction and continues to interact until its energy is depleted. The depth of the first collisions depends on the cross section of hadronic interactions in air, which in turn depends on the composition and energy of the colliding particles. For instance, in the 10^{15} eV energy range, the depth is close to 70 g/cm^2 for hydrogen nuclei and 15 g/cm^2 for iron nuclei. Particles that form part of the hadronic component may continue to interact with the atmosphere or decay. Either way, more hadrons come to existence through this processes, along with muons, electrons, neutrinos and gamma rays. Thence is the hadronic component that gives rise to the other two components of an EAS.

When the hadronic component is generated at the top of the atmosphere, the neutral pions that are part of it decay before interacting, given their very short life time. From these decays pairs of gamma rays, or less frequently electron-positron pairs, are created and in this manner the electromagnetic component is fed. This is the most well understood component of EAS. Because the majority of particles in the hadronic component re-interact before decaying, the electromagnetic component is constantly fed and most of the energy of the primary cosmic ray finds its way into it. When the first gamma rays, or electron-positron pairs, are created, a simplified picture of electromagnetic cascades allows to understand its development. Gamma rays generate electron-positron pairs after traveling a certain distance in the atmosphere. The electron and positron each carry away half the energy of the gamma ray and, after traveling a radiation length, they interact with the nuclei of the atmosphere to produce gamma rays by bremsstrahlung. This process continues until the particles in the electromagnetic component reach a critical energy. This energy is the turning point where energy losses by ionization become more important than bremsstrahlung, and hence, the number of particles reaches its maximum value. For gamma rays, it is instead Compton scattering that becomes the dominating mechanism over pair production. Thus it is easy to see that most of the primary energy is dissipated in the atmosphere through ionization.

The muonic component evolves differently from the electromagnetic one. Muons are produced high in the atmosphere mainly as products of the decays of pions and kaons. Because of their high energies, relatively long life time and small cross sections for interactions, muons are the only particles (disregarding neutrinos) that arrive at ground level and also reach underground detectors. For these reasons, the muonic component is the dominant component deep in the atmosphere and under-

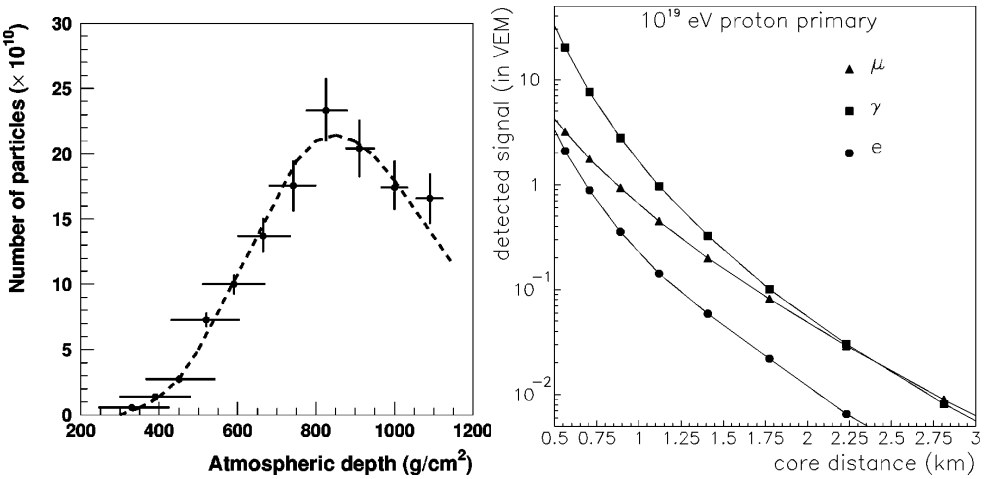


Figure 1.4: Left panel: The longitudinal profile of a shower as measured by the Fly's eye experiment. Figure taken from [10]. Right panel: The lateral distribution of a simulated shower for the Pierre Auger Observatory.

ground. It is this component, rather than the electromagnetic one, that it is used to study properties of hadrons in EAS, since muons travel virtually undisturbed from their production sites. Also, the number of muons in this component is sensitive to the tendency that a pion or kaon has to decay rather than interact, and it is therefore affected by the energy of hadrons and atmosphere properties at high altitudes, where the first interactions take place.

The development of a typical EAS is dominated by the evolution of the electromagnetic component, because most of the energy from the primary particle that initiated the shower is channeled into the production of electrons and gamma rays and it is further dissipated in the atmosphere by such particles. Thus, the electromagnetic component is sensitive to some properties of the primary cosmic ray.

The simplified model for the evolution of the electromagnetic component, illustrated in previous paragraphs, is known as the Heitler model [11]. It describes the development of purely electromagnetic showers, but has been successfully applied to EAS. Since the interaction length for pair production in air is quite close to the radiation length R of electrons, this two distances are considered the same in the model. In this way, it is possible to identify stages in the shower development every radiation length R or so. At each of these stages the number of particles doubles and, therefore, after n radiation lengths the number of particles in the shower is 2^n. The *longitudinal profile* of a shower is defined as the number of particles in it as a function of the slant depth. As previously stated, the number of particles will increase as long as the mean energy in the cascade is above the critical energy E_c, which is $\sim$ 83 MeV in air. When the mean energy drops below the critical energy, the number of particles will stop its increase and thus reach its maximum value. The slant depth at which this occurs is known as X_{max}. From this depth on, the number of particles will decrease. A function describing the longitudinal development of EAS was calculated, using an approach based on Monte Carlo techniques,

by Gaisser and Hillas [12]

$$N(X) = N_{max}\left(\frac{X - X_0}{X_{max} - X_0}\right)^{(X_{max}-X_0)/\lambda} \exp\left(\frac{X_{max} - X}{\lambda}\right) \tag{1.4}$$

where X_0 is the depth on the first interaction and λ represents the decay length of the shower. The measurement of a shower profile is shown, as an example, in figure 1.4 (left).

The *lateral distribution* of particles is a fundamental observable for experiments unable to assess the longitudinal development of EAS. The term 'lateral distribution' refers to the distribution of particles on a plane that is perpendicular to the shower axis. The latter is defined theoretically by the direction of the momentum vector of the incident primary. In fact, most experiments do not measure the longitudinal development of showers but its lateral distribution instead. This is usually done using an array of detectors to sample particle densities at ground level. Using the electromagnetic cascade theory Greisen [13], and later Nishimura and Kamata [14], derived a lateral distribution function (LDF) for electrons, which is fairly independent of the observation depth

$$f(x) \propto x^{s-2}(x+1)^{s-4.5}, \tag{1.5}$$

where $x = r/r_0$, r_0 is the Molière radius and s is the so called shower age parameter. This function is known as the NKG lateral distribution function. Experimentally, it is found that the LDF is very similar to what is expected from purely electromagnetic processes [15] when muons are not taken into account. Figure 1.4 (right) shows an example of the lateral distribution of an EAS for different particles.

1.3.2. EAS detectors

Ground arrays

Air shower ground arrays are sets of particle detectors intended to sample secondary particles that cross them. The area covered by these arrays is in direct proportion to the flux of cosmic rays in the region of the spectra of interest. A few thousands m^2 is enough for the region around the knee (10^{15} eV), while thousands of km^2 are necessary for studies near the spectral cut-off (10^{19} eV). The spacing of the detector is also a function of the energy range of interest. For cosmic rays of energy 10^{18} eV and above spacing is of the order of 1 km.

Scintillator arrays are usually made of flat pieces of plastic scintillators, laid on the ground and connected by cables. They are equally sensitive to all charged particles, thus measure mostly the electromagnetic component of the cascade. The aperture of flat scintillator arrays drops quickly with zenith angle because of the decrease of their effective surface and because of the absorption of the electromagnetic component. Water Cherenkov tanks have also been successfully used in large cosmic ray arrays. Requiring extra pure water with excellent protection against contamination, water Cherenkov detectors are not as easy to deploy as scintillators. However, since the Cherenkov light generated in the water is proportional to the path length of the particle, water tanks are sensitive to both the numerous electrons and photons,

and the shower muons. On the average, depending on the exact detector geometry, a muon will deposit about 10 times more light than a single 20 MeV electron. Because of their height, water tanks also offer a non zero effective surface for horizontal showers. Together with the muon sensitivities this extends the aperture of such arrays to nearly horizontal showers.

Examples for arrays in the knee region are the EAS-TOP experiment [16] and the KASCADE experiment [17]. Detectors that operate at higher energy are the scintillator array of the KASCADE-Grande experiment [18], or the AGASA experiment, which covered an area of nearly 100 km^2 [19]. The group responsible for the Haverah Park array [20] pioneered the use of water Cherenkov detectors.

Reconstruction of some of the primary particle parameters is based on timing for the geometry and on the distribution of particle densities as a function of the lateral distance to the shower axis for the energy. From the position of the different detectors and from the onset of the shower front signal recorded in each of them, it is possible to reconstruct the arrival direction of the original cosmic ray. The particle energy can be estimated using the detector position projected onto the plane transverse to the shower axis and a LDF, adjusted to the measured signals. In the early 1970s, Hillas proposed to use the signal at an optimal distance depending on the energy range and the array spacing [21]. At this distance, the sum of the fluctuations from shower to shower and of the statistical fluctuations from particle counting are minimum. Further details on the determination of the arrival direction and of the lateral distribution using a ground array are given in section 3.4.

Ground arrays do not have a direct access to the position of the shower maximum and this is a strong limitation for identification of the primary. Muon counting can be done with buried detectors or, in favorable conditions, when the electromagnetic to muon signal ratio is not too large, by counting muon spikes in the recorded traces of water Cherenkov tanks.

Atmospheric light detectors

At high energy ($E > 10^{17}$ eV), a technique based on the detection of the fluorescence light has been used by some experiments (e.g. Fly's Eye [22], HiRes [23]) to measure the longitudinal profile of air showers. The charged secondary particles in EAS produce ultraviolet light through nitrogen fluorescence. Nitrogen molecules, excited by a passing shower, emit photons isotropically into several spectral bands between 300 and 420 nm. A much larger fraction of UV light is emitted as Cherenkov photons, but this emission is strongly beamed along the shower axis and usually considered as a background to fluorescence detection.

The number of fluorescence photons emitted per unit length is known as the fluorescence yield. At ground level pressure, the fluorescence yield is 4 photons per electron per meter. On clear moonless nights, fluorescence photons can be collected using square-meter scale telescopes and sensitive photodetectors. Fluorescence photons reach the telescopes in a direct line from their source. Thus the recorded image reflects the development of the electromagnetic cascade. From the fluorescence profile it is in principle straightforward to obtain the position of the shower maximum and a calorimetric estimate of the primary energy. In practice a

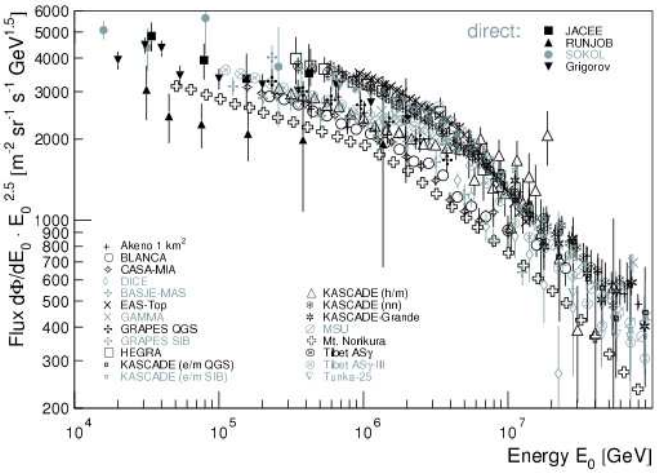


Figure 1.5: The differential cosmic ray flux in the knee region as observed by several experiments. Note that fluxes were multiplied by $E_0^{2.5}$. Direct measurements were performed as well as measurements with ground based instruments. Figure taken from [24].

number of corrections must be made to account for the scattering and the absorption of the fluorescence light in the atmosphere. Also pollution from other sources such as the Cherenkov component which can be emitted directly, or diffused by the atmosphere into the telescope, must be carefully evaluated and accounted for. A constant monitoring of the atmosphere and of its optical quality is necessary together with a precise knowledge of the shower geometry for a careful account for those corrections.

The energy estimate given by a fluorescence detector systematically underestimates the true primary energy. This is caused by neutrinos and energetic muons, which are invisible for a fluorescence detector. These particles carry away a fraction of the total energy, and the effect must be accounted for in order to properly estimate the cosmic ray energy. Employing detailed simulations with varying energy, composition and interaction models, the fraction of missing energy has been estimated. It is about 20% at 10^{18} eV for iron (10% for proton) and about 12% at 10^{20} eV (6% for protons).

1.4. Experimental results and interpretation

1.4.1. The knee energy region

The energy spectrum in the knee region, as measured by different experiments, is depicted in figure 1.5. Spectra obtained using direct measurements is noted in the plot. Although fluxes derived by different groups differ within a factor of two, they all share a common shape, with a knee close to 4 PeV. Results for the all-particle flux as obtained by direct observations above the atmosphere approach energies up to 1 PeV. In the region of overlap, the results from direct and indirect measure-

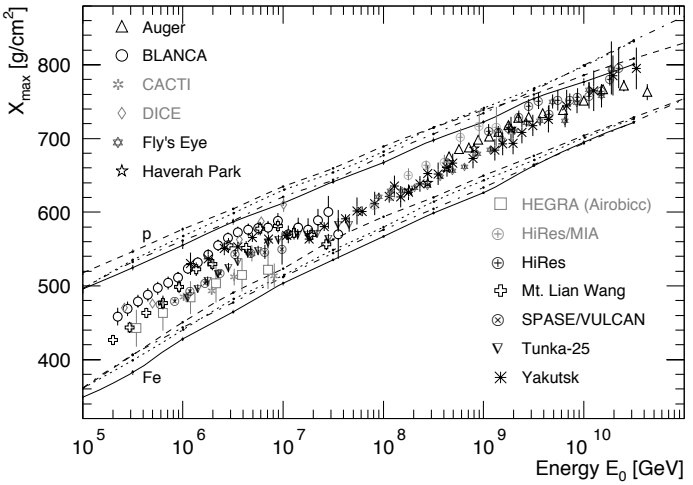


Figure 1.6: Average X_{max} values as obtained by several experiments. Lines represent results derived from Monte Carlo simulations for proton and iron induced showers. Figure extracted from [24].

ments are in reasonable agreement and the flux as obtained by direct measurements is smoothly continued to higher energies with the results of air shower measurements. Despite the experimental uncertainties and systematic differences between different experiments and different interpretations of air shower data using various air shower models, a clear picture of the spectra for elemental groups is evolving. It is evident that the knee in the all-particle spectrum is caused by a depression of the flux of light elements. The measurements follow power laws with a cut-off at high energies.

Regarding the mass composition of cosmic rays, at energies below 10^{14} eV, the abundance of individual elements has been measured with detectors above the atmosphere. At higher energies this is presently not possible due to the low flux values and the large fluctuations in the development of extensive air showers. Thus, in the past, mostly the mean mass has been investigated. A quantity used to characterize the cosmic ray composition is the mean logarithmic mass. This quantity is derived from the ratio of the number of electrons and muons registered at ground level or, alternatively, from the observed depth of the shower maximum, i.e., X_{max}. Recent measurements of average X_{max} values are shown in figure 1.6.

Below 4×10^6 GeV, the values obtained by different experiments exhibit a common trend, they seem to increase faster as function of energy than the simulations, which implies that the average composition would become lighter as function of energy. Above the knee the measured values flatten, indicating an increase of the average mass in this energy range. Above 4×10^7 GeV, measured data exhibit a nearly constant slope for X_{max} as function of energy. The slope is slightly steeper than the predicted slope for iron nuclei for all models shown.

Finally, anisotropies in the arrival directions of cosmic rays are also considered. Large-scale anisotropies are connected to the propagation process of cosmic rays in the galaxy, while small-scale anisotropies would hint towards cosmic ray sources.

The Super-Kamiokande experiment and the Tibet experiment reported anisotropies in the same region of the sky [25, 26]. For energies below 12 TeV the anisotropies show little dependence on energy, whereas above this energy anisotropies fade away, consistent with measurements of the KASCADE experiment in the energy range from 0.7 to 6 PeV [27]. A Compton Getting effect caused by the orbital motion of the solar system around the galactic center would cause an excess with an amplitude of 0.35%. However, the measurements at 300 TeV yield an anisotropy amplitude of $0.03\% \pm 0.03\%$, consistent with an isotropic cosmic ray intensity. Hence, a galactic Compton Getting effect can be excluded with a confidence level of about 5σ. This implies that galactic cosmic rays co-rotate with the local galactic magnetic field environment.

Several experiments have applied the Rayleigh formalism to the right ascension distribution of extensive air showers. Some experiments find anisotropies, however the phases do not agree between the different results. Hence, it seems to be more likely that all amplitudes derived should be considered as upper limits.

The measurements indicate that the knee in the all-particle energy spectrum is caused by a break in the spectra for the light elements, yielding an increase of the mean mass of cosmic rays in this energy region. One of the most popular explanations for the origin of the knee is that the spectra at the source exhibit a break. The bulk of cosmic rays is assumed to be accelerated in strong shock fronts of supernova remnants (SNR) [28]. The finite lifetime of a shock front limits the maximum energy attainable for particles with charge Z to $E_{max} \sim Z(0.1 - 5)$ PeV. A special case of SNR acceleration is the single source model [29], which predicts in the knee region pronounced structures in the all-particle energy spectrum, caused by a single SNR. Such structures can not be seen in the compilation of figure 1.5.

No point sources of charged cosmic rays were found in the knee region. Leaky-box models [30], with their extremely steep decrease of the path length as function of energy yield relatively large anisotropies even at modest energies below 10^{15} eV, seem to be ruled out by the measurements. The measured values are almost an order of magnitude smaller. On the other hand, a diffusion model [31] predicts relatively small values at low energies. The predicted Rayleigh amplitudes are compatible with the measured values. This may indicate that diffusion models are a realistic description of cosmic ray propagation in the galaxy at PeV energies.

In conclusion, it may be stated that a standard picture of the origin of galactic cosmic rays seems to emerge from the data. The measurements seem to be compatible with the assumption that cosmic rays are accelerated at strong shock fronts of supernova remnants. The particles propagate in a diffusive process through the galaxy. As origin for the knee a combination of the maximum energy attained in the acceleration process and leakage from the galaxy seems to be favored.

1.4.2. The top end of the spectrum

Since the 1960s, when the first cosmic ray above 10^{20} eV was detected, many EAS detectors have been operated to study ultra-high energy cosmic rays (UHECR), including the pioneer Volcano Ranch (New Mexico), Haverah Park (England), Yakutsk (Russia), SUGAR (Australia), Akeno/AGASA (Japan) and Flys' Eye (USA). By choice,

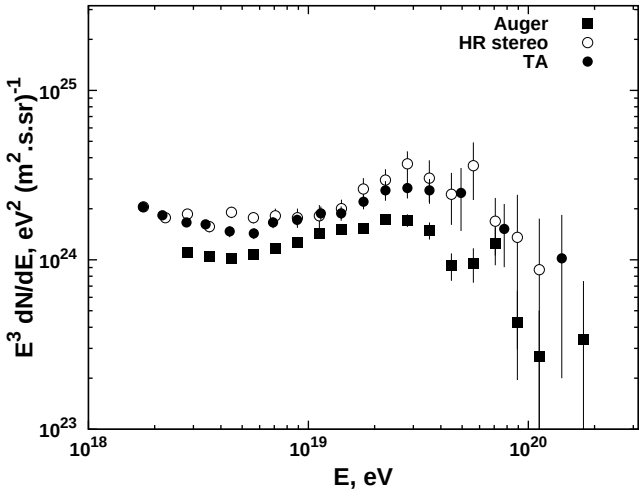


Figure 1.7: The cosmic ray spectrum for the highest energy region, as measured by HiRes, Auger and TA. Fluxes have been multiplied by E^3. Figure taken from [32].

the results presented here are those obtained by the last generation of UHECR detectors, namely HiRes, made of two fluorescence telescopes, and Auger and Telescope Array, both hybrid instruments.

The energy spectrum derived by the HiRes [33], Auger [34] and Telescope Array (TA) [35] experiments is shown in figure 1.7. Although systematic differences are apparent, all three measurements show the spectral flattening at the ankle, as well as a flux suppression at the highest energies. The latter is consistent with the GZK effect. The HiRes found two breaks in the cosmic ray spectrum, one at energy $10^{18.65}$ eV and another at $10^{19.75}$ eV: a cut-off. The spectral index between the two breaks is 2.81 ± 0.03 and after the last break is 5.1 ± 0.7. The spectrum is consistent with various models and in particular with a pure proton composition one [36]. The statistical significance of the cut-off is more than 5σ. The Auger spectrum has a slightly different shape. From 4×10^{18} eV to 4×10^{19} eV the slope of the spectrum is 2.68 ± 0.01 and above it is 4.2 ± 0.1. The Auger spectrum can be explained by several different models some of which include mixed chemical composition at acceleration in the sources. The fit to the spectrum measured by the TA experiment gives a slope between the breaks of 2.68 ± 0.04 and after the last break 4.2 ± 0.7. Since the TA measurement is based on smaller statistics its results should be considered preliminary. All three spectra are consistent with each other within systematic uncertainties, due to the energy scale.

Concerning primary mass composition, the HiRes Collaboration published an analysis based on X_{max} (the depth of the shower maximum) measurements [37]. The data sample included 553 events of energy above 10^{18} eV detected by both fluorescent detectors contemporarily during 20 months of data acquisition. The elongation rate between 10^{17} eV and $10^{18.5}$ eV was measured to be 93 ± 8.5 g/cm^2 with a systematic uncertainty of 10.5 g/cm^2. When compared to hadronic models predictions from primary protons and irons, this result suggested a quick transition from heavy

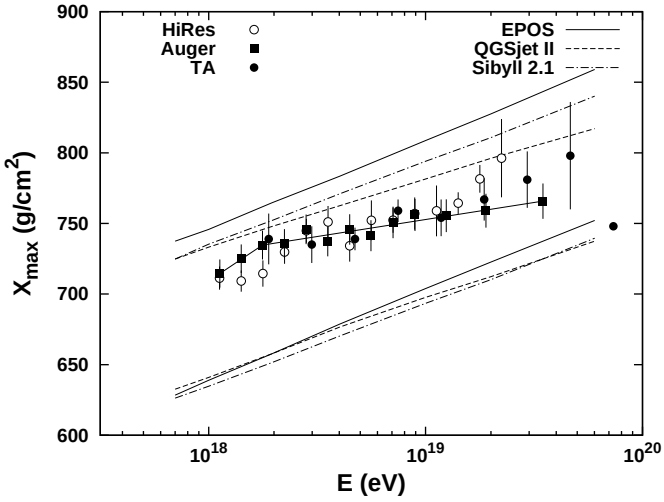


Figure 1.8: X_{max} measurements performed by the HiRes and Auger collaborations as a function of energy. Results are compared with the predictions of three different interaction models for protons and iron. Figure taken from [32].

to light cosmic ray composition. The HiRes points, which are derived from fitting the shower profile with the Gaisser-Hillas formula, are shown with empty circles in figure 1.8. The lines in the same figure show the expectations from three different hadronic interaction models: EPOS 1.99, QGSjet II, and SIBYLL 2.1. The results of the study presented by the Auger collaboration on X_{max} measurements have different implications with respect to those from those of HiRes [38]. This analysis is done using hybrid events, i.e. events detected by one or more fluorescent telescopes plus at least one surface detector. The fitting of the shower profiles is done using the Gaisser-Hillas function. The measured X_{max} values are shown in figure 1.8 with full squares. The elongation rate of the three points below $10^{18.25}$ eV is 106^{+35}_{-21} g/cm^2 and that above this point is 24 ± 3 g/cm^2. Results for the elongation rate seem to indicate that the cosmic ray composition becomes lighter up to $10^{18.25}$ eV and then consistently heavier up to the highest energy measured, provided that no drastic changes in the hadronic interactions occur. The heavy cosmic ray composition derived from the Auger data suggest that the strong decline of the cosmic ray flux may be caused by exceeding the maximum acceleration energy at the cosmic ray sources. In such a case only iron nuclei could be accelerated to energies exceeding 10^{20} eV. Similarly to HiRes, the interpretation of the X_{max} measurements by the TA experiment [39], shown as filled circles in the same figure, is that the cosmic ray composition composition is light, consisting mostly of protons and very light nuclei. It has to be noted that the procedure for the analysis is quite different between Auger and HiRes/TA.

The arrival directions of the highest energy events recorded by HiRes, Auger and TA are shown in figure 1.9. The Auger collaboration did a scanning analysis varying the angular distance, event energy and the source distances using the VCV catalog [40]. Using early data, the values of the energy threshold, maximum angular

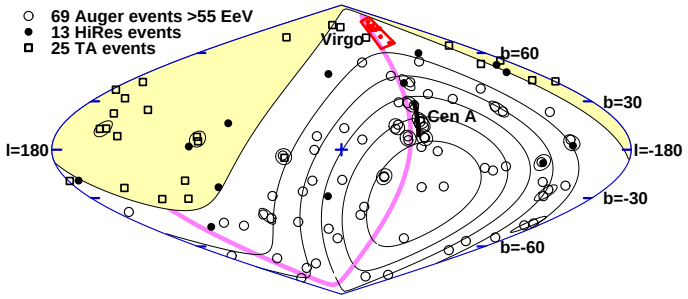


Figure 1.9: Arrival directions of the 69 Auger events, 13 HiRes events and the TA 25 events in galactic coordinates. The colored area shows the part of the galaxy that Auger does not see. The six areas defined within the Auger field of view have equal exposures. The events that form a pair at angular distance less than 5° are circled. Figure taken from [32].

separation, and maximum redshift were chosen as those that minimized the probability that the correlation with AGNs in the VCV catalog could occur by chance if the flux was isotropic. A test was then performed using data collected subsequent to the parameter specification by the exploratory scan. It measured the fraction of arrival directions that are less than 3.1° from the position of an AGN within 75 Mpc in the VCV catalog. The fraction expected under the isotropic hypothesis is 21%.The most recent update of the correlating fraction was obtained from an exposure of 20370 km^2 sr yr [41] which included 69 events with energy above 55 EeV. It is 38%, i.e., it has decreased from 69% reported in 2007. The analysis was repeated by the HiRes experiment [42]. There were only 2 out of 13 events with similar energy that correlated with the same AGNs. It has to be noted that the HiRes field of view is not the same as that of Auger and the VCV catalog has different coverage of the corresponding fields of view. Moreover, the two energy scales are different. The Telescope Array collaboration has also searched for this correlation. The TA exposure is peaked in the northern hemisphere so the AGNs visible to TA are not the same as the ones visible to Auger, though there is some overlap. When the distribution of nearby AGNs is taken into account, and assuming equal AGN luminosities, the correlating fraction would be 44%, rather consistent with the updated correlation fraction from Auger. In the full TA SD data set, there are 11 correlating events out of 25, while the expected number of random coincidences is 5.9. The probability of this correlation to occur by chance with isotropic distribution of arrival directions is 0.02. More data are necessary to show whether this correlation is statistically significant or not.

1.5. The transition from galactic to extragalactic cosmic rays

In the energy range between 10^{17} and 10^{19} eV the transition from galactic to extragalactic cosmic rays is expected to occur. Three key observables would provide hints

on the location of the transition region. The first of these observable is the energy spectrum: a change of slope in the energy spectrum should be observed due to the intersection of a steep galactic component with a flat extragalactic one. As will be discussed below, depending on the models such change might occur around a few 10^{17} eV or at $\sim 5 \times 10^{18}$ eV. While experimental results on the former feature are still scarce and partly controversial (although the KASCADE-Grande experiment has recently observed a second knee at about 10^{17} eV [43]), a flattening of the spectrum at about 4×10^{18} eV has been consistently observed by several experiments, as seen in the previous section. The second expected observable is a change of composition caused by different acceleration mechanisms and propagation effects. The last observable is a changing large scale anisotropy in the distribution of the arrival directions. Therefore, experimental measurements in this energy region of the cosmic ray flux, composition and arrival direction are important to discriminate between models. The infill array of the Pierre Auger Observatory, whose data are the subject of this thesis, is meant to extend Auger measurements down to energies that cover the range of the expected transition. In this section, as a conclusion of this chapter, the three most important models for the transition from galactic to extragalactic cosmic rays are briefly discussed: the ankle, dip and mixed composition models.

The ankle model

This model is based on the interpretation of the ankle as the feature in the spectrum where the transition occurs [44]. The beginning of the ankle, around 3–4 EeV, corresponds to the energy where fluxes of galactic and extragalactic cosmic rays are the same. The transition is given by the intersection of a flat extragalactic spectrum and a very steep galactic one. The extragalactic component is assumed to be pure proton, while the galactic one should be naturally represented by iron nuclei at energies above the iron knee. These models predict a transition from an iron-dominated composition to a proton-dominated one with a flat generation spectrum valid for non-relativistic shock acceleration. From the point of view of large-scale anisotropies, a large anisotropy would be expected below the ankle, due to the anisotropic distribution of galactic sources and to the motion of cosmic rays leaking from the galaxy. On the other side, above the ankle no anisotropy should be expected as at that energies cosmic rays would be isotropized by the galactic magnetic field.

The transition at the ankle is illustrated in figure 1.10 (top left panel). The solid curve presents the calculated extragalactic flux of protons and the dashed line gives the galactic cosmic ray spectrum. The latter is obtained by subtracting the extragalactic component from the total observed flux. A problem of the ankle model is the contradiction with the measured average depth of EAS maximum, X_{max} (E), in the energy range 1–5 EeV. While all data, including both HiRes and Auger, show proton or light nuclei composition here, the ankle model needs a heavy galactic component, predicting too small X_{max} in contradiction with observations.

The dip model

The dip model is based on the assumption that cosmic rays above $\sim$ 1 EeV are already mostly extragalactic protons, i.e., the transition occurs at a lower energy. This assumption is confirmed by the observation of the pair-production dip in the energy range 1–40 EeV and the beginning of the GZK cut-off in the energy range 40–100 EeV. Both features are signatures of a proton dominated spectrum. The shape of the dip allows an admixture of light nuclei, though not more than 15%. The transition from galactic to extragalactic cosmic rays occurs as the intersection at 0.5 EeV of the steep galactic component with the flat extragalactic proton component (see figure 1.10, top right panel). The flatness of the extragalactic spectrum is provided by the distribution of sources over the maximum acceleration or, in the case of diffusive propagation, by the 'magnetic flattening' [45]. The transition is completed at energy $\sim$ 1 EeV. This model consequently predicts an almost pure iron composition at the second knee, and a proton composition above. No anisotropy is expected above 1 EeV, where cosmic rays should be all extragalactic, while below a different anisotropy should be observed for different mass primaries.

The mixed composition model

The main concept of the mixed composition model is based on the argument that any acceleration mechanism operating in gas involves different nuclei in acceleration process and thus the primary flux must have a mixed composition [46]. Transition from galactic to extragalactic component in the mixed models depends on the choice of parameters. In most models transition occurs at the ankle. However, it is required that for strong source evolution and flat generation spectra the intersection of galactic and extragalactic components occurs between 0.5 and 1 EeV as in the dip model [47]. The transition in this model is shown in figure 1.10 bottom panel. The signature in terms of mass would be a change from iron primaries to a mixed composition of lighter nuclei. The composition of these extragalactic cosmic rays should be almost the same as for galactic ones, with protons and helium being the dominant components.

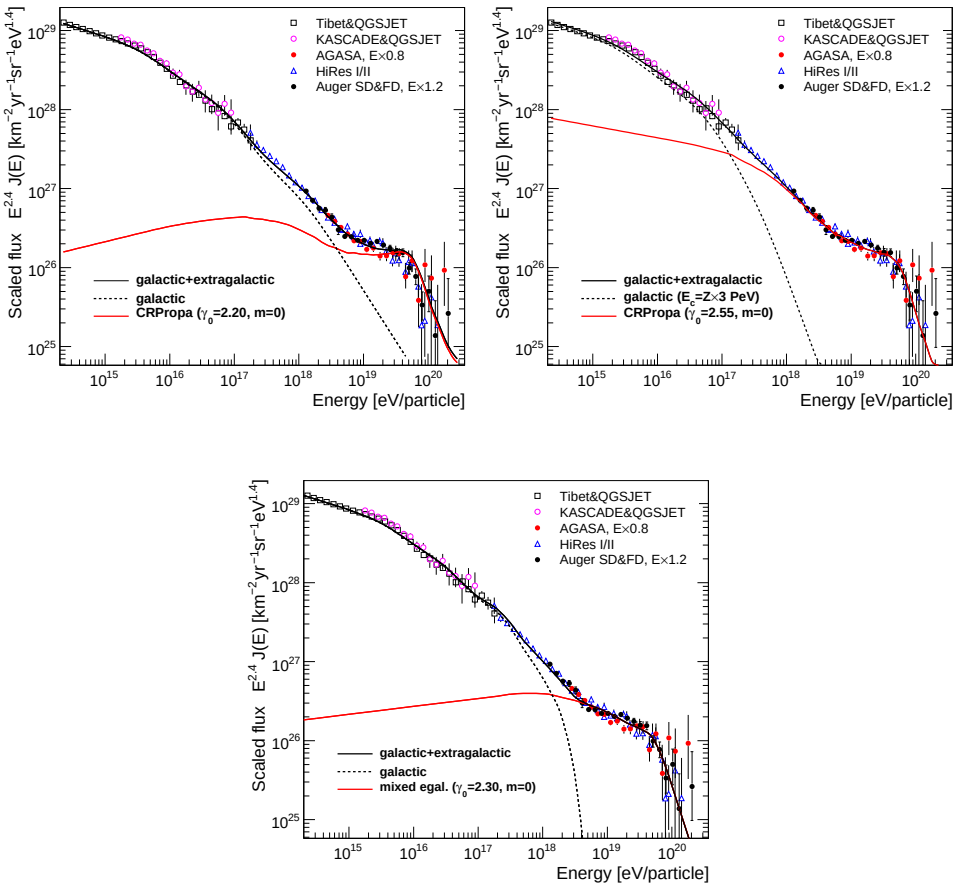


Figure 1.10: Models describing the transition between galactic and extragalactic cosmic ray component, together with data from several experiments. Ankle model (top left), dip model (top right) and mixed composition model (bottom). Extracted from [48].

Chapter 2

The Pierre Auger Observatory

The Pierre Auger Observatory is a cosmic ray observatory conceived to measure the properties of EAS generated by cosmic rays with energies higher than 10^{18} eV. Located at 1400 m.a.s.l. in a desert located near Malargüe, Province of Mendoza, Argentina, it started collecting data in 2004, and it was completed in 2008. The Auger Observatory performs hybrid measurements of air showers recorded by an array of more than 1600 water Cherenkov surface stations covering an area of 3000 km^2, together with 24 air fluorescence telescopes that observe the development of air showers in the atmosphere above the array during dark nights.

An infill array with half the grid size has been completed: its data extend down to 3×10^{17} eV, thereby covering the ankle of the primary energy spectrum with full detection efficiency. Moreover, the three high-elevation telescopes (HEAT) started operation and (together with the infill array in the FOV of the telescopes) will allow to extend the hybrid measurements further down to 10^{17} eV with unprecedented precision. This will enable the study of the transition from galactic to extragalactic cosmic rays. Construction of the buried muon detectors (AMIGA) in the infill area is in progress.

This chapter will focus on the description of the observatory as a whole, while the infill array, whose data are mostly used in the present work, will be discussed in more detail in the next chapter.

2.1. The surface detector

The surface detector (SD) includes over 1600 Water Cherenkov Detectors (WCD). These detectors are placed on an equilateral triangular grid spaced by 1.5 km.

Each WCD is composed of a cylindrical tank with a diameter of 3.6 m and a height of 1.55 m. The tanks are filled with 12 tons of pure water with a depth of 1.2 m. The walls of the tank are internally covered by a liner that diffuses and reflects the ultraviolet light emitted in water. The high water purity, with resistivity of 15 MΩ cm, is needed in order to reduce the absorption of propagating Cherenkov photons and to guarantee water stability for the twenty year life time intended for the detectors.

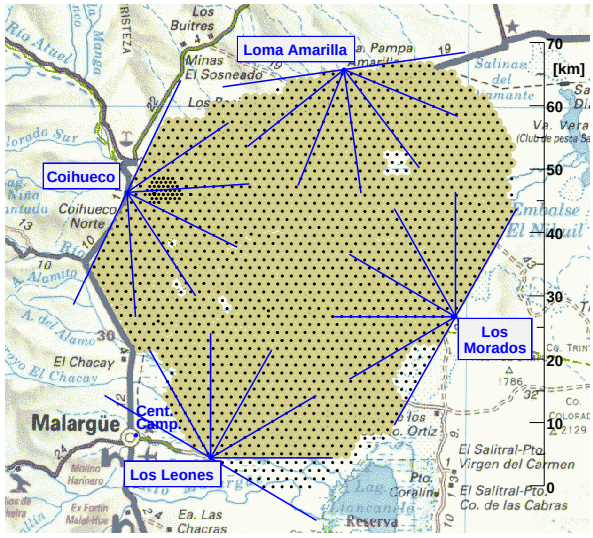


Figure 2.1: The Pierre Auger Observatory, located near Malargüe, Province of Mendoza, Argentina. Each black dot represents a WCD of the SD, while blue lines represent the field of view of the fluorescence telescopes positioned at four different locations.

The water volume is overlooked by three 9-inch Photonis PMTs. From each of these PMTs two signals are taken. One directly from the anode and the other from the last dynode, amplified by a factor of 32 and inverted. The signals from the three PMTs are obtained using Flash Analogue to Digital Converters (FADC) that process at a 40 MHz sampling rate. Once the digitized signals are acquired, they are sent to a programmable logic device that contains the implementation of the local station triggers. The signals are labeled with the time obtained from a commercial GPS unit with an accuracy close to 8 ns. All information that has to be exchanged with the Central Data Acquisition System (CDAS) is shared using commercial wireless technologies.

The power required for each station to function is provided by a solar panel located on top of the tank. This panel provides the 10 W power supply that is necessary for the electronics. A new type of lead acid battery is used to store energy collected by the solar panels. The combination of the wireless and power systems make each station fully autonomous. A photograph of a station of the surface detector of the observatory is shown in figure 2.2 (left). A detailed description of the stations of the surface array detector can be found in [49].

2.1.1. Calibration of the surface detectors

The SD obtains a measurement of the Cherenkov light produced by shower particles passing through the detector at ground, and reconstructs the air shower by fitting the observed signal as a function of lateral distance from the shower core. The recorded signal is converted into units of the signal produced by a vertical and central through-going (VCT) muon, termed a vertical-equivalent muon (VEM or Q_{VEM} when needed for clarity). Atmospheric muons passing through the detector

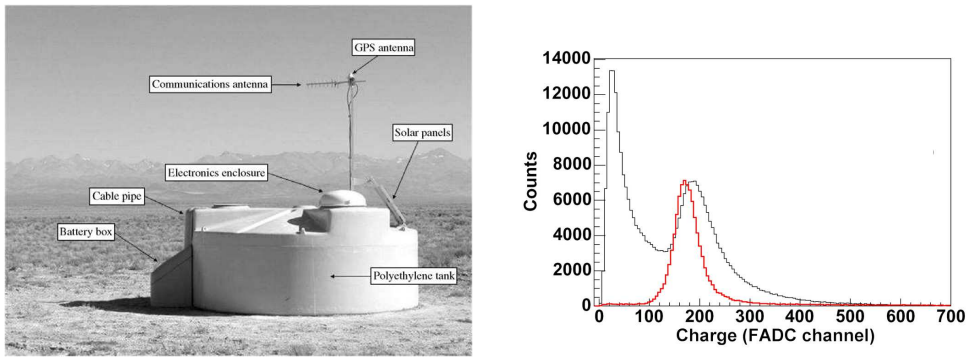


Figure 2.2: On the left panel, a water Cherenkov station from the surface detector. Figure extracted from [49]. On the right panel, the measured charge histograms for vertical equivalent muons (red) and all atmospheric muons (black). Taken from [50].

at a rate of approximately 2500 Hz give an excellent method for measuring 1 VEM precisely, but the SD in its normal configuration has no way to select only vertical muons. However, the distribution of the light of atmospheric muons produces a peak in the charge distribution, Q_{VEM}^{peak}, proportional to that produced by a vertical through-going muon. Q_{VEM}^{peak} is at approximately 1.09 VEM for the sum of the three PMTs measured with a muon telescope providing the trigger in a reference tank.

The plot in figure 2.2 (right) shows the charge histograms for vertical muons and atmospheric muons. The shift observed is caused by the convolution of photoelectron statistics on an asymmetric peak in the track length distribution and local light collection effects. Q_{VEM}^{peak} is measured with a high-statistics (150000 entries) charge histogram every minute: histograms are returned to the CDAS with each event and used to calibrate the data.

Atmospheric muons also produce a peak in a pulse height histogram, I_{VEM}^{peak}: the gain (i.e., the high voltage) of each PMT is set (when a station is first turned on) so that I_{VEM}^{peak} is at 50 ch. I_{VEM}^{peak} is also the reference value for determining the trigger threshold levels for each station (namely, 3.2 I_{VEM}^{peak} for the simple threshold trigger, and 0.2 I_{VEM}^{peak} for the time-over-threshold trigger, see next section). The drifts of the value of I_{VEM}^{peak} in electronics units for each detector are compensated to ensure that the array triggers uniformly. This compensation is done via adjusting the trigger levels based on a continual on-line calibration. Note that the PMT high voltage is not changed during normal operation.

More details on the calibration procedure for the stations in the surface array can be found in [51].

2.1.2. The surface detector triggers

The signal background in the surface array is primarily produced by surviving muons from low energy showers that died out in the atmosphere. In order to select high energy showers over the background, a set of conditions has been arranged in a trigger hierarchy comprising five levels. The first two levels in the hierarchy are

applied locally on each WCD. The third level is formed at the CDAS, based on timing and position of stations satisfying the prior two levels; if the required conditions in this level are met, data acquisition and storage is started. The last two levels in the trigger chain are software triggers applied offline on recorded data.

The same trigger hierarchy is used for the infill array. While the first two levels are unchanged for the local stations, the last three have been modified according to the characteristics of the infill array. In this section, the first two levels are described in detail while the last three levels are briefly reviewed. The next chapter contains a thorough explanation of the last three trigger levels in the hierarchy.

The first two trigger hierarchies are applied by the software running on each station from the array. Namely the *T1* and *T2* triggers. These, in turn, include two types of triggers, designed to measure in a complementary manner the muonic and electromagnetic components of showers. The *TH-T1* and *ToT-T1* for the T1 trigger and *TH-T2* and *ToT-T2* for the T2 trigger.

The TH-T1 trigger is a simple threshold trigger requiring at least one bin from the FADC traces, from all three PMTs (i.e. a 3-fold coincidence), to be above a threshold value of 1.75 I_{VEM}^{peak}. This trigger produces an event rate of $\sim$ 100 Hz in each station. In order to be promoted to a TH-T2 trigger a further threshold value of 3.20 I_{VEM}^{peak}, again with a 3-fold coincidence of the PMTs, is imposed. The TH-T2 triggers reduce the event rate from $\sim$ 100 Hz to $\sim$ 20 Hz. The TH triggers were designed to select large signals not spread in time: typical signatures of muons traversing the tank of the station. On the other hand, the ToT triggers are designed to select the traces produced by the electromagnetic component. They take advantage of the fact that the electromagnetic component of EAS produce signals well spread in time [52, 53]. The ToT-T1 trigger requires at least 13 bins from a sliding window of 120 bins to be above a threshold value of 0.20 I_{VEM}^{peak}, in at least two of the three PMTs. The bin values correspond to 325 ns and 3 μs, respectively. This trigger is quite effective in eliminating the signals produced by lonely muons that survive small showers absorbed in the atmosphere. Once a ToT-T1 is formed on a station it is automatically promoted to a ToT-T2, without any additional requirements.

The next level in the hierarchy is referred to as *T3*. This trigger is formed at the CDAS using information coming from stations satisfying the T2 trigger level. It is designed to select real showers triggering the stations, checking for relatively compact configurations. If satisfied it initiates data acquisition from the array, and FADC traces from all triggered stations are then sent to the CDAS for its storage. For a detailed description see section 3.3.1. The diagram in figure 2.3 (left) gives a graphic representation of the trigger hierarchy discussed so far.

The last two levels in the hierarchy, applied offline to recorded data, are intended to select physics events from the stored data set, and to guarantee a reasonable sampling of detected showers. The *T4* trigger (section 3.3.2) defines spatial and timing criteria to ensure the subsequent reconstruction of the event. Finally, the last component in the chain is a fiducial quality trigger *T5* (section 3.3.3). It requires that the station with the largest signal, usually referred as the hottest station, be surrounded by six working stations. It rejects showers falling too close to the array border, thus ensuring a good estimation of the shower observables in the reconstruction.

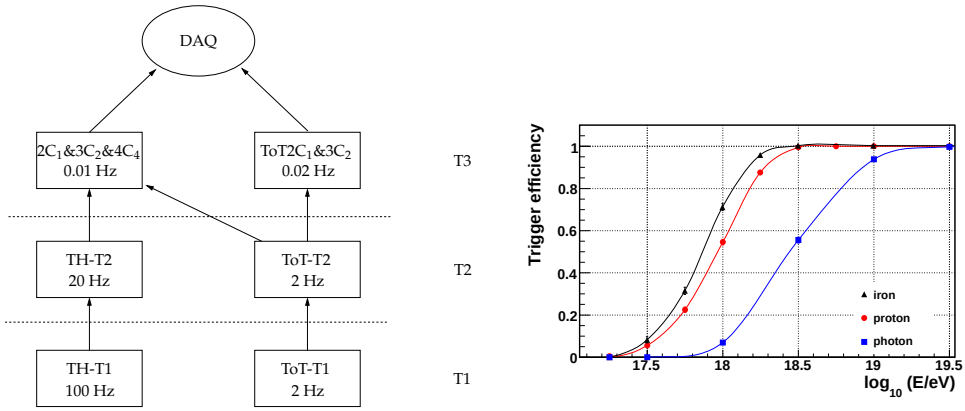


Figure 2.3: The panel on the left depicts a diagram illustrating the first three trigger levels in the hierarchy. The panel on the right shows the SD trigger efficiency as a function of Monte Carlo energy E for proton (circles), iron (triangles) and photon primaries (squares) and zenith angle integrated up to 60°.

The response of the surface detector array was simulated using Geant4 [54] within the framework provided by the Offline software [55]. The resulting trigger probability as a function of the Monte Carlo energy for proton, iron and photon primaries is shown in figure 2.3 (right) for showers with $0^\circ < \theta < 60^\circ$. Due to their larger muon content, at low energies iron primaries are slightly more efficient at triggering the array than protons. However, the trigger becomes fully efficient at 3×10^{18} eV, both for proton and iron primaries, in different intervals of zenith angles. It is important to notice that the trigger efficiency for photons is much lower. This is because photons tend to produce deeper showers that are poor in muons. The trigger efficiency versus energy has been checked with data too [56].

2.1.3. Event reconstruction

The main goal of the event reconstruction in the surface detector is to estimate the arrival direction and energy of the primary cosmic rays. The *geometrical* reconstruction relies on timing information from stations to apply a fit to estimate the arrival direction. The estimation of the *density of particles* at a fixed distance from the shower axis is based on a fit to the lateral distribution of signals in the event, using a given lateral distribution function (LDF). Knowledge of the particle density at a fixed distance is essential for the energy estimation of the primary particle that initiated the shower. This section gives an overview of the reconstruction chain, which will be discussed in greater detail in the next chapter, as the same strategy is used to reconstruct events recorded in the infill array.

Station selection

Besides disentangling accidental events, by imposing the T4 trigger requirements, there is also the need to identify, and reject, accidental WCD in events before reconstruction, i.e. detectors whose signals are by chance in time with the others, but

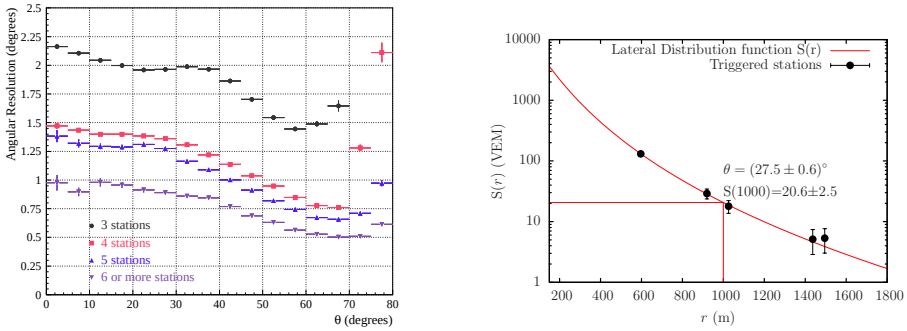


Figure 2.4: On the left panel, the angular resolution as a function of the zenith angle and for events with 3 stations (circles), 4 stations (squares), 5 stations (up-triangles), and 6 or more stations (down-triangles). On the right panel, the lateral profile of a shower as reconstructed with the SD.

that in fact are not part of the event. To this aim, a 'seed' was defined, made by three neighboring detectors in a non-aligned configuration. If there is more than one triangle of stations, the seed with the highest total signal is chosen. If the T4 trigger is a 3ToT, only ToT-T2 detectors can be considered to define the seed; if it is a 4C1, also TH-T2 detectors can be included. Once the triangle has been determined, the arrival direction is estimated by fitting the arrival times of the signals to a plane shower front moving with the speed of light. Subsequently, all other detectors are examined, and are defined as accidental if their time delay with respect to the front plane is outside a time window of $[-2\mu s, +1\mu s]$. Detectors that have no triggered neighbors within 3 km are always removed.

The geometrical reconstruction

When timing information is available in at least three stations in the surface detector, it is possible to apply a fit to the shower front assuming a plane front. Although this is a rough approximation, using it as a model for the fit has proven to be a robust estimator of the arrival direction. In this manner, the values of the zenith angle θ and azimuth angle φ in the local coordinate system can be estimated for each event. It is possible to refine the approximation of the shower front using a non-planar geometry, but knowledge about the position of the shower core is needed. After the LDF fit is performed on an event, the estimation of the core position is used to apply a spherical fit to the shower front and improve the determination of the arrival direction. Using the time variance model [57], the angular resolution of the surface array has been estimated as a function of the zenith angle and the station multiplicity. The plot in figure 2.4 (left) shows such dependence. The angular resolution is about 2.2° in the worst case of vertical showers with only 3 stations hit. This value improves significantly for 4 and 5 stations. For 6 or more stations, which roughly corresponds to events with energies above 10^{19} eV, the angular resolution

is in all cases better than 1°.

The lateral distribution of particles

The density of particles at a fixed distance from the shower axis provides a reliable estimator of the energy of the primary cosmic ray, as pointed out by Hillas [21]. Because it is not feasible to fully cover the effective area of the apparatus with detectors, it is unlikely to have an actual measurement of the density of particles at the desired distance from the EAS axis. Hence the signal a shower would have produced on a detector placed at that distance is estimated by means of a fit to a given LDF, using the signals coming from the stations that were triggered in the event. The fit allows to estimate two shower observables, the position of the core and the signal at a fixed distance (i.e. the particle density). On one hand, the estimation of the shower core allows to refine the determination of the arrival direction of the shower, modeling the shower front using a spherical geometry. On the other hand, the signal estimated at a given distance allows to determine the energy of the event, after the calibration procedure is applied. The optimal distance is determined primarily by the geometry of the array: for the main array of the SD it is 1000 m.

The lateral dependence of signals measured in stations is modeled as

$$S(r) = S_{1000} f_{LDF}(r), \tag{2.1}$$

where the function f_{LDF} satisfies $f_{LDF}(1000\text{ m}) = 1$. S_{1000} is called the size parameter and represents the signal that would have been measured by a detector placed at 1000 m from the shower axis. This value is then employed for estimating the energy of the event, via the calibration curve. An example of a fitted LDF is depicted in figure 2.4 (right).

2.1.4. Energy calibration

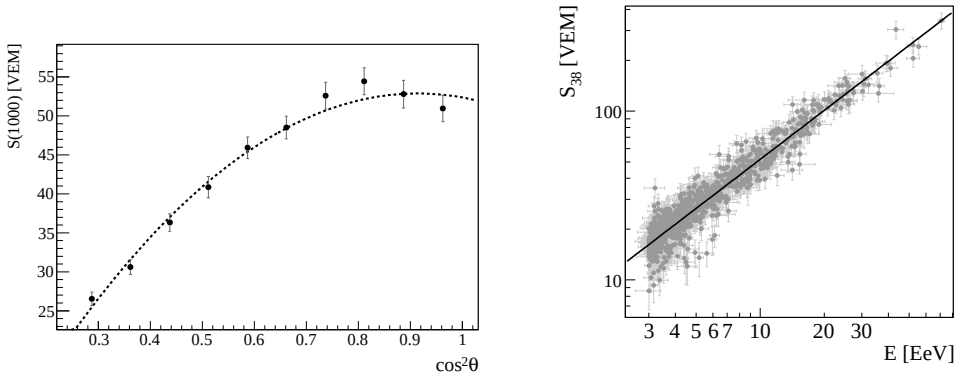
The reconstructed value of the signal at 1000 m from the shower axis cannot be used directly as the energy estimator of a shower. The estimator has a dependence on the zenith angle of the incoming shower that must be corrected beforehand. The correction is done by means of the attenuation curve of showers of fixed energy. The attenuation curve, shown in figure 2.5a, is derived by means of a constant intensity cut method (CIC) [58], which relies on the assumption that the flux of cosmic rays, above a certain energy threshold, is isotropic. The energy estimator, independent of θ is the S_{1000} that EAS would have produced had they arrived at the median zenith angle, 38°:

$$S_{38} = S_{1000}/p(x), \tag{2.2}$$

where $x = \cos^2\theta - \cos^2 35^\circ$ and p is a second order polynomial obtained from the CIC analysis. Then, the energy associated to each event is obtained by means of the formula

$$E = A S_{38}^B. \tag{2.3}$$

The coefficients A and B in the previous formula are determined by a fit applied to the correlation between the S_{38} values, provided by the surface detector, and the



(a) The attenuation curve for showers detected by the main array of the SD. The solid line represents the fit applied with a polynomial of second order.

(b) The correlation between S_{38} and E for 839 hybrid events used in the fit.

Figure 2.5: The curves illustrate the steps required to obtain an energy estimate for showers detected by the surface detector.

energy estimates given by the fluorescence detector. The fit is applied on a subset of high quality data measured by the instrument working in a hybrid detection mode. The correlation of S_{38} values and energy E is shown in figure 2.5b, along with the curve that provides the best fit. The values of the constants estimated by the fit are

$$A = 1.51 \pm 0.06(\text{stat}) \pm 0.12(\text{syst}) \times 10^{17}\ \text{eV}, \quad B = 1.07 \pm 0.01(\text{stat}) \pm 0.04(\text{syst}). \tag{2.4}$$

The overall energy resolution is about 20%.

2.2. The fluorescence detector

The fluorescence detector of the Pierre Auger Observatory is composed of twenty-four telescopes placed at four different locations in groups of six. These telescopes overview the stations of the surface array, allowing for a hybrid detection of EAS. The location of the fluorescence detectors relative to the surface array is shown in figure 2.1.

Each telescope has a 20×22 pixels camera placed at the focus of a 3.4 m $\times$ 3.4 m spherical mirror used to collect fluorescence light emitted by the atmosphere. Light enters the telescope through a diaphragm of 1.70 m of diameter. The aperture of the telescope is equipped with corrector lenses to reduce optical aberrations and a UV transmitting filter to improve the signal to background ratio. Each pixel in the camera is a Photonis XP-3062 PMT with a sky coverage close to $1.5\,^\circ$, giving the telescope a field of view of $30\,^\circ \times 30\,^\circ$. A schematic is shown in figure 2.6. A detailed account of the fluorescence detector can be found in [60].

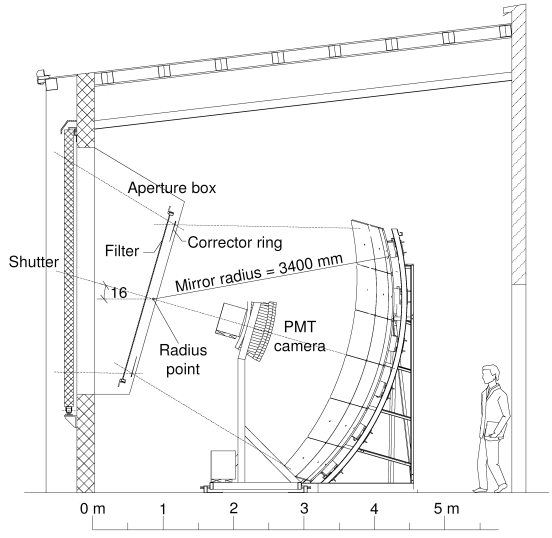


Figure 2.6: A schematic view of one of the telescopes of the fluorescence detector. Figure extracted from [59].

2.2.1. The calibration of the fluorescence detector

Two types of calibrations are performed on the fluorescence detector. The first is an absolute calibration [61] applied on a telescope using a diffuse light source. The light source consists of two pulsed UV LEDs (375 nm) embedded in a Teflon^{TM}cylinder placed on top of a 15 cm disk made of diffusively reflective Tyvek^{TM}. The source uniformly illuminates each PMT in the camera with a known intensity. This allows to readily transform an integrated electronic signal to the number of photons collected by every single pixel. This calibration procedure provides an end to end method that takes into account all the components of each telescope, from the optical system to the electronics.
The second type of calibration is a relative one. It is run on a nightly basis to monitor the response and stability of the detectors. Optical fibers are utilized to distribute light to different sections in every telescope, namely the center of the mirror, the edges of the PMT camera and the inner side of the shutters. Then, the total charge collected by each PMT is compared to that obtained in the absolute calibration procedure.

Finally, a cross-check of the calibration procedures is carried out using the Central Laser Facility (CLF), deployed at the center of the array, at distances that range from 26 km to 39 km from the FD sites. The CLF is able to fire laser shots of known energy and direction. The light from these shots scatters trough Rayleigh or aerosol scattering in the air, and a fraction of it arrives to the fluorescence detectors. This light produces a response similar to that generated by a shower initiated by a 10^{20} eV cosmic ray [62]. The observed difference between the reconstructed energy and the laser energy is 10-15%, consistent with the uncertainties in the absolute calibration and knowledge of the atmosphere. Further details on the calibration procedures and cross-check can be found in [63].

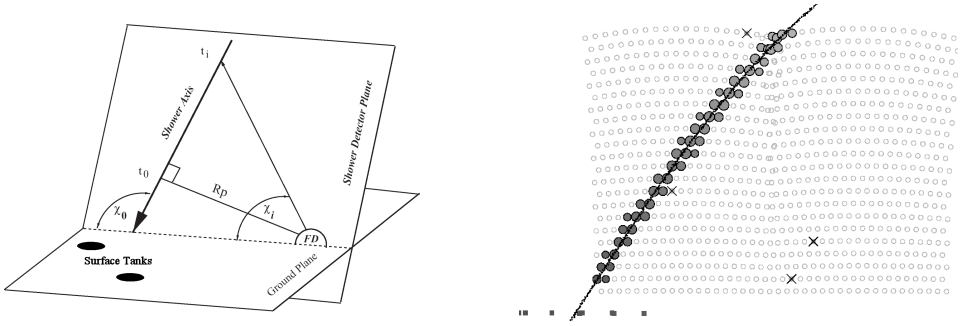


Figure 2.7: The schematic on the left shows the shower detector plane along the parameters estimated by the geometrical reconstruction [64]. The figure on the right shows an event as seen by the FD camera [65]. The squares at the bottom of the figure represent triggered stations in the surface array.

2.2.2. The fluorescence detector triggers

The PMT signals in the camera of a fluorescence telescope are collected by means of an analog board, that filters and amplifies them. This board is in turn connected to a digital front-end board that hosts the implementation of the first level (*FLT*) and second level trigger (*SLT*), that are applied on the digitized signals. The signals are sampled using a 10 MHz frequency and a 12 bit channel. The programmable logic chips (FPGA) apply tests on events in search of definite pattern configurations in a 1 μs time window.

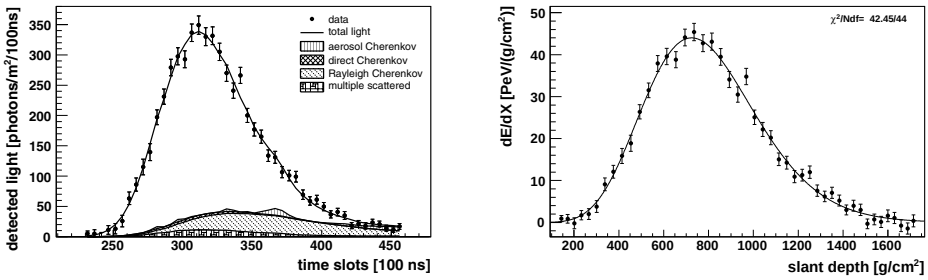
The FLT operates at the level of each single pixel. A threshold is set in order to keep the event rate close to 100 Hz. It is imposed on the PMT trace by requiring that the sum of the last 10 bins be above the adjusted threshold value. Next, the SLT searches for five adjacent pixels, with at most one of them below the threshold value, using a 5×22 sliding matrix. The requirement to have at most one pixel under threshold allows to include events detected with at most one bad pixel. The SLT event rate is close to 0.1 Hz.

The last step is performed on stations known as *mirror PCs*. A process in these stations loops through all SLT events looking for spurious triggers created by noise, by analyzing their time sequence. This process reduces the event rate by an order of magnitude.

2.2.3. Event reconstruction

The geometric reconstruction

In the FD, cosmic ray showers are detected as a sequence of triggered pixels in the camera (see figure 2.7, right). The first step taken in order to estimate the arrival direction of an EAS is to determine the Shower-Detector Plane (SDP). The SDP is the plane that includes the location of the eye and the line of the shower axis (see the sketch in figure 2.7 left). It is determined as the plane through the eye which most nearly contains the pointing directions of the FD pixels centered on the shower axis (figure 2.7 right). Next, the timing information of the pixels is used to reconstruct the



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

Figure 2.8: An example of a reconstructed shower profile.

shower axis. A fit is applied to the time labels of pixels illuminated by fluorescence light. Three parameters come out of the fit. The distance from the shower axis to the FD site, or R_p, the time t_0 at which that distance is reached and the angular orientation χ_0 of the shower axis on the SDP. Using these parameters it is possible to estimate the zenith angle θ and azimuth angle φ of the event in the local coordinate system.

The function used to fit the timing information follows

$$t_i = t_0 + \frac{R_p}{c} \tan\left(\frac{\chi_0 - \chi_i}{2}\right), \tag{2.5}$$

where t_i and χ_i stand for the time stamp and the direction on the SDP of the ith pixel.

The determination of the shower arrival direction can be greatly improved in two ways. One is to have an event observed by telescopes in at least two FD sites (i.e., stereo events). The other is to have the timing information in at least one station from the surface detector (i.e., hybrid events). Hybrid events achieve the best geometrical accuracy: the resolution on core location is 50 m, and the typical angular resolution is 0.6°.

The longitudinal profile

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 as a function of slant depth. For this purpose, the light attenuation from the shower to the telescope needs to be estimated and all contributing light sources need to be disentangled: fluorescence light. direct and scattered Cherenkov light as well as multiple-scattered light. An example of the measured light aperture and energy deposit profile is shown in figure 2.8 left and right, respectively. Note that the light collected by the telescopes must be corrected for the attenuation between the shower and the detector: this is done with the help of data from atmospheric monitoring devices [66]. Finally, the calorimetric energy of a shower is estimated by fitting a Gaisser-Hillas function [12]

to the reconstructed profile and integrating it:

$$f_{GH} = \left(\frac{\mathrm{d}E}{\mathrm{d}X}\right)_{max} \left(\frac{X - X_0}{X_{max} - X_0}\right)^{(X_{max}-X_0)/\lambda} \exp\left(\frac{X_{max} - X}{\lambda}\right) \tag{2.6}$$

The total energy is then obtained by correcting for the 'invisible energy' carried away by neutrinos and high energy muons. This correction is obtained by Monte Carlo simulations. The correction is of the order of 10% and has a systematic uncertainty close to 4% [67]. Overall, the energy resolution of the FD (defined as event-to-event statistical uncertainty) is 10%. The current systematic uncertainty on the energy scale amounts to 22% (for a details of the different contributions see e.g. [38])

The position of the shower maximum X_{max} is also inferred from the fit to the longitudinal profile, as can be seen from formula (2.6). The uncertainty associated with this estimate is around 20 g/cm^2 [68].

2.3. Overview of some of the results from the Pierre Auger Observatory

Auger results related to the measurement of the energy spectrum, the study of the mass composition and the search for anisotropies in the distribution of arrival directions are summarized in the next paragraphs. The detailed discussion on other results such as the search for photons [69], neutrinos [70] and neutrons [71], the measurement of proton-air cross section above 10^{18} eV [72] are not reported here and more details can be found in the corresponding papers.

The energy spectrum

The energy spectrum is measured by the Pierre Auger Observatory above an energy threshold of 10^{18} eV, by combining the independent measurements in hybrid and SD modes. The results presented here include only events with zenith angle smaller than 60°. The spectrum measured with hybrid events is determined using data collected between November 2005 and September 2010. The exposure of the hybrid detector has been calculated using a time-dependent Monte Carlo simulation [73] which takes into account the changing configurations of FD and SD, as well as the actual atmospheric conditions. Only events that satisfy strict quality criteria are selected ensuring an energy resolution of about 10%. The total systematic uncertainty on the derived exposure is estimated as 10% (6%) at 10^{18} eV ($> 10^{19}$ eV) and it is mainly due to the lack of knowledge on the mass composition. 3660 hybrid events have been selected [74] and have been used for the measurement of the energy spectrum above 10^{18} eV. The main source of systematic uncertainty on the energy spectrum is due to the energy scale (22%) [75]. The largest contribution (14%) is given by the absolute fluorescence yield. Other contributions are the uncertainties in the absolute calibration of FD (9%), the measurement of the atmospheric pressure, humidity and temperature (5%), the light attenuation (4-8%, depending on energy), the lateral width and hybrid reconstruction (9%) and the fraction of missing energy (4%).

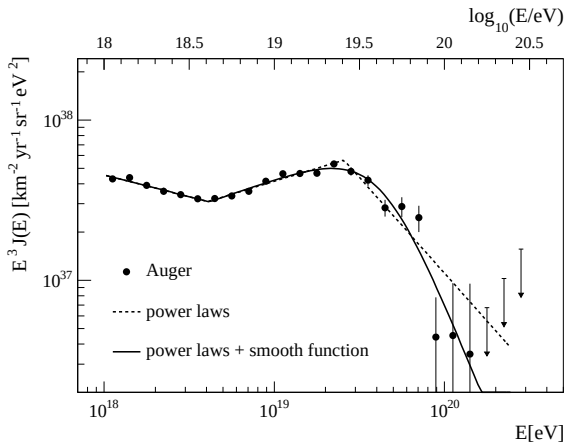


Figure 2.9: The energy spectrum obtained by combining the hybrid spectrum and the one measured with SD data. It is fitted with three power laws functions (dashed) and two power laws plus a smooth function (solid line). Only statistical uncertainties are shown. The systematic uncertainty on the energy scale is 22%.

The exposure for the SD is calculated integrating the number of active stations of the surface array over time and it is determined above 3×10^{18} eV, where the SD acceptance is saturated independently of the primary mass. It is about 21000 km^2 sr yr, as calculated between January 2004 and December 2010, with an associated uncertainty of about 3%. The total number of selected events above 3×10^{18} eV is about 64000. The energy resolution is about 16% at threshold, and is about 12% above 10^{19} eV. A forward-folding approach is thus applied to correct the flux for the energy resolution. This correction is mildly energy dependent but smaller than 20% over the entire energy range. The uncertainty on the normalization of the SD flux is about 6%.

In figure 2.9 the energy spectra derived with hybrid data and with the events collected by SD above $10^{18.5}$ eV are combined together using a maximum likelihood approach. The normalization uncertainties on the two spectra are used as additional constraints in the combination procedure and a flux scaling factor of $\sim 1\%$ has been derived to match the two spectra. Since the SD energy estimator is calibrated from the subset of golden hybrid events, the two input spectra have the same systematic uncertainty on the energy scale while the flux normalization uncertainties are independent. The characteristic features of the combined spectrum have been quantified with three power laws with free breaks between them (dashed line in figure 2.9) and with two power laws plus a smoothly changing function (solid line). A change of spectral index from 3.27 ± 0.02 to 2.68 ± 0.01 at an energy of $10^{18.61 \pm 0.01}$ eV has been observed. The hypothesis that the power law above the ankle continues to highest energies with the spectral index 2.68 ± 0.01 can be rejected with more than 20σ.

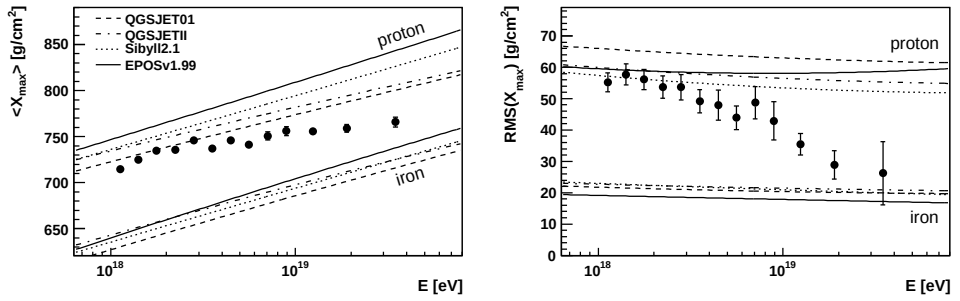


Figure 2.10: Values of $< X_{max} >$ and RMS(X_{max}) as a function of energy. A comparison with different air shower simulations is included. Figure taken from [38].

The mass composition

The atmospheric depth X_{max} at which a shower reaches its maximum development is related to the mass of the primary particle and to the characteristics of the hadronic interactions at very high energy. As the FD can directly observe the X_{max}, hybrid data collected between December 2004 and September 2010 are used for this study. Only events with zenith angle smaller than 60° and with a precise reconstruction of the geometry and of the longitudinal profile were selected. To ensure a reliable measurement of the longitudinal profile, the X_{max} has to be observed in the FD field of view and strict cuts on the aerosol content and cloud coverage are applied. The limited field of view of the fluorescence detector and the requirement of observing the shower maximum may introduce a different selection efficiency for different primary masses. To reduce a possible induced bias, a set of dedicated cuts have been defined to identify the geometrical volume which guarantees comparable selection efficiency to all nuclear primaries [38]. 6744 events (with $E > 10^{18}$ eV) fulfill these criteria and a resolution of about 20 g/cm^2 is obtained for X_{max}. This resolution is estimated with detailed simulations of the detector and cross-checked using events observed by two or more FD sites.

In figure 2.10 the average X_{max} is shown as a function of energy. An energy dependent correction has been applied to data to compensate for a small bias observed when reconstructing Monte Carlo simulated events. The expected $< X_{max} >$ for proton and iron primaries, assuming different hadronic interaction models, are shown for reference. As mentioned in the previous section, the systematic uncertainty on the energy scale is about 22%. The total systematic uncertainty on $< X_{max} >$ ranges between 10 g/cm^2 at low energy and 13 g/cm^2 at high energy. It includes contributions from the uncertainties in the calibration, the atmospheric data, the reconstruction and the event selection. The data are well described by a two slopes function, with a break point at $\log(E/\text{eV}) = 18.38^{+0.07}_{-0.17}$. The elongation rates derived below and above this energy are 82^{+48}_{-8} g/cm^2/decade and 27^{+3}_{-8} g/cm^2/decade, respectively. By comparing data and simulations, this result can be interpreted as a change in composition of cosmic rays from light to heavy primaries. However, this conclusion relies on the hadronic interaction models which are based on the extrapolation of accelerator data at low energies. Similar conclu-

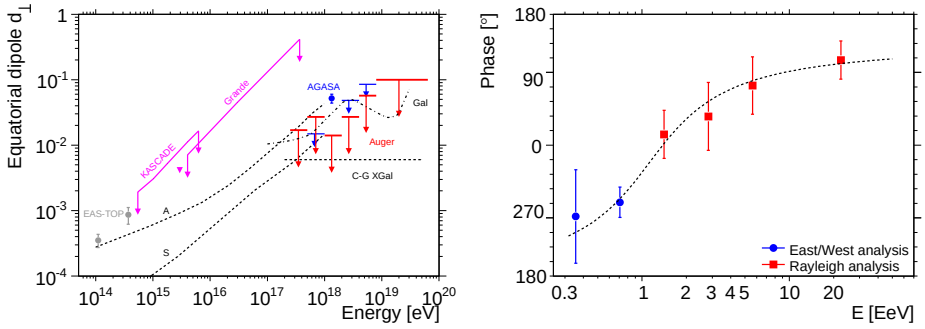


Figure 2.11: The figure on the left panel shows the upper limits on the anisotropy amplitude of first harmonic as a function of energy. The figure on the right panel shows the phase of the first harmonic as a function of energy. [76]

sions (with the same remark) can be derived by investigating the width of the X_{max} distribution. Indeed also the fluctuations in the shower development are connected to the mass of the primary type. The RMS(X_{max}), derived for the same sample of events, is shown in the same figure. The systematic uncertainty has been quoted as 5 g/cm^2. Moreover, it is remarkable that the decrease of the RMS(X_{max}) with energy become steeper around the same break point found for the elongation rate.

Anisotropies

Observation of the suppression of the cosmic ray flux at the highest energies (see above) and its possible interpretation in terms of the GZK effect suggests that the closest sources are situated within a volume in $d \leq 100$ Mpc. At these scales the matter distribution in the universe is inhomogeneous, and so must be the distribution of sources. If propagation of cosmic ray at these distances is quasi-rectilinear, anisotropies would be expected. The Pierre Auger Collaboration in fact reported [77] directional correlations at $E > 55$ EeV with AGN from the Veron-Cetty-Veron catalog within 75 Mpc on an angular scale of 3.1° at the 99% CL. The optimal parameters were found using an exploratory scan on an early data set. The latest update including data up to June 2011 yields a total of 28 of 84 events (excluding data used for the original scan) showing a correlation on a 3.1°-scale with a nearby AGN. The overall correlation strength thus decreased from the original (62 ± 10)% to (33 ± 5)%. The chance probability of observing such a correlation from a random distribution remains below 1%. Evidently, more data are needed to arrive at a definite conclusion.

Besides searching for point sources of charged cosmic rays, the large scale distribution of arrival directions of cosmic rays represents another important tool for understanding their origin. Using data from the SD array, upper limits below 2% at 99% C.L. have been reported for EeV energies on the dipole component in the equatorial plane [78]. Such upper limits are sensible, because cosmic rays of galactic origin, while escaping from the galaxy in this energy range, might generate a dipolar large-scale anisotropy with an amplitude at the % level as seen from the Earth.

Even for isotropic extragalactic cosmic rays, a large scale anisotropy may remain due to the motion of our galaxy with respect to the frame of extragalactic isotropy. The derived upper limits are shown in fig. 2.11 (left), together with predictions for anisotropies arising from models of both galactic and extragalactic cosmic ray origin. In models A and S (A and S standing for 2 different galactic magnetic field symmetries), the anisotropy is caused by drift motions due to the regular component of the galactic magnetic field, while in model Gal, the anisotropy is caused by purely diffusive motions due to the turbulent component of the field. Some of these amplitudes are challenged by Auger current sensitivity. For extragalactic cosmic rays considered in model C-GXgal, the motion of our galaxy with respect to the CMB induces a small dipolar anisotropy.

While the measurements of the amplitudes do not provide any evidence for anisotropy, it is interesting to note that the phase shown in fig. 2.11 (right) suggests a smooth transition between a common phase of $\sim 270^\circ$ below 1 EeV and another phase ($\sim 100^\circ$) above 5 EeV. This is interesting, because, with a real underlying anisotropy, a consistency of the phase measurements in ordered energy intervals is indeed expected with lower statistics than that required for the amplitudes to significantly stand out of the background noise. Since an a priori search for such a smooth transition in the phase measurements was not performed, no confidence level can be derived from this result. A prescription is currently running, by using SD and infill data.

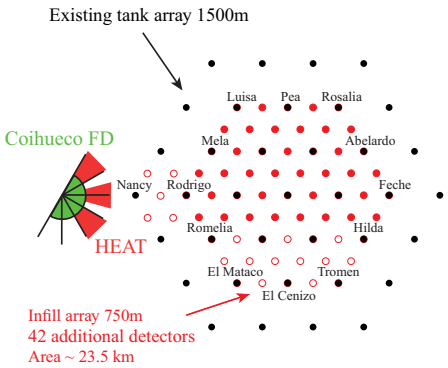
Chapter 3

The infill array

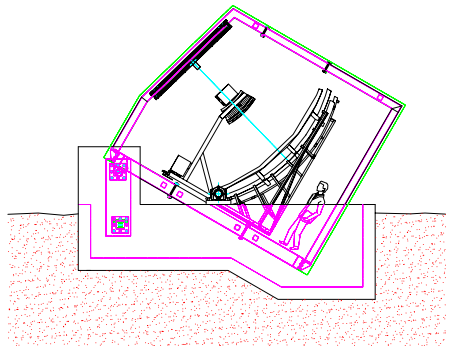
This chapter presents a general description of the infill array, with particular emphasis on the trigger and event reconstruction, as those will be extensively used in next chapters. As mentioned in the previous chapter, the Pierre Auger Observatory has been enhanced with new detectors, AMIGA (Auger Muons and Infill for the Ground Array [79]) and HEAT (High Elevation Auger Telescopes [80]) that aim at measuring the cosmic ray spectrum and its chemical composition components down to 10^{17} eV by exploiting the hybrid technique. Both extensions started in 2008 after the construction of the observatory was completed. The infill array is part of the AMIGA enhancement, which will be introduced in the first section of this chapter, along with HEAT, for the sake of completeness.

3.1. HEAT and AMIGA

As the fluorescence light signal is roughly proportional to the primary particle energy, low energy showers can be detected only at short distances from the telescopes. In addition, as these showers develop earlier in the atmosphere, their shower maximum lies higher in the atmosphere and thus is not accessible to the standard FD telescopes due to their limited field of view in elevation (30°). This was the motivation to build **HEAT**: its three telescopes have now been constructed, and they are located 180 m north-east of the Coihueco FD building (see fig. 3.1). They are similar to the 24 standard ones except for the ability to tilt the telescopes upwards by 29°: this provides the extension of the field of view to larger elevation angles. Data taking with HEAT is possible in horizontal position as well as in the tilted position. The horizontal position is used for installation, commissioning and maintenance of the hardware, as well as for the absolute calibration. With HEAT in the tilted position, the combined HEAT-Coihueco telescopes cover an elevation range from the horizon to 58°. This extended field of view enables the reconstruction of low energy showers for close-by events with improved resolution in energy and X_{max} determination. The telescopes of HEAT have been operated since 2010. First calibrated data are being released at the time of writing. In this thesis no data from HEAT were used: the analyses presented here rely exclusively on regular FD data.

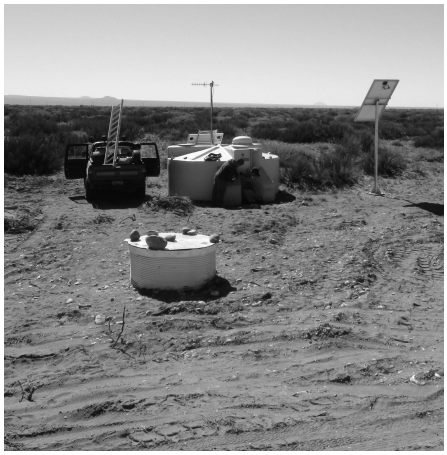


(a) Layout of the infill array and the Coihueco and HEAT fluorescence telescopes. The additional infill stations are marked red.

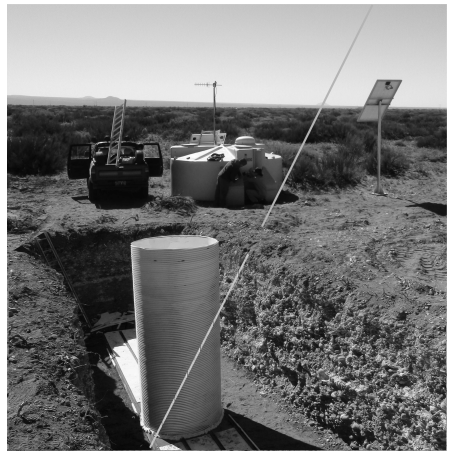


(b) Schematic view of the cross-section of one of the HEAT telescopes: data-taking mode in tilted orientation.

Figure 3.1: The HEAT enhancement [80].



(a) Surface infill SD station with its associated muon counter already buried.



(b) The detector concept: a muon of sufficiently high energy is capable of reaching the buried scintillator.

Figure 3.2: The AMIGA enhancement [79].

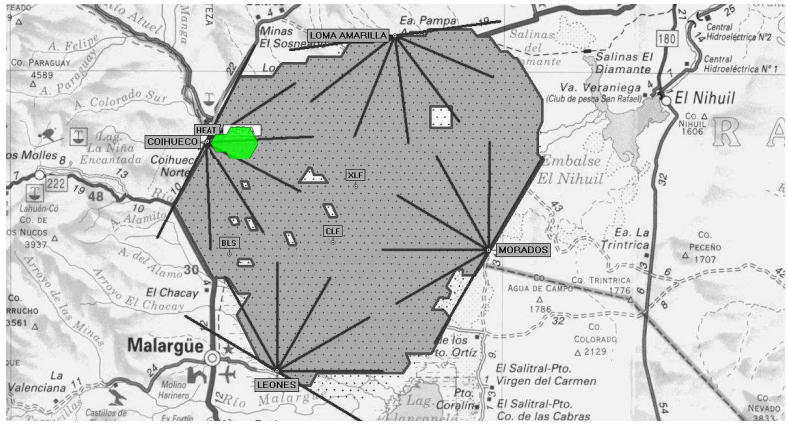


Figure 3.3: The position of the infill array (green) at the Pierre Auger Observatory.

AMIGA consists of an array of water-Cherenkov detectors (the same as in SD) set out on a hexagonal spacing with sides of 750 and 433 m and an associated set of muon detectors (MDs) each of 30 m^2 buried at a depth of 2.3 m corresponding to 540 g/cm^2. Any impinging muon with energy $\geq$1 GeV propagates in the soil and is capable of reaching the buried muon detectors that are located near to the WCDs as shown in fig. 3.2. Each counter has three 10 m^2 modules with 4.1 cm wide per 1.0 cm high per 400 cm long strips. Each module consists of 64 strips made of extruded polystyrene doped with fluor and co-extruded with a reflecting coating. The strips have a groove where a wavelength shifter optical fiber (WLS) is glued and covered with a reflective foil. The manifold of fibers ends in an optical connector matched to a 64 multi-anode PMT from the Hamamatsu H8804 series (2 mm per 2 mm pixel size). Scintillators are grouped in two sets of 32 strips on each side of the PMT. The muon detector is under deployment: first data associated to EAS have been detected in summer 2012 from the first four modules.

Concerning the WCD array, the 750 m infill (completely deployed and subject of this work) is laid out over an area of 23.5 km^2. The planned 433 m array will cover only 5.9 km^2 within the larger area. They will allow cosmic rays to be detected with full efficiency down to an energy of 3×10^{17} eV and 10^{17} eV, respectively.

3.2. Description and history of the infill array

The infill array is contained within the main array. Its relative location is presented in figure 3.3. The grid used for the infill follows the same triangular pattern as for the main array, using a tighter spacing of 750 m. Since the cosmic ray flux increases rapidly with the decreasing energy, the 750 m infill covers an area of only 23.5 km^2: as it will be shown in chapter 6, its spacing allows cosmic rays to be detected with full efficiency down to an energy of 3×10^{17} eV. The decision to use additional stations identical to those employed in the regular array allows one to take advantage of existing knowledge about the infrastructure required for the operation of such instruments. Moreover, the use of identical detectors allows one to do relevant cross-

check analyses of events recorded simultaneously with the main and infill arrays (as it will be shown in chapter 5).

The first completed hexagon started operation in September 2007. The infill array kept changing configuration until September 2011: not only the number of operating stations increased, but also the characteristics of some of the stations changed. In particular, few stations were equipped, for a limited period of time, with different radio receivers as those in the SD. These radio receivers, called Xbee, have caused serious communication problems with the stations that were equipped with them. The receivers were eventually replaced by the standard ones.

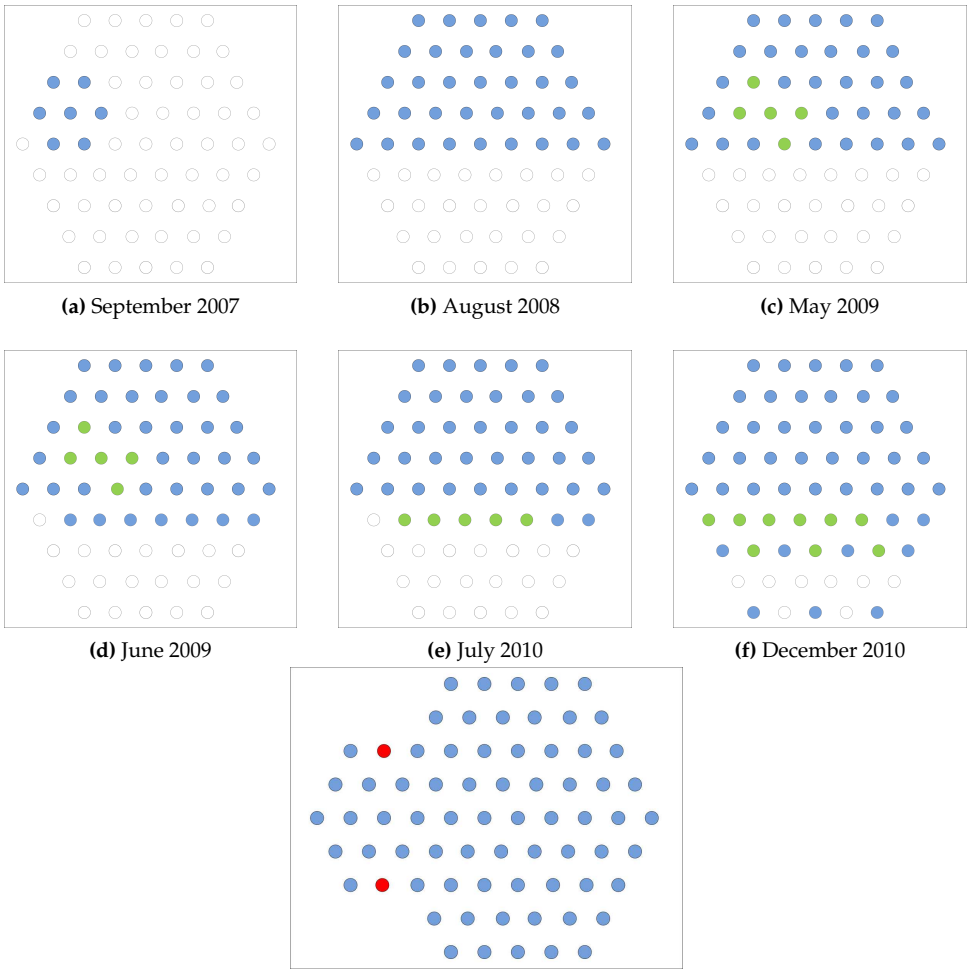
The full infill array presently has 69 working stations, forming 39 hexagons over a total area of 25 km^2. The effective collection area for of the apparatus is 19 km^2. Two more stations were deployed in mid-2012 to complete four additional hexagons: these are not considered in the analyses presented in this work. The evolution of the configuration of the array is shown in figure 3.4.

3.3. EAS triggers and selection

The trigger system implemented in the infill array is based on the same strategy used in the main array. The two first levels of the trigger chain, applied locally on each stations, are exactly the same and have already been described in section 2.1.2. In this section, the last three trigger levels are described in detail, as they have been modified to properly function with the infill array. The algorithms and the general structure of the trigger hierarchy is identical to the one used in the main array, and changes have been applied only to scale them to the 750 m spacing used in the infill array. The implementation of the T4 and T5 trigger levels for the infill array used in this thesis is based on the official Offline [55] code used in the main array of the SD. This section is propaedeutic to chapter6 where knowledge of high-level triggers is the basis to derive the exposure and its uncertainty.

3.3.1. The array trigger (T3)

The stations in the array are constantly transmitting their T2 triggers to the Central Data Acquisition System (CDAS). It is there, at the CDAS, that the occurrence of the *T3* trigger condition is checked based on the T2 trigger types, the position of the triggered stations and the time label from the signals. Two T3 conditions are implemented. The first requires at least three stations satisfying the ToT-T2 with a minimum compactness. That is, one of the ToT-T2 stations must have one of its neighbors within the first crown and the other within the second crown. This trigger type is called $ToT2C_1\&3C_2$. The term 'crown' is used to refer to the set of stations surrounding any given station. For example, the first crown is the set composed of the six nearest stations to a particular station. The second crown is the set of twelve stations which are the second nearest and so on (see figure 3.5). The second type of T3 trigger is more relaxed in terms of compactness and signal spread. It is satisfied by at least four stations with any kind of T2 trigger, in which one of the stations has one neighbor within the first crown, the second neighbor within the



(a) September 2007 **(b)** August 2008 **(c)** May 2009

(d) June 2009 **(e)** July 2010 **(f)** December 2010

(g) September 2011

Figure 3.4: The evolution in time of the configuration of the infill array. Stations with Leeds radios are denoted in blue, while stations with Xbee radios in green. Red circles represent two stations that were deployed in 2012, but are not used in the analyses presented in this work.

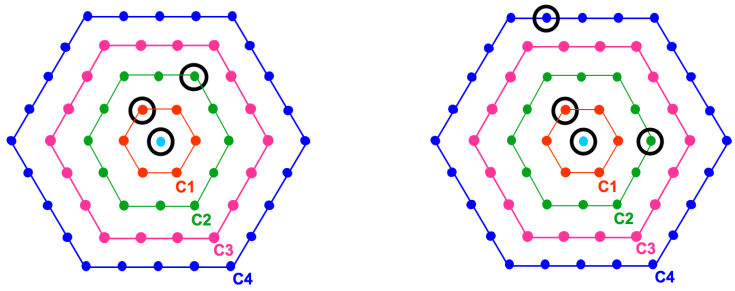


Figure 3.5: The two types of T3 trigger conditions are illustrated. On the left side the ToT2C$_1$&3C$_2$ trigger and on the right side the 2C$_1$&3C$_2$&4C$_4$ trigger. Note the definition of crowns.

T3 trigger	Infill array	Regular array
ToT2C$_1$&3C$_2$	0.04 Hz	0.02 Hz
2C$_1$&3C$_2$&4C$_4$	0.03 Hz	0.01 Hz

Table 3.1: The T3 trigger rate in the infill and main arrays of the surface detector.

second crown and the last neighbor must be within the fourth crown. This trigger is named 2C_1&3C_2&4C_4. Figure 3.5 illustrates the two types of T3 triggers.

Once the spatial requirements of any of the T3 trigger types are met, checks of the timing information are performed. A simple formula is used to define the size of a time window in which the triggered stations must be contained. The formula is $(6 + 5\mathrm{C}_n)$ μs where C$_n$ stands for the largest crown in the event. Once the size of the window is given, all the triggers must be contained in the time window defined by the first triggered station and the window size. If this timing condition is also verified, the event passes the T3 level in the hierarchy and data acquisition is started in the infill array. Thus the FADC traces of all the triggered stations are recorded, plus any station satisfying the T1 trigger, provided it is within 30 μs of the T3.

As an example, the number of T3 events recorded on November 3rd 2011 was determined, according to their type. Note that by this time the infill configuration was already complete. The ToT2C$_1$&3C$_2$ trigger selected 64% of the 6046 T3 events recorded during that day, while the 2C$_1$&3C$_2$&4C$_4$ trigger selected the remaining 36%. The trigger rates derived from the same day of data are $\sim$ 0.045 Hz (i.e. about 3/min) for the ToT2C$_1$&3C$_2$ trigger and $\sim$ 0.025 Hz (i.e., about 1.5/min) for the 2C$_1$&3C$_2$&4C$_4$ trigger. Table 3.1 compares the T3 rates in the infill array to those corresponding to the main array.

3.3.2. The physics trigger (T4)

The next step in the trigger hierarchy is a physics trigger applied to recorded data in order to select real showers. In a similar fashion as for the T3 level, two types of *T4* trigger conditions are defined. The first is the *3ToT* that requires the presence of three stations satisfying the T2-ToT, where at least one of the stations has the other two within the first crown. Furthermore, the stations should not be aligned

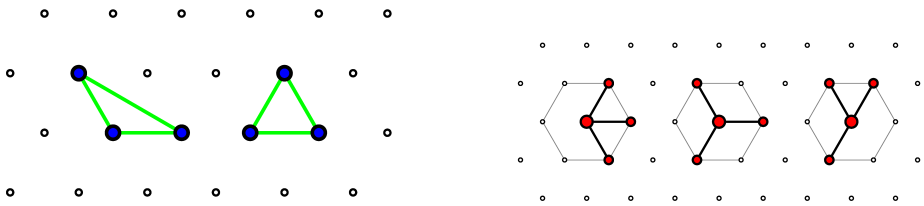


Figure 3.6: The geometry of both kinds of T4 triggers is illustrated. On the left the 3ToT trigger and on the right the $4C_1$.

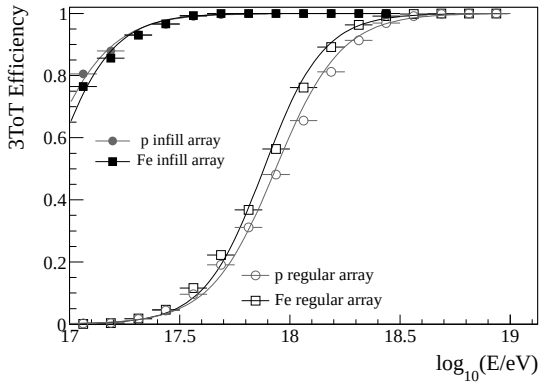


Figure 3.7: The trigger efficiency for the infill array as a function of Monte Carlo energy E for proton (circles), iron (squares) and zenith angle integrated up to 55°. The trigger efficiency of the main array, integrated up to 60°, is shown as a reference. Figure taken from [81].

but should form a triangular pattern. The timing of the stations has to be such that they all fit a shower plane moving at the speed of light. The second type of T4 trigger requires the presence of four stations, with any kind of T2 trigger, where the compactness condition demands that one of the stations should have the other three stations within the first crown. This trigger condition is known as the $4C_1$. Again, the time label of the signals should conform a shower front moving at the speed of light. See figure 3.6 for an illustration of the geometrical configurations of both types of T4 triggers.

The total rate of T4 events in the infill array in its present configuration is close to 0.017 Hz (about 1/min). Out of the T3 ToT2C_1&3C_2 events 39% are T4 ones (3ToT), while out of the T3 2C_1&3C_2&4C_4 events only 0.8% pass the T4 condition (4C1). The total number of T4 events is thus mostly due to the 3ToT condition (98%): this is expected as that one is the most efficient one for vertical events, i.e., events up to 60°. The 4C1 condition is relevant mostly for inclined showers.

The trigger efficiency as a function of energy for 3ToT events, dominant in the infill and main arrays is shown in figure 3.7: events with zenith angles below 55° (infill) and 60° (main) are used. The subject of trigger efficiency is more thoroughly discussed in section 6.1.

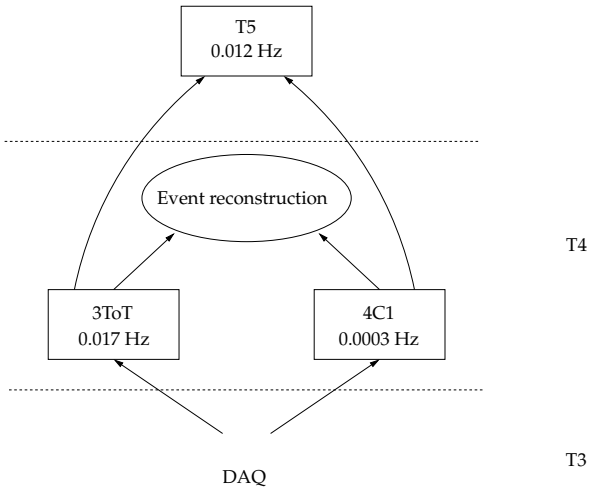


Figure 3.8: A diagram illustrating the last two trigger levels in the hierarchy. These are applied off-line.

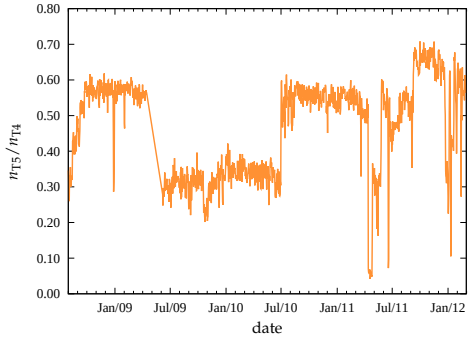


Figure 3.9: The time evolution of the ratio of T5 events over T4 events in the infill array.

3.3.3. The fiducial trigger (T5)

Showers landing on the borders of the array may satisfy all trigger conditions in the hierarchy described so far. These showers have a significant portion of the particles falling outside the array and thus not detected. Such showers might be badly reconstructed because a considerable part of the information is missing. The last level of the trigger hierarchy (*T5*) is designed to select only showers that are well contained within the boundaries of the array. Moreover, it is also meant to reject showers that fall on parts of the array where some stations are not active, although deployed. The T5 condition requires that the station with the highest signal be surrounded by six, not necessarily triggered, working stations. Besides ensuring a reasonable sampling of EAS, the T5 trigger level allows to define the effective area of the array in terms of the geometrical area only, above a certain energy threshold. This simplifies the determination of the exposure of the array, that will be discussed in later chapters. Figure 3.8 illustrates the hierarchy of the last two trigger conditions.

The rate of T5 events for the full infill array is around 0.012 Hz (about 0.7/min).

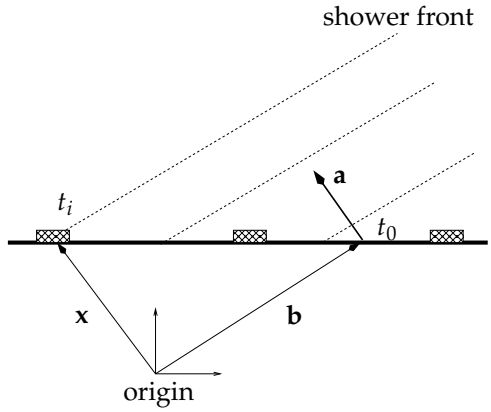


Figure 3.10: A depiction of the shower front.

The evolution in time of the ratio of T5 to T4 events n_{T5}/n_{T4} is shown in figure 3.9. The irregularities in the ratio n_{T5}/n_{T4} are due in part to the deployment (the array was completed in summer 2011) and in part to problems during data acquisition, discussed in chapter 6 and appendix A.

3.3.4. Station selection

The last step for the definition of an event to be reconstructed is the rejection of accidental stations (i.e. stations whose signals are in time with others only by chance). To do so, the same approach as the main array is used (see section 2.1.3). The time window used to reject stations has been modified in order to optimize it for events in the infill array [82]. Stations with a time delay outside the $[-0.5\mu s, +1.4\mu s]$ interval are discarded as accidental stations. Moreover, stations with no neighbors within the first crown or only one neighbor within the third crown are discarded as lonely stations.

3.4. EAS reconstruction

As in the case of the trigger system, the event reconstruction in the infill array is largely based on the reconstruction used in the main array. The geometrical reconstruction used to determine the arrival direction of showers was not modified at all from the original implementation. Also, the strategy employed for fitting the lateral distribution of showers is the same as the one used in the main array, and only the LDF has been modified to adapt it to the infill array. The LDF will be described in the next chapter. The implementation of the event reconstruction chain is based on the official Offline software [55].

3.4.1. The geometrical reconstruction

The geometrical reconstruction in the infill array is applied to all events satisfying the T4 condition, either as 3ToT or 4C1. The output of this procedure are the zenith

and azimuth angles in local coordinates, that define the arrival direction of each shower. The first step to determine the arrival direction of an EAS is to apply a fit to the shower front using a plane as an approximation [83]. The following relation

$$\mathbf{x}(t) - \mathbf{b} = -c(t - t_0)\mathbf{a} \tag{3.1}$$

describes the movement of a point $\mathbf{x}$ with the speed of light c along the straight line defined by the normal vector $\mathbf{a}$ that points to the source of the shower. Here $\mathbf{b}$ stands for the signal weighted barycenter of the event.

The time $t(\mathbf{x})$ when the shower front passes trough point $\mathbf{x}$ is inferred by projecting $\mathbf{x}$ to the shower axis $\mathbf{a}$,

$$ct(\mathbf{x}) = ct_0 - (\mathbf{x} - \mathbf{b}) \cdot \mathbf{a}. \tag{3.2}$$

Assuming that the positions of the stations are known with infinite precision, the deviations from (3.2) occur only due to uncertainties in the signal start times. This allows to write a χ^2 function

$$\chi^2 = \frac{1}{\sigma_t} \sum_i \left[t_i - t(\mathbf{x}_i') \right]^2 = \frac{1}{c^2 \sigma_t^2} \sum_i \left[ct_i - ct_0 + \mathbf{x}_i' \cdot \mathbf{a} \right]^2, \tag{3.3}$$

where $\mathbf{x}_i' = \mathbf{x}_i - \mathbf{b}$ is the position of the i-th station as measured from the barycenter, and σ_t is the time uncertainty in the signal start time. This function is then minimized in order to obtain an estimation of the normal vector $\mathbf{a}$. The schematic in plot 3.10 illustrates the situation.

The plane fit can be extended with a parabolic term to account for the shower curvature R_K near the impact point $\mathbf{K}$. Setting $\mathbf{x}' = \mathbf{x} - \mathbf{K}$ it is possible to write an extended form of (3.2)

$$ct(\mathbf{x}) = ct_0 - \mathbf{a} \cdot \mathbf{x}' + \frac{\rho(\mathbf{x}')^2}{2R_K} \tag{3.4}$$

with the distance $\rho(\mathbf{x}') = (\mathbf{a} \times \mathbf{x}')^2 = x^2 - (\mathbf{a} \cdot \mathbf{x})^2$. In order to apply the curvature fit, an estimation of the shower core on the detection plane is needed beforehand and, therefore, this method is applied after a core estimation is given by the fit to the lateral distribution. Further details of the curvature fit procedure can be found in [83].

The resolution on the determination of the arrival direction has been estimated as a function of the zenith angle θ and the station multiplicity [81], as shown in figure 3.11. The resolution is obtained from the single station time variance, modeled to take into account the size of the total signal and its time evolution [84]. For showers triggering more than three stations, the resolution is better than 1.3°.

3.4.2. The lateral distribution function

Two different lateral distribution functions (LDF) were used to model the dependence of the signal measured by the stations in the infill array. In chapter 4 the performances of both functions will be studied in details. The main focus of this section is to describe in general the fitting procedure and the used treatment of signals.

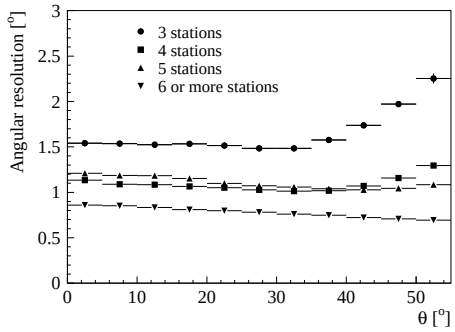


Figure 3.11: Angular resolution as a function of the zenith angle and station multiplicity in the infill array.

A maximum likelihood method is implemented for fitting the LDF $f(r)$ [83]. The function has to be such that $f(r_{opt}) = 1$. This allows to write the lateral dependency of the signal as

$$S(r) = S_{opt}\, f(r). \tag{3.5}$$

Here r_{opt} stands for the optimum ground parameter [85]. This parameter is weakly dependent on the primary energy and arrival direction and it is mainly determined by the array topology. It is at this distance that fluctuations in the expected signal, determined through a LDF, are minimized. Studies carried out using data from the infill array [86] and Monte Carlo simulations [87] established the value of $r_{opt} =$ 450 m as a reasonable estimate for the optimum ground parameter.

The choice of a maximum likelihood method for the fitting procedure permits the consideration of non-triggered stations (usually referred as silent stations), stations with small signals, station with large signals and saturated stations. The signal uncertainty is parameterized as [88]

$$\sigma_S = (0.32 + 0.42\ \sec\theta)\ \sqrt{S}. \tag{3.6}$$

The total number of particles n in the detector is

$$n = p(r, \theta | E, A) S, \tag{3.7}$$

where $p(r, \theta | E, A)$, known as the Poisson factor, is assumed to be a constant [83].

Four different cases are considered in describing the variations of n in each station:

Small signals are modeled using Poisson's statistics

$$f_P(n, \mu) = \mu^n e^{-\mu} / n! \tag{3.8}$$

Large signals are approximated by a Gaussian

$$f_G(n,\mu) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{n-\mu}{\sigma}\right)^2\right] \tag{3.9}$$

No signal in a non-triggered station is modeled by Poisson's statistics. A threshold value of at least three particles hitting a tank in order to trigger the station is assumed, hence $F_{zero}(3,\mu)$ gives the probabilities of having three or less particles in a tank. Since Poisson's statistics is used, F_{zero} is the Poisson distribution.

Saturated signals may be recovered using a technique described in [89]. If the recovery is successful the signal is treated as a large signal with a parameterized value for σ. But if the recovery fails, the saturated signal is treated as a lower limit, and thus the probability of having a signal larger that the one observed is

$$F_{sat}(n,\mu) = \int_n^\infty f_G(n,\mu)\mathrm{d}n = \frac{1}{2}\mathrm{erfc}\left(\frac{n-\mu}{\sqrt{2}\sigma}\right), \tag{3.10}$$

where erfc is the complementary error function.

Using the functions just defined it is possible to build a likelihood function

$$L = \prod_i f_P(n_i,\mu_i)\prod_i f_G(n_i,\mu_i)\prod_i F_{zero}(3,\mu_i)\prod_i F_{sat}(n_i,\mu_i) \tag{3.11}$$

that results in a log-likelihood function

$$\ell = \sum_i \ln f_P(n_i,\mu_i) + \sum_i \ln f_G(n_i,\mu_i) + \sum_i \ln F_{zero}(3,\mu_i) + \sum_i \ln F_{sat}(n_i,\mu_i) \tag{3.12}$$

that it is then maximized in the fitting routine.

Two important observables come out as estimates from the fit. The first is the position of the shower core on the detector plane, and the second is the value of S_{opt}, as introduced in (3.5). Since the value of r_{opt} has been established to be 450 m [86,87], the signal measured at that distance is $S_{opt} \equiv S(450\text{ m}) \equiv S_{450}$, which is utilized as the energy estimator of the event.

3.4.3. Constant intensity cut and energy calibration

The steps required to obtain an energy estimate for each shower detected by the infill array are identical to those employed in the main array of the SD (section 2.1.4). That is, the signal obtained at 450 m from the shower axis is corrected for its dependence on the zenith angle, by means of the attenuation curve derived with the constant intensity cut method. The conversion from the corrected energy estimator to the primary energy is then obtained from hybrid events. A complete treatment of these two parts of the procedure, are given in chapter 7.

3.5. Conclusions

In this chapter the main characteristics of the infill array were introduced. Particular emphasis was given to the trigger definitions used in the array, since the calculation

of the exposure of the apparatus and its related uncertainties largely depend on such definitions. The event reconstruction algorithms were also detailed in order to give a more complete picture of the steps necessary to derive the energy spectrum using data from the array. As already mentioned, both the trigger definitions and event reconstruction used in the infill array are largely based on those employed in the main array of the surface detector.

Chapter 4

The lateral distribution of particles

The distribution of particles on a plane perpendicular to the shower axis is known as the lateral profile, and it is modeled with a lateral distribution function (LDF). The water Cherenkov detectors (WCD) of the infill array of the Pierre Auger Observatory only sample the lateral distribution of showers at a limited number of points at different distances from the shower axis. The signals recorded at the WCDs are fitted using an assumed LDF to evaluate the position of the shower core on the detection plane and the signal at an optimal distance, r_{opt}. A correct estimation of such observables is critical since the signal at the optimal distance is sensitive to the energy of the primary cosmic ray that initiated the EAS [21]. r_{opt} is such that fluctuations in the expected signal, due to lack of knowledge of the true lateral distribution, are minimized. The choice of the LDF or the energy or zenith angle of the detected shower have little influence on r_{opt}: it is the array geometry that primarily determines it [85]. Studies carried out using data from the infill array (e.g., [86]) and Monte Carlo simulations (e.g., [87]) concluded that the optimal value of r_{opt} for the infill array is about 450 m.

In spite of its limited impact on the signal at r_{opt}, the LDF has to be chosen in order to describe at best the observed lateral distribution. An accurate parametrization of the LDF is necessary and must be studied by means of data. Within the Auger collaboration two functional forms for the LDF have been considered for data from the infill array: a NKG-like function and a log-log parabola (LLP) function. In the first part of this chapter, the two forms are presented. The NGK-like LDF is extensively studied and discussed (as this is the one officially adopted by the collaboration) in particular for what concerns the parametrization of the slope. In the second part, the performances of the two LDFs are presented, showing their agreement with data recorded in the infill array. This is done by studying the residuals of the LDF fit and the analysis of the average curves of the experimental lateral distribution.

The data set employed includes T5 events recorded by the infill array from August 2008 to February 2012, excluding bad periods of data acquisition, 'fake T5'

Parameter	Value	Uncertainty
a_0	-2.91	0.02
a_1	-0.94	0.15
a_2	5.97	0.30

Table 4.1: The values of the constants estimated for the parameterization (4.2). Values taken from [91].

events (section 6.3.1) and a subset of events collected during the communications crisis.

4.1. The lateral distribution functions

In this section, the NKG-like function and its parametrization is studied. As the LLP LDF is used too in the last part of the chapter, it is briefly outlined here for convenience.

The **LLP function** takes the form given by

$$S(r) = \begin{cases} S_{450} \left(\frac{r}{r_{450}}\right)^{\beta - \frac{2}{5}\ln\frac{r}{r_{450}}} & r \geq 150 \text{ m} \\ S_{450} \frac{\left(\frac{r}{r_{150}}\right)^{\beta - \frac{4}{5}\ln\frac{r}{r_{450}}}}{\left(\frac{r_{450}}{r_{150}}\right)^{\beta - \frac{2}{5}\ln\frac{r}{r_{450}}}} & r < 150 \text{ m} \end{cases} \tag{4.1}$$

where $r_{150} = 150$ m, $r_{450} = 450$ m. The slope parameter β is related to the shower age, and it has been parameterized as

$$\beta(\theta) = a_0 + a_1(1 - \cos\theta) + a_2(1 - \cos\theta)^2, \tag{4.2}$$

where θ represents the zenith angle in local coordinates. In this parameterization the dependence on the energy has been neglected. The estimation of the constants in (4.2) was performed using a subset of events of high station multiplicity. A fit has been applied to the correlation of the slope parameters, found by the event reconstruction, with the zenith angle of the incoming showers. Table 4.1 contains the estimated values for the constants. For further details on the parameterization of the LLP slope, see for example [90, 91].

4.1.1. The NKG-like LDF

One of the functional forms considered for the LDF is the modified Nishimura-Kamata-Greisen, that can be written as

$$S(r) = S_{450} \left(\frac{r}{r_{450}}\right)^{\beta} \left(\frac{r + r_{700}}{r_{450} + r_{700}}\right)^{\beta}, \tag{4.3}$$

where $r_{450} = 450$ m, $r_{700} = 700$ m and the slope parameter β is set using the following parameterization [92]

$$\beta(S_{450}, \theta) = a_0 + a_1 \log S_{450} + (b_0 + b_1 \log S_{450}) \sec\theta + (c_0 + c_1 \log S_{450}) \sec^2\theta, \tag{4.4}$$

where θ is the zenith angle in local coordinates and S_{450} is the signal estimated at 450 m from the shower axis. The vector of constants $(a_0\ a_1\ b_0\ b_1\ c_0\ c_1)$ will be inferred from data in the following section. Note that the function (4.3) is such that the size parameter S_{450} corresponds to the predicted signal at 450 m.

The slope parameterization

The age of a shower has a significant influence on the value of the slope parameter β. As showers get older, its electromagnetic component becomes fainter and the muonic component starts to dominate the lateral distribution, particularly when the distance to the shower axis increases. The relatively small interaction probability of muons with molecules in the atmosphere, compared to that of electrons and gamma rays, explain this effect. The differences in the development of these components result in flatter shower profiles for older showers. The parameterization in (4.4) roughly models this dependency.

The main reason for using a parameterization like (4.4) is easy to understand when considering the energy spectrum. The fact that cosmic rays follow an energy spectrum that is well described by a steep inverse power law implies that the number of events produced by low energy cosmic rays will be much larger than the number of events produced by cosmic rays of high energy. As a consequence, a low station multiplicity in the majority of the recorded events is expected. For every event, the number of parameters to be estimated by the reconstruction procedure is three, namely, the core location on the detection plane and the size parameter of the LDF, i.e. S_{450}. Therefore, a parameterization of the slope coefficient is needed for events with less than five stations, that represent the majority of the EAS triggering the array. There are other reasons to use a parameterization for the slope β. When the particle density in a shower produces a signal that is close to the detector threshold, upward fluctuations around this value will be preferentially selected and, as a consequence, the measured LDF will be flatter than the true LDF. When considering stations with high signals, close to the saturation of FADCs and PMTs, a similar problem arises. In this case, downward fluctuations in the signal values are favored and the measured LDF will be flatter than the true LDF, once again. Using a fixed slope value in the LDF fit avoids these inconveniences.

In order to estimate the values of the vector of constants $(a_0\ a_1\ b_0\ b_1\ c_0\ c_1)$ in (4.4), a subset of events satisfying a number of constraints has been selected. For this subset of events, the fit procedure of the lateral profile was applied with β as a free parameter (that is, not determined by a parameterization). The constraints imposed ensure that selected events provide a lever arm for fitting the slope parameter β. An event is included in the subset if one or more of the following requirements are verified (these were adapted from those for the regular 1500 m array [93]):

- Two or more stations are located between 225 m and 675 m away from the shower axis, and at least a pair of them is separated by more than 225 m.

- Three or more stations are located between 225 m and 675 m away from the shower axis, and at least a pair of them is separated by more than 180 m.

Cut	Population size
T5	465992
Lever arm req.	139283
$S_{35} > 25$ VEM	12430
$\theta < 55^\circ$	8236
No saturated stations	7805
$\chi^2/N_d < 7$	7517
$\sigma_{core} < 75$ m	7454

Table 4.2: The impact that different cuts have on the population of events.

- Four or more stations are located between 225 m and 675 m away from the shower axis, and at least a pair of them is separated by more than 135 m.

When these conditions are imposed, the original sample of T5 events in the infill array is reduced from 465992 events to 139283 events. After the LDF fit, with β as a free parameter, has been performed on this subset of events, further quality cuts are applied:

- The event must fall in the regime of full trigger efficiency for the detector, that is, the estimated values should satisfy $S_{35} > 25$ VEM and $\theta < 55^\circ$.
- The event must contain no saturated stations.
- The value for the reduced χ^2 of the fit must be less than 7.
- The uncertainty of the position of the estimated shower core given by the fit must be less than 75 m.

The effects these cuts have on the size of the sample are summarized in table 4.2. This set of quality cuts is applied in order to minimize undesired effects that might bias the estimated value of the constant vector $(a_0\ a_1\ b_0\ b_1\ c_0\ c_1)$. Events badly reconstructed (i.e. with high χ^2 values) are a good example of the type of events that should not be included in the fit, or events whose reconstruction might be dominated by stations with low signals. A cut on the shower energy, or S_{35} for that matter, reduces the impact of low energy showers in the parameterization of β. Another example is given by events with saturated stations. If stations with saturated signals are present, a significant fraction of the particle density might be missing from the recorded signal, leading to large uncertainties in the estimated values given by the fit procedure. The uncertainties are such because saturated stations are near the shower core, and should provide the lever arm necessary to constrain the LDF fit. For this reason, events with saturated stations are also excluded from the set. Finally, the number of events that survive all cuts is 7454. Those events were used to estimate the coefficients in (4.4).

Although the applied cuts are intended to extract the population of events that minimizes biases in the estimated values of the parameters, they are by no means perfect. For example, when events with saturated stations are rejected, EAS generated by cosmic rays of the highest energies are systematically excluded from the fit,

because the probability of having a saturated station among the triggered stations increases with primary energy. Nevertheless, the alternative approach, i.e., the inclusion of events with saturated stations, adds a set of badly reconstructed events to the sample and, thus, a compromise has to be taken.

The dependence of β on the estimated value of S_{450} and θ is shown in plots in figure 4.1. Those plots include events with energies lower than the saturation energy for the full trigger efficiency in the infill array (i.e. removing the $S_{35} > 25$ VEM requirement). Below this energy threshold, the trigger probability depends on the arrival direction of the shower and on the intrinsic fluctuations present in the evolution of the components. From figure 4.1a a preference for flatter profiles is observed for events below the trigger saturation threshold. This is more evident for vertical showers. As the mean value of $\log(S_{450}/\text{VEM})$ decreases, and hence the primary energy, the value of β increases (i.e. approaches zero). Therefore, showers with flatter lateral profiles are selected preferentially. There are two effects that explain this behavior. Firstly, showers with flatter lateral profiles, due to their intrinsic fluctuations, are more likely to fulfill the trigger conditions for the array (section 3.3.1), as the area hit by its secondary particles is larger than the area hit by showers with steeper profiles, and the mean energy of those secondary particles is high enough to generate signals that satisfy the local station triggers (section 2.1.2). Secondly, the lever arm requirements previously described are more likely to be fulfilled for the very same reason: showers with flatter profiles are likely to have a higher multiplicity (i.e. more stations with signal) than showers with steeper profiles. The conclusion that showers with flatter profiles are preferentially selected below the saturation threshold is further supported by the change of slope of β as a function of $\sec\theta$, shown in figure 4.1b. For bins in $\log(S_{450}/\text{VEM})$ that are above threshold, the slope of β is apparently the same, as showers arriving closer to the horizon are considered. The case for events below threshold is different. Vertical showers have a mean value of β significantly larger than what is expected from a function with a constant slope, implying that showers with flatter profiles are favored in the population of events. From these considerations, requiring events above the full efficiency for the array is needed in order to avoid such selection biases.

A linear regression has been applied to estimate the values of the constants in the parameterization of the slope. Rewriting formula (4.4) using matrix notation leads to

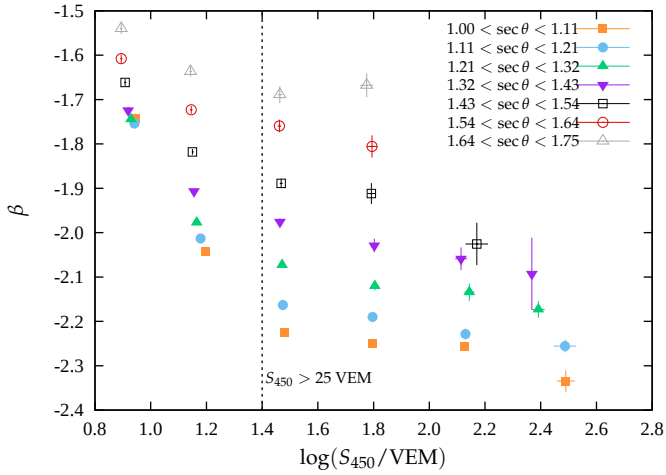
$$B = X\alpha + \epsilon, \tag{4.5}$$

where B is a $n \times 1$ vector of observed values of the slope β of the LDF, $\alpha' = (a_0\, a_1\, b_0\, b_1\, c_0\, c_1)$ is the $1 \times p$ vector of constants to be estimated, ϵ is a $n \times 1$ vector of independent random variables, usually referred as random error, with an expectation of $E[\epsilon_i] = 0$ and variance $V[\epsilon_i] = \sigma_i^2$, and X is a $n \times p$ matrix given by

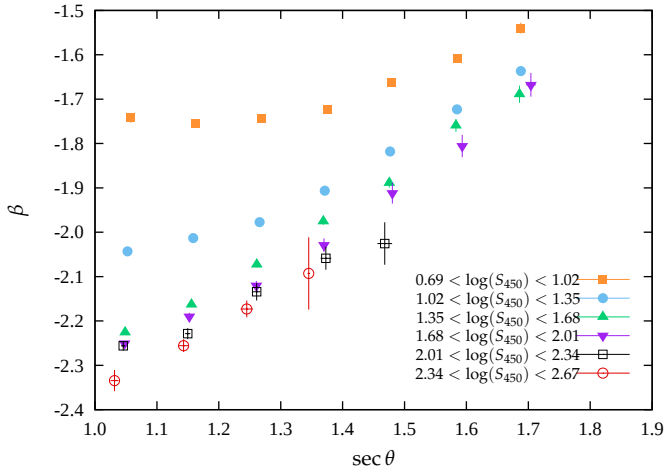
$$X = \begin{pmatrix} 1 & \log S_{450_1} & \sec\theta_1 & \log S_{450_1}\sec\theta_1 & \sec^2\theta_1 & \log S_{450_1}\sec^2\theta_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \log S_{450_n} & \sec\theta_n & \log S_{450_n}\sec\theta_n & \sec^2\theta_n & \log S_{450_n}\sec^2\theta_n \end{pmatrix}. \tag{4.6}$$

In this case $n = 7454$ and $p = 6$. The best linear unbiased estimator for α is given by

$$\hat{\alpha} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}B, \tag{4.7}$$



(a) The slope parameter β as a function of S_{450}.



(b) The slope parameter β as a function of $\sec\theta$

Figure 4.1: The dependence of the slope parameter β on the reconstructed values of S_{450} and θ, for events fulfilling the conditions listed in table 4.2, except $S_{35} > 25$ VEM. Thus events below the full trigger efficiency regime have been included in these plots. Low energy events with steep shower profiles (i.e. β with relatively large absolute values) are less likely to trigger the apparatus and to fulfill the lever arm conditions imposed to fit the slope parameter β. The effect is more evident for vertical showers.

Parameter	Value	Uncertainty
a_0	-3.40	0.37
a_1	0.48	0.22
b_0	1.35	0.56
b_1	-0.77	0.34
c_0	-0.06	0.21
c_1	0.21	0.13

Table 4.3: The values of the constants estimated for the parameterization (4.4).

which has an estimated covariance matrix given by

$$\widehat{\mathrm{cov}}(\hat{\alpha}|X) = (X'\Sigma^{-1}X)^{-1}. \tag{4.8}$$

Here Σ stands for the estimated covariance matrix of measurements, which is diagonal in this case: measurements are independent of each other. Note that Σ itself is an estimator of the true covariance matrix of measurements, which is unknown. It is given by

$$\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2), \tag{4.9}$$

where σ_i^2 is the estimated variance of the i-th value of β. To a first approximation, the value of σ_i^2 is assumed to be

$$\sigma_i^2 = \sigma_{\beta_i}^2, \tag{4.10}$$

where σ_β^2 is given by the event reconstruction procedure. In other words, the errors in the estimates of S_{450} and θ, which are also provided by the reconstruction procedure, were neglected for estimating α: the explanatory variables are treated as fixed entities (i.e. not random). This assumption was verified a posteriori, employing a more accurate approximation of the variance

$$\sigma_i^2 = \sigma_{\beta_i}^2 + \left(\frac{\partial\beta(S_{450_i}, \theta_i)}{\partial S_{450}}\,\sigma_{S_{450_i}}\right)^2 + \left(\frac{\partial\beta(S_{450_i}, \theta_i)}{\partial\theta}\,\sigma_{\theta_i}\right)^2. \tag{4.11}$$

Using (4.11) to approximate the value of the variance, the average change in the value of σ_i^2 with respect to the one obtained using (4.10) is close to 0.4%. Therefore it is reasonable to neglect the contribution of $\sigma_{S_{450}}$ and σ_θ to σ_i.

The estimated values for the constants, given by (4.7), are listed in table 4.3. The fit produces $\chi^2/N_d = 1.21$. The inspection of the fit residuals provide further information on the goodness of fit for the parameterization of $\beta(S_{450}, \theta)$. A residual of the fit is defined to be

$$R_i = \frac{B_i - \beta(S_{450_i}, \theta_i)}{\sigma_i}, \tag{4.12}$$

for the i-th observed value of β. If the model is a good approximation of the behavior of the underlying process, a priori knowledge of the measurement errors is correct and the estimated values for the constants are close to the true values, the residuals should follow a standard normal distribution (that is $\sim N(0,1)$). Plots in figure 4.2 show the fit residuals as function of S_{450} and θ. A small systematic shift from the expected value of the residuals is visible over a large range of S_{450} and θ

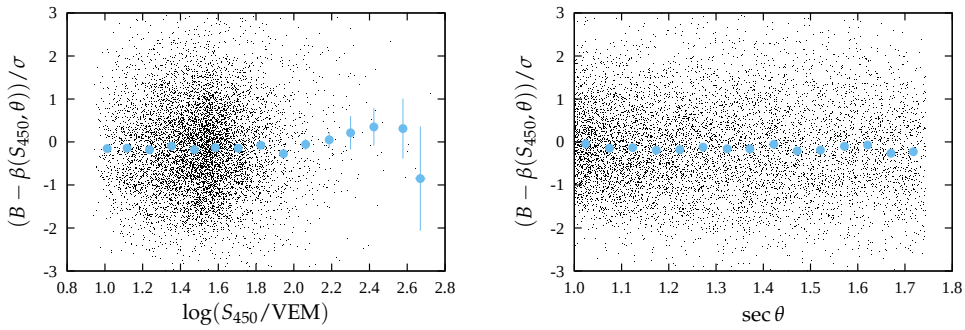


Figure 4.2: The residuals obtained from the fit, using parameterization (4.5), as a function of $\log(S_{450}/\mathrm{VEM})$ (left) and $\sec\theta$ (right). A systematic deviation from the expected value of zero can be seen in both plots. The plot on the left shows a trend to underestimate the values of β for large values of $\log(S_{450}/\mathrm{VEM})$.

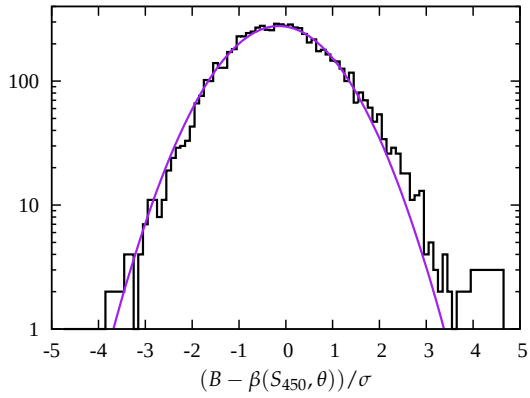


Figure 4.3: A histogram of the fit residuals as defined in (4.12). The values for the sample mean and standard deviation are $\overline{R} = -0.14 \pm 0.01$ and $\mathrm{SD}_R = 1.09 \pm 0.02$, respectively. A fit to a normal pdf is shown as a reference.

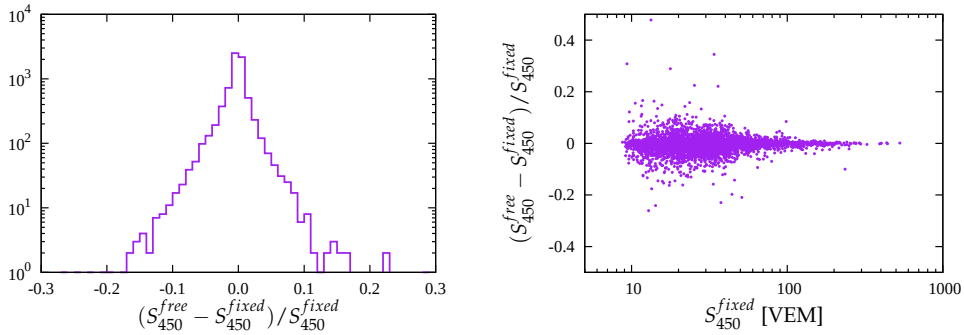


Figure 4.4: Relative differences between S_{450} values obtained for a parameterized β slope (fixed) and estimations of β given by the LDF fit (free). The plot on the left panel shows a histogram of the relative differences built for all events satisfying the required conditions. The mean value of the relative differences is 0.00 with a standard deviation of 0.03. The plot on the right panel shows the relative differences as a function of S_{450} found using a fixed slope β.

values. Moreover, even with a relatively low number of events, a trend to underestimate the values of β for large values of S_{450} is observable. These effects could be caused by the inability of the model to accurately describe available data, but a wrong estimation of the errors of the observed values of β would also cause the distribution of the residuals to deviate from a standard normal distribution. Figure 4.3 shows a histogram of the residuals obtained from the fit, with a normal pdf fitted to the histogram as a reference. Although the histogram resembles a normal pdf, the values for the sample mean and standard deviation are $\overline{R} = -0.14 \pm 0.01$ and $\mathrm{SD}_R = 1.09 \pm 0.02$, significantly different from 0 and 1, respectively. This suggests that the main reason for the deviations are deficiencies in the model to explain data accurately over the possibility of wrong estimations of the errors for β. When errors are badly calculated, the standard deviation of the distribution of residuals in the fit largely deviates from 1, which is not the case here.

The values for S_{450} obtained for showers reconstructed using the parameterization for the slope parameter (4.4) are in good agreement with those reconstructed with β as a free parameter, as can be seen from figure 4.4. The distribution of the relative differences has a mean value of 0.00 and a standard deviation of 0.03. Furthermore, no dependence on the magnitude of S_{450} is found.

In spite of the fact that a number of issues arise when estimating the values of the constants in the β parameterization, the resulting estimates for $(a_0\ a_1\ b_0\ b_1\ c_0\ c_1)$ allow to characterize data in a reasonable manner. In figure 4.5, the parameterization curves are compared to data using bins in $\log(S_{450}/\mathrm{VEM})$ and $\sec\theta$. The curves have been drawn using the average value for the corresponding bin in each plot. As the regression applied was performed using a weighted least squares method, the points in the plots are calculated using a weighted mean for β

$$\overline{\beta}_w = \frac{\sum_i \beta_i \sigma_{\beta_i}^{-2}}{\sum_i \sigma_{\beta_i}^{-2}}. \tag{4.13}$$

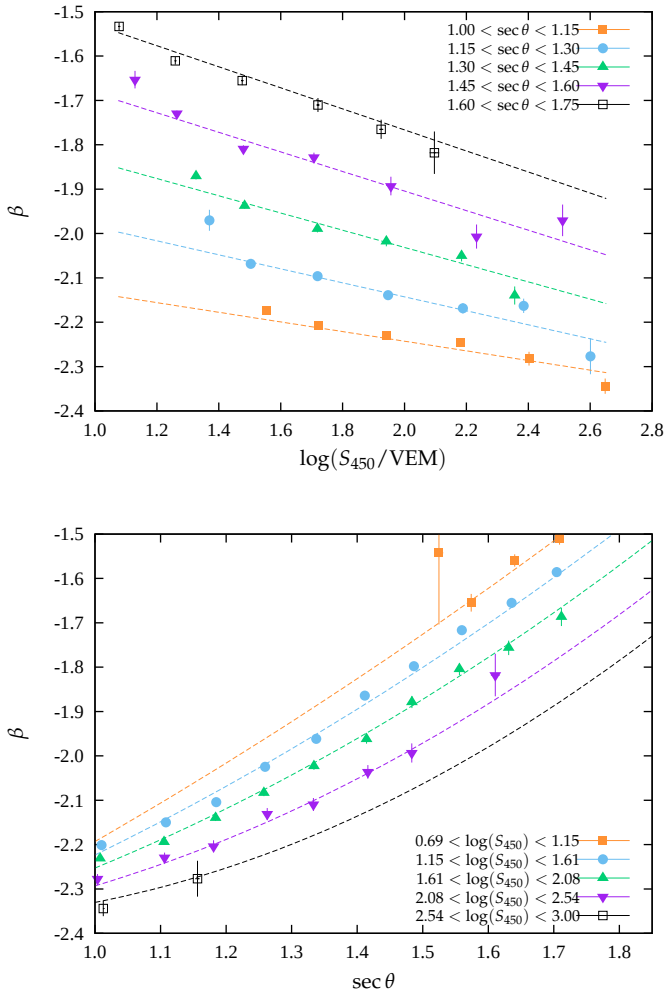


Figure 4.5: The weighted average value of the slope parameter as a function of $\log(S_{450}/\mathrm{VEM})$ and $\sec\theta$. Dashed lines represent the parameterization curve determined for the average value in the corresponding bin.

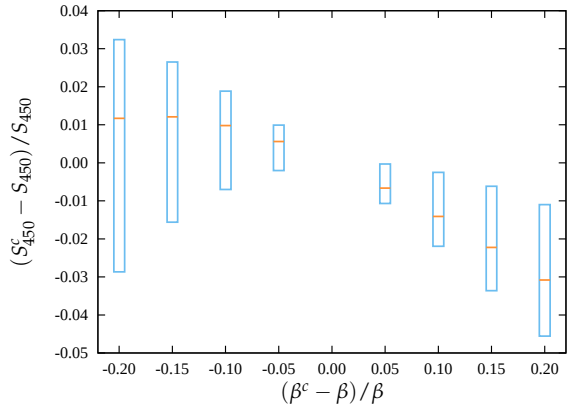


Figure 4.6: The impact of artificially modifying the parameterized value of β on the reconstructed S_{450}. The three quartiles of the distributions obtained are shown as a function of the variations on the slope parameter. Orange lines represent the median value, while the extremes of the box plots represent the lower and upper quartiles.

This illustration shows a good agreement between the parameterization $\beta(S_{450}, \theta)$ and the underlying data set.

Influence on S_{450}

One way to assess the cost of fixing the β slope using a parameterization such as (4.4), is to modify the value predicted for the slope by a fixed relative amount $\delta\beta$ and quantify the induced variation δS_{450} in the reconstructed value of S_{450}. This is easily accomplished by replacing the original parameterized slope β by $\beta^c = \beta \pm \delta\beta$ for each shower that undergoes the reconstruction procedure. The variation $\delta\beta$ is a relative one, and reconstructions for values of $\delta\beta = 0.05\beta, 0.10\beta, 0.15\beta, 0.20\beta$ have been performed, for all reconstructed T5 events collected in the period spanning from August 2008 to April 2011. The ultimate impact these variations have on the energy estimator is rather small, as shown in figure 4.6. In that figure, the quartiles of the distribution of the relative S_{450} differences are shown as a function of the artificial variation introduced in the slope parameter β.

4.2. The performance of the lateral distribution functions

The energy estimation critically relies on the quality of the estimated particle density at a fixed distance from the shower axis. A number of approaches to verify the agreement between infill data and the LDFs (both the NKG-like and LLP functions are considered in the following) are exploited in this section, based on methods employed for the regular 1500 m array of the surface detector [92] . These checks are not intended to quantify the quality of the LDF or to perform a rigorous goodness-of-fit test on the fit, as this would require the treatment of non-triggered stations,

since they are included in the LDF fit procedure.

4.2.1. Analysis of residuals

A useful approach to asses the suitability of a model to describe a data set is to study the residual distribution; that is, the differences between the predicted values given by the model and the actual measurements. In the particular case of fitting an assumed LDF to signals recorded in water Cherenkov detectors, the residual for each station may be defined as

$$R = \frac{S - S_{LDF}(r)}{\sigma_S}, \tag{4.14}$$

where S is the station signal, $S_{LDF}(r)$ is the signal predicted by the fitted LDF and σ is the uncertainty in the measured value of the signal in the station. This signal uncertainty has been found to be $\sigma_S = 0.99\sqrt{S}$ [88] from SD data. If the assumed LDF is a good approximation of the lateral profile of the shower, then the residuals should have a mean value of zero and a standard deviation equal to one; moreover, this should be the case for all distances from the shower axis.

In practice, the required conditions for the distribution of the residuals are not met. One difficulty arises because residuals can only be defined for stations with a measured signal. The local station trigger definitions prevent to assess low particle densities that would produce a signal less than about 3 VEM, and thus it is impossible to verify the agreement between the LDF and the shower profiles for low particle densities. As the density of particles decreases with increasing distance to the axis, this effect plays a major role at high distances, where the local trigger definitions in the stations systematically select upward fluctuations in the particle densities. Thence, an increase in the mean of the residuals is expected for large distances from the axis. As for stations without signal, a residual cannot be defined for saturated stations. Although saturated stations do provide a measured signal, a significant part of the information is missing, and the signal can be used as a lower limit at best. Hence the residual definition in (4.14) is not applicable. By utilizing the signal recovery method [89], it is possible to calculate residuals for saturated stations according to (4.14), as these signals are used in the fitting procedure of the LDF. If this is the case, the signal S in the station and its uncertainty σ_S, indispensable to define a residual, are estimated by the recovery method.

Even though the recovery method generally improves the shower reconstruction procedure, in some cases the estimated signal for saturated stations is far off the predicted value given by the LDF. If these residuals are included, the calculation of the mean and standard deviation of the sample may be strongly biased when stations near the shower core are considered. In order to decrease the impact of these stations, a cut on the recovered signal is imposed: if the recovered signal deviates more than twenty times from the predicted signal given by the LDF, the residual is discarded from the calculations.

Figure 4.7 shows the relative deviation (right panel) and residuals (left panel), defined according to (4.14), as a function of the distance to the shower axis. The predicted rise of the residuals due to the selection bias introduced by the local station

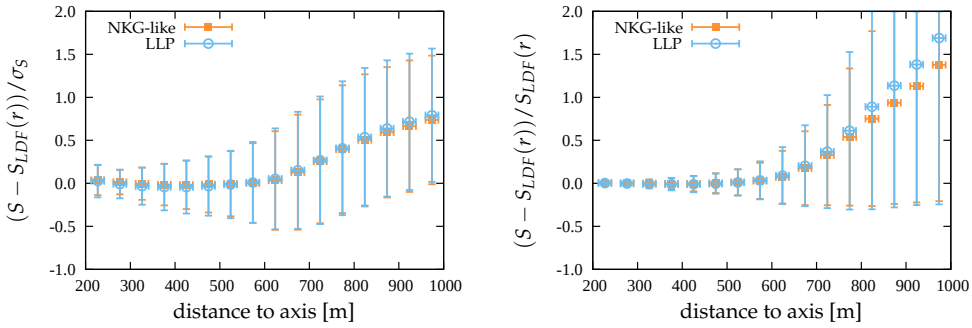


Figure 4.7: On the left panel, the mean of the residuals is plotted as a function of the distance to the shower axis, for both LDFs considered in this work. On the right panel, the mean of the relative deviations as a function of the distance is included. In both plots, error bars represent the standard deviation of the residual sample.

trigger definitions is clearly seen for distances greater than $\sim$ 700 m. The effect is more dramatic when the relative deviation is inspected, because the relative change becomes more important as the predicted signal diminishes.

The analysis of residuals shows a good agreement between data and both LDFs considered in the distance range spanning 200–700 m. For distances larger than 700 m, a discrepancy is found. Note that such discrepancy is caused by limitations of the detector to measure small particle signals, and not by the functional form of the LDFs themselves.

4.2.2. χ^2 statistic

Another useful method for quantifying the degree of agreement between data and the fitted LDF is the reduced χ^2. It allows to directly compare the performance of the LDFs considered. The χ^2 is defined by the squared sum of the residuals

$$\chi^2 = \sum_i R_i^2 = \sum_i \left(\frac{S_i - S_{LDF}(r_i)}{\sigma_{S_i}} \right)^2 . \tag{4.15}$$

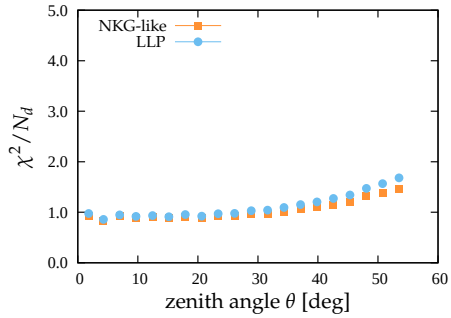
The reduced χ^2 statistic is obtained by simply dividing the χ^2 by the number of degrees of freedom N_d in the model. In the case of the LDF fit, the number of degrees of freedom is given by the number of candidate stations in the event minus the number of parameters determined by the fit, which is three. For the same reasons explained in the case of the residuals, the χ^2/N_d statistic does not take into account the influence of non-triggered stations. As before, some of the signals estimated by the signal recovery method [89] are very far off from the signal value predicted by the LDF. A cut of $\chi^2/N_d < 50$ has been applied to reject events with very high χ^2 values, that would otherwise distort the analysis of the LDF performance. Generally, such events produce high χ^2 values due to the presence of saturated stations whose signals cannot be recovered by the method or might be badly recovered. By imposing this cut in the χ^2/N_d, the size of the original population of events is reduced by 0.6%. Moreover, a cut on the value of $S_{450} > 8$ VEM was imposed to avoid

low energy showers to dominate the sample mean. The population size is reduced by 36% by the application of this cut.

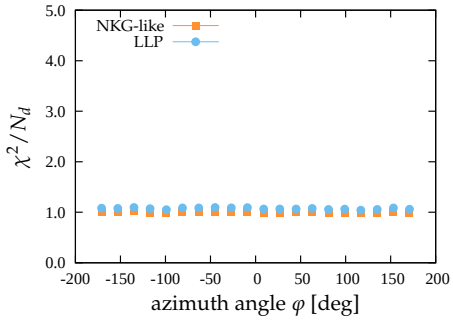
It is instructive to inspect the behavior of the χ^2/N_d statistic as a function of different observables in the events. Several plots are included in figure 4.8 for this purpose. The first of these is shown in figure 4.8a, where the average χ^2/N_d is plotted against the zenith angle, for events reconstructed using the NKG-like and LLP LDFs. As can be seen, the agreement between predicted values of the LDF and measured signals deteriorate as the zenith angle increases, for both functions. This behavior can be explained in terms of signal asymmetries: as showers become more horizontal, a station that lies directly underneath the axis will measure the particle density at an earlier stage in the shower development than a station located above it, even if they are placed at the same distance from the axis. If the maximum of the shower has been reached above ground, the signal measured by the former station will be larger than the signal recorded at the later, because the electromagnetic component will be more attenuated, as the amount of matter traversed is larger. The relative difference between the asymmetric signals depends mostly on the ratio of particles in the electromagnetic and muonic components, which is sensitive to the zenith angle θ. Another effect that contributes to the asymmetries is a geometrical one: stations have different effective areas for different zenith angles. This geometric effect is also present for stations measuring the particle densities of the same EAS, as the effective area varies according to the position with respect to the shower core. Since no correction for such effects is applied during the LDF fit, the signal asymmetries cause that the agreement between the assumed LDF and signals in stations deteriorate for increasing zenith angles. This is reflected in figure 4.8a. Nevertheless, the overall effect is by no means dramatic.

While the effect of the geomagnetic field on very inclined showers is strong, it is usually ignored for showers arriving with zenith angles less than $\sim 60^\circ$. The field bends the trajectories of charged particles, and its influence depends on the arrival direction of the EAS, that is, on its zenith and azimuth angles (θ, φ). The strength of the perpendicular component of the geomagnetic field could break the radial symmetry exhibited in the LDFs, and the performance of the LDF could thus vary as a function of the azimuth angle φ. Figure 4.8b shows the average χ^2/N_d as a function of the estimated value of φ for events reconstructed using both LDFs. From that figure, it is reasonable to conclude that geomagnetic effects have a low impact on the performance of the LDFs, as no significant azimuth modulation can be observed.

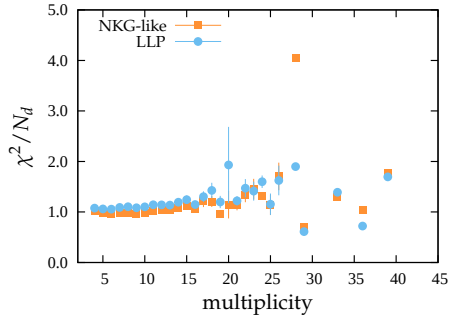
The performance of the LDFs is rather flat as a function of the station multiplicity, as can be seen from figure 4.8c. Nonetheless, it is possible to appreciate a slight trend to higher χ^2/N_d values as the number of triggered stations rises, in particular for the reconstructions based on the LLP function. If figure 4.8d is inspected a similar trend is found, this time as a function of the estimated value of S_{450}. Since the station multiplicity and estimated signal at 450 m are positively correlated (both increase with primary energy), a possible cause of the drift in the χ^2/N_d values for the LLP LDF is that the influence of the primary energy on the steepness of the lateral profile has been ignored in formula (4.2), that models β as a function of the zenith angle θ only. Since the influence of the shower energy on the steepness



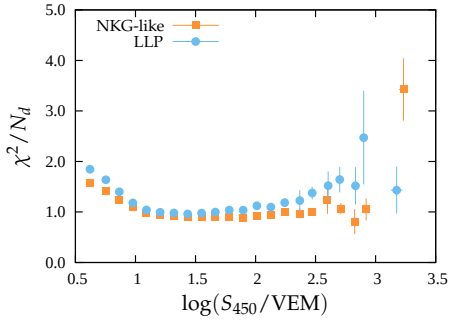
(a) The reduced χ^2 as a function of zenith angle. The degradation of the LDF performance with increasing zenith angle can be attributed to the growth of shower asymmetries.



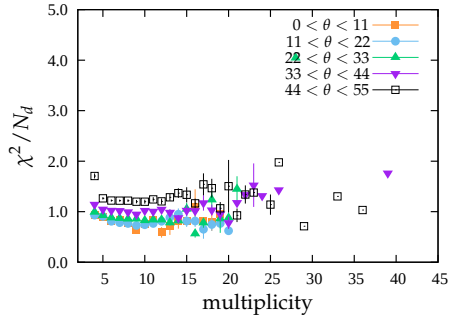
(b) The reduced χ^2 as a function of azimuth angle. The LDF performance is rather flat across azimuth angles.



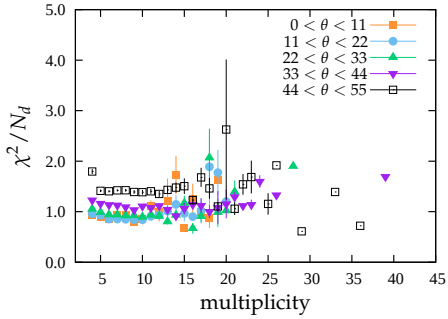
(c) The reduced χ^2 as a function of event multiplicity. A slight deterioration of the agreement can be seen for increasing values of the multiplicity, particularly for the LLP LDF.



(d) The reduced χ^2 as a function of S_{450}. As in the case of the multiplicity, a slight deterioration of the performance can be seen for larger values, particularly for the LLP function.



(e) The reduced χ^2 as a function of event multiplicity for several angular bins for the NKG-like LDF. The behavior is similar across angular bins.



(f) The reduced χ^2 as a function of event multiplicity for several angular bins for the LLP LDF. The behavior is similar across angular bins.

Figure 4.8: The reduced χ^2 statistic as a function of different observables.

of its lateral profile was neglected, it is reasonable to expect a worsening of the LDF performance as more energetic showers are included. However further studies should be carried out to arrive to definite conclusions. Finally, the dependence on station multiplicity has been decoupled for different zenith angle bins for both the NKG-like function (figure 4.8e) and LLP function (figure 4.8f). Besides the average increase of the χ^2/N_d with larger zenith angles, the behavior is found to be similar across zenith angles, for both LDFs. From those last two plots, it is interesting to note that events with highest station multiplicity arrive preferentially from larger zenith angles, because the shower foot print scales with increasing values of θ.

4.2.3. The experimental mean lateral distribution

Lastly, for demonstration purposes, the experimental mean lateral distribution of showers detected with the infill array is derived. With this aim, data have been divided in zenith angle bins in such a way that showers that belong to the same bin have a difference in the amount of matter traversed no larger than a radiation length. In this manner, any two showers with similar S_{450} values would have been generated by primary particles of roughly the same energy. The edges of the zenith angle bins are determined by means of the slant depth $X_s(h,\theta)$ using the following formula

$$X_s(h,\theta) = X(h,\theta=0)\sec\theta = X_0\, e^{-h/h_s}\sec\theta, \tag{4.16}$$

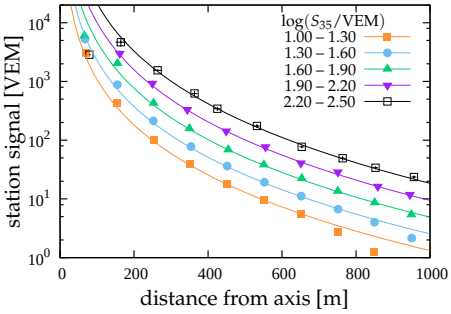
which does not take into account the curvature of the Earth. Here h is the observation level of the detector, h_s is the scale height of the atmosphere and X_0 is the vertical column density at sea level. Hence (4.16) allows to write

$$e_i^\theta = \arccos\frac{X_s(h,\theta=0)}{X_s(h,\theta=0)+iR} \tag{4.17}$$

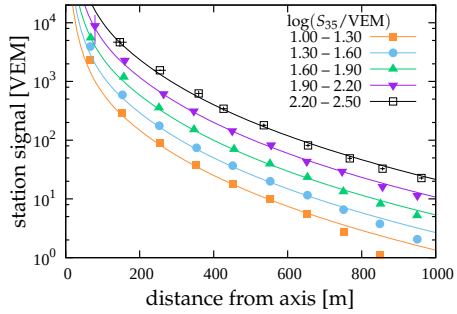
for the values of the bin edges, where e_i^θ is the value of the i-th bin edge and R stands for the radiation length of electrons in air. The observation level h was set to 1400 m, the vertical column density X_0 is 1030 g/cm^2 and the radiation length R for electrons to 36 g/cm^2.

For determining the average lateral distribution, the signal for non-triggered stations was set to zero and, therefore, a significant presence of these stations at any distance range would have an impact on the signal average, as it would be clearly below the predicted value given by the mean LDF. Stations with saturated signals have also been included, regardless of the quality of the estimated signal given by the recovery method, or even if the method failed to recover the signal. By doing so, the effect of saturated stations as a function of the distance might be visualized in a similar fashion as for non-triggered stations: significant deviations of the predicted values given by the LDF could be identified.

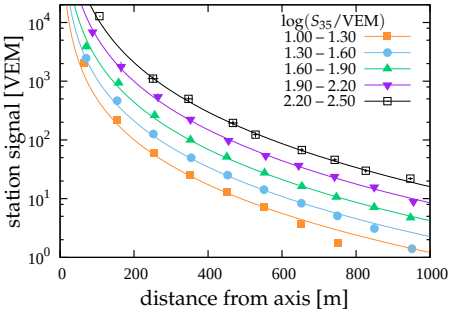
Plots for three angular bins are shown in figure 4.9: $\theta < 15^\circ$, $36^\circ < \theta < 38^\circ$ and $53^\circ < \theta < 55^\circ$. For each angular bin and choice of LDF, showers have been divided according to their size, using the estimated value of the S_{35} signal (section 7.1). The lines represent the LDF with the average values of its size parameter S_{450} and slope β, determined for all showers belonging to the size bin. Stations have been grouped



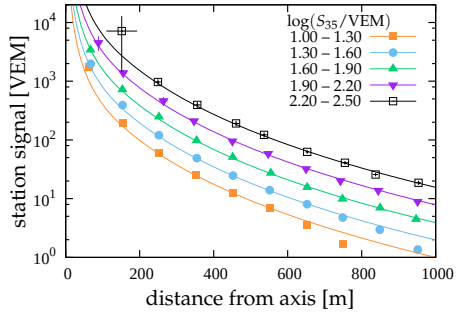
(a) NKG-like LDF and $\theta < 15\,^\circ$



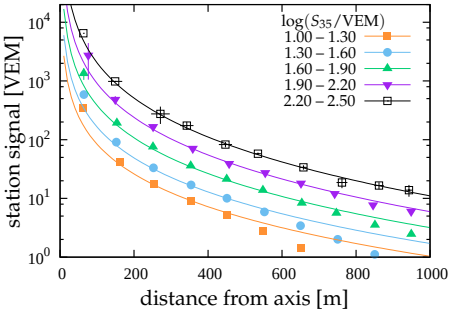
(b) LLP LDF and $\theta < 15\,^\circ$



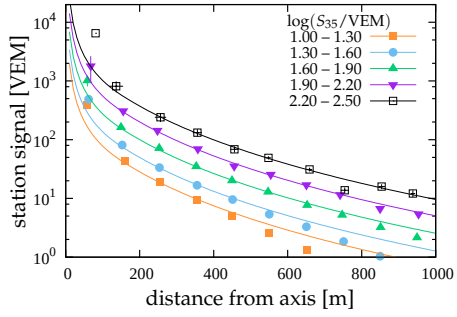
(c) NKG-like LDF and $36\,^\circ < \theta < 38\,^\circ$



(d) LLP LDF and $36\,^\circ < \theta < 38\,^\circ$



(e) NKG-like LDF and $53\,^\circ < \theta < 55\,^\circ$



(f) LLP LDF and $53\,^\circ < \theta < 55\,^\circ$

Figure 4.9: The mean lateral distributions for three angular bins and two different LDFs. Points represent signal averages, whereas lines were determined using the corresponding LDF with average values of S_{450} and β in the corresponding size interval.

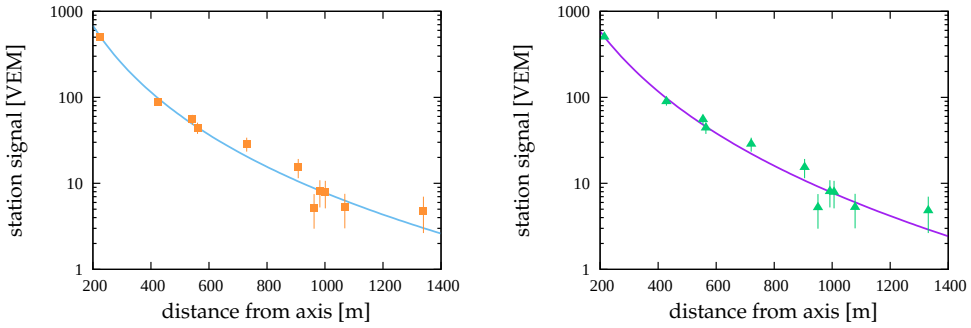


Figure 4.10: An example of an event recorded in the infill array, arriving with a zenith angle of $\theta = 44°$. Points represent the signal as measured in detectors, while the solid line represent the applied fit using an assumed LDF. The event was reconstructed using the NKG-like LDF (left) and the LLP LDF (right). The reconstructed core positions differ by 12 m, while the estimated signals are $S_{450}^{NKG} = 83$ VEM and $S_{450}^{LLP} = 86$ VEM.

in distance bins of 100 m in length, using the average distance in the bin instead of its center.

It can be seen from all plots in figure 4.9 that data are well represented by both LDFs, in the distance range extending from 200 m to 500 m, for any and all angular bins, even if events below the full trigger saturation are included (i.e. events with $\log(S_{35}/\text{VEM}) < 1.3$ in the $53° < \theta < 55°$ zenith bin). As more energetic showers are considered, the agreement at large distances from the axis improves. This effect is caused by the influence non-triggered stations have on the average signal, because the fraction of such stations increases with the distance from the axis. Nevertheless their effect on the average lateral distribution becomes apparent only around 500 m in the worst case, which is to say, for low energy showers in the most horizontal bins. On the other hand, the influence of saturated stations in the average signal becomes important only for stations closer than 200 m. In this case, the effect is more evident if vertical showers are inspected. The attenuation of the electromagnetic component is the cause of such effect, since showers arriving from larger zenith angles are older on average, and therefore it is possible to measure the lateral profile closer to the axis, as the probability of having saturated stations diminishes.

4.3. Conclusions

After verifying the performance of the assumed LDFs (i.e. NKG-like and LLP) on data recorded in the infill array, a good agreement was found in both cases, particularly in the distance range where the energy estimator S_{450} is determined. For larger distances, the selection bias introduced by the trigger definitions in the detectors cause a discrepancy between data and the LDFs. The effect is not related to the functional forms of the LDFs but to the limitations of the instrument employed.

Both LDFs describe data well. An illustrative example is shown in figure 4.10. An event recorded in the infill array was reconstructed using the NKG-like LDF (left panel) and the LLP LDF (right panel). Note that, as expected, the impact of the

choice of the LDF on the signal at 450 m is negligible.

The analysis of the residuals and the χ^2 statistics show that overall the NKG-like function performs slightly better than the LLP. For this reason, the NGK-like form will be used in the event reconstruction in the analyses done in the rest of this work.

Chapter 5

Infill array reconstruction: comparison with regular array and hybrid events

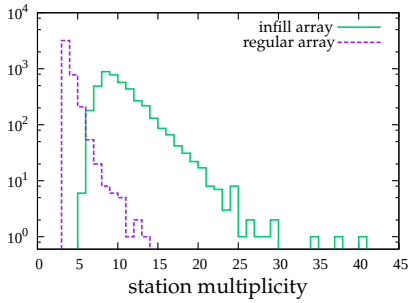
The water Cherenkov detectors used in the infill array are identical to those used in the regular array. Moreover, the infill is embedded in the regular array and, therefore, the same selection and reconstruction strategies may be applied to observed showers. The trigger efficiency, exposure calculation, event selection, geometry reconstruction and lateral distribution functions used (LDFs), as well as the energy estimator and the energy calibration, all benefit from algorithms tested successfully over the past years for the regular surface detector. Besides the uniform treatment of data from the highest energies down to 3×10^{17} eV, an additional advantage of having the infill within the main array is that it is possible to make cross-checks on the reconstruction of showers detected in the overlapping region by both arrays. In the first part of this chapter, a comparison between arrival direction, core position and energy estimator of events detected at the same time by the infill and regular arrays is presented. Further cross-checks of the event reconstruction in the infill array can be performed by exploiting the so-called golden hybrid events, i.e, events that have triggered simultaneously the infill array and the fluorescence detectors. This event sample is used in the second part of the chapter for a study of the performances of the infill geometrical reconstruction (arrival direction and core position). The reconstruction of events on the regular array and hybrid events was performed using the Offline framework [55].

5.1. Comparison between infill and regular events

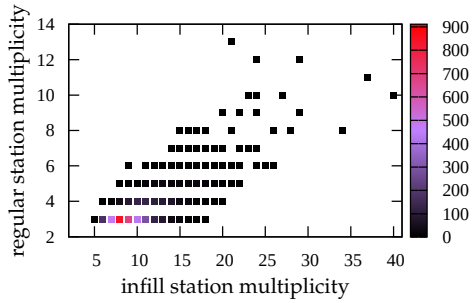
The data set includes all showers satisfying the physics trigger (T4) and the fiducial trigger in the infill array (T5), with the same requirements imposed on the main array. Events have been included starting from August 2008 to February 2012, excluding bad periods of data acquisition and fake T5 events (see section 6.3.1). Only

Cut	Population size
All common events	16298
Fiducial cut in the infill array (T5)	5651
$\theta^I < 55°$	5264
$S_{35} > 20$ VEM	4619
Bad period & fake T5 rejection	4267

Table 5.1: The impact of the cuts on the population size, for events recorded using both infill and regular arrays.



(a) The distribution of the station multiplicity for the infill and regular array.



(b) The regular array multiplicity versus the infill array one. The color scale denotes the density of events.

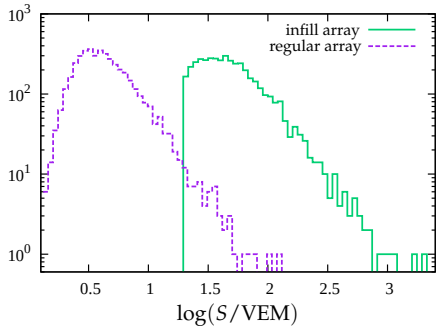
Figure 5.1: The multiplicity of events recorded using the regular and infill arrays.

showers with reconstructed zenith angles below 55° and energy estimator S_{35} above 20 VEM are considered, i.e., only events in the infill trigger-saturation regime (see chapter 6). The effects these cuts have on the population size of showers are summarized in table 5.1. The number of surviving common events is 4267.

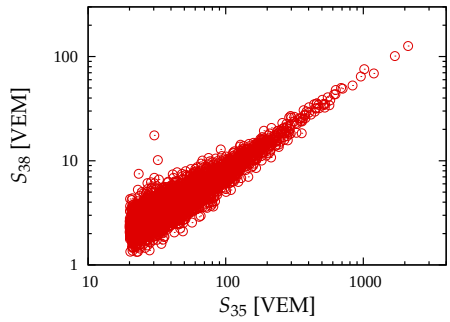
Due to the different structure of the two arrays, the common data set has different characteristics when it is inspected from the point of view of one of the two. It is important to keep those differences in mind in the rest of the section.

The first difference is the obvious result of performing a measurement of the lateral profile of showers with two arrays of different spacing: the event multiplicity (the number of triggered stations) is larger in the denser array. This fact can be appreciated from figure 5.1a, where the distribution of station multiplicity is shown for infill and regular arrays. As expected, the mean multiplicity for the infill array is larger than the mean multiplicity for the regular one, the latter being dominated by 3-fold events. To further illustrate that, the multiplicity of events recorded with the regular array is plotted against the multiplicity for the infill one, in figure 5.1b, where the number of events per point is denoted using a color scale. As the performances of the event reconstruction (arrival direction and LDF fit) depend on the station multiplicity, the larger multiplicity of the infill array implies a better accuracy in the reconstruction of common events.

The second difference is in the distribution of the signals used as energy estima-



(a) Histograms of the values of S_{35} (infill array) and S_{38} (regular array).



(b) Correlation of S_{38} and S_{35}.

Figure 5.2: Energy estimators for showers recorded by both arrays, as provided by their respective reconstructions.

tors. The distributions of S_{35} and S_{38} (the zenith-corrected energy estimator for the regular array, see chapter 2) are shown for the common data set in figure 5.2 (left). Due to the described cut in S_{35} (corresponding to $\sim 3 \times 10^{17}$ eV) and due to the steep spectrum of cosmic rays, most of the common events correspond to energies well below the trigger-saturation for the regular array (corresponding to $S_{38} \approx 15$ VEM). As will be seen below, that must be taken into account in the comparison of the energy estimators determined by the two arrays.

5.1.1. Comparison of the arrival directions

Both in the infill and in the regular arrays, the arrival direction of showers is estimated exploiting the timing information of the stations triggered by secondary particles at ground. In both cases, the angular resolution depends on the event multiplicity and zenith angle (see chapters 2 and 3).

In this section, the arrival directions reconstructed by the two arrays are compared. To identify the possible existence of biases in the reconstructions, an approach based on a linear regression was taken, for the zenith and azimuth angles separately. As the event multiplicity for the infill array is larger than for the regular one (see fig. 5.1b), the angular resolution of the infill array is better than that of the regular one. Consequently, the zenith angle θ^R, estimated by the reconstruction procedure of the regular array, is regarded as a linear function of the estimated value θ^I, obtained from the reconstruction of the same event in the infill array

$$\theta^R = a_\theta + b_\theta \theta^I + \epsilon_\theta, \tag{5.1}$$

where ϵ_θ is the usual disturbance term. For this model, the uncertainties on the estimates are neglected. Note that there is a correlation between θ^I and ϵ_θ, since the angular resolution is not independent of the zenith angle. Given the minimum multiplicity for infill events in the data set (five stations) the maximal difference in angular resolution between vertical and inclined events is 0.2° (see figure 3.11). Such small dependence is neglected in the linear regression, implying that the estimates

for the standard errors of the constants will be biased. Nevertheless, the estimates of the constants themselves remain unbiased. Instead, a source of bias in the estimated values of a_θ and b_θ is introduced by the cut $\theta^I < 55°$: this systematically includes downward fluctuations in the zenith angle estimates for values of $\theta > 55°$, while at the same time, excludes upward fluctuations for values $\theta < 55°$. This effect is also neglected, given the relatively high resolution of the infill array for inclined events (close to $1°$ for 5 or more stations, as shown in figure 3.11).

The case for the azimuth angle φ is identical. The values obtained from both reconstructions are assumed to follow

$$\varphi^R = a_\varphi + b_\varphi \varphi^I + \epsilon_\varphi, \tag{5.2}$$

and the constants a_φ and b_φ are estimated by means of a linear regression. The independence of the azimuth angle and the disturbance term seems more plausible for postulating (5.2), since the azimuth angle has a negligible impact, if not null, on the angular resolution, and thus φ^I can be regarded as independent of ϵ_φ.

If the reconstruction of the arrival direction in the regular array is consistent with the reconstruction in the infill array, the values of the constants $a_{\theta,\varphi}$ and $b_{\theta,\varphi}$ in (5.1) and (5.2) should be 0 and 1, respectively. For this reason, it is expected that $\hat{a}_{\theta,\varphi}$ be compatible with 0 and $\hat{b}_{\theta,\varphi}$ be compatible with 1, for both the zenith and azimuth angle estimations.

The correlation of reconstructed zenith and azimuth angles is shown in figures 5.3a and 5.3b, respectively. Dots in those plots represent the reconstructed values for the angles, while the solid line represents the curve given by the fit procedure. The fraction of variance explained is defined as

$$R^2 = \frac{\mathrm{var}(\hat{a}_\theta + \hat{b}_\theta \theta^I)}{\mathrm{var}(\theta^R)} \tag{5.3}$$

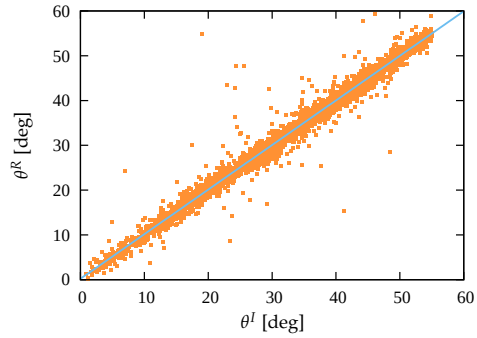
in the case of the zenith angle. Analogous definitions are used in all other regressions, and it is used as a measure of the goodness of fit.

The angular difference ψ in arrival directions is calculated according to

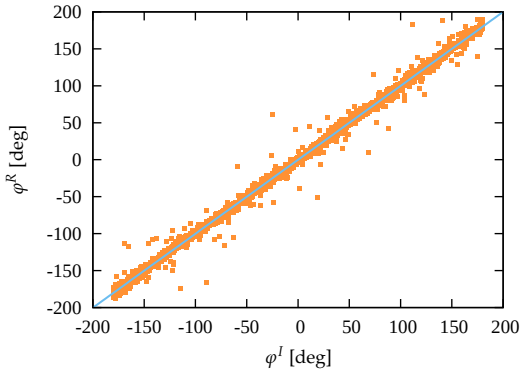
$$\psi = \arccos\left[\cos\theta^R \cos\theta^I + \sin\theta^R \sin\theta^I \cos(\varphi^R - \varphi^I)\right]. \tag{5.4}$$

The histogram of ψ values resulting from the common data set is shown in figure 5.3c. The arrow signals the quantile of the distribution of ψ for 68% of the events, that corresponds to $1.5°$. For the vast majority of events ($\sim$ 98%), the arrival direction determined by the regular array reconstruction lies within $5°$ of the value given by the infill reconstruction. Events with differences above $10°$ have been checked one by one, and it has been found that the problem originates in the regular array. Most of the discrepancies are caused by casual stations reporting a signal induced by a lonely muon, just before the shower signal is detected. As a consequence the arrival direction is badly estimated in the main array, due to the low station multiplicity available for reconstruction.

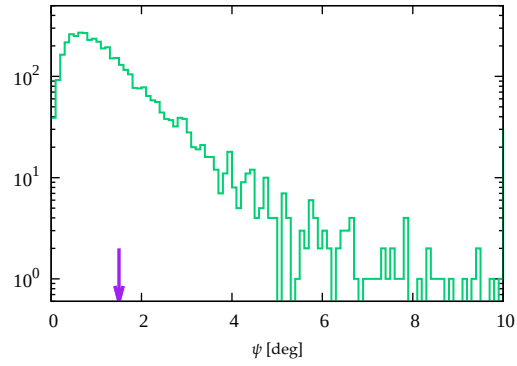
The values of the constants in (5.1) and (5.2) given by the fit procedure are listed in table 5.2. As can be seen, the estimated values of $a_{\theta,\varphi}$ and $b_{\theta,\varphi}$ are very close to 0 and 1, respectively, for both angular coordinates. In spite the fact that the error



(a) The correlation of reconstructed zenith angles. The solid line represents the applied fit.



(b) The correlation of reconstructed azimuth angles. The solid line represents the applied fit.



(c) Histogram of the values obtained for the angular difference ψ between the arrival direction estimations. The arrow points to the 68% quantile, corresponding to 1.5°.

Figure 5.3: Comparison of the arrival direction as determined by the regular array reconstruction and the infill reconstruction.

Parameter	Value	Unc.
a_θ [deg]	0.21	0.08
b_θ	1.00	0.01
a_φ [deg]	-0.11	0.09
b_φ	1.00	0.01

Table 5.2: The values of the estimated constants for (5.1) and (5.2). The fraction of variance explained in the case of the zenith angle is $R^2 = 0.98$, while for the azimuth angle is $R^2 = 1.00$.

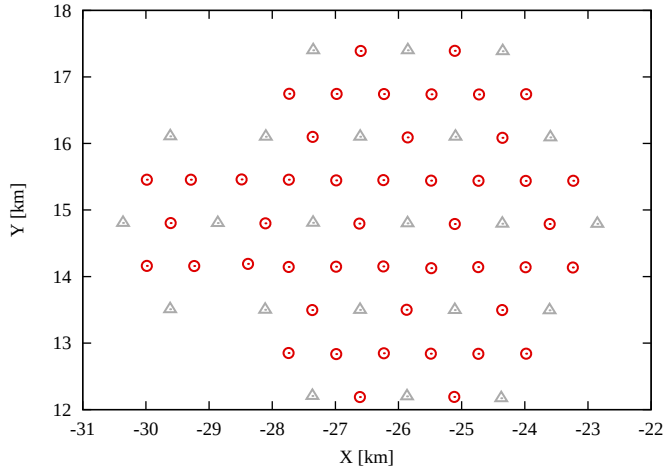


Figure 5.4: The position of the stations in the infill array in the coordinate system used to compare the estimated core positions given by different reconstructions. Red circles represent stations that are part of the infill array only, while gray triangles represent stations that are shared between the infill and the main arrays.

estimates are not dependable, as the angular resolution is regarded as constant for all zenith angles, the values of the slope b in both cases are compatible with 1. The intercept a values show a small bias for the zenith and azimuth angles, well within the angular resolution of both instruments (note that events recorded by the regular array are mostly 3-fold, for which the angular resolution is around 2°).

The results given by this analysis show no significant differences between the geometrical reconstructions of the two arrays. The infill and regular arrays are rather consistent in terms of the determination of the arrival direction.

5.1.2. Comparison of the core positions

The determination of the core position in the surface detector relies on the fit procedure that estimates the LDF used to describe the lateral profile of each shower (see chapter 4). A bad estimation of the core position can induce a poor determination of the energy estimator. Thence, it is interesting to compare the estimated positions, given by regular and infill array separately, looking for anomalies that may hint to problems in the reconstruction procedures, specifically in the fit of the lateral distribution of showers. The position of the stations in the coordinate system used to

Parameter	Value	Unc.
a_X [m]	53	41
b_X	1.00	0.01
a_Y [m]	-43	30
b_Y	1.00	0.01

Table 5.3: The values of the estimated constants in (5.5), for both Cartesian components of the core position on the detection plane. The fraction of variance explained in both cases is $R^2 = 0.99$.

compare the estimated core positions is shown in figure 5.4.

The approach taken is similar to the one used in the comparison of the arrival directions. The core position on the detector plane is determined by its two Cartesian coordinates (X, Y) and, therefore, the model used to describe the relation between the coordinates of the shower core estimated by the regular reconstruction and the infill reconstruction is

$$Q^R = a_Q + b_Q Q^I + \epsilon_Q, \tag{5.5}$$

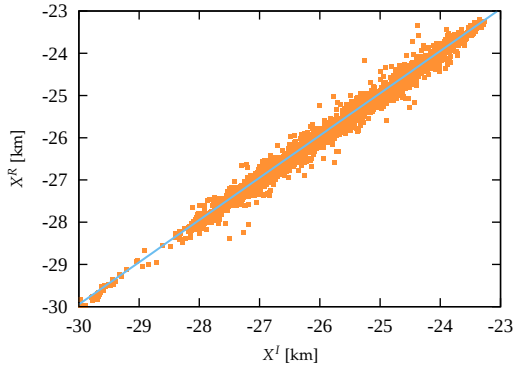
where Q is the Cartesian coordinate (i.e. X or Y) and the superscript specifies the estimated value given by the regular (R) or infill (I) reconstruction. As in the case of the arrival directions, the uncertainties in the values of the estimates are neglected. Moreover, given the regularity of the grid in the main and infill arrays, it seems reasonable to assume that the resolution on each coordinate is independent of its actual value: that is to say that the exogeneity assumption in the fit holds.

The correlation of the estimated X values is shown in figure 5.5a, while figure 5.5b depicts the correlation between the estimated values of Y. The higher density of points in certain regions of those figures is caused by the larger number of showers detected in some areas of the infill array. This is due either to the non-uniform location of the stations (see figure 5.4), or to the fact that data acquisition started before the completion of the array, implying that some areas of the instrument have a larger exposure than others.

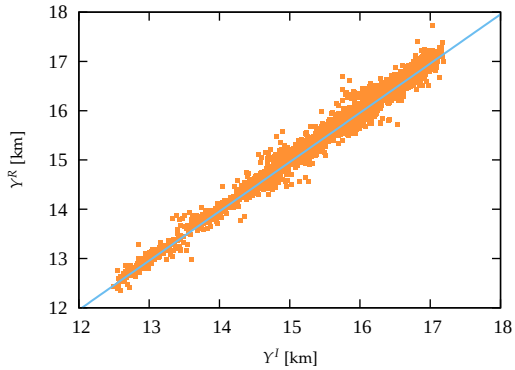
If the reconstruction procedures for the infill and regular arrays are consistent, the values of a_Q and b_Q in (5.5) should be 0 and 1, respectively, and the estimates provided by the fit are expected to be compatible with those values. Such estimates are listed in table 5.3: the intercept values are mildly incompatible with zero, their values being however within the uncertainties of the two arrays (note once again that most events from the regular array are 3-fold, for which the core position uncertainty is around 100 m).

A histogram has been built using the distances between the core estimates given by the infill reconstruction and the regular reconstruction, shown in figure 5.5c. 68% of the events studied have a difference in core position less than 150 m, as signaled by the arrow included in the plot.

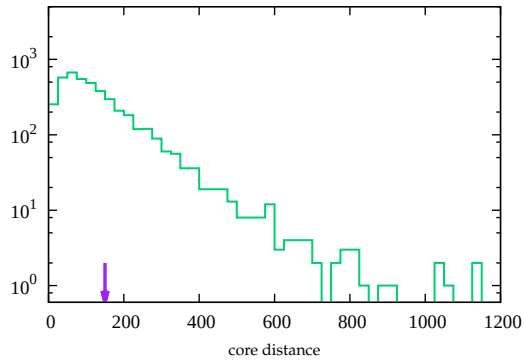
As for the arrival directions, the reconstructions from the two arrays show a relatively good agreement in terms of the core positions determination.



(a) The correlation between the X coordinates of the core positions, as obtained from the regular and infill reconstructions. The solid line represents the applied fit.

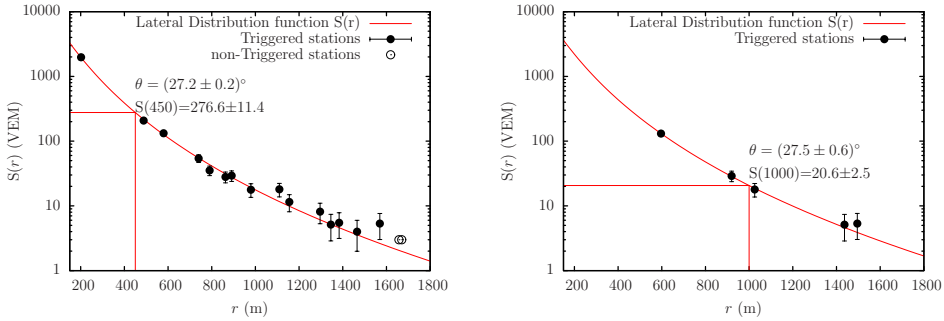


(b) The correlation between the Y coordinates of the core positions, as obtained from the regular and infill reconstructions. The solid line represents the applied fit.



(c) Histogram of the values obtained for the distances between core estimations given by both reconstructions. The arrow points to the 68% quantile, corresponding to 150 m.

Figure 5.5: Comparison of the core positions as determined by the regular and infill reconstructions.



(a) Infill array reconstruction.

(b) Regular array reconstruction.

Figure 5.6: The lateral distribution of a shower simultaneously recorded by the infill array and the regular array. The signal at 450 m is produced directly by the infill reconstruction, while in the regular reconstruction it is determined by means of the LDF of the event.

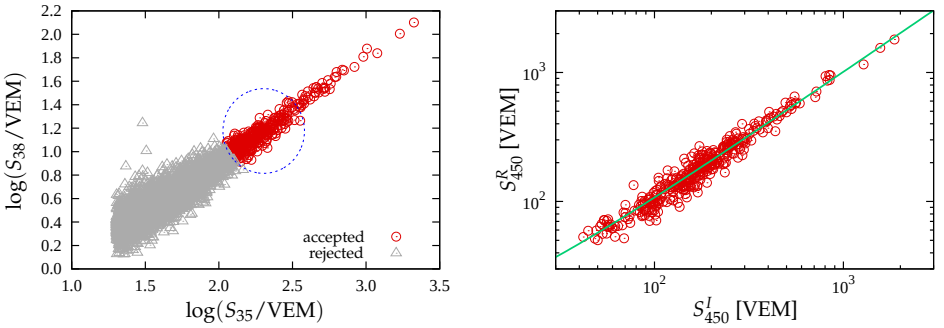
5.1.3. Comparison of S_{450}

As a final check on the reconstruction procedures, the estimates of S_{450} given by the infill and main arrays are investigated. The event reconstruction in the infill array produces S_{450} directly as a fit estimate, while the regular reconstruction produces S_{1000} instead (that is, the signal estimated at 1000 m from the shower axis). Consequently, the value of S^R_{450} given by the regular reconstruction is determined by means of the LDF function found. As an illustrative example, the reconstructed LDFs for a single event, detected at the same time by the infill array and the main array, are shown in figure 5.6: black dots correspond to signals recorded in stations, while the red solid line represents the LDF found by the applied fit. On the left, the event is reconstructed by means of the infill stations, on the right by regular stations only. The values ofS_{450} (figure 5.6a) and S_{1000} (figure 5.6b) as directly obtained by the two fits are indicated in the left and right plots for the infill and regular array, respectively. The value of S_{450} from the regular array can be obtained by the LDF in the right plot, as the fitted signal at 450 m.

The estimated values of S_{450} given by the reconstruction in both arrays are assumed to be related to each other as

$$S^R_{450} = a_S + b_S S^I_{450} + \epsilon_S, \tag{5.6}$$

where the constants a_S and b_S were determined using a linear regression. Here S^I_{450} stands for the signal estimated with the infill reconstruction, while S^R_{450} represents the signal estimated with the regular reconstruction. There is an important caveat here. As anticipated at the beginning of the section, most of the showers detected by the regular array correspond to energies below the trigger-saturation regime. This would introduce a strong bias in the estimates of the constants in 5.6: upward fluctuations on a steep cosmic ray spectrum are more likely to be detected by the regular array in this energy range and, consequently, this introduces a selection bias in the estimates provided by the regression analysis. To circumvent this problem, showers outside the trigger-saturation regime of the regular array must be excluded from



(a) The correlation of the S_{38} and S_{35} given by the regular and infill reconstructions, respectively. The population of events surviving the elliptic cut is represented by red circles.

(b) The correlation of the S_{450} values given by the reconstructions for the subset of selected events. The solid line represents the applied fit.

Figure 5.7: Comparison of the energy estimator S_{450} as determined by the regular and infill event reconstruction procedures.

Parameter	Value	Uncertainty
a_S [VEM]	6.87	2.37
b_S	1.01	0.01

Table 5.4: The values of the constants estimated for the correlation of S_{450} modeled in (5.6). The fraction of variance explained is $R^2 = 0.98$. The fit used to determine the constants was applied to a high quality subset of detected showers.

the regression. To do so, an elliptic cut is employed based on the hypothesis that measurements follow a Gaussian distribution with known deviations. The original method is presented in [75]: it aims at including only showers whose energy (or energy estimator) is above or at the verge of the full trigger efficiency of the array. Showers near the verge are included provided that they lie inside the 95% CL ellipse centered in the cut value. It is thus necessary to establish a reasonable value for the cut point.

In order to determine the cut point, the correlation between the energy estimators S_{1000} and S_{450} is considered after correcting for their dependency on zenith angle (section 2.1.4 and 7.1). Such correction gives S_{38} values for the main array and S_{35} for the infill; their correlation is shown in figure 5.2b. The regular array is fully efficient to detect showers between 0° and 60° and $S_{38} > 15$ VEM, thus the cut value S^c_{38} is set to 15 VEM. The corresponding S^c_{35} is set by means of the correlation between S_{38} and S_{35}, which is assumed to be linear. Initially, the constants in the formula relating S_{35} and S_{38} are determined by a fit using all events in the sample. By means of this formula the corresponding value of $S^c_{35_1}$ is set during this first stage. Then, a cut to events is applied using an ellipse centered in $(S^c_{35_1}, S^c_{38})$, and their uncertainties $(\sigma_{S^c_{35_1}}, \sigma_{S^c_{38}})$ are set to the measured uncertainties in S_{35} and S_{38} respectively. The population of events that survive the cut is then used to refine the constants in the formula relating S_{35} and S_{38}. The refined formula is employed

to set a new value of $S^c_{35_2}$ that is then used to set the center of a new ellipse used to perform a new cut, and so on and so forth. The process is repeated until stability is reached (i.e. the i-th iteration applies a cut that results in the same population of events than the previous $i-1$ iteration). The values produced for $S^c_{35_i}$ and S^c_{38} are used as the ellipse center for the cut ultimately used to select the population of events that is employed to determine a_S and b_S in (5.6). This is illustrated in figure 5.7a, where the correlation of S_{38} and S_{35} is shown for all events in the sample of showers detected by both arrays, together with the ellipse utilized to apply the cut. The population size of surviving events is 353, which correspond to 8% of the original population size. Although the cut rejects most of the showers detected by both arrays, the remaining population represents the highest quality subset available to perform a comparison of the signal estimates given by both event reconstructions.

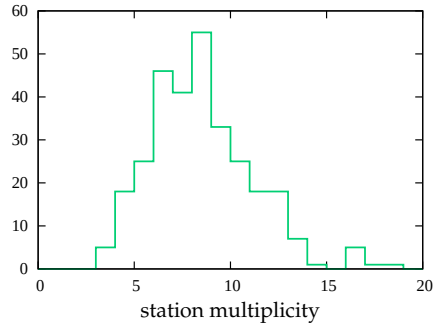
The fact that the disturbance term ϵ_S in (5.6) is correlated to S^I_{450} (the uncertainty depends on the signal itself) causes the covariance matrix of the constant estimates to be biased, even though the estimates of the constants themselves remain unbiased. For this reason, the uncertainties related to S_{450} are ignored.

Once the elliptic cut has been applied, remaining events are used for the regression analysis used in formula (5.6). The estimated values of the constants a_S and b_S are listed in table 5.4. Differences found are rather small when compared to the values of the signals, as can be observed. The S_{450} values used for the fit, along with the curve that provided the best fit, are shown in figure 5.7b.

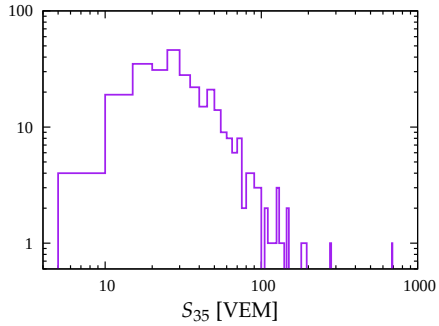
5.2. Comparison between infill and hybrid events

A unique capability of the Pierre Auger Observatory is that of measuring air showers with two complementary instruments: the fluorescence telescopes and the surface array of WCD. One of the major benefits of this hybrid detection method is that the resolution on shower observables, such as the arrival direction or core position, improves dramatically with respect to the resolutions associated with each apparatus taken separately. The hybrid reconstruction is mainly based on the fluorescence detector reconstruction algorithm (section 2.2.3), but further exploits information available from one station from the surface array: the timing provided by such station is used to constrain the fit employed in the geometrical reconstruction of the FD. The gain on the angular resolution can be great, as analyses from hybrid events established that the spread in angular measurements decrease from $\sim 8^\circ$ to $\sim 0.3^\circ$ [94] for the fluorescence detector. The hybrid technique is able to locate the shower core on the detection plane of the surface array without prior knowledge of the LDF, with a resolution close to 20 m [94], providing an excellent handle on the quality of the infill array reconstruction. This section explores the performance of the reconstruction procedure of the infill array by carrying out a comparison with the hybrid reconstruction.

The hybrid data set was selected from the period spanning from August 2008 to September 2011. Several quality cuts are applied on hybrid events, imposed by stringent conditions required on the fluorescence detector, intended to ensure a robust hybrid reconstruction: the water Cherenkov detector with the highest signal



(a) The distribution of the station multiplicity in the infill array.



(b) The distribution of S_{35} values as estimated by the infill reconstruction procedure.

Figure 5.8: The station multiplicity and energy estimator distributions of hybrid events used in this work.

must be within 750 m of the shower axis [95]. The reduced χ^2 of the longitudinal profile fit to the Gaisser-Hillas function has to be less than 2.5. Furthermore the χ^2 of a linear fit to the longitudinal profile has to exceed the Gaisser-Hillas fit χ^2 by at least 4 units [96]. The depth of shower maximum, X_{max} , must be within the field of view of the telescopes and the fraction of the signal detected by the FD and attributed to Cherenkov light must be less than 50%. The uncertainties on the energy estimate E_{FD} and on X_{max} are required to be less than 20% and 40 g/cm^2, respectively. The selection criteria include a measurement of the vertical aerosol optical depth profile (VAOD) made using laser shots generated by the central laser facility (CLF) [97] and observed by the fluorescence telescopes in the same hour of each selected hybrid event; the VAOD value must be smaller than 0.1. Furthermore the cloud fraction in the field of view, measured from the information provided by the LIDAR systems of the observatory [97], is required to be less than 25%. On the other hand, the usual cuts on showers detected in the infill array are applied, namely, T5 events with a reconstructed zenith angle comprised between 0° and 55°, excluding bad periods of data acquisition and fake T5 events. For the sake of completeness, the distributions of the station multiplicity and the signal estimator values for the population of surviving events are shown in figure 5.8.

5.2.1. Comparison of arrival directions

In order to compare the reconstructed zenith angles obtained by the hybrid reconstruction and the infill reconstruction, the approach based on a linear regression from the previous section is reproduced. In this case, as the hybrid system is more accurate than the infill array alone, it is the zenith angle θ^I, estimated by the infill array which is regarded as a linear function of the value θ^H obtained from the hybrid reconstruction

$$\theta^I = a_\theta + b_\theta \theta^H + \epsilon_\theta. \tag{5.7}$$

The uncertainties on θ^I were not taken into account in the fit. Furthermore, given the high resolution associated to the hybrid estimates, any correlation between the

Parameter	Value	Unc.
a_θ [deg]	-0.33	0.15
b_θ	1.00	0.01
a_φ [deg]	0.00	0.11
b_φ	1.00	0.01

Table 5.5: The values of the estimated constants for (5.1) and (5.2). The fraction of variance explained in both cases is $R^2 = 1.00$.

zenith angle and the angular resolution is also neglected. As already discussed in the previous section, the constant estimates should remain unbiased even if the covariance matrix does not.

The case for the azimuth angle φ is identical. The values obtained from both reconstructions are assumed to follow

$$\varphi^I = a_\varphi + b_\theta \varphi^H + \epsilon_\varphi, \tag{5.8}$$

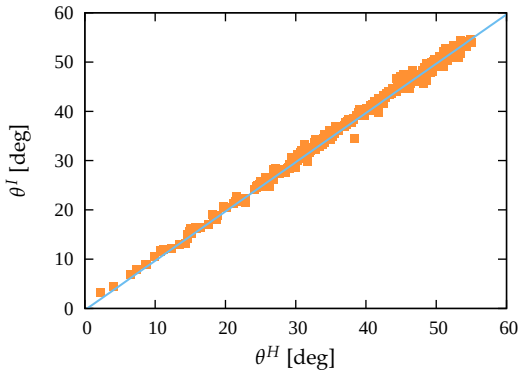
and the constants a_φ and b_φ are estimated by means of a linear regression. The independence of the azimuth angle and the disturbance term in the case of the hybrid reconstruction is not obvious, since the performance of the geometrical reconstruction of the fluorescence detectors depends on the relative direction of showers with respect to the field of view of the telescopes. However, the effect is ignored in the fit applied using the model in (5.8), thus resulting in a covariance matrix that is not dependable. This implies that the standard errors for the constants are not reliable, although the estimated constants themselves should remain unbiased.

If both geometrical reconstructions are consistent with each other, the values of the constants a and b in (5.7) and (5.8) should be 0 and 1, respectively.

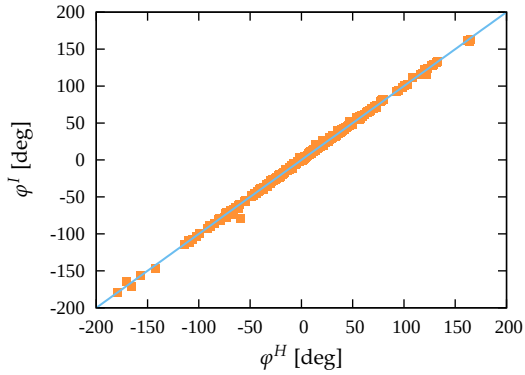
The correlation of reconstructed zenith and azimuth angles is shown if figures 5.9a and 5.9b, respectively. Dots in those plots represent the estimated values for the angles, while the solid line represents the curve given by the fit procedure.

The angular difference ψ in arrival directions is calculated using the formula in (5.4), replacing the regular array estimates with the hybrid estimates. The histogram of ψ values resulting from the data set utilized is shown in figure 5.9c. The arrow in that plot signals the quantile of the distribution of ψ for 68% of the events, that corresponds to 1.1°. No events with an angular difference larger than 5° degrees were found.

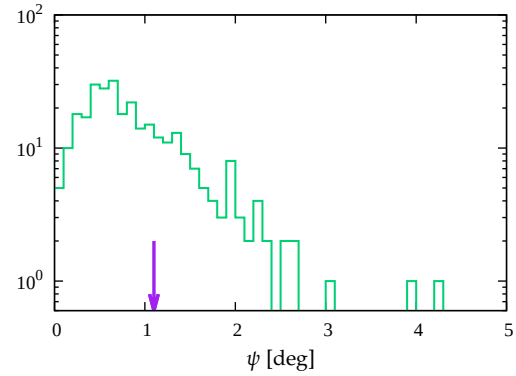
The constants in (5.7) and (5.8) given by the fit procedure are included in table 5.5. As can be seen, the estimated values of a and b are close to 0 and 1, respectively, for both angular coordinates. In spite of the fact that the error estimates are not dependable, the values of the slope b in both cases are near unity. The fact that the intercept in (5.1) is slightly below zero may be a hint on some systematic bias in the reconstruction procedure of the infill array, but further studies should be carried out. Nevertheless, the infill reconstruction shows a very good agreement with the hybrid one in terms of arrival direction of showers.



(a) The correlation of reconstructed zenith angles. The solid line represents the applied fit.



(b) The correlation of reconstructed azimuth angles. The solid line represents the applied fit.



(c) A histogram of the values obtained for the angular difference ψ between the arrival direction estimations. The arrow points to the 68% quantile, corresponding to 1.1°.

Figure 5.9: Comparison of the arrival direction as determined by the hybrid reconstruction and the infill reconstruction.

Parameter	Value	Unc.
a_X [m]	-78	125
b_X	1.00	0.01
a_Y [m]	44	95
b_Y	1.00	0.01

Table 5.6: The values of the estimated constants in (5.9), for both Cartesian components of the core position on the detection plane. The fraction of variance explained in both cases is $R^2 = 0.99$.

5.2.2. Comparison of the core position

The fact that the estimation of the core position, given by the hybrid reconstruction algorithm, does not depend on an assumed LDF is quite convenient to assess the performance of the fit of the lateral profile applied for showers reconstructed in the infill array. If the core location is poorly determined by the infill reconstruction, or if it is systematically biased, would bring to a poor estimation of the signal at a distance of the shower axis, as previously discussed.

The approach taken is identical to the one used in the comparison of the core positions employed in the comparison between the infill and regular reconstruction. The components of the core position on the detector plane are treated as

$$Q^I = a_Q + b_Q Q^H + \epsilon_Q, \tag{5.9}$$

where Q is the Cartesian coordinate (i.e. X or Y). As in the case of the arrival directions, the uncertainties in the values of the estimates are neglected.

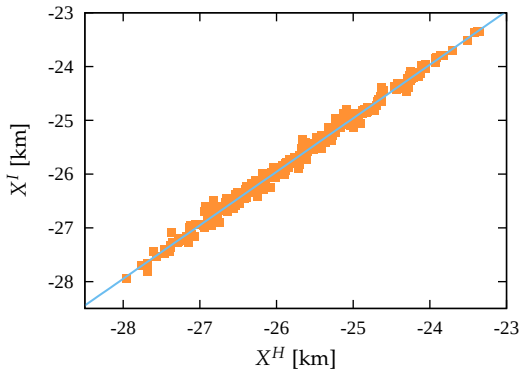
The correlation of the estimated X values is shown in figure 5.10a, while figure 5.10b depicts the correlation between the estimated values of Y. As already pointed out in a previous section, the higher density of points in certain regions of those figures is caused by the larger number of showers detected in some areas of the infill array, either due to its geometry or to the exposure of its different parts.

Also in this case, a histogram has been built using the distances between the core estimates given by the infill reconstruction and the hybrid one, shown in figure 5.10c. 68% of the events studied have a difference in core position less than 150 m, as indicated by the arrow included in the plot.

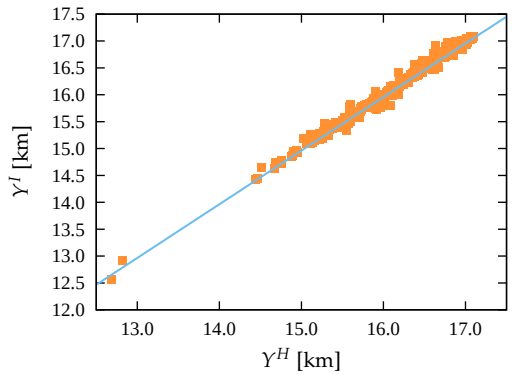
Once again, if the reconstruction procedures for the infill array and the hybrid instrument are consistent, the values of a_Q and b_Q in (5.9) should be 0 and 1, respectively, and the estimates provided by the fit are expected to be compatible with those values. Such estimates are listed in table 5.3: no strong biases seem to be present. Furthermore, no large deviations have been found for any of the events considered. Thence, the infill reconstruction shows a good agreement with the hybrid one in terms of the reconstruction of the core positions.

5.3. Conclusions

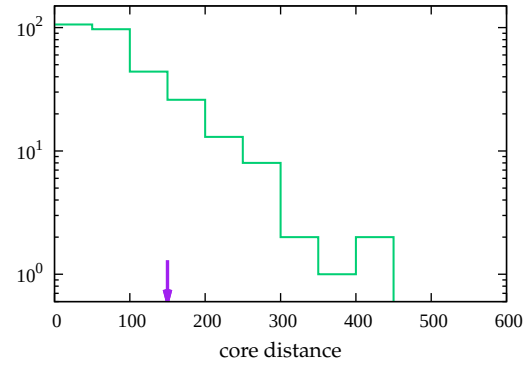
Although the comparisons made in this chapter are simple, they provide useful information on the reconstruction of events in the infill array. The fact that no sys-



(a) The correlation between the X coordinates of the core positions. The solid line represents the applied fit.



(b) The correlation between the Y coordinates of the core positions. The solid line represents the applied fit.



(c) A histogram of the values obtained for the distances between core estimations given by both reconstructions. The arrow points to the 68% quantile, corresponding to 150 m.

Figure 5.10: Comparison of the core positions as determined by the hybrid and infill reconstructions.

tematic disagreements have been found with the reconstruction of the regular array, in terms of the geometrical reconstruction and energy estimator, nor with the hybrid one, in terms of the geometrical reconstruction, suggests that the performance of the infill reconstruction is under control.

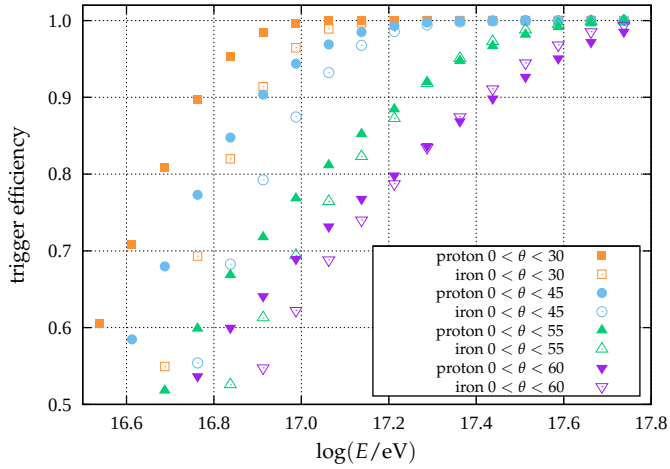
Chapter 6

The exposure of the infill array

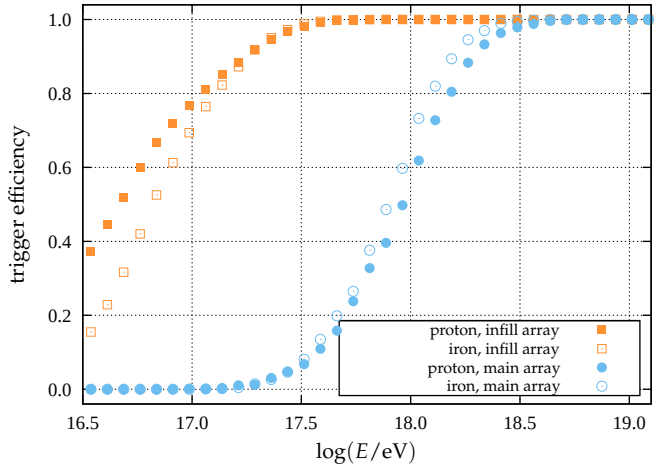
This chapter presents the techniques employed to determine the saturation energy for the trigger efficiency and then describes the methods used to estimate the exposure of the infill array, as well as an evaluation of the systematic impact that different effects in the detector have on the exposure estimate. The selected data set consists of showers recorded by the infill array, from August 2008 to February 2012. All T5 events were included, with the exception of fake T5 events (section 6.3.1) and a subset of events collected during the communication crisis. Bad periods of data acquisition were also excluded. Furthermore, only events with $\theta < 55^\circ$ were used. After applying these cuts a population of 448340 events survived.

6.1. The saturation of the trigger efficiency

In order to estimate accurately the acceptance of the infill array, its efficiency to detect showers must be known precisely. Since the central trigger is given by a combination of trigger states of neighboring stations, it all comes down to calculating the probability that an individual station reports a trigger when a shower falls at some distance from it. This probability depends on the physics of showers, the characteristics of the detector and the choice of the trigger definitions. The physics dependence is related to the particle density as a function of the distance to the shower core, together with their energy and angular distributions. The dependence on the detector characteristics involves the tank design (e.g. shape, water quality) and the electronics and signal processing algorithms employed. As for the trigger definitions, it is possible to select from several types of definitions and modify the parameters related to any particular choice (e.g. the signal threshold level). With the purpose of modeling such dependencies, a function known as the lateral trigger probability (LTP) has been defined [98]. When the energy E, mass A and zenith angle θ of the primary particle are given, as well as the trigger definition T for the local station, the LTP parameterization determines the probability $P(r, \phi | E, A, \theta, T)$ the generated air shower has to trigger a single water Cherenkov detector in the array, located at coordinates (r, ϕ) in the shower reference frame. Monte Carlo simulations of EAS were used extensively to derive the LTP functions in the energy



(a) The trigger efficiency in the infill array for different zenith angle ranges.



(b) Zenith angle integrated up to 55° for the infill array (orange squares) and up to 60° for the main array (blue circles).

Figure 6.1: The trigger efficiency for 3ToT events in the infill array as a function of the simulated energy E. Filled symbols are used for proton primaries and empty symbols for iron primaries.

range comprised between 10^{17} to 10^{19} eV and zenith angles less than 65°.

The trigger efficiency for 3ToT events in the infill array has been determined using LTP functions [81]. For any shower core position and arrival direction, it is possible to compute the distance of each station in the shower reference frame and deduce its trigger probability from LTP functions. The probability to pass the global T4 trigger then follows from a simple combination of these local LTP values. From the plot included in figure 6.1a, it can be seen that the trigger efficiency in the infill array looses its dependence on primary mass and energy from $\sim 3 \times 10^{17}$ eV, that corresponds to $S_{35} \sim 20$ VEM (section 7.2), for showers between 0° and 55°. Figure 6.1b shows the estimated dependence of the 3ToT trigger efficiency as a function of primary energy, for proton and iron primaries and showers generated with incoming zenith angles between 0° and 55°. The trigger efficiency for the main array is also shown, for reference purposes. Thus, the array is fully efficient for energies above $\sim 3 \times 10^{17}$ eV, an order of magnitude below the saturation energy of the 1500 m main SD array.

6.1.1. Experimental verification of the trigger efficiency

Two experimental tests were employed to verify the full trigger efficiency regime in the infill array, for showers arriving with zenith angles less than 55°. The main purpose of such tests was to validate, by means of data, the results obtained from shower simulations, as described in the previous paragraph. The first test performs a check on the distribution of the arrival direction of events, while the second investigates the distribution of the shower cores on the detection plane.

Test based on the isotropy of arrival directions

The first test relies on the assumption that the flux of cosmic rays is, to a good approximation, isotropic. This is the same basic idea behind the constant intensity cut (section 7.1). The assumption implies that the flux in local coordinates during a certain time interval depends only on the array acceptance and not on the zenith angle. Therefore, the number of detected air showers in any two angular bins of equal exposure should be the same, if the detector is fully efficient for both bins.

Angular bins of equal exposure for the surface detector are obtained if bins of equal size in $\sin^2\theta$ are employed, where θ represents the zenith angle in local coordinates. This is so because the effective area of the array scales with the cosine of the zenith angle of the incident shower. As a consequence of the isotropy of the cosmic ray flux, the number of events per unit of $\sin^2\theta$ should be constant. As an illustration, showers recorded by the infill array have been binned according to their zenith angles, using different cut values on the energy estimator S_{35}. The resulting plot is shown in figure 6.2. When the cut in S_{35} is too low, the number of events detected is not uniform across $\sin^2\theta$. The effect is caused by the lack of events for the most inclined angular bins, since the instrument is not fully efficient to detect inclined showers with low energy. As the cut value is enlarged, the efficiency of the array improves and the distribution of showers becomes more uniform. This suggests a simple test that can be used to determine the cut value S_{35}^c of the S_{35} energy

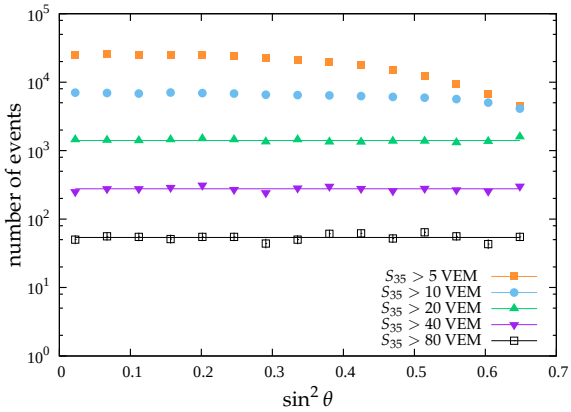


Figure 6.2: The number of showers above a specific S_{35} value. Angular bins of equal exposure were used.

estimator: starting from a relatively low value, S_{35}^c is increased, step by step, using a fixed increment δS_{35}^c. At each step, the resulting distribution of the number of events over $\sin^2\theta$ is tested to verify if it agrees with a uniform distribution. When the agreement is found, the full efficiency regime of the apparatus has been reached for the corresponding value of S_{35}^c.

The Kolmogorov-Smirnov test was used to compare the sample of $\sin^2\theta$ realizations against a uniform distribution. The sample is given by the subset of showers that satisfy the requirement $S_{35} > S_{35}^c$. The null hypothesis for the test is that the sample was drawn from a uniform distribution. The basic idea is to determine the maximum distance between the empirical distribution function (EDF) and the reference uniform distribution. The empirical distribution function Γ of a random variable X is defined by means of the formula

$$\Gamma(x) = \begin{cases} 0, & \text{if } x < x_1 \\ i/n, & \text{if } x_i \leq x < x_{i+1} \\ 1, & \text{if } x \geq x_n \end{cases} \tag{6.1}$$

where the n values in the sample must be arranged in order of size beforehand; that is $x_1 \leq x_2 \leq \ldots \leq x_n$. The EDF is to be compared with the reference distribution

$$F(\sin^2\theta) = \frac{\sin^2\theta}{\sin^2 55^\circ}, \quad \text{for } 0^\circ \leq \theta \leq 55^\circ, \tag{6.2}$$

which is the uniform distribution function of $\sin^2\theta$ values for $0^\circ \leq \theta \leq 55^\circ$.

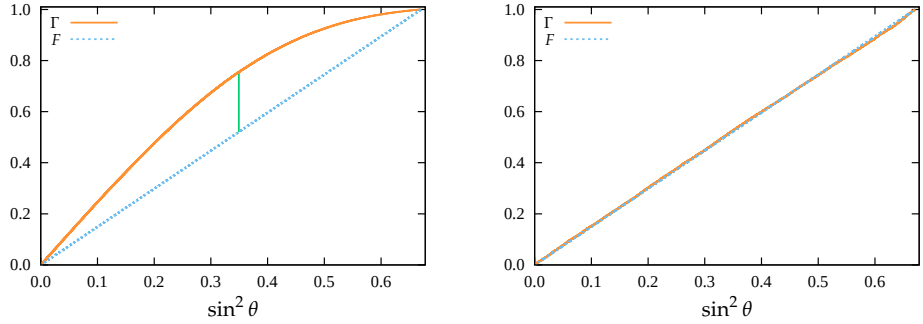
In practice, the maximum distance D between the EDF and F is given by the maximum value in the finite set of distances

$$D = \max(D_1, \ldots, D_n), \tag{6.3}$$

where

$$D_i = \max(|\Gamma(\sin^2\theta) - F(\sin^2\theta_i)|, |\Gamma(\sin^2\theta) - F(\sin^2\theta_{i+1})|). \tag{6.4}$$

Finally, once the value of D is determined, it is compared to a critical value D_c, and the null hypothesis is accepted if $D \leq D_c$, otherwise it is rejected. An example



(a) Empirical distribution built for events fulfilling $S_{35} > S_{35}^c = 2$ VEM.

(b) Empirical distribution built for events fulfilling $S_{35} > S_{35}^c = 25$ VEM. The maximum distance is so small that it is not visible in the figure.

Figure 6.3: Comparison of the empirical distribution with a uniform distribution. The maximum distance D is drawn in green.

is shown in figure 6.3. The figure contains a couple of plots comparing $F(\sin^2\theta)$ and $\Gamma(\sin^2\theta)$. The plot in figure 6.3a shows the EDF Γ, for events having $S_{35} > S_{35}^c = 2$ VEM, and the uniform distribution F. The maximum distance is represented by a green vertical line. The critical value D_c for a significance level of 5% is given by the formula $1.22/\sqrt{n}$, valid for $n > 50$. In this particular case the number of elements in the sample is $n = 446649$, and the null hypothesis is rejected, since the maximum distance is 0.234 and the critical value is 1.83×10^{-3}. Instead, the plot in figure 6.3b shows Γ and F once more, but for events satisfying the $S_{35} > S_{35}^c = 25$ VEM condition. In this case, the overlap between both functions is almost perfect, hence the maximum distance is very small and, therefore, the null hypothesis is accepted.

By using the D/D_c ratio it is possible to visualize how the EDF approaches a uniform distribution, as the value of S_{35}^c is increased. The plot is included in figure 6.4. Note that the condition required to accept the null hypothesis is $D \leq D_c$ or, equivalently, $D/D_c \leq 1$. The plot was created using 5% significance level for the Kolmogorov-Smirnov test. When the ratio D/D_c goes below 1, it is reasonable to consider the distribution of the arrival directions of events as uniform over the solid angle, for the associated S_{35}^c value. This implies that the trigger probability in the detector is independent of the shower arrival direction over the range of zenith angles considered ($0^\circ - 55^\circ$ in this work).

From the plot in figure 6.4 the conclusion is that the full trigger efficiency for the infill array, for a zenith angle range between $0^\circ - 55^\circ$, is reached for values of $S_{35} > S_{35}^c \approx 20$ VEM. This result is in good agreement with the one obtained using simulations, described in the previous section.

Test on the core distribution

The second test performed is based on the distribution of the shower cores on the detection plane of the apparatus. The shower cores should fill the effective collection

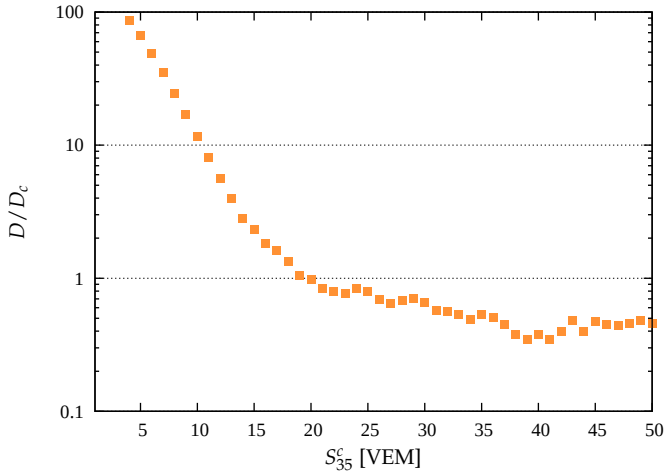
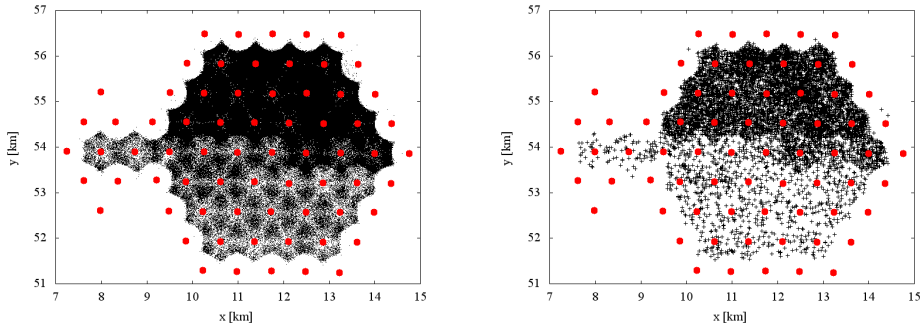


Figure 6.4: The ratio of the maximum distance D over its critical value D_c is shown for the empirical distribution function of events, versus S_{35}^c. The critical region is defined by the area above 1. The values of D_c were obtained for a 5% significance level.

area of the detector in a uniform manner, when events detected in the full trigger efficiency regime are considered.

When checking for the uniformity of the core distribution, it should be noted that data acquisition started before the construction of the array was complete. As can be seen from figure 6.5 the core distribution over the effective area of the instrument is far from being uniform, even for events well beyond the nominal trigger saturation threshold (fig. 6.5b). The problem arises because some hexagons in the infill array are older than others, and thus the exposure of the former is larger than the one for the latter. As a consequence, the number of cores falling on the effective area of different hexagons might be very different, but for a reason that has nothing to do with the trigger efficiency of the detector. One possibility to overcome this obstacle is weighting each hexagon by its exposure and then check for the uniformity of the core distribution. Another simpler possibility is applying a coordinate transformation in order to translate the cores into a unique ideal hexagon of detectors. The latter approach was taken.

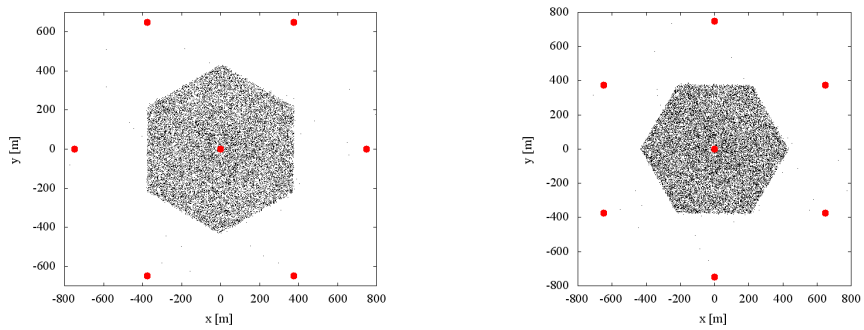
The method employed translates all the considered shower cores into a single ideal hexagon of detectors by means of coordinate transformations. This ideal hexagon is built in such a way that the nearest station to a shower core is always placed in its center. The reason for picking the nearest station to the core instead of the hottest station is that when the particle density of a shower is similar in the hottest and second hottest stations (or even a third station), the signal produced will thus be similar; when the event is reconstructed, it is possible that the LDF fit places the estimated core position closer to the second hottest station rather than to the hottest one. This causes slight deformations in the distribution of the cores near the border of the effective area of the ideal hexagon, if the hottest station is placed on its center. After translating the cores in to the ideal hexagon, it is easy to identify its effective area (figure 6.6a). Finally, a thirty degree rotation in the clockwise direction



(a) All T5 events.

(b) Events fulfilling $S_{35} > 30$ VEM.

Figure 6.5: The estimated position of the shower cores (black symbols) on the detector plane of the infill array. The stations are represented by red circles. The exposure of individual stations is not accounted for.



(a) The result of the translation of all the shower cores into the ideal hexagon of detectors (red circles).

(b) The ideal hexagon after a thirty degree rotation in the clockwise direction.

Figure 6.6: The positions of the estimated shower cores in the ideal hexagon, for events with $S_{35} > 20$ VEM. The effective area is clearly demarcated by the core positions.

is applied for convenience. This leaves the hexagon of detectors and its effective area as seen in figure 6.6b.

Once the positions of the cores have been translated into the ideal hexagon, it is interesting to test the hypothesis that these positions follow a uniform distribution over the effective area of the hexagon. Because the modifications of the Kolmogorov-Smirnov test necessary for multivariate data are not obvious, the choice was to apply a binned method. The effective area of the hexagon was thus divided using bins shaped as equilateral triangles, since this geometrical configuration fits perfectly inside a regular hexagon. Figure 6.7a contains an example plot of the two-dimensional histogram filled with the core positions in the ideal hexagon. The histogram was built using a low S_{35}^c value of 5 VEM and, therefore, the shower core distribution on the effective area is far from uniform. From the same figure, it is possible to observe six regions that are more densely populated by shower cores than the rest of the effective area. These regions are located at and around the triangle center of the equilateral triangle that constitutes the grid of the array. When low energy showers fall near the triangle center, the probability of triggering the array increases with respect to showers that fall near the triangle vertexes or edges.

In order to test for the uniformity of the core distribution, two approaches were taken. The first approach is based on a fit to a plane and on the value obtained for the χ^2. The second approach compares the sample mean and sample variance of the number of events per bin.

A fit to a plane was performed using the number of events per bin n_i in the histogram, along with their (x_i, y_i) position on the detection plane. The function used was

$$f_p(x, y) = a + bx + cy. \tag{6.5}$$

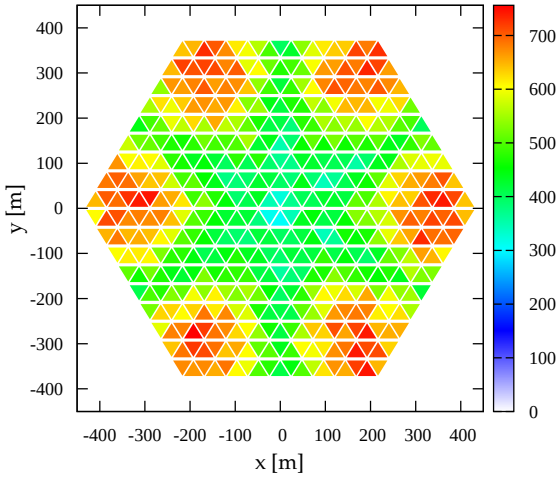
The number of events per bin n_i is treated as a Poisson variable and, therefore, the uncertainty associated to each bin is $\sqrt{n_i}$. Thus, the χ^2 function is given by the formula

$$\chi^2 = \sum_i^m \frac{(n_i - f_p(x_i, y_i))^2}{n_i}, \tag{6.6}$$

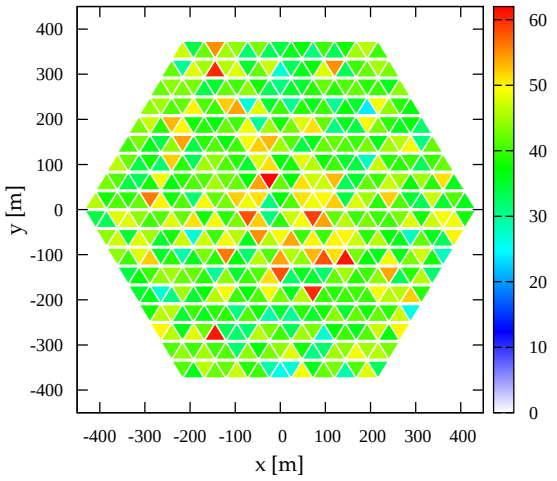
and is then minimized in order to estimate the values of the constants in (6.5). Here m stands for the number of bins in the histogram. The goodness of the fit is determined by the value of the reduced χ^2.

It is thus possible to check the agreement of the core densities on the detector plane with a uniform distribution by determining the reduced χ^2 value of the fit to the plane, as a function of the cut value S_{35}^c for the energy estimator. The resulting plot is shown in figure 6.8a. A good agreement between the model and data is found starting from $S_{35}^c \sim 15$ VEM and above, slightly less than what has been found previously. The estimated values for the constants in (6.5) given by the fit procedure are shown in table 6.1, for showers fulfilling $S_{35} > 20$ VEM : the plane is fairly compatible from what is expected for a uniform distribution.

The second approach taken determines the sample mean and sample variance of the number of events per bin n_i. As previously stated, this number is expected to follow a Poisson distribution. Hence, a simple test is to compare for equality the mean $\bar{n}$ and variance $s_{\bar{n}}^2$ of the population of events considered. To verify the



(a) The density of shower cores for events with $S_{35} > 5$ VEM.

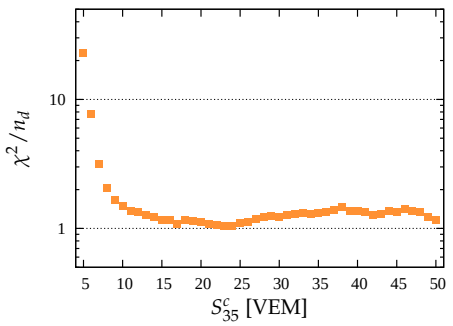


(b) The density of shower cores for events with $S_{35} > 20$ VEM.

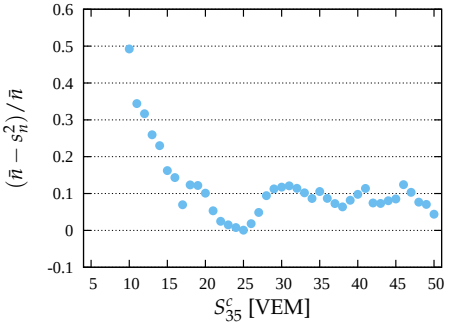
Figure 6.7: Histograms of the number of showers falling on the effective area of the ideal hexagon.

Parameter	Value	Uncertainty
a	40.3	0.3
b	$-1.35 \cdot 10^{-3}$	$1.54 \cdot 10^{-3}$
c	$2.85 \cdot 10^{-4}$	$1.52 \cdot 10^{-3}$

Table 6.1: The estimated values of the parameters in (6.5), using events with $S_{35} > S_{35}^c = 20$ VEM. Note that the confidence intervals include the value of 0, as expected for b and c.



(a) The reduced χ^2 value of the fit of the core densities to the model given in (6.5).



(b) The relative difference between the sample mean and sample variance for the core densities found for the effective area of the ideal hexagon.

Figure 6.8: The agreement between the core densities on the effective area of the ideal hexagon and a uniform distribution, as a function of the cut value S_{35}^c of the energy estimator.

convergence of the values of $\bar{n}$ and s_n^2 as the value of S_{35}^c is increased, the relative difference is shown as a function of S_{35}^c in figure 6.8b. The relative difference initially drops below 10% around 20 VEM and stabilizes around that value close to 27 VEM and above. When events with $S_{35} > S_{35}^c = 20$ VEM are used, the values for $\bar{n}$ and s_n^2 are 41.4 ± 0.3 and 45.5 ± 2.9, fairly compatible from what it is expected from a Poisson distribution.

The results given by simulations supported by the tests applied utilizing data allow to determine the value of $S_{35}^c \approx 20$ VEM as a reasonable estimate of the minimum signal required to be in the full trigger efficiency regime for the infill array, for showers with zenith angle between $0\,^\circ - 55\,^\circ$. The estimated value of S_{35}^c corresponds to an energy of $\sim 3 \times 10^{17}$ eV (section 7.2).

6.2. The calculation of the exposure

When the surface detector is fully efficient, the determination of its exposure is based solely on the geometrical acceptance and on the observation time. The acceptance of the entire array is obtained as a multiple of the geometrical acceptance of a single hexagon, a_h, defined by a station surrounded by six active neighbors, as illustrated in figure 6.9. In that figure, the effective area of the hexagon is represented by the shaded region. Such a simple calculation takes advantage of the regularity of the array as well on the definition of the fiducial trigger, discussed in

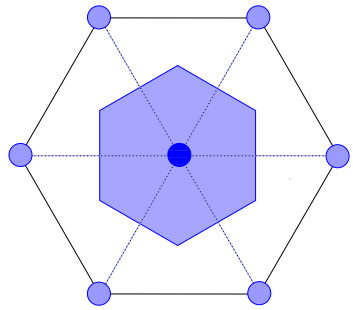


Figure 6.9: A single hexagon in the infill array, composed of seven stations. The shaded region represents its effective area of 0.49 km^2.

section 3.3. The fiducial trigger allows to determine the collection area of a single hexagon as the area defined by all points that lie closest to the central station. In this manner, the total collection area of the array is calculated using the number of hexagons that compose it, times the effective area of the single hexagon. Similarly, the acceptance of the array is obtained from the acceptance of the single hexagon times the number of hexagons.

The acceptance a_h of a single hexagon is given by

$$a_h = \int A_e \, \mathrm{d}\Omega \tag{6.7}$$

where A_e is the effective area for showers arriving from a given direction, integrated over the solid angle. Since the effective area scales with the cosine of the zenith angle, that is $A_e = A_h \cos\theta$, the acceptance is given by

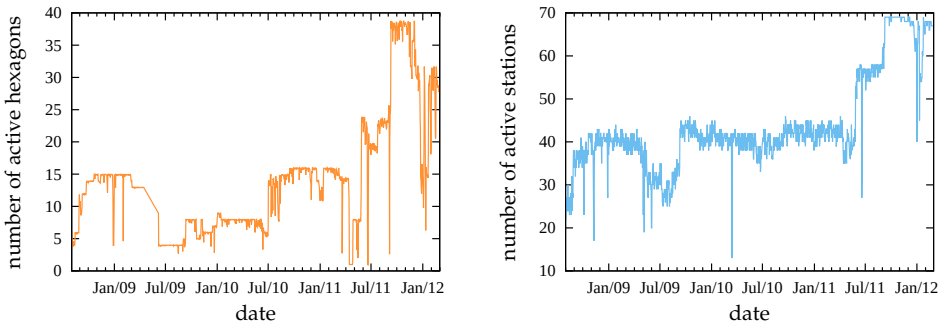
$$a_h = A_h \int_0^{2\pi} \int_{\theta_1}^{\theta_2} \cos\theta \sin\theta \mathrm{d}\theta \mathrm{d}\phi = A_h \pi (\sin^2\theta_2 - \sin^2\theta_1). \tag{6.8}$$

Knowing that the collection area of a single hexagon in the infill array is $A_h = 0.49$ km^2, and integrating θ between $0\,^\circ$ and $55\,^\circ$, the formula above produces $a_h \approx 1.03$ km^{2}sr.

For the evaluation of the exposure, it is essential to have a knowledge on the number of active hexagons N_h as a function of time. The number of working stations in the array is monitored on a second by second basis, based on the T2 trigger rate that each station reports to the CDAS [56]. This technique allows to finely track the detector configuration as a function of time and, besides providing a way to estimate the exposure, it makes possible to study the performance of single stations in the infill array of the surface detector. Hence it is possible to determine the number of working hexagons N_h at any given second using this information along with the position of the stations. Therefore, the integration of the number of active hexagons N_h in the time period of interest provides an estimate of the exposure of the detector

$$\varepsilon = a_h \sum N_h \Delta t. \tag{6.9}$$

Because data acquisition started before the construction of the array was complete, the number of working hexagons N_h changes as a function of time, as new



(a) Number of active hexagons as a function of time. **(b)** Number of active stations as a function of time.

Figure 6.10: The temporal evolution of the number of working hexagons and stations in the infill array for the time period considered in this work. Active hexagons are inferred from the T2 triggers each station reports, while the number of stations are extracted from data.

stations are deployed in the field. Moreover, N_h does not only vary because more stations are added, but also because stations already in acquisition may temporarily cease to work properly, due to problems in electronics, power supply, communication systems, etc. When a station ceases to function, all the hexagons it makes part of enter a dead time period. These dead times have an impact in the exposure calculation and thus monitoring each single station provides a way to take them in to consideration. The plot in figure 6.10 shows the evolution in time of N_h and the number of active stations for the period considered in this work.

Dead times may also appear at a later stage, during data acquisition, and are thus not taken in to account by the previous method. For example, communication problems may cause the loss of PMT traces belonging to stations that are part of an event that passed the T3 trigger condition: such problems may result in a loss of physics events (see the following section). Another source of event loss are problems that may occur in the data storage system. An empirical method has been developed in order to identify the dead times generated by such losses [99]. The method depends on the distribution of the arrival times of events and assumes they follow a Poisson distribution. The final result of this method is the identification of time periods where the difference of arrival times between events is unusually high. This periods are labeled as bad data acquisition periods and data collected during this time are not used in further analyses.

The integrated exposure for the infill array, from July 2008 to February 2012, excluding bad periods of data acquisition, is shown in figure 6.11. By the end of February 2012 it amounts to 48.1 km^2 sr yr.

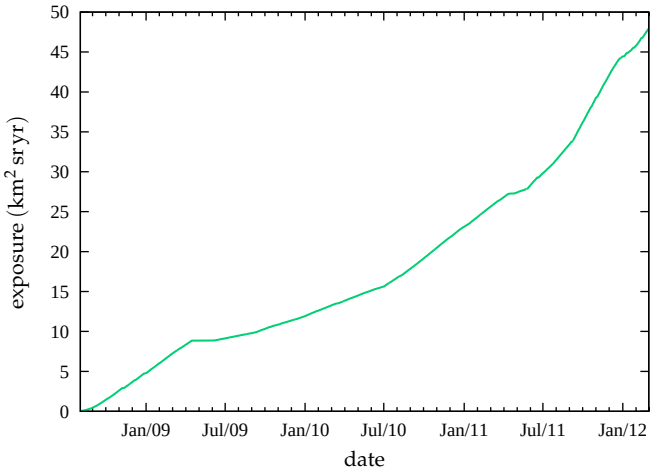


Figure 6.11: The integrated exposure for the infill array as a function of time.

6.3. Systematic errors on the calculation of the exposure

Although the second by second monitoring of stations in the array provides a very reliable way of estimating the exposure of the detector, problems in the communication between the CDAS and stations introduce uncertainties in the calculations. It is possible to estimate the impact these problems have on the exposure in two different ways: first by checking the consistency between T2 triggers and SD data in the infill array and, second, by studying the effect communication errors (referred as T3 errors) have on events. These errors may produce a loss of physics events. If the number of errors induces too many lost events in a period of time, that period is identified [99] and labeled as a bad period of data acquisition, and recorded data are discarded from further analyses.

6.3.1. Consistency between T2 triggers and the event rate

Since the calculation of the number of working hexagons relies entirely on the information on T2 triggers of each single station, it is important to check the consistency between such triggers and recorded data. The check on consistency was performed in two ways. First, it was verified that stations that were part of any event should have reported T2 triggers during the second when the event took place. And second, it was checked that stations that normally reported their T2 triggers participated to events regularly.

Status of stations participating to events

An inconsistency between recorded data and T2 trigger rates is found if a station, or stations, is part of an event but in fact it was not sending T2 triggers to the CDAS during the second the event took place. This means a station may appear in data as a

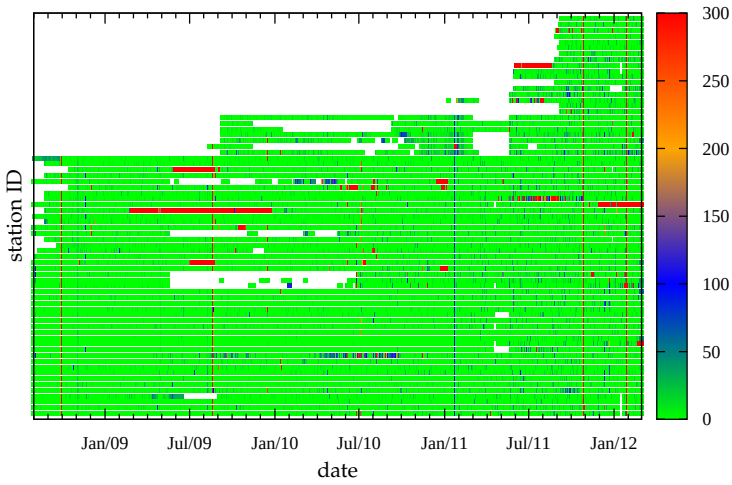


Figure 6.12: Number of inconsistencies per station as a function of time.

silent one (i.e. active but not triggered), while in fact the station was not operational at that particular time. The number of such inconsistencies per station as a function of time is shown in figure 6.12. Note that the effect is rather uniform over all stations in the infill array: for limited periods of time a significant increase can be observed in some stations.

T5 events in the infill array were analyzed. For each of these events it was checked that the hottest station and the six surrounding stations were actually sending T2 triggers to the CDAS, at the time of the event. If at least one of those stations was actually inactive, the event is labeled as a 'fake T5' and thus discarded. An example of a fake T5 event in the infill array is shown in figure 6.13. The origin of these inconsistencies between data and T2 rates it is unknown at the time of writing.

To ensure the consistency between the exposure calculation and the number of events, the correct procedure is to discard fake T5 events from the data set. Because the effect of the described inconsistencies can be compensated by discarding fake T5 events, they have no impact on the uncertainty of the exposure. The total number of fake T5 events for the data set utilized in this work is $\sim$ 17000. The plot in figure 6.14a shows the number of fake T5 events as a function of time. As the total number of events that satisfy the T5 trigger is $\sim$ 492000, the percentage of fake T5 events is $\sim$ 3.4%. Figure 6.14b shows the evolution in time of the fraction of fake T5 events.

Participation of stations to events

A test was performed in order to check if the participation of stations to events is consistent with their running time as derived from T2 files. This effect is a source of

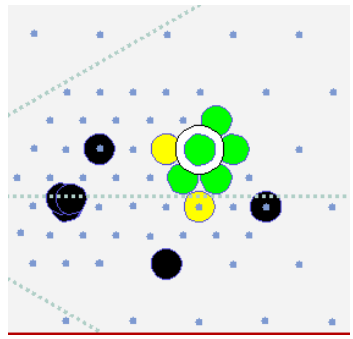
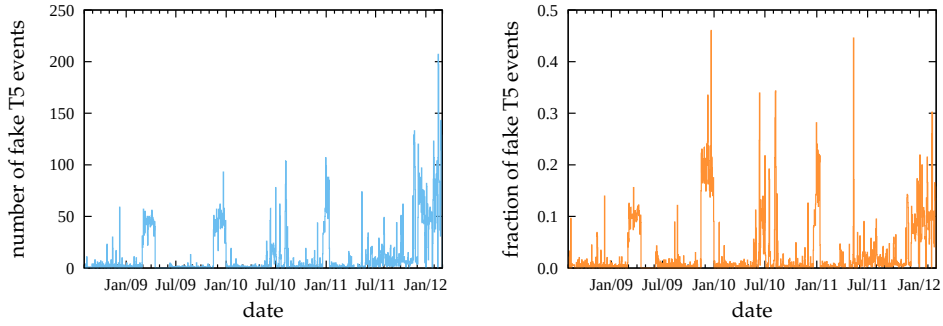


Figure 6.13: Example of a fake T5 event caused by an inconsistency between data and T2 trigger rates. Stations with signal are represented by green and yellow circles. According to recorded data, the hottest station, noted by a white circle, is surrounded by five neighbors with signal and a non-triggered station (silent). In reality the silent station was not functional and thus the inconsistency causes a misclassification of the event.



(a) Number of fake T5 events

(b) Fraction of fake T5 events

Figure 6.14: The evolution in time of fake T5 events.

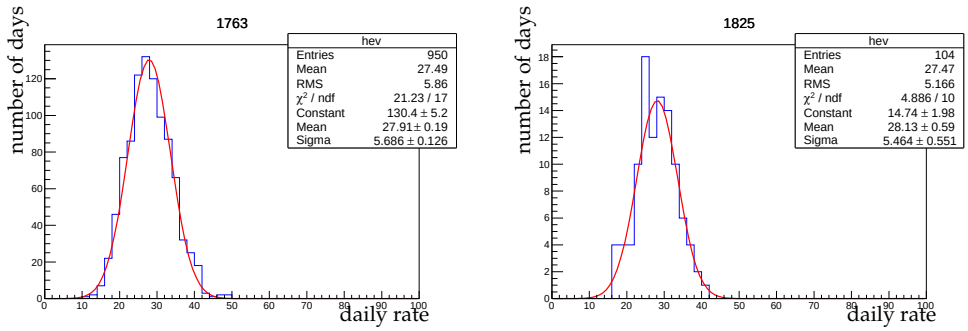


Figure 6.15: Histograms of the daily rate for stations 1763 (left) and 1825 (right). Red lines represent Gaussian fits.

uncertainty in the exposure calculation.

Possible anomalies in the T5 event rate of infill stations were investigated. As this rate depends on the number of neighbors (the higher the number of neighbors, the higher the probability that the station participates to an event), only stations for which the number of neighbors is six were used, that is, stations surrounded by a full hexagon of detectors. The event rate depends also on the up-time of the station. To take this into account, it is possible to either re-normalize the number of events to the up-time, or to use days with constant up-time only. The re-normalization is not trivial, as it should be applied to the neighbors too. Thus, it was decided to exploit stations with a constant up-time, by considering only those stations with more than 90% up-time on a daily basis. This allows to study almost 90% of the total number of days.

A histogram of the daily rates for each station satisfying the described up-time criteria was built. Two examples for two different stations are given in figure 6.15. As the average daily rate per station is always above 20, it is reasonable to use a Gaussian distribution to describe data. Station 1763, whose event rate distribution is shown in the left panel, has been in activity for about 3 years, while station 1825, shown in the right panel, has only 104 days of activity at the time of the analysis. As can be seen by the values of the reduced χ^2 shown in the figures, both fits are good. The lower quality of the fit for station 1825 can be explained by the smaller statistics.

Figure 6.16, shows the mean value of the daily rates as a function of the reduced χ^2 given by the Gaussian fit, for all infill stations considered. Overall, out of the 39 stations studied, none of them shows any abnormal behavior.

Mean values of the daily rates range between 23 and 31 and there is no correlation between low rates and bad values of reduced χ^2. Note that there is only one station, 1761, that has 15 days out of 646 when it did not participate to any event, while it was active. All together, this lack of events amounts to 15/14327 (0.1%) of studied days per station. A possible loss of events was estimated by comparing, for each station considered, the actual distribution to the Gaussian fit: the deficit of events with respect to the number of events expected from the fit are counted and added. It is found that this number is less than 0.5%.

This study shows that overall hexagons work in a way consistent with the up-

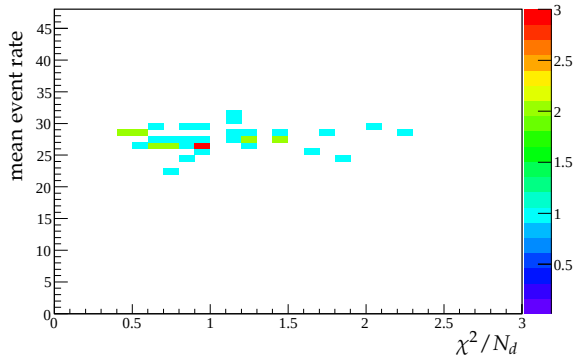


Figure 6.16: Results of the Gaussian fit to the daily number of events for 39 stations in the infill: mean value as a function of reduced χ^2.

time of the stations. The associated uncertainty to the exposure is 0.5%, that represents a conservative upper limit.

6.3.2. T3 errors in the infill array

T3 errors

When the CDAS forms a T3 trigger in the array, it interrogates all T2 stations that participated in the original SD trigger, as well as those stations passing the local T1 trigger only, if they are within a time window of 30 μs. When this query occurs, some of the polled stations may return an error. This type of error is called a T3 error. If a T3 error is returned, information that might be present in the station, such as the FADC trace, timing information, or local trigger type are lost and this may result in the loss of physics events.

It is possible to monitor the occurrence of T3 errors for each station from data, and also to estimate the rate of event loss due to these errors. There are two relevant parameters, related to each single station, that allow to quantify the impact of T3 errors in the array; these parameters, summarized below, are described in greater detail in [100]. On one hand, the *trigger window* (W), whose value is greater than 0 if the station participated in the formation of the T3 trigger, otherwise it is equal to 0. On the other hand, the station *error flag* (E) allows to determine the status of a station in any given event. There are eight types of error flags:

$E = 0$**: No error.** The station properly returned all its data.

$E = 1$**: T3 lost.** The CDAS request arrived when the LS buffer was already overwritten.

$E = 2$**: T3 not found.** No data found. It is not an error if $W > 0$.

$E = 3$**: T3 too young.** T3 request arrived later than the current time in the LS.

$E = 4$**: T3 already.** The station already sent its data, thus data are lost if not recorded at this point.

$E = 5$**: T1 stopped.** The station went off after the T3 trigger and before the CDAS request.

$E = 6$**: Bad compressed data.** CDAS was not able to decompress the data received.

$E = 7$**: Data lost.** The station has not been able to transfer its data to CDAS. It is not an error if the station is turned off. It is always an error when $W = 0$.

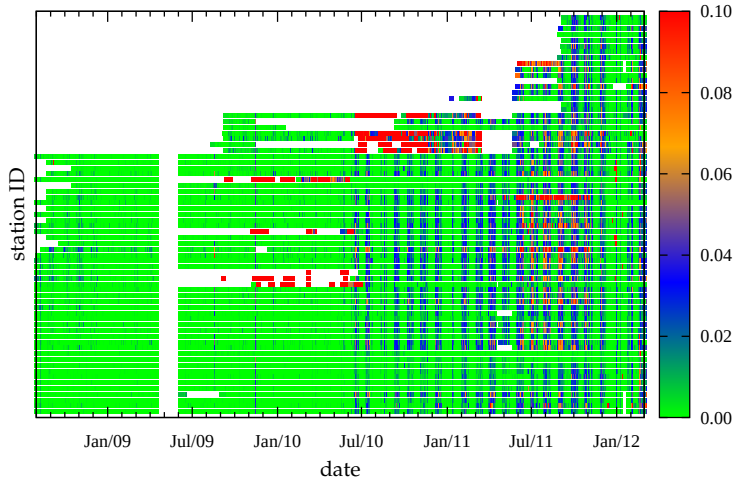
A study was performed on the fraction of T3 errors with respect to the total number of T3 messages on a daily basis, for all the stations that are part of the infill array. For stations that contributed to the trigger formation in the array (i.e. $W = 0$), any error flag $E > 0$ is considered as a T3 error. Otherwise, if $W > 0$ only values of $E = 1, 3, 4, 5, 6$ are treated as errors. Bad periods of data acquisition were excluded for this analysis. The daily fraction of T3 errors is shown in figure 6.17.

From the plots in figure 6.17, two interesting issues were found. First, during certain time periods some stations responded with T3 errors virtually every time the CDAS polled them for data. These stations were equipped with a different kind of transmission antennas, that turned out to be problematic for the communication systems. Such stations are removed from physics analyses during those specific time periods. The second issue discovered was that there exists a modulation in time of the fraction of T3 errors for most of the infill stations in the data set used in this work. The modulation can be seen as an increase and decrease in the fraction of T3 errors, starting from $\sim$ July 2010. The effect appears as blue vertical bands, particularly visible in the plot that includes all stations independently from their window size W (fig. 6.17a). The rate of errors started to increase in July 2010 and has further increased from June 2011. The origin of this problem is related to an excessive rate of events in the area where the infill array is located. The issue becomes manifest when the FD is taking data, as well as in the period where the ToTd trigger was activated. In order to understand this effect a study is presented in appendix A.

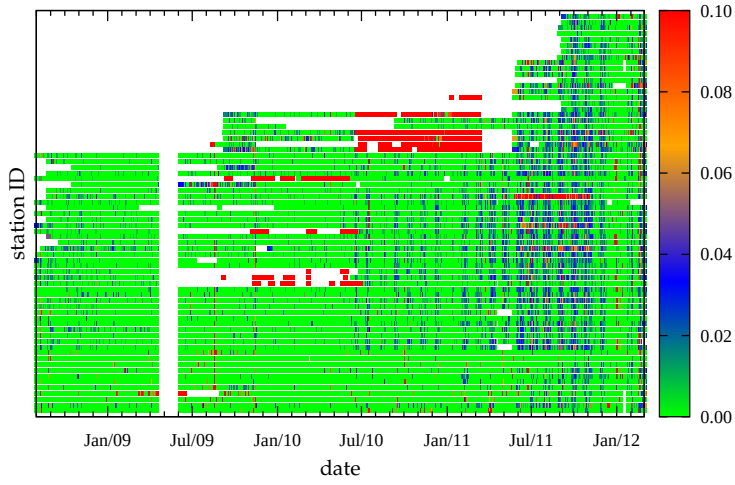
Impact of T3 errors on T3, T4 and T5 events

Before evaluating the effect of T3 errors on potentially lost events at the T4 and T5 trigger levels, the impact of T3 errors on T3 events is presented. After discarding bad periods of data acquisition and removing stations with an excess of T3 errors, as discussed in the previous paragraph, the fraction of T3 events with at least one station with a T3 error was derived, on a daily basis. Plots in figure 6.18 show this fraction as a function of time. The trend in time is consistent with the previous analysis on the T3 error rates on single stations: a first increase in July 2010 and then a much larger increase starting from June 2011 can be seen. Because T3 errors are the source of lost events, the same trend is expected in the number of lost events in the infill array.

Two types of lost events are defined. Events lost at the *T4* trigger level, that comprise the population of events that have triggered the array at the T3 level, but then, due to one or more stations returning T3 errors, they do not fulfill any of the T4 conditions (3ToT or $4C_1$, see section 3.3). An example is shown in figure 6.19a. The event has passed the third trigger level T3 as a ToT2C1&3C2, being thus a T4

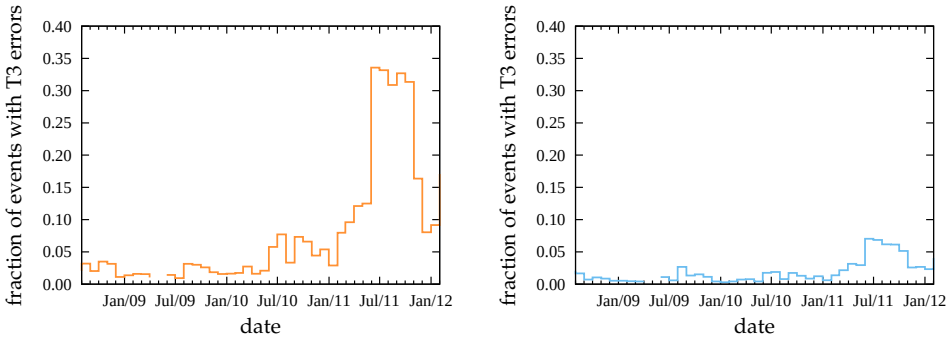


(a) Any W.



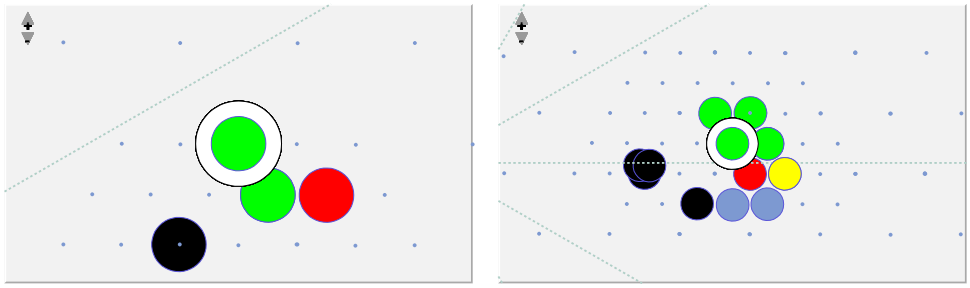
(b) $W = 0$ only.

Figure 6.17: Fraction of T3 error messages for each station in the infill array, obtained on a daily basis.



(a) Any W. **(b)** $W = 0$ only.

Figure 6.18: Fraction of T3 events with at least a station returning a T3 error.



(a) T4 level. **(b)** T5 level.

Figure 6.19: Typical configurations for events lost at the T4 and T5 trigger levels. Stations with a recorded signal are represented by green and yellow circles, while red circles are used for stations returning T3 errors. The hottest station is signaled with a white circle.

event with a 3ToT trigger with high probability. However, the red station is in error and thus data are lost. It is thus said that the event is lost at the T4 level.

The second type of lost events comprise the population of events that have triggered the array at T3 level and, in spite of one or more stations returning T3 errors, they also pass the T4 level. However, they are lost at T5 level, as one of the erroneous stations is close to the hottest station. Events lost in this way are referred as lost *T5* events. An example is shown in figure 6.19b.

To estimate the number of lost events at the T4 level, the approach taken in [100] was used and improved to include 4C1 events. A loop through all events that satisfy the T3 trigger was set, looking for candidates that could have satisfied the 3ToT or 4C1 trigger requirements, but did not because at least one of the stations, that were part of the original SD trigger, returned a T3 error. Such events are labeled as lost T4 events. For both types of trigger, 3ToT and 4C1, the approach is conservative. Indeed, the station(s) that had the T3 error could have been out of time and the event would not have passed the T4 level in any case. As data are lost, the timing information is not available, so it is impossible to determine if the station was in time. This overestimates the number of lost T4 events. Moreover, for 3ToT events the number of lost events is overestimated further, because it is assumed that stations with T3 errors actually satisfied the ToT trigger requirements. All it is known is that those stations satisfied the T2 trigger level, but it is impossible to differentiate between stations with a TH or a ToT trigger. In this sense the estimate is conservative.

To estimate the number of lost events at the T5 level, the following approach was taken. A search through all events that passed the T4 level is performed. If there is a station with a T3 error, it is included into the event as a non-triggered station. If by doing so the T4 event also passes the T5 level that did not satisfied previously, then the event is labeled as lost at the T5 trigger level.

The time evolution of lost events at both T4 and T5 levels is shown in figure 6.20. There is an increase in the fraction of lost events from June 2011, as expected from previous analyses. As seen before, figure 6.18 showed an increase in the fraction of T3 events with errors around this date too. This fact comes as no surprise, as T3 errors are the source of lost events; their correlation is shown in figure 6.21. Note that dead times for the array have been excluded to build these plots. The fraction of lost events at the T4 and T5 level for the entire data set are 16447/948087 (1.7%) and 10535/508158 (2.1%) respectively. To a first approximation, these fractions are added directly, resulting in a net loss of events of 3.8%.

6.4. Conclusions

The exposure of the infill array has been derived based on the information provided by T2 rates on each single station. Its uncertainty comes from two sources: inconsistencies between the rate of each detector and its up-time (0.5%) and the fraction of lost events (3.8%). Thence, the exposure of the apparatus between August 1st 2008 and February 29 2012, integrated for $0^\circ < \theta < 55^\circ$, is 48.1 ± 1.9 km^2 sr yr.

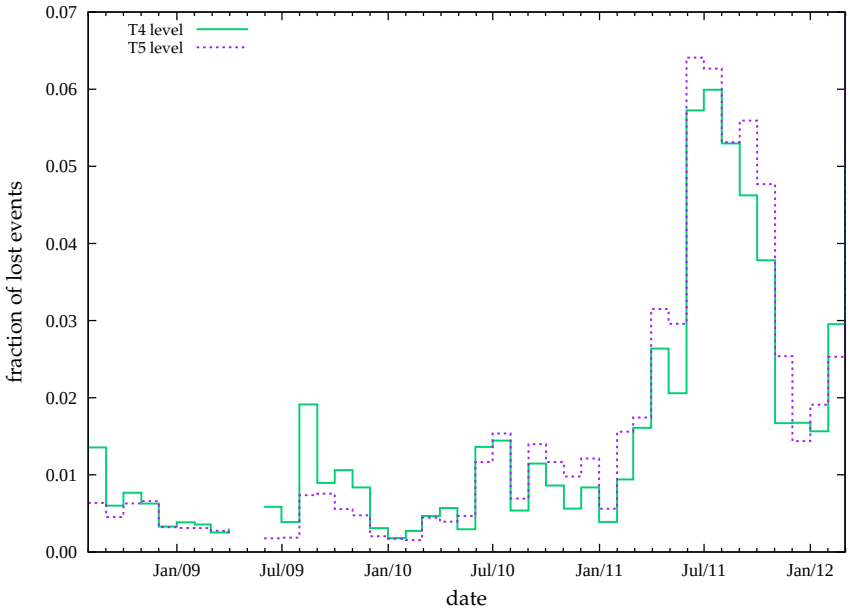
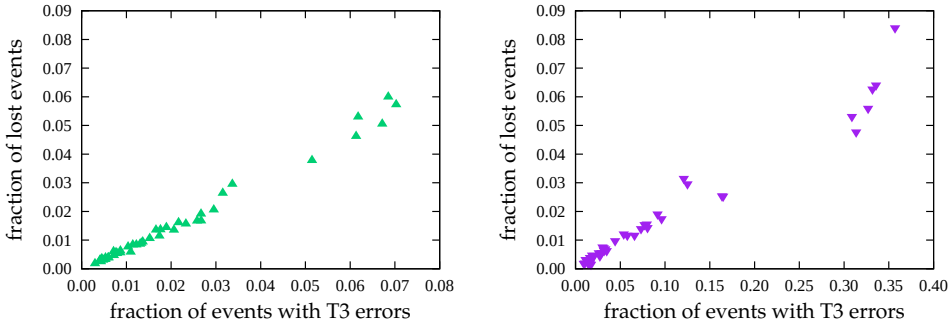


Figure 6.20: Fraction of lost events as function of time.



(a) T4 level.

(b) T5 level.

Figure 6.21: Correlation between the fraction of T3 errors in T3 events and events lost at the T4 and T5 trigger levels.

Chapter 7

The energy spectrum

This chapter presents the methods used to estimate the cosmic ray flux with the infill array. The first section describes the technique employed to correct the S_{450} energy estimator for its dependence on zenith angle. This technique is known as the constant intensity cut. The second section explains the calibration procedure used to assign an energy to each event in the infill array, based on their corrected energy estimator S_{35}. In the last section, the derived cosmic ray spectrum is presented.

The data set employed for the determination of the energy spectrum consists of showers recorded from August 2008 to February 2012. All T5 events were included, with the exception of fake T5 events (section 6.3.1) and a subset of events collected during the communication crisis. Bad periods of data acquisition were also excluded. Furthermore, only events with $\theta < 55^\circ$ were used. The data set selected for the energy calibration is composed by hybrid events (i.e. events simultaneously detected by the FD and the infill array) where several quality cuts are imposed. These will be discussed in the section that describes the energy calibration.

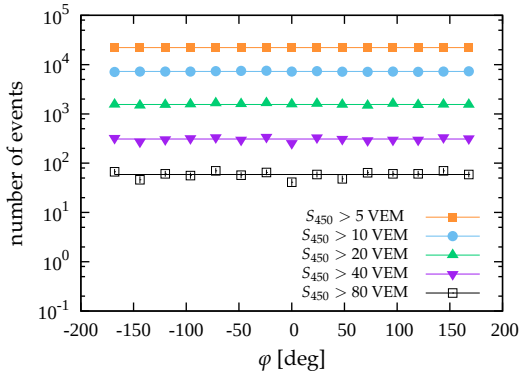
7.1. The constant intensity cut

The method, originally introduced by the MIT group [58], relies on the basic assumption that the flux of cosmic rays outside the Earth's atmosphere is isotropic above a given energy threshold. The assumption means that the number of particles per unit area is independent of arrival direction, and that spectral features are shared among all particles. Even though anisotropies have been measured by a number of different instruments over a wide energy range, the effect is very small and hence immaterial to the main ideas discussed in this section.

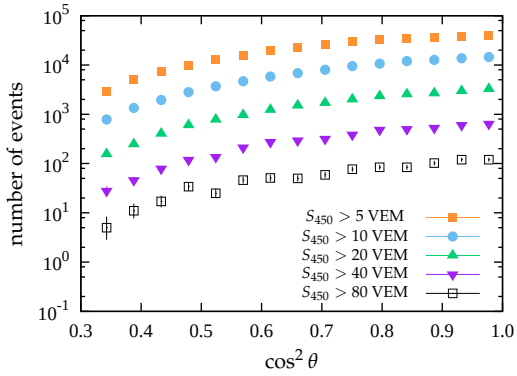
If the flux of cosmic rays in the interplanetary medium can indeed be considered isotropic, then the number of particles N hitting the area A of a surface, during a fixed time interval, can be expressed as

$$\frac{\mathrm{d}N}{\mathrm{d}A} = \text{constant}, \tag{7.1}$$

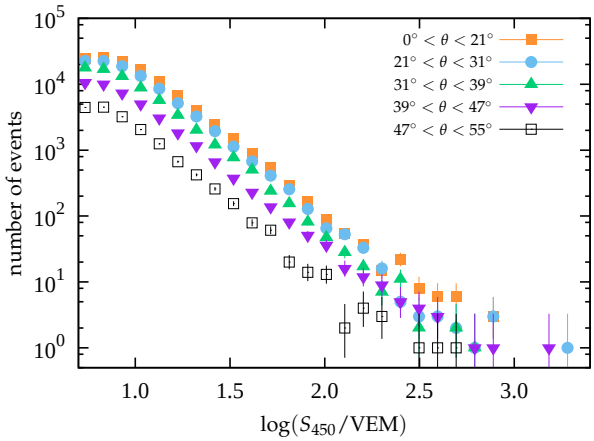
independent of how the surface is oriented in space. When an array of particle detectors deployed on ground is considered, its effective area A_e decreases with



(a) Number of events as a function of the azimuth angle φ.



(b) Number of events as a function of $\cos^2\theta$.



(c) The S_{450} spectra derived for angular bins of equal exposure.

Figure 7.1: The rate of events as function of the zenith and azimuth angles is shown in the top panels. Showers have been classified according to the measured shower size on ground. The S_{450} spectra for different angular bins is shown in the bottom panel.

increasing zenith angle θ in local coordinates as

$$A_e = A \cos\theta, \tag{7.2}$$

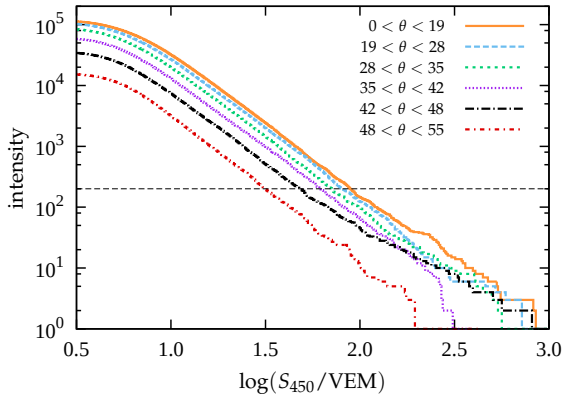
and thus, from (7.1), it is possible to write

$$\frac{\mathrm{d}N}{\mathrm{d}\cos^2\theta} = \text{constant}. \tag{7.3}$$

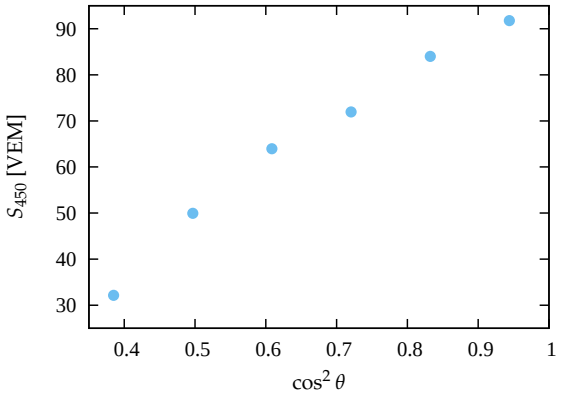
The previous equation implies that the rate of showers initiated by primaries above a given energy must be the same when events are grouped in bins of equal size in $\cos^2\theta$. It is interesting to note that the azimuth angle φ plays no role in the determination of the effective area presented to incoming showers, and thus the rate of events is expected to be uniform as a function of azimuth angles. The plot in figure 7.1a shows the total number of events for the time period considered, as a function of azimuth angle. Showers have been divided according to their size by means of the estimated value of the signal at 450 m away from the shower axis, i.e. their S_{450} value. From that plot it is clear that the number of events is fairly uniform versus azimuth angles, as expected. Instead, figure 7.1b reflects a different situation. In this case, the number of events is far from being uniform versus $\cos^2\theta$ values, even if relatively large showers are considered, thus suggesting a dependence of the shower size estimator S_{450} on the zenith angle θ. The origin of such dependence becomes apparent when considering the attenuation of showers in the atmosphere.

As said, if above a given energy cut the flux of cosmic rays is isotropic the rate of showers generated in the atmosphere should be the same for all zenith angles. Nonetheless, in order to reach a detector placed on ground, showers must traverse different paths in atmosphere according to their inclination, and thus arrive at detection level at different stages in the development of their components. If, for fixed primary energy and mass composition, the altitude of the array is considerably higher than the altitude at which the maximum development of vertical showers occurs, the measured shower size will increase with zenith angle until the slant depth corresponding to the maximum development of inclined showers is reached, and from that point on the shower size will decrease. On the other hand, if the array is located below the altitude of the maximum development of vertical showers, the measured size of showers will steadily decrease with increasing zenith angles. Either way, there is a dependence of the observed shower size on the zenith angle, which is reflected in figure 7.1b for showers recorded with the infill array. For the bulk of observed EAS, the infill array is below the altitude where the maximum development is reached, therefore the number of events above fixed S_{450} values diminishes with increasing zenith angles. As a consequence of the dependence of the shower size estimator on the zenith angle in local coordinates, the measured size spectra will be different for different angular bins, as shown in figure 7.1c.

By exploiting the dependence of the observed shower size on zenith angle, it is possible to study the average development of showers, using the measured signal at ground level and the arrival direction. For this purpose, the definition of *intensity* is useful. The intensity is defined as the number of events above a size threshold



(a) Intensity curves.



(b) Shower signals derived using the constant intensity cut.

Figure 7.2: The plot on the top panel shows the intensity of events as a function of the lower cut value of S_{450}, derived for angular bins of equal exposure. The horizontal line represents a constant cut corresponding to 200 events. The plot on the bottom panel shows the S_{450} signals found by the constant cut applied in 7.2a, versus the $\cos^2\theta$ of the corresponding angular bin.

S_{450}, that arrive per unit of $\cos^2\theta$:

$$I(S_{450}, \theta) = \frac{\mathrm{d}N(S_{450})}{\mathrm{d}\cos^2\theta}. \tag{7.4}$$

Figure 7.2a shows the intensity of events as a function of the shower size threshold S_{450}, for different zenith angle bins of equal dimension in $\cos^2\theta$. If a horizontal line is drawn in that figure, that is, a cut of constant intensity, the attenuation curve of showers can be derived by determining the corresponding shower signal for the $\cos^2\theta$ bin considered. Events with equal intensity in different angular bins correspond to the same primary energy, which allows to deduce the attenuation of showers as previously explained. For example, the attenuation curve for an intensity of 200 events (corresponding to an energy of $\sim 7 \times 10^{17}$ eV) is shown in figure 7.2b.

The attenuation curve allows one to correct the shower size estimator S_{450} for

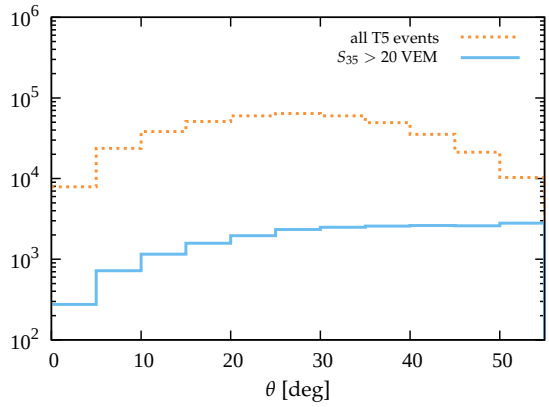


Figure 7.3: Histograms for the distribution of zenith angles for all events in the data set (orange) and events with $S_{35} > 20$ VEM (blue), which correspond to showers belonging to the full trigger efficiency regime for the instrument. The median values for each distribution are 27° and 35°, respectively.

its dependence on the zenith angle θ. It does so by scaling the signal of a shower measured at a given zenith angle, to another value taken at reference angle. In other words, it estimates the signal a shower would have produced if it would have arrived at the reference zenith angle instead of its true zenith angle. The reference angle for showers recorded with the infill array is taken as the value where the zenith angle distribution has its median, for showers detected at the full trigger efficiency. This is 35°, as can be seen in in figure 7.3. The figure shows the distribution of zenith angles for all events and events in the full trigger efficiency regime ($S_{35} > 20$ VEM). The median of the distribution is 27° for all events and 35° for events with $S_{35} > 20$ VEM. The choice of 35° as the reference angle minimizes the impact of the signal correction, as most events arrive with zenith angles close to that value. After the correction is performed, the dependence on the zenith angle should be removed. The estimated signal S_{450} translated to the reference value of zenith angle $\theta = 35°$ is referred as S_{35}, and it is used as the energy estimator of the event.

Several studies on the attenuation curve have been carried out with data recorded using the 1500 m array of the surface detector [101–105], and a few considering data recorded with the infill array [91, 106]. In particular, [105] suggests a model for the attenuation curve of showers that will be used in this work. Such a model postulates a second degree polynomial in $\cos^2\theta$ to describe the measured signal of showers as a function of the zenith angle

$$S_{450} = S_{35} f_{CIC}(\theta) \tag{7.5}$$

where

$$f_{CIC}(\theta) = 1 + a(\cos^2\theta - \cos^2 35°) + b(\cos^2\theta - \cos^2 35°)^2. \tag{7.6}$$

Once the constants in (7.6) are determined, the function $f_{CIC}(\theta)$ is used to correct the S_{450} signal of each single shower as

$$S_{35} = S_{450}/f_{CIC}(\theta), \tag{7.7}$$

which produces the energy estimator of showers independent of their zenith angle.

By applying a constant intensity cut on the S_{450} signal spectra, for angular bins of equal exposure, it is possible to obtain S_{450} signals as function of zenith angle. In practice, this is achieved by classifying events according to their estimated zenith angle, and then, for each zenith angle bin, events are sorted based on their S_{450} signals. Once showers are ordered according to their size, the signal at the k-th position starting from the highest signal is selected in each angular bin, and then paired with the corresponding value of zenith angle. The operation is analogous to that of estimating the quantiles of a distribution. Therefore, the average attenuation curve is built by doing so, at a fixed intensity of k events.

In order to estimate the standard errors on the signals measured for each $\cos^2\theta$ bin, an explicit estimation of the errors has been deduced in [105]. In this work, a simpler approach is taken, originally suggested in [106]. The approach is based on the bootstrap method, that will be briefly discussed in the following paragraphs.

The bootstrap method [107] is a powerful tool for approximating the bias and standard error of an estimator in a complex statistical model. It relies on an approximation of the underlying distribution of an observed sample of a random variable X. The distribution is approximated by means of the empirical distribution of the sample $\mathbf{X} = \{X_1, \ldots, X_n\}$, and thus such an approximation is reasonable if the value of n is relatively large. Otherwise the approximation is not good and the results provided by the method are not dependable.

Given a statistic $T(\mathbf{X})$ depending on the realizations of the random variable X, the main idea behind the bootstrap method is to simulate n draws, made at random with replacement, from the population of n realizations of X in the sample. The fact that draws are made at random with replacement implies that each realization of X in the sample has a probability equal to $1/n$ to be select in any particular draw. Although a single bootstrap sample, or resample, might not contain much information, it is possible to obtain many bootstrap resamples in order to get the sample distribution of the statistic $T(\mathbf{X}^*)$; here the asterisk superscript denotes the bootstrap estimator (although $T(\mathbf{X}^*)$ is referred as the bootstrap estimator, it is not a new estimator but the original estimator T applied on the bootstrap resample $\mathbf{X}^*$). Hence, N bootstrap replicates provide a sample of a N-dimensional vector $\mathbf{T} = \{T_1, \ldots, T_N\}$. If the size of the original sample is large, the distribution of $T(\mathbf{X}^*) - T(X)$ is a good approximation of the distribution of $T(X) - E[T(X)]$. In particular, the standard deviation of $T(\mathbf{X}^*)$ is a good approximation of the standard error of $T(X)$. It is thus possible to apply the bootstrap method to determine the errors on the estimated signals for the attenuation curve.

For the estimation of the standard error on the signals in the attenuation curve, signals on each angular bin are considered as realizations of a single random variable following a certain distribution. This is necessary because in order for the bootstrap method to work properly, it is required that the resampled realizations come from independent and identically distributed random variables, since their empirical distribution is used to approximate the real, but unknown, underlying distribution common to all realizations. Then, for each of the zenith angle bins, signals are resampled, at random with replacement, and the resulting bootstrap sample is used to estimate the value of the signal, as in the original sample. The procedure is

Parameter	Value	Uncertainty
S_{35}	47.4	0.4
a	1.55	0.05
b	-1.12	0.16

Table 7.1: The estimated values of the constants in (7.5), using an attenuation curve obtained for an intensity cut of 300 events and 10 angular bins.

repeated until N bootstrap samples are obtained in each angular bin, and therefore, N values of the bootstrap estimator for the signal. Finally, the standard deviation of these N bootstrap estimators is used as an approximation of the standard error of the signal for the angular bin considered. The sample standard deviation is calculated by means of

$$\sigma = \sqrt{\sum_{i=1}^{N} \frac{\left(X_i - \overline{X}\right)^2}{N-1}} \tag{7.8}$$

for the bootstrap estimators. The number of bootstrap resamples N for each angular bin is set to 5000.

Once the errors on the signal estimates have been determined, it is possible to apply a fit to the attenuation curve, using the model in formula (7.5). The fit is performed by means of the minimization of a χ^2 function defined as

$$\chi^2 = \sum_i \left(\frac{S_i - S_{450}(\theta_i)}{\sigma_i} \right)^2, \tag{7.9}$$

where S_i is the signal value estimated for the attenuation curve in the i-th angular bin, $S_{450}(\theta_i)$ is the predicted value given by (7.5) and σ_i is the estimated error of S_i given by the bootstrap method. The cut on the intensity was applied at 300 events (corresponding to an energy of 5.7×10^{17} eV) using 10 zenith angle bins. The resulting values for the constants are shown in table 7.1.

The derived attenuation curve is shown in figure 7.4 along with the estimated values for the signals and their uncertainties as estimated by the bootstrap method. The value of the reduced χ^2 for the fit is 1.01, reflecting the good agreement that has been found. The relative systematic uncertainty introduced by the correction, due to the uncertainties in the estimates of the constants in (7.5), is at most 4%.

It is instructive to check the effect the correction has on the estimated signal. Since no correction based on the azimuth angle has been introduced, the number of events across different values of φ should remain uniform, as in the case of uncorrected values (figure 7.1a). After examining figure 7.5a, that shows the number of events per bin of azimuth angle, this seems to still be the case. Instead the distribution of the number of events versus $\cos^2\theta$ values is expected to change: above the trigger saturation threshold for the array (i.e. $S_{35} > 20$ VEM) events appear to be uniformly distributed, as can be seen from figure 7.5b. The figure shows the number of events per bin of $\cos^2\theta$ for different cut values of S_{35}. For events below the saturation threshold the distribution is still not uniform, due to the fact that more inclined showers have a probability significantly less than unity to trigger the apparatus and, therefore, the number of events observed decreases with zenith angle.

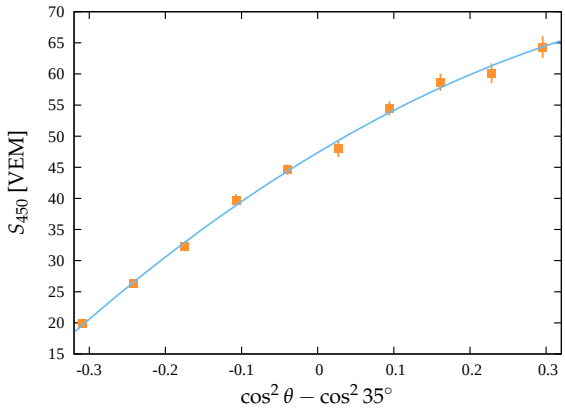
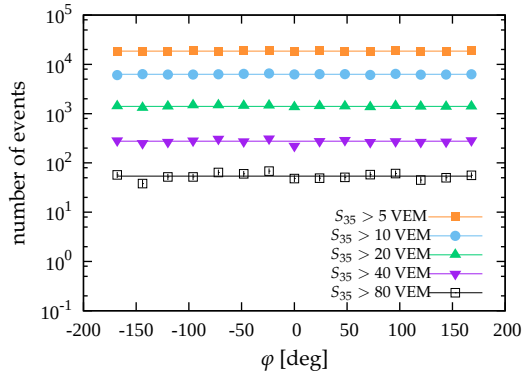


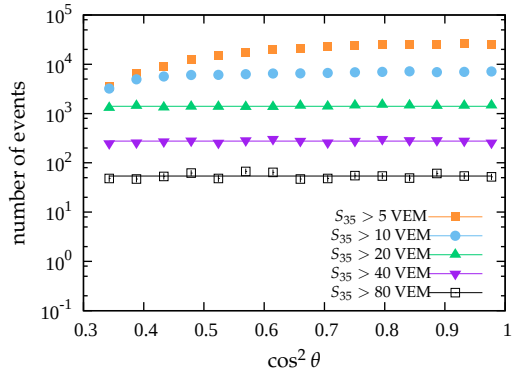
Figure 7.4: The attenuation curve derived for a constant intensity of 300 events. Points represent shower signals, with their uncertainties estimated with the bootstrap procedure. The curve was obtained by a χ^2 minimization using the model given in (7.5).

As a final check, the spectra of the corrected size estimators obtained for different zenith angles fairly agree across angular bins, as shown in figure 7.5c. This plot should be compared with figure 7.1c. The effects of the trigger efficiency explain the large differences among the spectra, for $S_{35} < 20$ VEM.

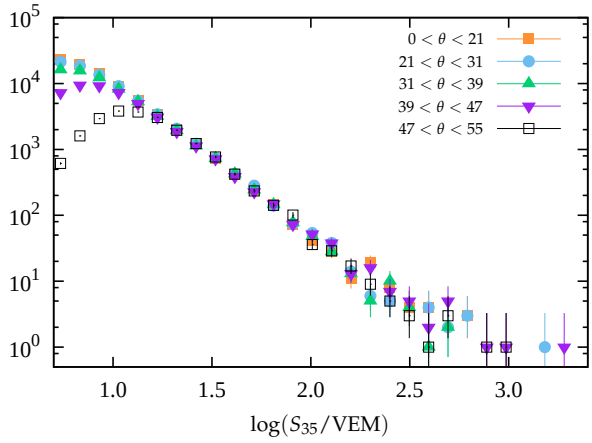
The choice of the specific value for the intensity, on which the constant cut is applied, is not unique. Some issues have to be considered in order to select an adequate value: if the value of the intensity cut is too low, a small sample bias could be introduced due to fluctuations in the intensity curves. Such fluctuations are clearly visible in figure 7.2a, as deviations can easily be observed for low intensity values. On the other hand, setting a value for the intensity that is too high could introduce undesired effects on the attenuation curve due to the acceptance of the array. If such issues are borne in mind, a wide range of possible intensity values still remain as potential candidates to derive an attenuation curve. In order to assess the impact of arbitrarily choosing the intensity at which to apply the cut, attenuation curves were determined for several intensities, between 100 and 1000 events. As an illustration, several intensity curves are shown in figure 7.6. The constants in (7.5) were estimated repeating the minimization procedure already described. The only difference is that, for practical reasons, the number of bootstrap resamples for estimating the uncertainties on the signals was reduced to $N = 1000$. The resulting estimated constants are shown in figure 7.7, as function of S_{35}. The values obtained for different intensities, for S_{35} greater than $\sim$ 20 VEM, are in agreement among them, albeit the estimated errors are correlated. The increase on the uncertainties for larger S_{35} values is caused by the decrease in the intensity, which introduces larger uncertainties in the signals.



(a) Number of events as a function of the azimuth angle φ.



(b) Number of events as a function of the zenith angle θ.



(c) S_{35} spectra derived for angular bins of equal exposure.

Figure 7.5: The rate of events as function of the zenith and azimuth angles is shown in the top panels, for the shower size estimator corrected for its zenith angle dependence. The S_{35} spectra for different angular bins is included in the bottom panel.

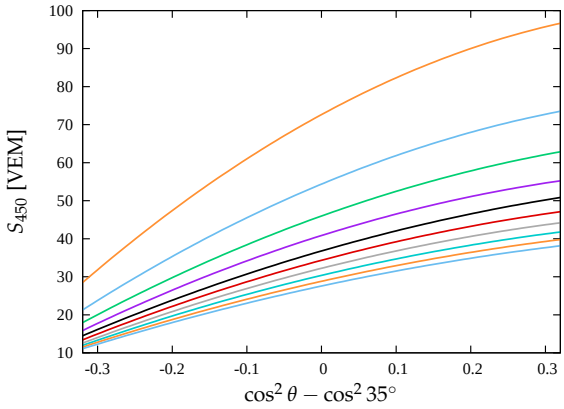


Figure 7.6: The attenuation curves derived for constant intensities. The range of intensities used runs from 100 to 1000 in steps of 100.

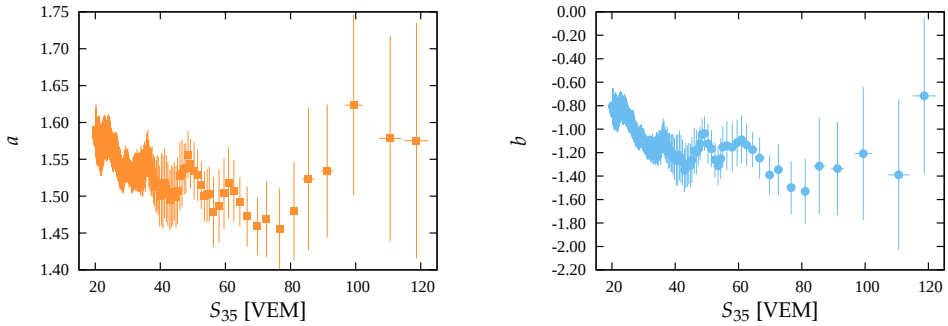


Figure 7.7: Estimated values of the constants in (7.5) vary according to the intensity selected for the constant cut. Values for a (left) and b (right) and S_{35} with their corresponding uncertainties were determined by the fitting procedure. Both a and b become stable for S_{35} values larger than $\sim$ 25 VEM.

7.2. Energy calibration

The energy calibration of infill data is obtained from showers that were simultaneously measured with the fluorescence detector (i.e. hybrid events). The capability of the fluorescence detector to perform a quasi-calorimetric energy measurement of showers, almost independent of hadronic interaction models, allows to assign an energy to each event recorded by the infill detector, based on the estimated signal corrected for its dependence on arrival direction (i.e. S_{35}, see previous section). This is achieved by means of a calibration procedure that will be described in this section.

It is thus necessary to select a subset of high quality hybrid events in order to minimize potential biases in the calibration estimates. On the infill detector side, showers detected by the array are required to fulfill the T5 fiducial trigger and to have an estimated zenith angle $\theta < 55^\circ$. On the fluorescence detector side, several constraints are imposed [75, 108]:

- Variations in the atmospheric conditions significantly affect the measurement of the longitudinal profile. Hence time periods with cloud coverage are rejected, as well as requiring a reliable measurement of the vertical optical depth of aerosols.
- In order to suppress spurious triggers from the surface detector, the station on ground used in the hybrid reconstruction must lie within 750 m from the reconstructed shower axis.
- The depth of shower maximum, X_{max} , must be within the field of view of the telescopes.
- The uncertainty on the reconstructed X_{max} must be below 40 g/cm^2 and the uncertainty on energy below 20%.
- The reduced χ^2 of the profile fit should be less than 2.5.
- The reduced χ^2 of the profile fitted with a Gaisser-Hillas function must be lower than the reduced χ^2 of a straight line by at least four units.
- The amount of Cherenkov light in the shower signal must not exceed a threshold value of 50%.
- If showers are detected by more than one telescope, some of the fluorescence light might be lost at the boundaries between telescopes. If the 'hole' in the shower track is larger than 20% of the total length, the event is rejected.
- The cuts imposed on X_{max} introduce a selection bias caused by a systematic undersampling of the true X_{max} distribution. In order to correct for this effect a subset of showers is selected using fiducial volume cuts based on the shower geometry, dismissing geometries for which an unbiased measurement is not guaranteed. The fiducial volume cuts are described in detail in [38].

The number of events that survive the imposed cuts is 50.

The energy estimate given by the fluorescence detector has an overall systematic uncertainty of $\sim$ 22%. The contributions to this uncertainty come from several different sources, which are then added in quadrature to obtain the total systematic uncertainty. The sources include the fluorescence yield [109], the calibration of the telescopes, the invisible energy correction [110], the reconstruction algorithm used to determine the longitudinal profile and atmospheric effects [97].

The intensity of fluorescence light with respect to the total energy deposit in the atmosphere is known as the fluorescence yield, and presently has an associated systematic uncertainty close to 14%. The fluorescence light emission along the track of the EAS is converted into energy deposit by using the absolute fluorescence yield in air in the 337 nm band of 5.05 ± 0.71 photons/MeV of energy deposited [109]. The yield is known to be dependent on atmospheric conditions such as the air humidity, pressure and temperature. The pressure dependence has been recently measured by the AIRFLY experiment [111].

The absolute optical calibration of the telescopes is another source of systematic uncertainties in the energy estimates. The uncertainty associated with the calibration performed with the 'drum' technique is around 9.5%.

The fluorescence detector underestimates the energy of each shower it observes. This 'lost' energy is carried away by high energy muons and neutrinos, which are invisible for the telescopes. The fraction of energy carried away by those particles is estimated using simulations of showers in the atmosphere, and it is thus dependent on hadronic interaction models. The systematic uncertainty due to the missing energy correction was determined to be close to 4%.

Different reconstruction algorithms have been used to determine the longitudinal profile of showers of a set of high quality hybrid events. The standard algorithm was compared against the so called 'spot method' algorithm, and as a result energy differences of $\sim$ 10% have been found. Therefore a systematic uncertainty of 10% has been assigned to the reconstruction procedure.

The final source of systematic uncertainties is related to atmospheric conditions. These uncertainties include those related to the measurement of the aerosol optical depth (5% - 7.5%), phase function (1%), wavelength dependence (0.5%), atmospherical variability (1%) and the residual uncertainties of the estimation of pressure, temperature and humidity dependence of the fluorescence yield (1.5%). The systematic uncertainty associated to these effects is 6% - 8%.

7.2.1. Event selection

As said, the calibration curve is inferred from the correlation between the S_{35} values as estimated for the infill array, and the energy estimates given by the fluorescence detector. Note that for $S_{35} < 20$ VEM, the infill array is not fully efficient. This has to be taken into account in order not to bias the constants determined by the calibration procedure. Several approaches have been considered inside the collaboration [75, 104, 105, 112], and for this work the elliptic cut, originally proposed in [75], was applied.

As previously discussed (section 5.1.3), the cut relies on the hypothesis that measurements follow a Gaussian distribution with known deviations. The basic idea is

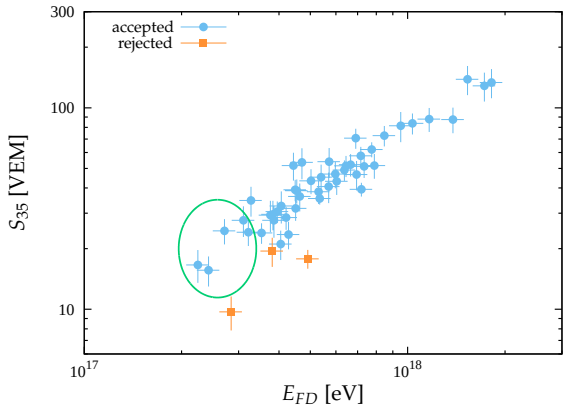


Figure 7.8: The correlation between S_{35} and energy for hybrid events. Showers rejected by the ellipse cut are denoted by orange squares, while accepted ones are denoted by blue circles.

to include showers near or at the verge of the full trigger efficiency of the array, provided they lie inside the 95% CL ellipse centered at the cut value (E^c_{FD}, S^c_{35}); low energy showers outside the ellipse are rejected. Once the value of the cut has been determined, the uncertainties on each component are taken as the uncertainties for S^c_{35} and E^c_{FD}, separately.

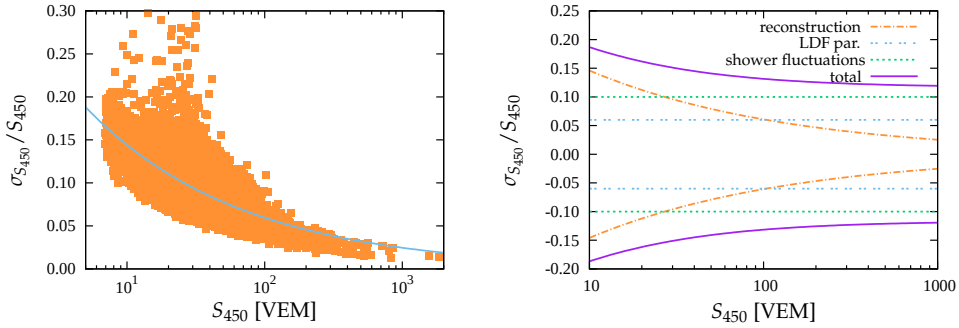
The actual value of (E^c_{FD}, S^c_{35}) is estimated following an iterative approach. As a first step, the calibration curve is fitted to data without applying any cut on events, thus obtaining a strongly biased set of constants. From studies employing pure Monte Carlo techniques, as well as studies relying on data only (section 6.1), the value of $S_{35} \approx 20$ VEM has been established as the minimal shower signal for the infill array to be fully efficient in the angular range comprised between 0° and 55°, and, therefore, the cut value for the signal is set to $S^c_{35} = 20$ VEM. By means of the calibration curve that was obtained in the first step, the value of E^c_{FD} is set providing an initial estimate of the cut point (E^c_{FD}, S^c_{35}). Then, the elliptic cut is imposed using the respective uncertainties associated to each of the point components, and a second calibration curve is obtained from the surviving events. The process is repeated until the estimated values of the constants in the calibration curve settles, which occurs after very few iterations. Events rejected by this procedure are shown using orange squares in figure 7.8, while events used in the energy calibration are represented by blue circles. The total number of events used in the calibration is 47.

The contributions to the uncertainty on the S_{35} signal are not constant, and come from three different sources: the event reconstruction, the parametrization of the lateral profile and shower to shower fluctuations [105]. The contributions to the uncertainty coming from the zenith angle correction (section 7.1) are of the order of a few percent and are neglected.

The uncertainty emerging from the event reconstruction is parametrized using the following model

$$\sigma_{S_{450}} / S_{450} = \alpha S^{\beta}_{450}. \tag{7.10}$$

The constants are determined by a fit applied to data, where the values of S_{450} and



(a) The relative uncertainties on S_{450} as given by the reconstruction procedure for the infill array. The solid line represents the best fit for model (7.10).

(b) The relative uncertainty on the S_{450} energy estimator as considered in this work.

Figure 7.9: The relative uncertainties on the energy estimator in the infill array, as used in the calibration procedure.

$\sigma_{S_{450}}$ are taken directly from the reconstruction procedure. In figure 7.9a, the relative uncertainty is plotted against the value of S_{450}, along with the curve that gave the best fit. The resulting values of the constants are $\alpha = 0.35 \pm 0.01$ and $\beta = -0.38 \pm 0.01$.

The relative uncertainty coming from the parametrization of the lateral distribution is assumed constant

$$\sigma_{LDF}/S_{450} = 0.06, \tag{7.11}$$

while the shower to shower fluctuations are taken as

$$\sigma_{sh}/S_{450} = 0.1, \tag{7.12}$$

based on results obtained using detailed simulations of shower development in the atmosphere, studying fluctuations of the electromagnetic and muonic components separately [113]. The relative uncertainties on S_{450} are depicted in figure 7.9b. Finally, the relative uncertainty for the energy estimate given by the fluorescence detector is considered to be 0.08 ± 0.03 and constant with energy [75].

7.2.2. The fitting procedure

The model used to explain the correlation between the FD energy estimate E_{FD} and S_{35} is

$$S_{35} = 10^a E_{FD}^b, \tag{7.13}$$

where the constants a and b are estimated using a linear regression. For this purpose, the formula (7.13) can be written as

$$S = X\beta + \epsilon, \tag{7.14}$$

using matrix notation. Here S stands for a $n \times 1$ vector of $\log(S_{35})$ estimates given by the infill reconstruction, $\beta' = (a\ b)$ is the $1 \times p$ vector of constant parameters to

Parameter	Value	Uncertainty
a	-1.07	0.15
b	0.984	0.052

Table 7.2: The estimated values of the constants in (7.13).

be estimated, ϵ is the $n \times 1$ vector of disturbance terms with expectation $E[\epsilon_i] = 0$ and variance $V[\epsilon_i] = \sigma_i^2$ and X is the $n \times p$ design matrix given by

$$X = \begin{pmatrix} 1 & \log E_{FD_1} \\ \vdots & \vdots \\ 1 & \log E_{FD_n} \end{pmatrix}, \tag{7.15}$$

where E_{FD_i} is measured in PeV. In this case, $n = 47$ and $p = 2$. The vector of constants is estimated employing

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}S \tag{7.16}$$

with a covariance matrix given by

$$\widehat{\text{cov}}(\hat{\beta}|X) = (X'\Sigma^{-1}X)^{-1}. \tag{7.17}$$

Here Σ stands for the diagonal covariance matrix of measurements, since shower measurements are assumed to be independent of each other. For the purpose of determining the calibration curve, Σ is the true covariance matrix, which is given by

$$\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2), \tag{7.18}$$

where σ_i^2 is the variance of the i-th measurement. Since generalized least squares methods do not regard the independent variable as a random variable, they do not have an associated variance. One way to circumvent this obstacle is to include the variance of the independent variable into the variance of the dependent variable by means of the model and the Taylor approximation. Henceforth, the value of σ_i^2 is regarded as

$$\sigma_i^2 \approx \left(\frac{\sigma_{E_{FD}}}{E \ln 10}\right)^2 + \left(\frac{\sigma_{S_{35}}}{S_{35} \ln 10}\right)^2, \tag{7.19}$$

where the constant b has been assumed equal to unity.

The estimated values of the constants in (7.13) are listed in table 7.2, while the correlation between S_{35} and E_{FD} is shown in figure 7.10, together with the curve that provided the best fit. The fit produced $\chi^2/N_d = 1.10$. The inspection of the fit residuals provide further information on the goodness of fit of the calibration curve to recorded data. Residuals are defined as

$$R_i = \frac{S_i - (X\hat{\beta})_i}{\sigma_i} \tag{7.20}$$

for the i-th observed values of S_{35} and E_{FD}. Their dependence on E_{FD} is shown in figure 7.11a, while their distribution is depicted in figure 7.11b. If the uncertainty related to each measurement is correct and (7.13) models data accurately it is expected

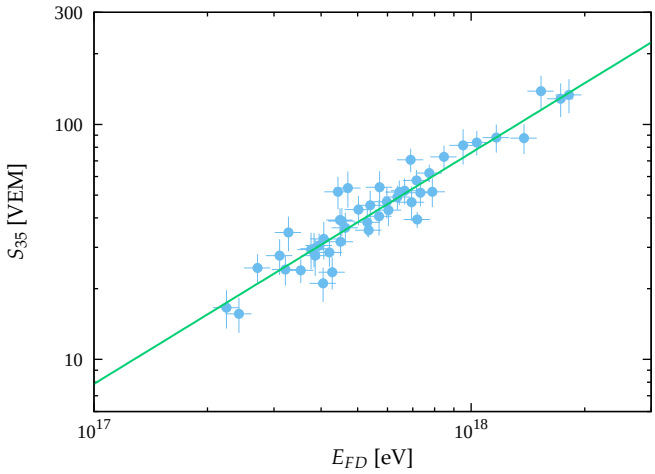
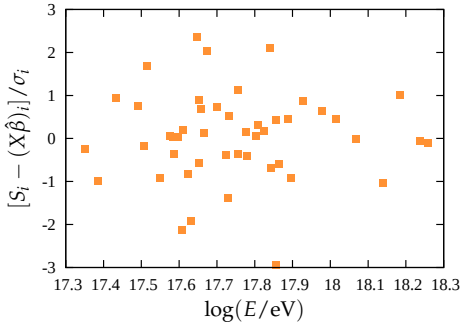
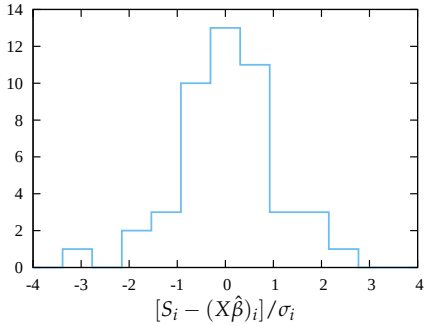


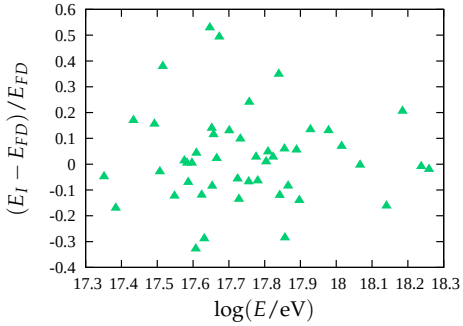
Figure 7.10: The correlation between S_{35} and energy for hybrid events along with the calibration curve estimated by the fit procedure. The reduced χ^2 value of the fit is 1.10.



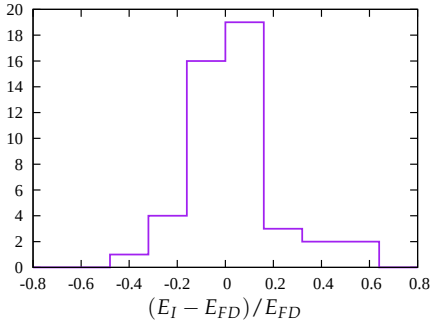
(a) Residuals as a function of E_{FD}.



(b) A histogram of the fit residuals. The mean value is 0.04 ± 0.15 and the standard deviation 1.04 ± 0.26.



(c) Relative differences as a function of E_{FD}.



(d) A histogram of relative differences. The mean value is 0.026 ± 0.026 and the standard deviation 0.178 ± 0.008.

Figure 7.11: Fit residuals (top panels) and the relative differences in energy as obtained from the FD and the infill array (bottom panels).

that the residuals follow a standard normal distribution. This seems to be the case, since the mean value of the fit residuals is 0.04 ± 0.15 and their standard deviation 1.04 ± 0.26 . Furthermore, the Shapiro-Wilk test does not reject the hypothesis of normality with a p value of 0.05.

The scope of the calibration is to transform S_{35} values to energy estimates. The energy estimates E_I given by the infill reconstruction are obtained from

$$E_I = AS_{35}^B. \tag{7.21}$$

The values of A and B are deduced from the estimates of a and b in (7.13) along with their uncertainties, resulting in $A = 12.29 \pm 3.65$ PeV and $B = 1.016 \pm 0.054$. The relative difference between predicted energy values given by the infill reconstruction and the energy estimates given by the FD are shown in figure 7.11c, as a function of the FD energy E_{FD}. The distribution of these relative differences can be observed from plot 7.11d. The standard deviation of the relative differences in energy is regarded as the energy resolution of the infill array, being close to 18%.

7.2.3. Check on the fit estimates

The bootstrap technique, previously used to estimate the uncertainties of the signals for the constant intensity cut, might also be applied to the fit procedure that determines the calibration curve. However, in this case, it will be employed to look for the existence of biases in the estimates of the constants and their uncertainties, given in the previous section, and correct them, if such biases are indeed found.

The main idea behind the method is to perform a sampling with replacement of a set of identically and independently distributed (IID) random variables. Although, at first sight, it is not apparent what this set of variables might be for (7.14), since heterosdatic variables do not have identically distributed disturbance terms, it is possible to transform (7.14) by multiplying it by $\Sigma^{-1/2}$:

$$\Sigma^{-1/2}S = \Sigma^{-1/2}X\beta + \underbrace{\Sigma^{-1/2}\epsilon}_{\delta}. \tag{7.22}$$

This transformed model has $\Sigma^{-1/2}S$ as the response variable, $\Sigma^{-1/2}X$ as the design matrix and δ as the vector of IID disturbance terms. The vector of parameters is still β. The fit residuals in this framework are defined as

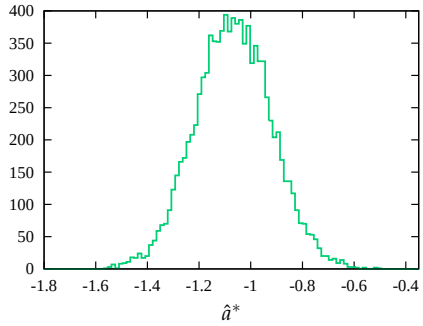
$$R = \Sigma^{-1/2}(S - X\hat{\beta}). \tag{7.23}$$

Note how this definition of residuals matches definition (7.20). The point of applying a transformation to (7.14), to then obtain (7.22), is to modify the disturbance terms so they follow the same underlying distribution. Moreover, the following assumption holds

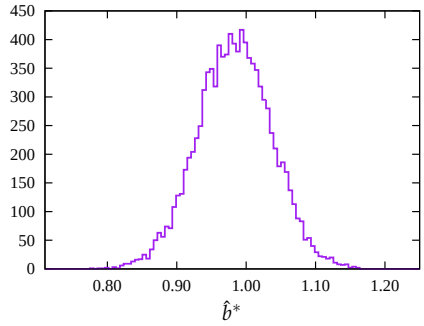
$$E[\delta|\Sigma^{-1/2}X] = 0, \quad \text{cov}(\delta|\Sigma^{-1/2}X) = I, \tag{7.24}$$

where I represents the identity matrix.

Event though the disturbance terms δ are unobservable variables, their value can be estimated by the residuals as defined in (7.23). Since such residuals are IID it is



(a) A histogram of the bootstrap values of $\hat{a}$. The sample mean is -1.0674 ± 0.0015 and the sample standard deviation is 0.1484 ± 0.0003.



(b) A histogram of the bootstrap values of $\hat{b}$. The sample mean is 0.9832 ± 0.0005 and the sample standard deviation is 0.05362 ± 0.00004.

Figure 7.12: Histograms obtained for the bootstrap estimates of a and b in (7.14).

possible to apply the bootstrap principle on them. Therefore, the residuals of the calibration fit $R_1, \ldots, R_n$ can be resampled n times at random with replacement to get the bootstrap disturbances $\delta_1^*, \ldots, \delta_n^*$, in which each δ_i^* behaves like δ_i. The next step is to apply the inverted transformation on the bootstrap vector δ^*

$$\epsilon^* = \Sigma^{1/2}\delta^* \tag{7.25}$$

in order to obtain the corresponding disturbances for (7.14).

Once the bootstrap disturbances are available, the values of the energy estimates are regenerated according to

$$S^* = X\hat{\beta} + \epsilon^*, \tag{7.26}$$

where the design matrix X does not change. Note how S^* follows the regression model postulated in (7.14), because each ϵ_i^* is assumed to follow the same underlying distribution as ϵ_i. Thence, the purpose of the method is to imitate the original model on the computer, with a difference: the 'true' parameter vector is known, as well as the disturbance terms, they are $\hat{\beta}$ and ϵ^*, respectively. In this way, it is possible to obtain a set of bootstrap estimates of $\hat{\beta}$ by repeating the procedure described several times. On each iteration of the algorithm, the bootstrap estimate is calculated as

$$\hat{\beta}^* = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}S^*. \tag{7.27}$$

Following these lines, the bootstrap principle for regression is that the distribution of $\hat{\beta}^* - \hat{\beta}$ gives a good approximation of the distribution of $\hat{\beta} - \beta$, and in particular that the empirical covariance matrix of β^* is a good approximation of the true covariance matrix of $\hat{\beta}$. This is how the presence of biases in the original fit estimates can be investigated.

The bootstrap technique was applied to the data set used in the calibration procedure. The number of bootstrap resamples was set to 10000. The histograms of the values obtained for the bootstrap estimates of $\hat{a}^*$ and $\hat{b}^*$ are shown in figure 7.12, while, in table 7.3, the fit estimates are compared to the results given by the bootstrap technique. As can be seen from that table, no significant biases have been

Parameter	Weighted least squares		Bootstrap	
	Value	Uncertainty	Mean	SD
a	-1.072	0.145	-1.067	0.001
b	0.9838	0.0525	0.9832	0.0005
σ_a	0.1453	-	0.1484	0.0003
σ_b	0.0525	-	0.05362	0.00004

Table 7.3: The fit estimates of the calibration constants compared to the bootstrap results.

found in the fit estimates. The uncertainties on $\hat{a}$ and $\hat{b}$ are slightly biased (around 2% in both cases). Since these biases are quite small they are neglected.

7.3. The energy spectrum

The differential energy flux of cosmic rays J as measured with the infill array is calculated according to

$$J = \frac{\mathrm{d}N}{\mathrm{d}E\mathrm{d}\varepsilon}, \tag{7.28}$$

where $\mathrm{d}N$ is the number of particles arriving with energies $(E, E + \mathrm{d}E)$ and ε is the exposure of the instrument, determined by its acceptance a and up time t (see chapter 6). Since the energy spectrum is derived for the full efficiency regime of the infill array, a is independent of primary energy and mass composition.

As explained in previous chapters, the exposure calculation is based on the determination of the geometrical aperture and of the observation time. T2 files provide information on both elements. For the determination of the observation time, dead times are taken into account: bad periods are evaluated through an empirical technique based on the distribution of arrival times of events [99]. Finally, the uncertainties in the aperture and observation time have been studied and are well known (see chapter 6 for further details).

Showers have been included in the spectrum provided they satisfy the fiducial trigger in the infill array (T5), an estimated energy above 3×10^{17} eV ($S_{35} \sim 25$ VEM) and a zenith angle below 55°. The resulting energy spectrum in the energy range spanning $10^{17.5}$ - $10^{19.5}$ eV is shown in figure 7.13. The number of events in each energy bin is noted in the figure. The uncertainties depicted are statistical only, under the assumption they follow Poisson distribution. Corrections to the estimated spectrum due to the finite energy resolution of the array were not applied in this thesis.

To examine the shape of the spectrum, two functions were adjusted to data employing a maximum likelihood approach. The first function is a pure power law with a single spectral index, and the second function is a broken power law with a dynamic break point. The fit results are listed in table 7.4 and the spectrum, together with both functions, is shown in figure 7.14. The likelihood ratio test for these two models gives

$$D = -2 \ln \frac{L_{pure}}{L_{broken}} = 4.34. \tag{7.29}$$

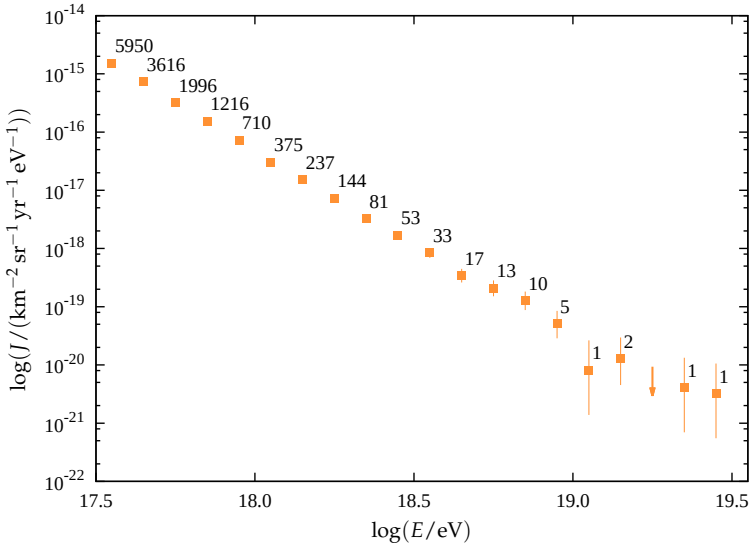


Figure 7.13: The energy spectrum as derived for the full trigger efficiency regime in the infill array. The number of events per bin is indicated.

Pure power law		
Parameter	Value	Uncertainty
Spectral index γ	3.32	0.02
Broken power law		
E_{break}	2.07×10^{18} eV	0.79×10^{18} eV
Spectral index γ_1 ($E < E_{break}$)	3.34	0.02
Spectral index γ_2 ($E > E_{break}$)	3.07	0.15

Table 7.4: The estimated values of the spectral indexes for different models.

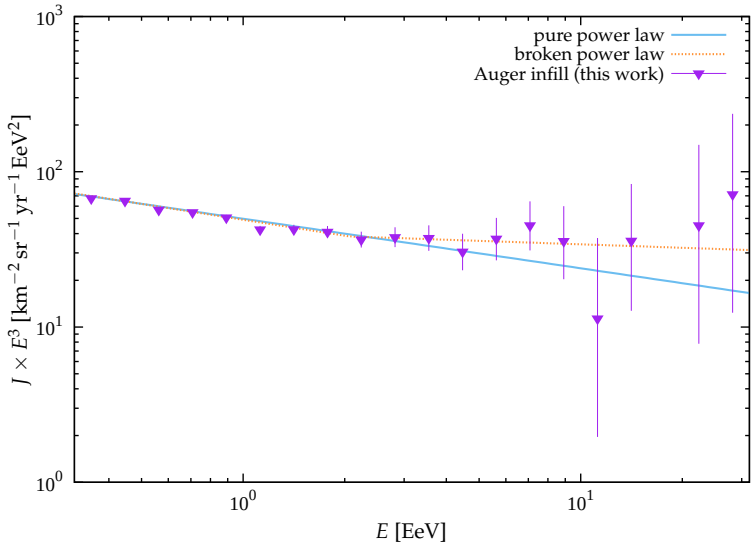


Figure 7.14: The energy spectrum as derived in this work. Note that fluxes have been multiplied by E^3. Fits have been applied using a pure power law and a broken power law, shown in the figure as lines.

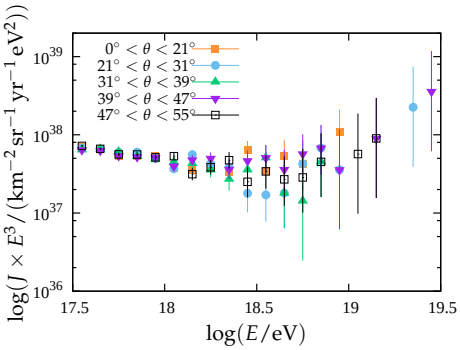
Angular range (θ)	Mean	SD	Normality test ($p = 0.05$)
0° - 21°	0.10 ± 0.28	1.06 ± 0.31	not rejected
21° - 31°	0.01 ± 0.31	1.25 ± 0.49	not rejected
31° - 39°	-0.34 ± 0.25	0.94 ± 0.26	not rejected
39° - 47°	0.10 ± 0.22	0.93 ± 0.25	not rejected
47° - 55°	0.22 ± 0.30	1.21 ± 0.53	not rejected

Table 7.5: The mean and standard deviation of the residual population for spectra derived using different angular bins. The results of the Shapiro-Wilk normality test are shown in the last column.

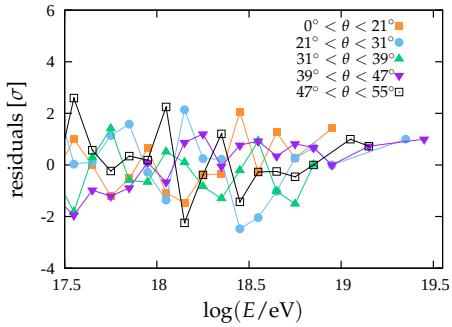
Since the difference in the number of free parameters between models is 2, D is expected to approximately follow a χ^2 distribution with 2 degrees of freedom. Therefore the pure power law hypothesis is not rejected, using a p value of 0.05.

7.3.1. Consistency checks

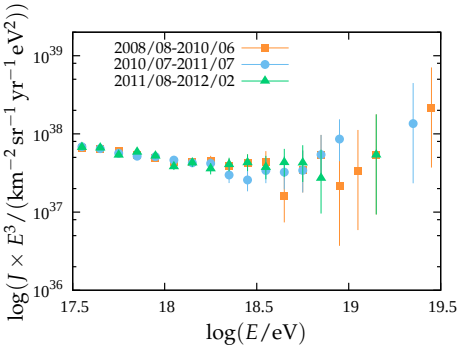
It is interesting to check the consistency between spectra derived using different angular bins of equal exposure. The obtained spectra for five angular bins are shown in figure 7.15a. No large deviations are apparent from that plot. Residuals between these spectra and the one calculated for $0° < \theta < 55°$ are shown in figure 7.15b. Since the population of events in any of the angular bins is a subset of the population of events with $0° < \theta < 55°$, their correlation has to be considered while determining such residuals. As can be seen from the residual plot, no significant deviations have been found. For the residuals in each angular bin, the sample mean and standard deviation have been determined. Moreover, the Shapiro-Wilk test for



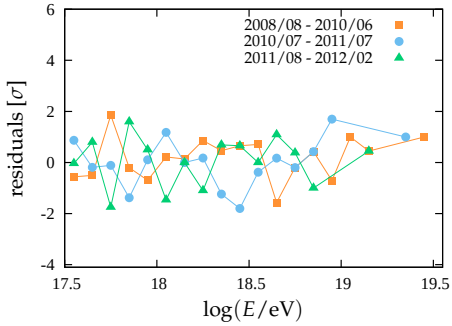
(a) Spectra derived using five angular bins of equal exposure.



(b) Residuals of derived spectra, using different angular bins, as a function of energy.



(c) Spectra derived using three non-overlapping time periods in the entire data set.



(d) Residuals of derived spectra for different periods of time.

Figure 7.15: Spectra has been derived using different angular bins with data from the infill array. Their residuals with respect to the spectrum depicted in figure 7.13 are shown in the panels on the right.

Time period	Mean	SD	Normality test ($p = 0.05$)
2008/08 - 2010/06	0.19 ± 0.19	0.81 ± 0.21	not rejected
2010/07 - 2011/07	0.02 ± 0.23	0.94 ± 0.28	not rejected
2011/08 - 2012/02	0.05 ± 0.25	0.97 ± 0.27	not rejected

Table 7.6: The mean and standard deviation of the residual population for spectra derived using non-overlapping time periods. The results of the Shapiro-Wilk normality test are shown in the last column.

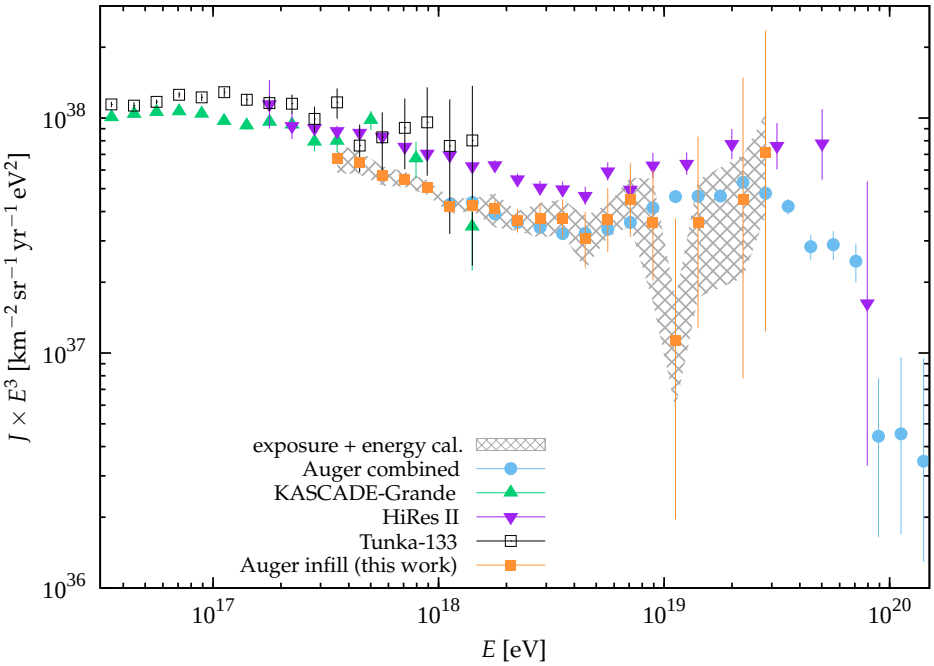


Figure 7.16: The Auger infill spectrum as derived in this thesis is shown, together with spectra obtained by the Tunka-133 experiment [114], KASCADE-Grande [43], HiRes [115] and the combined spectrum from the Auger collaboration [74]. Note that fluxes have been multiplied by E^3. Systematic uncertainties for the Auger infill spectrum, arising from the energy calibration and exposure calculation, are represented by a filled band.

normality has been applied on each sample, using a p value of 0.05. The results are listed in table 7.5. As can be observed, the residual distributions are fairly consistent with the standard normal distribution.

A further consistency test was performed. The data set used to derive the spectrum was divided in three periods of time in such a way that the number of events in each subset was the same. Using these subsets, spectra has been determined for each time period. These are shown in figure 7.15c. The residuals obtained with respect to the spectrum determined with the entire data set are depicted in figure 7.15d. Once again, the sample mean and standard deviation were calculated for the residuals of each time period and the Shapiro-Wilk test for normality has been applied too. The results are listed in table 7.6. As can be observed, the residual distributions are fairly consistent with the standard normal distribution.

7.4. Conclusions

After performing the energy calibration for showers falling in the infill array of the Pierre Auger Observatory, the cosmic ray spectrum has been obtained. Using this spectrum, two models were used to fit data: a pure power law and a broken power law. Within the current statistics, both models are found to fit data well. Moreover,

the broken power law does not provide a significantly better fit.

For comparison purposes, the energy spectrum derived in this work is shown in figure 7.16, together with the spectrum obtained by the Tunka-133 experiment [114], KASCADE-Grande [43], HiRes [115] and the combined spectrum from the Auger collaboration [74]. The filled region in that figure represents the systematic uncertainties related to the Auger infill spectrum arising from two sources: the exposure determination and the energy calibration procedure. The former has been estimated to be 3.8% (see chapter 6) while the latter was calculated using the estimated values given by the calibration procedure, together with their covariance matrix, by means of a simple Monte Carlo algorithm. Several spectra were generated using different values for the calibration constants, where such constants were sampled from a bidimensional Gaussian distribution. This way the spread in the differential spectra obtained was taken as an estimate of the uncertainties related to the calibration procedure.

In the overlapping energy region, the spectrum obtained in this work and the combined Auger spectrum are consistent with each other within statistical and systematic uncertainties. In the low energy region, the transition from the spectrum obtained by the KASCADE-Grande experiment and the infill spectrum is rather smooth. On the other hand, systematic differences between the infill spectrum and those derived by HiRes and Tunka-133 are evident. However, such differences are within the systematic uncertainties claimed by groups responsible of the experiments. In the case of Auger, and in particular the infill array, systematic uncertainties on the FD energy used to calibrate data are around 22%, as discussed in section 7.2. Note that this uncertainty is related to the energy estimates given by the FD and it is independent of the uncertainties arising from the energy calibration procedure, described few lines above. The systematic uncertainty on the energy estimates given by HiRes fluorescence detector is close to 17%, arising from several sources. Experiments such as Tunka-133 and KASCADE-Grande rely on Monte Carlo methods to calibrate their data, which in turn depend on the hadronic interaction models used to simulate EAS. This introduces uncertainties on their energy estimates since hadronic interactions at these energies are poorly understood. Although systematic differences are found between spectra, it is presently believed that these are mainly caused by differences in the energy scale. If an energy rescaling is applied, the spectral normalizations and shapes are very consistent between experiments.

Chapter 8

Conclusions

The combined spectrum published by the Auger collaboration extends from 10^{18} eV to the highest energies. The main goal of this thesis is the measurement of the energy spectrum of cosmic rays down to $10^{17.5}$ eV. This is the first step to extend the Auger spectrum in a region where the transition from galactic to extragalactic cosmic rays component is expected to occur. For this purpose, data from the new 750 m infill array have been employed. To get to the spectrum, different aspects have been studied in the thesis, from the reconstruction of the events, through the precise determination of the aperture of the array, up to the determination of the primary energy for each event.

From the event reconstruction viewpoint, the lateral distribution function (LDF) used to characterize the lateral profile of showers has been studied (chapter 4). It has the same functional form as the one used for showers detected in the main SD array. It has, however, been modified in accordance to the optimum ground parameter for the infill array, i.e, the particle density at an optimal core distance that is then used as energy estimator. The optimal core distance for the infill is 450 m. The slope parameterization employed in the LDF has been optimized to better describe data, employing a subset of high quality showers registered in the infill array. Using this new parameterization, the performance of the LDF was assessed. The analysis of residuals, χ^2 statistics and experimental mean lateral distribution showed a good agreement between data and the LDF utilized, particularly in the range of distances where the energy estimator is determined, i.e., at 450 m from the shower axis.

As a check of the quality of the event reconstruction in the infill array, two comparisons were performed (chapter 5). The first comparison was done using the set of showers that were simultaneously detected by both the infill and the main arrays. Three observables were considered: the arrival direction, the core position on the detection plane and the signal measured at 450 m from the shower axis. In spite of the fact that the great majority of showers had very low station multiplicity in the main array, no systematic disagreements were found for any of the observables studied. The second comparison followed the same lines as the previous one, but it was performed employing the hybrid data set: showers simultaneously detected by the infill array and the fluorescence telescopes. In this case two observables were considered: the arrival direction and the core position on the detection plane. Since

the hybrid reconstruction is mainly based on the FD reconstruction algorithm, and only exploits timing information from one surface station, estimates given by both reconstructions are fairly independent of each other, providing an excellent handle on the quality of the infill reconstruction. As in the previous comparison, no systematic disagreements were found. The results obtained from the performed checks thus suggest that the event reconstruction in the infill array is under control.

To measure the flux of cosmic rays, it is essential to define a set of events and the associated aperture of the array. In order to select showers detected by the infill array, the trigger definitions applied offline on data were adapted from the original ones employed in the main surface detector. Both the physics trigger (T4) and the fiducial trigger (T5) as well as the station selection algorithms were scaled down to work properly with the tighter spacing of the new array (750 m versus 1500 m in the main array). The adopted trigger chain makes the array fully efficient above a certain energy: through Monte Carlo simulations, this is found to be 3×10^{17} eV for events with zenithal angle less than 55°. These values fix the data set. In chapter 6, two experimental tests were employed to verify this finding. The first test was based on the isotropy of the arrival directions of showers, relying on the same idea behind the constant intensity cut. The second test was based on a study of the distribution of the core positions on the detector plane. The results provided by both tests are in agreement with what has been found with Monte Carlo methods.

The procedure used to determine the exposure of the infill array has also been explained in chapter 6. The calculation relies entirely on the trigger rates reported by each single station in the array. In spite of the fact that this technique provides a reliable way to estimate the exposure of the detector, errors that occur while stations communicate with the Central Data Acquisition System (CDAS) introduce uncertainties in the calculations. These errors result in the loss of physics events. The impact of such losses on the exposure calculation was evaluated, and a conservative estimate of the associated uncertainty was derived. This uncertainty has been found to be close to 3.8%. Moreover, a number of inconsistencies between station trigger rates and their participation to data were found. Such inconsistencies were found to cause a misclassification of events to be used in the data set. These events have been identified and discarded from physics analyses. The study done in this chapter has thus provided the ingredients for the determination of the spectrum, i.e., the data set on one side and the exposure (with its uncertainty) on the other one.

Finally, in the last chapter of this thesis the cosmic ray energy spectrum from infill data has been derived. The first part of the chapter has been devoted to the correction that must be made on the energy estimator before proceeding to the energy calibration. The correction has been performed by means of the attenuation curve that was derived using the constant intensity cut method. Once the correction has been applied, the calibration curve has been determined, using a subset of high quality events simultaneously detected by the infill array and the fluorescence telescopes. The signal at 450 m from the shower axis has been used as the energy estimator of events in the infill array, while the FD telescopes provided the energy estimate for showers in a way almost independent of hadronic interaction models. Once the calibration curve has been determined, the spectrum has been derived. Two models were used to fit the shape of the spectrum: a pure power

law and a broken power law with a dynamic break. Within the current statistics, both models are found to fit data well. In particular, the broken power law used to fit the spectrum does not provide a significantly better fit than the pure power law. The spectral index estimated for the latter is $\gamma = 3.32 \pm 0.02$. After performing some consistency checks on the spectrum, it is compared with results given by other experiments. The spectra are in good agreement, within systematic uncertainties. In the overlapping energy region, the spectrum obtained in this work and the combined Auger spectrum are fairly consistent with each other. In the low energy region, the transition from the spectrum obtained by the KASCADE-Grande experiment and the infill spectrum is rather smooth. The fact that the spectrum derived with infill data is consistent both in the low and high energy regions with previous measurements is encouraging and sets the stage for future measurements that will include the HEAT. This will enable the observatory to perform a reliable measurement of the cosmic ray energy spectrum from $\sim 10^{17}$ eV up to $\sim 10^{20}$ eV using four instruments (the SD and its infill array plus the FD and the HEAT). This thesis is a first step to make that measurement possible.

Agradecimientos

Primeramente quiero mencionar al profesor Oscar Saavedra, quién me dió la oportunidad de continuar mis estudios en el campo de los rayos cósmicos. Por esto le estoy profundamente agradecido.

Habría sido imposible culminar esta tesis sin la colaboración y guía de Lino Miramonti y Piera Ghia, los supervisores de mi trabajo de doctorado, quienes siguieron atentamente mis pasos a lo largo de los tres años en los que estuve involucrado con el observatorio Pierre Auger. Ambos compartieron conmigo su experiencia y conocimientos de física experimental. Junto a Andrea Chiavassa y Viviana Scherini, me brindaron el apoyo técnico y soporte anímico necesarios para concluir este trabajo. Sus comentarios y observaciones fueron imprescindibles. Muchas gracias a todos ellos.

Agradezco también a Antoine Letessier Selvon y a la gente del grupo Auger del LPNHE por recibirme en su laboratorio y permitirme participar en un ambiente de trabajo tan dinámico y productivo.

Finalmente, mi cariño y agradecimientos más sinceros van para mis padres, hermanos, familia y amigos, dentro y fuera de Bolivia, que estuvieron a mi lado siempre que los necesité y me dieron tantos buenos momentos juntos.

Appendix A

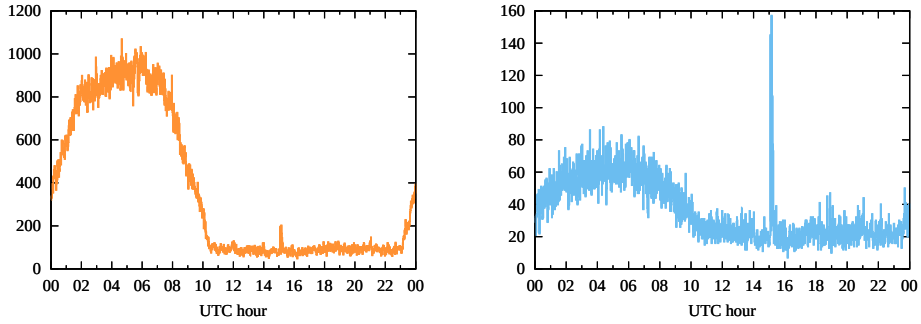
T3 errors and lost events

A.1. T3 errors

The analysis of the fraction of T3 errors in section 6.3.2 showed an increase for most stations in the infill array, for certain periods of time. A general trend can be seen from figure 6.17, where the relative number of returned T3 errors rises with time. From June 2010, the increase in the fraction of T3 errors is apparent and shows a modulation in time from that point on. That date roughly corresponds with the start of data acquisition of one of the Auger extensions, the High Elevation Auger Telescopes (HEAT, see section 3.1) To understand if the effect is actually due to the increase of the event rate, the first step taken was to check the number of T3 messages between the CDAS and infill stations. The load of T3 messages has increased firstly in June 2010, and then even further in June 2011. Hence the conclusion was that the increased rate of T3 errors is due to the increased event rate.

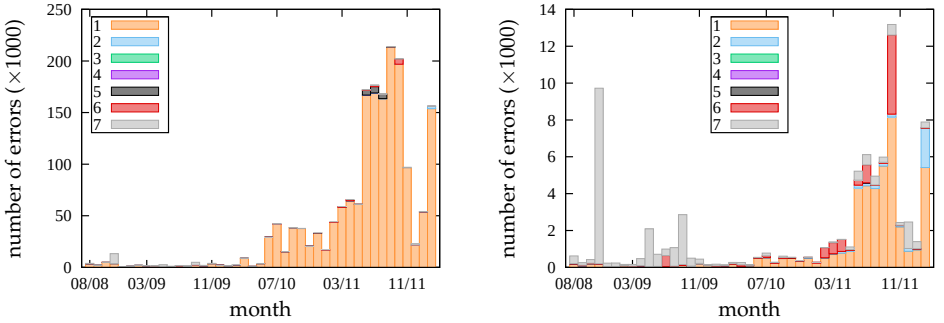
The previous conclusion is supported by the check on the distribution of T3 errors over the time of day. A 24 hour histogram has been filled using the time of day of all the T3 errors registered in the data set. Figure A.1 shows such histogram. It is evident that T3 errors happen mostly between 23:00 and 10:00 GMT, that correspond to 20:00 and 07:00 local time, that is, during night time in Malargüe. This explains the accumulation of T3 errors observed in figures 6.17. The fluorescence detector, and HEAT in particular, increases the T3 message load in the infill area. Note that the peak visible in the plot around 15:00 is due to a single day (7/Oct/2011) when errors clustered around that particular time of day. It is not related to the main problem found.

With the aim of identifying the origin of the problem, the evolution of error types has been studied. The time evolution is shown in figure A.2, dividing events according to their window size. Different colors indicate different kinds of errors (refer to section 6.3.2 for a description of the types of errors). Note how the 2009 communication crisis is clearly visible in the right panel, giving an excess of errors of type 7. By the end of 2011, the situation presents errors of type 1 as the dominating type of error. This type of error is returned when the CDAS request arrives too late to the station, and the late arrival of the request implies that the trace has been



(a) Any W. **(b)** $W = 0$ only.

Figure A.1: Distribution of the number of T3 errors over the time of day, for the infill array.



(a) All W values. **(b)** $W = 0$ only.

Figure A.2: Time evolution of the error types in the infill array.

overwritten. If the station was part of the original SD trigger ($W = 0$) the trace is lost, while if it was an additional station ($W > 0$), it is impossible to tell if the station had a trace or not. An excessive delay in polling the stations is due to an excessive rate of requests. This in turn can be due to an overload in the communication system.

Tests in the field, as of December 2011, seem to indicated that this kind of error may be due to the LS software, tuned for a much smaller rate of requests. Indeed, if the observed errors were due to the communication system, it is expected that such errors would consistently be returned by all stations belonging to the same BSU, as these are broadcast at the same time. The number of stations that belong to the same BSU is typically 30–40. The plot on figure A.3 shows the distribution of the number of stations returning T3 errors per event in the data set under consideration. The distribution shows an excess at 1, and a mean value of 3. This seems to indicate that the origin of the errors is not due to communication system. The tail of the distribution is given by few events in which most of the polled stations returned T3 errors. An example is shown in the right plot of figure A.4.

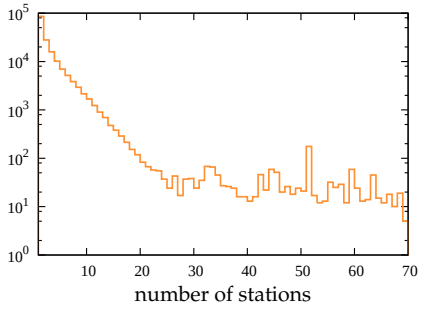


Figure A.3: The distribution of the number of stations returning T3 errors per event. The mean of the distribution is 3.

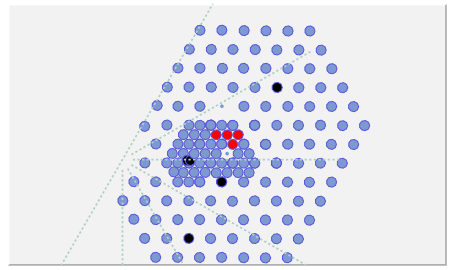
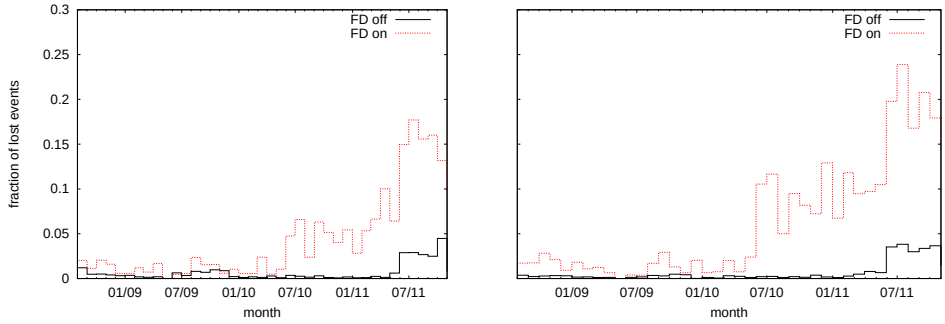


Figure A.4: Event with SDID 11786500, that had most of the stations returning a T3 error.



(a) T4 level.

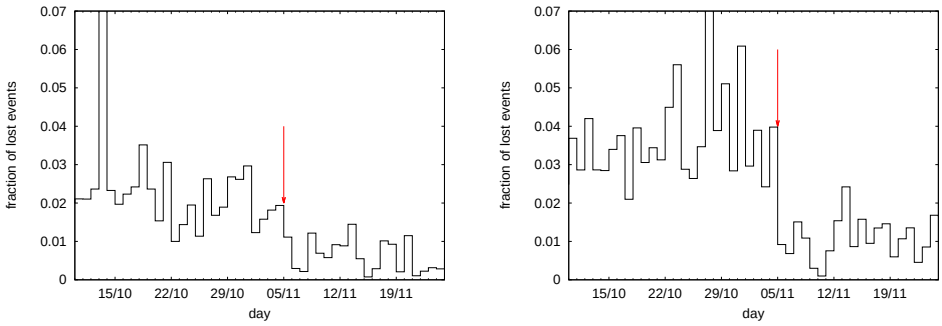
(b) T5 level.

Figure A.5: Fraction of lost events as a function of time. The effect of the FD operation on the event loss is evident.

A.2. Lost physics events

The last check applied in order to indicate the origin of the problem was to study lost events in two different conditions of data taking. When the FD is in operation or when it is not. The resulting plots are shown in figure A.5. The fraction of lost events during FD operation is much larger than when the FD is off.

It was previously noted, from plots in figure 6.18, an increase in the fraction of T3 errors from June 2011. This increase is also present in the fraction of events lost from that date on (figure 6.20). After dividing the population of lost events in two, when the FD is in operation or when it is not, the increase is still present in both cases, that is, there is a non negligible loss of events even when FD is not taking data. June 2011 corresponds with the start of the new local ToTd trigger in most of the infill stations. This alone induces a relevant increase in the event rate, by nearly a factor 2, since the new trigger is more efficient at lower energies than the standard ToT. Thus, the ToTd trigger was removed in early November 2011, to observe the effect it had on lost events. Indeed, figure A.6 shows a decrease in the fraction of lost events from the day the ToTd trigger was turned off, further supporting the hypothesis that the excessive event rate in the infill is the source of the excess of T3 errors.



(a) T4 level.

(b) T5 level.

Figure A.6: Fraction of lost events as a function of time. The date when the ToTd trigger was turned off is marked with an arrow.

A.3. Conclusions

While evaluating the uncertainty in the calculation of the exposure, an anomalous increase in time of the number of T3 errors, and hence in the fraction of lost events, was found. The studies done point to a too high rate of events in the infill area. The issue becomes manifest when the FD is taking data, as well as in the period where the ToTd trigger was activated. The issue seems to be related to software running on the stations, tuned for a lower rate of events. At the time of writing, a number of software updates were applied on the stations in the infill array, effectively reducing the number of errors. Further efforts are being directed to completely eliminate this problem.

Bibliography

[1] M. Nagano and A. A. Watson. *Rev. Mod. Phys.*, 72:689, 2000.

[2] E. Fermi. *Phys. Rev.*, 75:1169, 1949.

[3] A. R. Bell. *Mon. Not. Roy. Astron. Soc.*, 182:443, 1978.

[4] R. D. Blanford and J. P. Ostriker. *Astrophys. J.*, 221:L29, 1978.

[5] F. K. Lamb, editor. *High Energy Astrophysics*. Benjamin-Cummings, Menlo Park, CA, 1944. 277.

[6] P. Bhattacharjee and G. Sigl. *Phys. Rept.*, 327:109, 2000.

[7] G. A. Medina-Tanco and A. A. Watson. *Astropart. Phys.*, 12:25, 1999.

[8] J. Abraham et al. *Astropart. Phys.*, 29:243, 2008.

[9] J. W. Cronin. *Nucl. Phys. B (Proc. Suppl.)*, 138:465, 2005.

[10] D. Bird et al. *Astrophys. J.*, 441:144, 1995.

[11] W. Heitler. *Quantum theory of Radiation*. Oxford University Press, 2nd edition, 1944.

[12] T. K. Gaisser and A. M. Hillas. *Proc. 15th ICRC, Plovdiv, Bulgaria*, 10:353, 1977.

[13] K. Greisen. *Prog. Cosmic Ray Physics*, 3(1), 1956.

[14] K. Kamata and J. Nishimura. *Prog. Theoretical Phys.*, 6:93, 1958.

[15] K. Greisen. *Annual Review of Nuclear Science*, 10:63, 1960.

[16] M. Aglietta et al. *Nucl. Instr. and Meth. A*, 336:310, 1993.

[17] T. Antoni et al. *Nucl. Instr. and Meth. A*, 513:490, 2003.

[18] G. Navarra et al. *Nucl. Instr. and Meth. A*, 518:207, 2004.

[19] N. Chiba et al. *Nucl. Instr. and Meth. A*, 311:338, 1992.

[20] D. M. Edge et al. *J. Phys. A*, 6(10):1612, 1973.

[21] A. M. Hillas. *Acta Phys. Acad. Sci. Hung., Suppl. 3*, 29:355, 1970.

[22] D. J. Bird et al. *Phys. Rev. Lett.*, 71:3401, 1993.

[23] R. U. Abassi et al. *Astrophys. J.*, 622:910, 2005.

[24] J. Blümer et al. *Prog. Part. Nucl. Phys.*, 63:293, 2009.

[25] Y. Oyama et al. arXiv: astro-ph/0605020.

[26] M. Amenomori et al. *Science*, 314:439, 2006.

[27] T. Antoni et al. *Astrophys. J.*, 604:687, 2004.

[28] W. Axford et al. *Proc. 15th ICRC, Plovdiv, Bulgaria*, 11:132, 1977.

[29] A. D. Erlykin and A. W. Wolfendale. *J. Phys. G*, 27:1005, 2001.

[30] S. Swordy. *Proc. 24th ICRC, Rome, Italy*, 2:697, 1995.

[31] J. Candia et al. *JCAP*, 0305:003, 2003.

[32] T. stanev. arXiv:1210.0927 [astro-ph].

[33] R. Abbasi et al. *Phys. Rev. Lett.*, 100:101101, 2008.

[34] J. Abraham et al. *Phys. Rev. Lett.*, 101:061101, 2008.

[35] D. Ivanov and others (Telescope Array Collaboration). *Proc. 32nd ICRC, Beijing, China*, 2:258, 2011.

[36] V. Berezinsky et al. *Phys. Lett. B*, 612:147, 2005.

[37] R. U. Abbasi et al. *Astrophys. J.*, 622:910, 2005.

[38] J. Abraham et al. *Phys. Rev. Lett*, 104:091101, 2010.

[39] Y. Tsunesada. *Proc. 32nd ICRC, Beijing, China*, 12:58, 2011.

[40] M. P Véron-Cetty and P. Véron. *Astron. Astrophys.*, 455:773, 2006.

[41] P. Abreu et al. *Astropart. Phys.*, 34:314, 2010.

[42] R. U. Abbasi et al. *Astropart. Phys.*, 30:175, 2008.

[43] W. D. Apel et al. *Astropart. Phys.*, 36:183, 2012.

[44] D. De Marco and T. Stanev. *Phys. Rev. D*, 72:081301, 2005.

[45] M. Lemoine. *Phys. Rev. D*, 71:083007, 2005.

[46] D. Allard et al. arXiv: astro-ph/0703633.

[47] D. Allard et al. *Astropart. Phys.*, 27:61, 2007.

[48] M. Unger. arXiv:0812.2763.

[49] I. Allekotte et al. *Nucl. Instr. and Meth. A*, 586:409, 2008.

[50] M. Aglietta et al. *Proc. 29th ICRC, Pune, India*, 7:83, 2005.

[51] X. Bertou et al. *Nucl. Instr. and Meth. A*, 568:839, 2006.

[52] J. Linsley and L. Scarsi. *Phys. Rev.*, 128:2384, 1962.

[53] A. A. Watson and J.G. Wilson. *J. Phys. A*, 7:1199, 1974.

[54] S. Agostinelli et al. *Nucl. Instr. and Meth. A*, 506:250, 2003.

[55] S. Argirò et al. *Nucl. Instr. and Meth. A*, 580:1485, 2007.

[56] J. Abraham et al. *Nucl. Instr. and Meth. A*, 613:29, 2010.

[57] C. Bonifazi et al. *Astropart. Phys.*, 28:523, 2008.

[58] S. Scott et al. *Phys. Rev. Lett.*, 6:22, 1961.

[59] C. Wileman. *The Spread in the Arrival Times of Particles in Air-Showers for Photon and Anisotropy Searches above 10 EeV*. PhD thesis, School of Physics & Astronomy, The University of Leeds, 2008.

[60] J. Abraham et al. *Nucl. Instr. and Meth. A*, 620:227, 2010.

[61] P. Bauleo et al. *Proc. 29th ICRC, Pune, India*, 00:101, 2005.

[62] A. A. Watson. *Journal of Physics Conference Series*, 39:365, 2006.

[63] R. Knapik et al. arXiv:0708.1924v1 [astro-ph].

[64] J. Abraham et al. *Nucl. Instr. and Meth. A*, 523:50, 2004.

[65] M. Mostafá. *Nucl. Instr. and Meth. B Proc. Suppl.*, 165:50, 2007.

[66] The Pierre Auger Collaboration. *JINST*, 7:P09001, 2012.

[67] T. Pierog et al. *Proc. 30nd ICRC, Mérida, Mexico*.

[68] M. Settimo. *Hybrid detection of ultra high energy cosmic rays with the Pierre Auger Observatory*. PhD thesis, Department of Physics, The University of Salento, 2010.

[69] J. Abraham et al. *Astropart. Phys.*, 31:399, 2009.

[70] The Pierre Auger Collaboration. *Astrophysical Journal Letters*, 755:L4, 2012.

[71] The Pierre Auger Collaboration. *ApJ*, 760:148, 2012.

[72] The Pierre Auger Collaboration. *Phys. Rev. Lett.*, 109:062002, 2012.

[73] The Pierre Auger Collaboration. *Astropart. Phys.*, 34:368, 2011.

[74] F. Salamida for the Pierre Auger Collaboration. *Proc. 32nd ICRC, Beijing, China*, 2:145, 2011.

[75] C. Di Giulio. *The cosmic rays flux from the Pierre Auger Observatory data*. PhD thesis, Department of Physics, the University of Rome "Tor Vergata", 2009.

[76] The Pierre Auger Collaboration. *Astropart. Phys.*, 34:627, 2011.

[77] The Pierre Auger Collaboration. *Science*, 318:938, 2007.

[78] The Pierre Auger Collaboration. *Astrophysical Journal Supplement*, 203:34, 2012.

[79] F. Sanchez for the Pierre Auger Collaboration. *Proc. 32th ICRC, Beijing, China*.

[80] H. J. Mathes for the Pierre Auger Collaboration. *Proc. 32th ICRC, Beijing, China*.

[81] I. C. Mariş for the Pierre Auger collaboration. *Proc. 32th ICRC, Beijing, China*, 1:267, 2011.

[82] J. Aublin. *Auger collaboration mailing list*.

[83] D. Veberič and M. Roth. *GAP note*, 2005-035.

[84] C. Bonifazi for the Pierre Auger Collaboration. *Nucl. Phys. B Proc. Suppl.*, 190:20, 2009.

[85] D. Newton et al. *Astropart. Phys.*, 26:414, 2007.

[86] I. Sidelnik et al. *GAP note*, 2011-013.

[87] A. Tapia et al. *GAP note*, 2012-004.

[88] M. Ave et al. *Nucl. Instrum. Meth. A*, 578:180, 2007.

[89] M. Aglietta et al. *GAP note*, 2008-030.

[90] D. Ravignani et al. *GAP note*, 2009-116.

[91] D. Ravignani et al. *GAP note*, 2011-010.

[92] T. Schmidt et al. *GAP note*, 2007-106.

[93] P. Billoir. *Auger collaboration mailing list*.

[94] D. Newton. *Measuring the Lateral Distribution of Extensive Air Showers to Characterise Properties of Cosmic Rays Above 1 EeV*. PhD thesis, School of Physics & Astronomy, the University of Leeds, 2005.

[95] J. Abraham et al. *Phys. Lett. B*, 685:239, 2010.

[96] M. Unger et al. *Nucl. Instr. and Meth. A*, 588:433, 2008.

[97] The Pierre Auger Collaboration. *Astropart. Phys.*, 33:108, 2010.

[98] P. Abreu et al. *Astropart. Phys.*, 35:266, 2011.

[99] C. Bonifazi and P. L. Ghia. *GAP note*, 2006-101.

[100] P. L. Ghia and I. Lhenry-Yvon. *GAP note*, 2009-092.

[101] A. Tripathi et al. *GAP note*, 2005-061.

[102] D. Allard et al. *GAP note*, 2006-024.

[103] T. Schmidt et al. *GAP note*, 2006-070.

[104] N. G. Busca. *THE ULTRA HIGH ENERGY COSMIC RAY FLUX FROM THE SOUTHERN PIERRE AUGER OBSERVATORY DATA*. PhD thesis, Department of Physics, The University of Chicago, 2008.

[105] I. C. Mariş. *Measurement of the Ultra High Energy Cosmic Ray Flux using Data of the Pierre Auger Observatory*. PhD thesis, Department of Physics, University of Karlsruhe, 2008.

[106] H. P. Dembinski and M. Roth. *GAP note*, 2011-074.

[107] B. Efron. *The Annals of Statistics*, 7:1, 1979.

[108] The Pierre Auger Collaboration. *Phys. Rev. Lett.*, 104:091101, 2010.

[109] M. Nagano et al. *Astropart. Phys.*, 22:235, 2004.

[110] The Pierre Auger Collaboration. *Nucl. Instr. and Meth. A*, 620:227, 2010.

[111] AIRFLY Collaboration. *Astropart. Phys.*, 28:41, 2007.

[112] H. P. Dembinski. *Proc. 32nd ICRC, Beijing, China*, 2:101, 2011.

[113] M. Ave et al. *GAP note*, 2007-052.

[114] S.F. Berezhnev et al. *Proc. 32nd ICRC, Beijing, China*, 1:209, 2011.

[115] R. U. Abassi et al. *Phys. Rev. Lett.*, 100:101101, 2008.