Measurement of the Ultra High Energy Cosmic Ray Flux using Data of the Pierre Auger Observatory

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik der Universität (TH)
Karlsruhe

genehmigte
DISSERTATION

von Ioana Codrina Mariş

aus Baia Mare (Rumänien)

Tag der mündlichen Prüfung: 15.02.2008
Referent: Prof. J. Blümer
Korreferent: Prof. G. Quast

Bestimmung des Energiespektrums ultra-hochenergetischer kosmischer Strahlung mit Daten des Pierre-Auger-Observatoriums

Das Thema dieser Arbeit ist die Bestimmung des Energiespektrums kosmischer Strahlung bei höchsten Energien unter Verwendung von Daten des Oberflächendetektors des Pierre-Auger-Observatoriums. Ein wichtiger Beitrag zu den Unsicherheiten der Messung der lateralen Verteilung von Teilchen mit dem Oberflächendetektor ist das fehlende Signal der dem Zentrum des Luftschauers am nächsten liegenden Station. Dieses Problem wird hervorgerufen durch die Sättigung der Auslese-Elektronik und den Übergang zu einem nichtlinearen Verhalten der Photovervielfacher-Antwort. Davon betroffen sind 50 % der Ereignisse in dem untersuchten Energiebereich. Eine Methode zur Bestimmung dieses Signal wird entwickelt. Die Abschwächung des Signals in der Atmosphäre wird mittels der Annahme räumlich konstanten Intensität der Kosmischen Strahlung abgeleitet. Besonderer Wert wird auf die Energiekonvertierung vermittels Fluoreszenzdetektor gelegt. Dazu werden Luftschauern Daten verwendet, welche so wohl mit dem Fluoreszenz- als auch mit dem Oberflächen-Detektor gemessen wurden. Um den Effekt der Detektor-Auflösung auf den gemessenen Fluss abzuschätzen, werden realistische Simulationen der Detektorantwort durchgeführt. Die Akzeptanz des Oberflächedetektors ist nur aus Daten bestimmt und die systematischen Unsicherheiten sind mit Hilfe von Simulationen berechnet. Dies ermöglicht eine Erweiterung des Energiespektrums aus Oberflächendetektordaten zu niedrigeren Energien hin, außerhalb des Bereiches voller Trigger-Effizienz. Außer dem durch vertikale Ereignisse des Oberflächen-detektors erhaltenen Energiespektrum können zwei weitere unabhängige Spektren bestimmt werden: Ein Hybridspektrum, basierend auf Luftschauern, die den Fluoreszenzdetektor und mindestenz eine Station des Oberflächendetektores getriggert haben, und ein Spektrum mit Oberflächendetektor-Ereignissen mit Zenitwinkeln größer als 60 Grad. Diese Spektren werden in einem einzigen Spektrum kombiniert, das sich über den breitesten Energiebereich der aktuellen Daten des Pierre-Auger-Observatoriums erstreckt und dabei die kleinsten Unsicherheiten hat. Das daraus resultierende Spektrum wird in einen astrophysikalischen Kontext gesetzt. Die Fortsetzung des Spektrums oberhalb 40 EeV als Potenzgesetz wird mit einer Signifikanz von 5.2σ verworfen. Die Anzahl der erwarteten Ereignisse für ein Potenzgesetz ist 167 ± 3 über 40 EeV und 35 ± 1 über 100 EeV, wobei die Daten nur je 64 Ereignisse beziehungsweise 1 Ereignis im jeweiligen Bereich enthalten.

Abstract

The main objective of this thesis is the measurement of the cosmic ray energy spectrum above 1 EeV based on the data recorded at the Pierre Auger Observatory. From the reconstruction of the lateral distribution of particles on ground up to the final flux determination from vertical surface detector events, all steps are investigated in detail. A contribution of the surface detector measurement to the uncertainties of the estimate of the lateral distribution of particles on ground is the missing signal of the station closest to center of the air shower. This effect is present for 30% of the events in the energy region of interest. It is caused by the overflow of the readout electronics and the turn over of the photomultiplier response from a linear to non-linear behavior. A method to recover the signal is developed. The attenuation of the signal in the atmosphere is deduced from a constant intensity assumption. An emphasis is put on the energy conversion obtained from the fluorescence detector measurements. A new method to deduce the energy conversion, based on coincident measurements with the fluorescence and surface detectors of the same air shower is used as a cross check. To estimate the detector resolution effect on the measured flux the detector response is obtained from realistic simulations. The acceptance of the surface detector is obtained from data only, and the systematic uncertainties are checked with simulations. This allows an extension of the energy spectrum deduced from surface detector data to lower energies below the range of full trigger efficiency. Besides the energy spectrum obtained from the vertical surface detector measurements, two other independent spectra can be estimated from the Pierre Auger Observatory using showers triggering the fluorescence detector and at least one station and inclined surface detector events. The vertical, hybrid and horizontal spectrum are combined with a novel maximum likelihood method taking into account the systematic uncertainties of each measurement. Two spectral features are observed: The ankle at $E = 3.3 \pm 1\,\mathrm{EeV}$ and a flux suppression starting at $E = 32.4 \pm 1.1\,\mathrm{EeV}$. The measured spectral index changes from $\gamma_1 = 3.26 + -0.03$, to $\gamma_2 = 2.65 \pm 0.02$ and at highest energies to $\gamma_3 = 3.9 \pm 0.3$. The continuation of the spectrum in the form of a power law is rejected with a $5.2\,\sigma$ significance. The number of events expected if a power law holds, above 40 EeV are 167 ± 3 and 35 ± 1 above 100 EeV whereas in the measured flux there are 64 events and 1 event.

Contents

Introduction

Several important observations were reported in the 1960's that had a large impact on astroparticle physics. It was discovered that cosmic ray particles can reach macroscopic energies of more than 10^{20} eV, as being measured by Linsley, Rossi and Scarsi [1]. A few years after Penzias and Wilson established the existence of the cosmic microwave background [2] with a mean temperature of 2.7 K [3,4]. Then Greisen [5], Zatsepin and Kuzmin [6], predicted that a cutoff of the cosmic ray flux at the highest energies is expected due to the interaction of the ultra high energy cosmic rays with the cosmic microwave background photons.

This flux suppression, which can be also interpreted as a drop of acceleration power at the sources, was first observed by the HiRes collaboration [7, 8]. They were using optical telescopes to observe the air showers initiated by ultra high energy cosmic rays (UHECRs): the charged particles from the air showers excite nitrogen molecules of the atmosphere which emit fluorescence light. The longitudinal development of the air shower is measured and the primary particle energy is proportional to the total light collected at the aperture. The HiRes results were challenged by the measurement of the AGASA collaboration which observed a continuation of the cosmic ray flux [9]. The latter used a surface detector array, sampling the particles from the air showers reaching the ground. The energy is calibrated with help of Monte Carlo simulations.

The event statistics and the acceptance of a surface detector depends only on the collection area. The duty cicle is nearly 100%. In the case of a fluorescence detector the energy dependent efficiency calculation must be deduced from detector simulations and duty cycle is only 10% of an array experiment. The Pierre Auger Observatory was designed to measure the ultra high energy tail of the spectrum with an accuracy far better than any forerunner experiment [10]. Combining high statistics and geometrical acceptance of the surface detector with a reliable energy assignment from the fluorescence measurement the cosmic ray flux is established without the need of air shower simulations.

A continuation of the cosmic ray flux at extreme energies would require different scenarios beyond the standard physics like a violation of Lorentz invariance [11–14], or the existence of super-massive stable particles, related to grand unified theories, produced in the early universe and associated with spontaneous symmetry breaking. The flux excess might be caused by decay products of these particles [15–18].

The main objective of this thesis is the measurement of the cosmic ray energy spectrum above 1 EeV based on the data recorded at the Pierre Auger Observatory.

From the reconstruction of the lateral distribution of particles on ground up to the final flux determination from vertical surface detector events, all steps are investigated in detail.

A contribution of the surface detector measurement to the uncertainties of the estimate of the lateral distribution of particles on ground is the missing signal of the station closest to center of the air shower. This effect is present for 30% of the events in the energy region of interest. It is caused by the overflow of the readout electronics and the turn over of the photomultiplier response from a linear to non-linear behavior. A method to recover the signal is developed.

The attenuation of the signal in the atmosphere is deduced from a constant intensity assumption. An emphasis is put on the energy conversion obtained from the fluorescence detector measurements. A new method to deduce the energy conversion, based on coincident measurements with the fluorescence and surface detectors of the same air shower is used as a cross check. To estimate the detector resolution effect on the measured flux the detector response is obtained from realistic simulations. The acceptance of the surface detector is obtained from data only, and the systematic uncertainties are checked with simulations. This allows an extension of the energy spectrum deduced from surface detector data to lower energies below the range of full trigger efficiency.

Besides the energy spectrum obtained from the vertical surface detector measurements, two other independent spectra can be estimated from the Pierre Auger Observatory using showers triggering the fluorescence detector and at least one station and inclined surface detector events. The spectra are combined in a single spectrum that extends over the largest energy range obtainable from the current data of the Pierre Auger Observatory and with smallest uncertainties.

The resulting spectrum is put in astrophysical context and implications due to the shape of the cosmic ray flux are discussed.

Chapter 1

Ultra High Energy Cosmic Rays

Almost 30 years after the discovery of cosmic rays [19], it was realized [20–23] that particles reaching ground are correlated in time, which lead to the detection of extensive air showers (EAS). A cosmic ray entering the atmosphere interacts with nuclei generating a subsequent shower of particles which can be detected on ground.

Even now, more than 90 years after, some pieces from the puzzle of comic rays are missing [24]. The energies of these particles reach several orders of magnitude above the ones achievable in man made accelerators. The mechanism to accelerate particles to these high energies is not yet understood. The interpretation of the air shower development is based on Monte Carlo simulations, which rely on hadronic interaction models and extrapolations of measured cross sections.

Finding new particles in the early times, like the positron [25], the muon [26] and the pion [27–29], has been one of the main achievements in the early cosmic ray physics studies. The new era of precision measurements in the extreme energy range, above 1 EeV, has begun with the Pierre Auger Observatory.

1.1 Cosmic ray flux

The energy spectrum of the cosmic rays is almost featureless. Extending from 10^9 eV up to 10^{20} eV the spectrum follows a power law. This behavior is expected in the case of stochastic acceleration of charged particles at astrophysical shocks as proposed by Fermi [30]. It was shown that this mechanism works most efficiently in the case of diffuse shock acceleration when particles encounter several times the shock wave front. This mechanism explains for example acceleration during supernova explosions, where material of several solar masses are ejected at a speed larger than the speed of sound in the interstellar medium creating a strong shock wave propagating radially out.

The acceleration mechanism together with propagation processes through the medium can explain qualitatively the whole energy range of the energy spectrum observed on Earth starting with a few GeV up to the highest energies. One missing piece of the puzzle are the sites of acceleration (see Fig. 1.1).

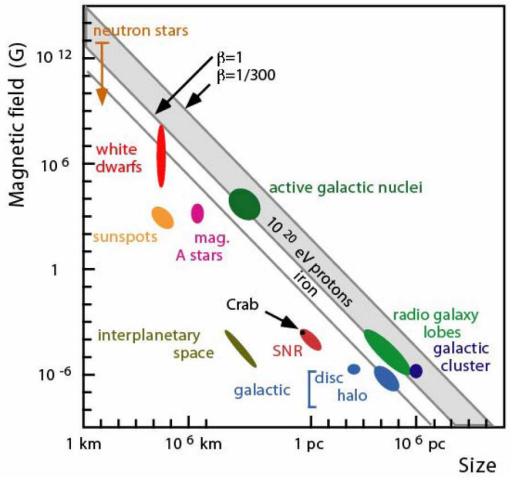


Figure 1.1: Hillas diagram.

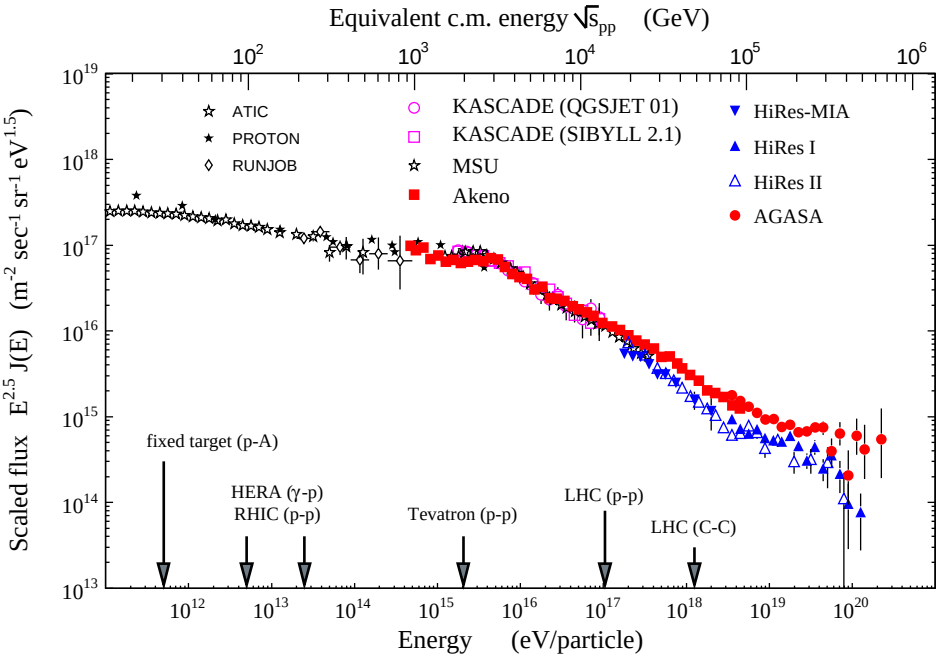
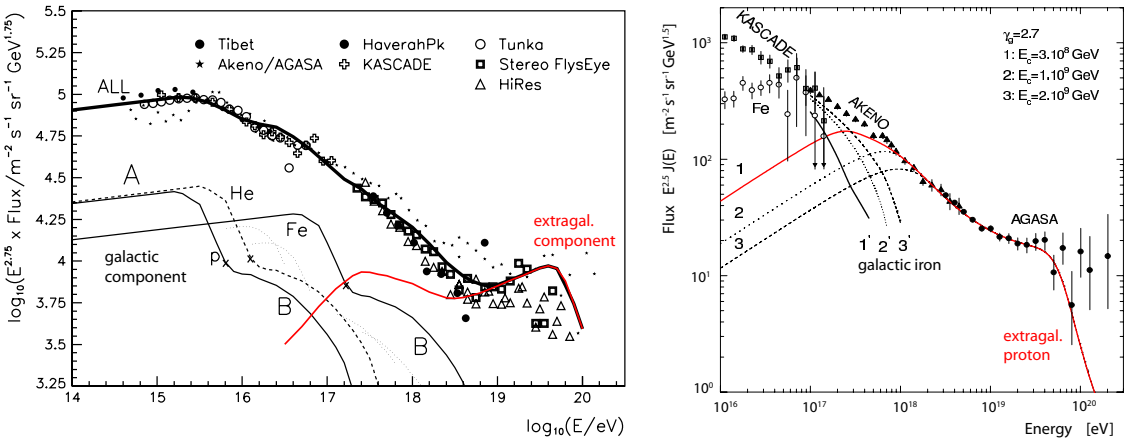


Figure 1.2: Cosmic rays flux. Data from [7, 9, 31–36]

In order to point out deviations from the power law falloff, the flux, shown in Fig. 1.1, has been multiplied by $E^{2.5}$. The two visible changes of the spectral index, one at $\approx 3 \cdot 10^{15}$ eV [37] and the other at $\approx 10^{18}$ eV [38] are referred within the astrophysics community as the *knee* and the *ankle* in analogy to a leg shape. In the highest energy region (only data published before the measurements of the Pierre Auger Observatory are shown) the shape of the spectrum is not conclusive. All the spectral features might be interpreted either as a change of the acceleration mechanism at sources, either as a propagation effect or as a change of the hadronic interaction



(a) *mixed composition* (b) *dip model*

Figure 1.3: (a) the extragalactic mixed composition scenario (taken from [24] and modified). (b) The *dip* model with the transition to a proton dominated extragalactic component at lower energies [47]

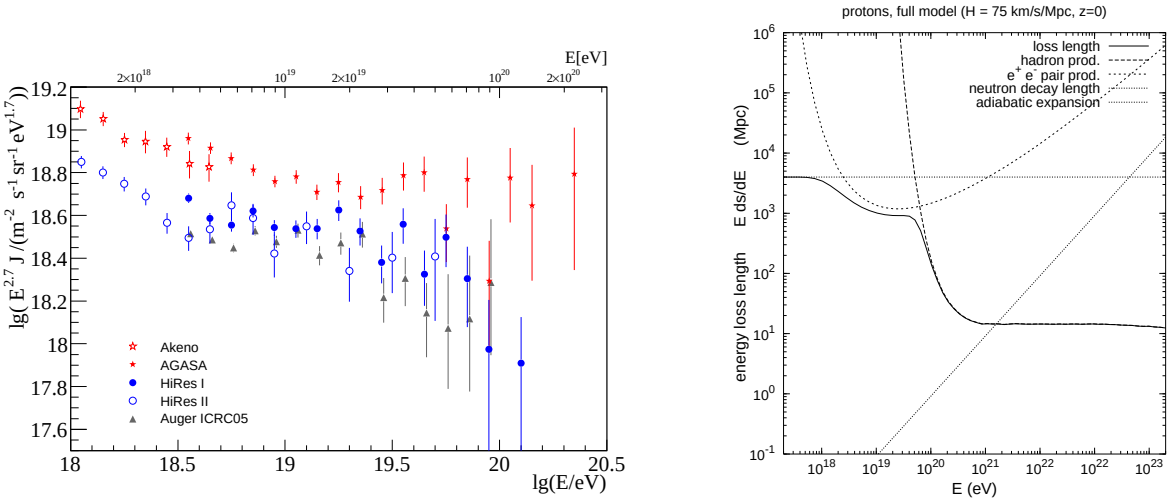
cross sections with increasing energy.

The knee is considered to be an acceleration feature. The energy of the cosmic ray is proportional to the charge of the particle. The maximum energy that can be reached, determined by the condition that the Larmor radius should be smaller than the size of the acceleration region. The knee is thus explained as a drop of the acceleration possibilities for different particles. Possible acceleration sites are supernova remnants [39]. The knee energy would be proportional to the charge, Z.

Another concurrent scenario assumes that the knee might be caused by a sudden change of the hadronic interactions at these energies [40, 41]. In this case the knee observed is not a characteristic of the spectrum itself, but of its observation at Earth. As a result of the changes in the high energy interaction, a new type of heavy particles might be produced and escape unseen. In this scenario the energy of the knee for individual elements would scale with their mass number A and not their nuclear charge Z. The knee can be interpreted also as a propagation result due to a change in the regime of diffusion in the galactic magnetic field [42, 43]. The KArlsruhe Shower Core and Array DEtector (KASCADE) experiment has shown [44] that the contribution to the flux from protons falls off earlier than for helium, giving a big support for these possibilities.

In the region between the knee and the ankle there are two peculiarities of the cosmic ray expected: A drop of the heavy components at an energy scaled with the charge or with the mass and a transition region from the galactic to extragalactic origin of cosmic rays.

Around 0.1 EeV there might be a second steepening of the spectral index observed [31, 33, 45], the second knee. The experimental status does not allow to be conclusive on the energy where this occurs, but the KASCADE-Grande and the IceTop [46] experiments will clarify this energy range of the cosmic ray flux in the near future.



(a) *comic ray flux* (b) *energy loss*

Figure 1.4: (a)The cosmic ray energy spectrum in the highest energy region. Only data before 2007 are shown.(b) Energy loss for extragalactic cosmic ray propagation [48].

The ankle One model assumes that a transition from the galactic to the extragalactic origin of the cosmic rays occurs at this energies [24, 49, 50]. The extragalactic component is thought to have a pure proton composition. The position of the ankle would be the energy where the two components contribute equally to the total flux.

In another model [51, 52] the tail of the flux has a mixed composition and the ankle is still a transition of the origin of the cosmic rays (Fig. 1.3(a)). The nuclei with energies greater than 10 EeV are disintegrated while propagating through the medium.

Other model [47, 53] for this region is build from the hypothesis that the extragalactic component composed mainly of protons starts to dominate at lower energies and the transition from the galactic to extragalactic component takes place at around 0.5 EeV (Fig. 1.3(b)). In the ankle region the galactic component already vanished. The spectral index change is then just a propagation effect: protons passing through the cosmic microwave background loose energy via e^+e^- production and this causes a flux suppression at higher energies and an accumulation at a slightly lower energy. This is called the dip model.

The end of the cosmic ray energy spectrum Observations on this were without result until 2007 [8, 54], mainly due to the rapid increase with energy of systematic uncertainties of the energy and of the statistical uncertainties due to the extremely statistics dominating the previous measurements [7, 9, 33]. Only ≈1 particle per km^2 sr year is expected at 100 EeV. The cosmic ray flux above 1 EeV, from measurements until 2007 is shown in Fig. 1.4(a).

Data are compatible within the energy systematic uncertainty of each experiment, but two distinct trends were observed: A continuation of the cosmic ray flux in form of a power low for AGASA measurement, and a flux suppression above ≈ 50 EeV in case of HiRes. The detection technique is different for the two experiments: The HiRes Collaboration has built telescopes to observe the longitudinal development of air shower generated by cosmic rays

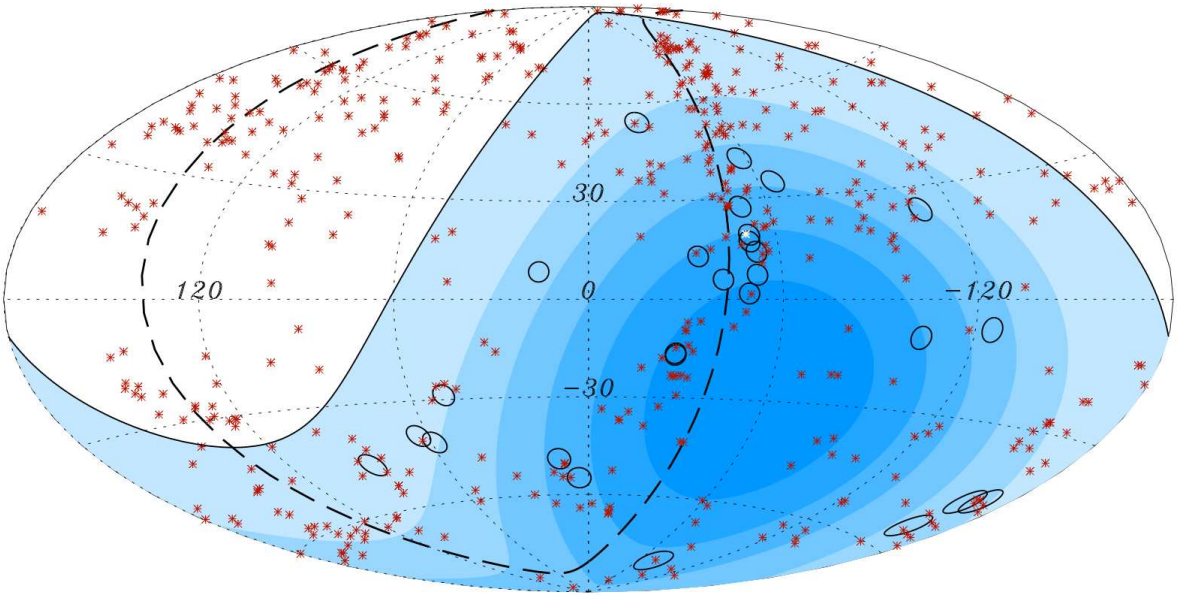


Figure 1.5: Arrival direction of the cosmic rays(circles) with energies greater than 57 EeV measured by the Pierre Auger Observatory and the position of the close by AGN objects (stars) in the Aitoff projection of the galactic coordinates. In blue the integrated exposure of the detector is shown [55].

in atmosphere, while the AGASA apparatus detected the lateral distribution of particles on the ground. The dispute has been settled down by the recent results of the HiRes and Pierre Auger Observatory collaborations, partly based on the results presented in this thesis. The flux suppression is determined with a $5-6\sigma$ significance.

The standard physics can explain the flux suppression as an energy loss through Δ resonances of cosmic rays on their path to Earth by interacting with the cosmic microwave background (i.e. $N+\gamma_{2.7K} \rightarrow \Delta \rightarrow N+\pi$), known as the Greisen-Zatsepin-Kuzmin (GZK) cutoff [5, 6]. In Fig. 1.4(b) the main energy losses suffered by the cosmic rays on their path to Earth are shown. The hadron production is responsible for the GZK-effect while the $e^{\pm}$ production is causing the dip explained above. For example, a proton with an energy of 100 EeV must have the source within 100 Mpc to reach the Earth.

It can also be interpreted as a drop of the acceleration power at sources. A continuation of the cosmic ray flux can be explained only by exotic scenarios or by a high density of nearby powerful sources.

The acceleration and propagation of cosmic rays at high energies is one of the puzzles that have not been solved in cosmic rays physics. Up to energies of 10^9 eV they are of solar origin, but above this energy due to the leakage of day-night variations of the measured flux and of the maximum acceleration power in the Sun, they must have another source, outside of the solar system.

Tracking back particles is not an easy task. The first evidence that they possess a charge was given by East-West asymmetries caused by their deflection in the magnetic field of Earth. Up to energies of a few 10 EeV they are completely isotropic. In the highest energies they

are affected by the galactic and extragalactic magnetic fields. Only above 1 EeV, particles cannot be anymore confined by the $\approx 3\,\mu$G magnetic field of our galaxy. Their origin can be solved depending on the intergalactic magnetic field and the distance to the source, but the deflection angle at a few times 10 EeV is less than 5° for the nearby astrophysical objects.

Finding the sources of the UHECRs has a long history. Even before they were identified it was assumed that very good accelerating sites are nearby radio galaxies [56], which are containing active galactic nuclei, super-massive black holes with a mass 6 orders of magnitude larger than the solar mass. In active galactic nuclei (AGN) a big amount of matter is accreting, part of it being released in the form of jets.

The maximum energy that can be reached in these astrophysical objects is 10^{21} eV [57] (this theory being supported by the cutoff in the non-thermal emission spectrum produced by electrons, observed in many radio galaxies). The acceleration sites for the UHECRs can be the beam dumped AGN: Galaxies where the jet hits an intergalactic cloud of matter [58], or the region very close to the black hole [59] and even in the remains of the fossil jets of old AGNs [60].

After lots of trials to find significant anisotropies at the highest energy the recent results of the Pierre Auger Observatory have proven a correlation of the arrival direction of cosmic rays at energies above 57 EeV with close by AGNs [55]. The data illustrated in Fig. 1.5(a) show the arrival direction together with AGN locations. 22 events out of 27 are within a 3.1° separation angle from AGNs, and the majority of the ones that do not correlate are near the galactic plane, where the expected cosmic ray deflections are largest and the used catalog is incomplete The data does not identify AGNs as the sources of cosmic rays, other astrophysical object distributed similar to nearby AGNs could be also possible candidates.

1.2 Extensive air showers

The cosmic rays entering the atmosphere interact with the nuclei of air molecules producing secondary particles that lead to the formation of extensive air showers. The development in the atmosphere for a simulated shower of 140 EeV initiated by a proton is shown in Fig. 1.6. Three main components are displayed: electromagnetic (e^+, e^- and γ), muonic (μ^+ and μ^-) and hadronic (mainly $\pi^\pm$, π^0, n, p, $K^\pm$ and K^0). Electromagnetic particles are the most numerous carrying the main part of the total energy (98%) the rest being shared by the other components (the muonic part having 1.7%).

Electromagnetic component The main processes for the electromagnetic part of the shower are pair production, when photons interact with an atomic nucleus producing $e^\pm$, photons are produced by $e^\pm$ through Bremsstrahlung processes and below 80 MeV the ionization process starts to dominate. The critical energy, where energy losses through collision processes start to predominate over pair production and bremsstrahlung, is the main ingredient for a simple cascading model, developed by Heitler [63].

Starting with an energy, E, larger than the critical energy, E_c, the result of one interaction is always two particles assumed to be equally energetic. After a number of interactions, n, the total number of particles is 2^n, and the total energy $E_c \cdot 2^n$. One interaction occurs after

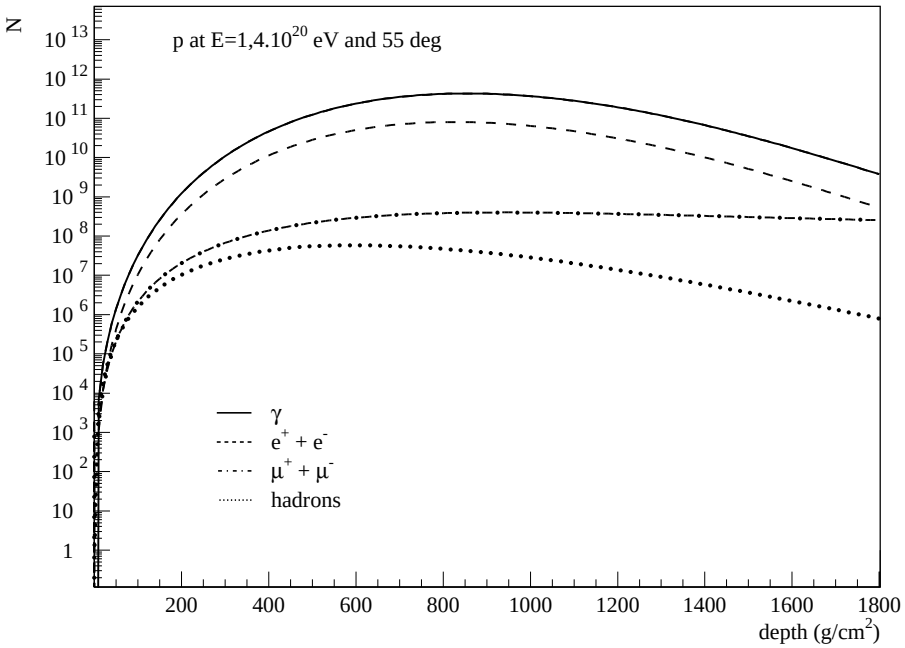


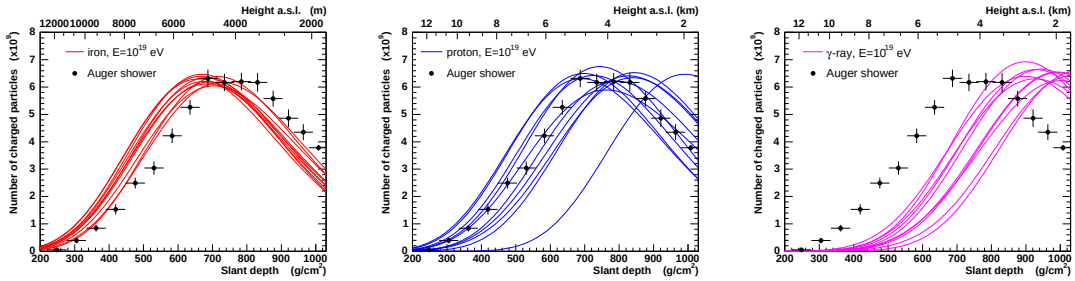
Figure 1.6: The particle content of a proton induced air shower as a function of the depth in the atmosphere [61] simulated with CONEX [62].

a splitting length, $\lambda = X_0 \ln 2$, the distance after which the electromagnetic particle has lost half of its energy. The maximum number of interactions is proportional to $\ln(E/E_c)$. The air shower develops reaching its maximum at a depth $\propto \ln(E)$ with a number of particles $\propto E$. The electromagnetic radiation length for air is $X_0 \approx 37\,\mathrm{g/cm^2}$.

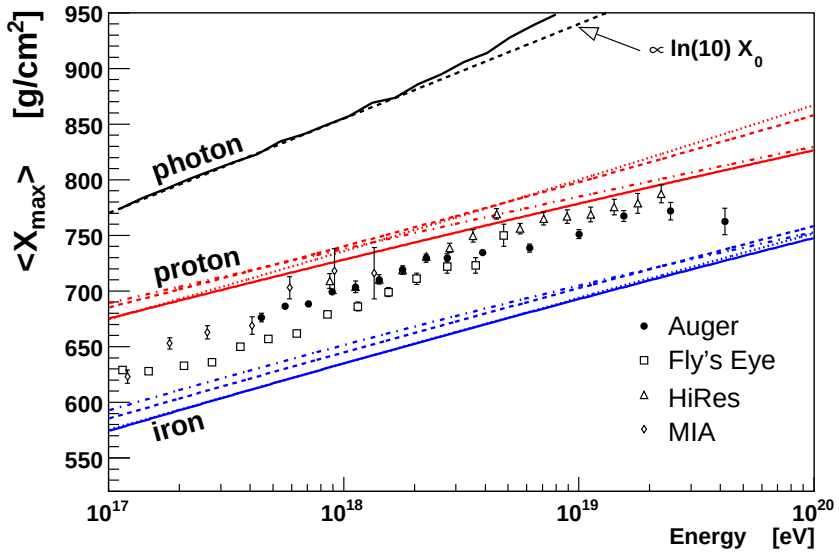
Heitler model for the hadronic component The hadronic component is more sophisticated. In analogy with the electromagnetic cascade, it can be described by assuming that in each hadronic interaction the main part of products consists of pions out of which one third are neutral. The π^0 will decay and fuel the electromagnetic part. Charged pions carry on the hadronic cascades.

The critical energy, when pions start to decay into muons ($\pi^{\pm} \rightarrow \mu^{\pm} + \nu_{\mu}$), determines the end of the cascade. The cascade stops with a total number of muons, identical with the number of charged pions,which is proportional to E/E_c^{α}. $\alpha = \ln(n_{\mathrm{charged}})/\ln(n_{\mathrm{total}}) = 0.85..0.96$. E_c depends on the atmospheric density and on the pion decay length and can be approximated to be 20 GeV [74]. Neglecting the subsequent electromagnetic cascades produced by the π^0 decays within this approach the shower maximum is proportional to $X_{\max} = \lambda + X_0 \ln(E_0/(n_{\mathrm{total}} E_c^{\mathrm{e.m.}}))$, still depending on the electromagnetic critical energy, the primary energy, but also on the multiplicity, n_{total}.

If the primary is heavier, having an atomic number A, the superposition theory supposes that it acts like A proton induced air showers with an energy of E/A. An iron initiated shower has the maximum of the development higher in atmosphere than one produced by a proton



(a) *iron* (b) *proton* (c) *photon*

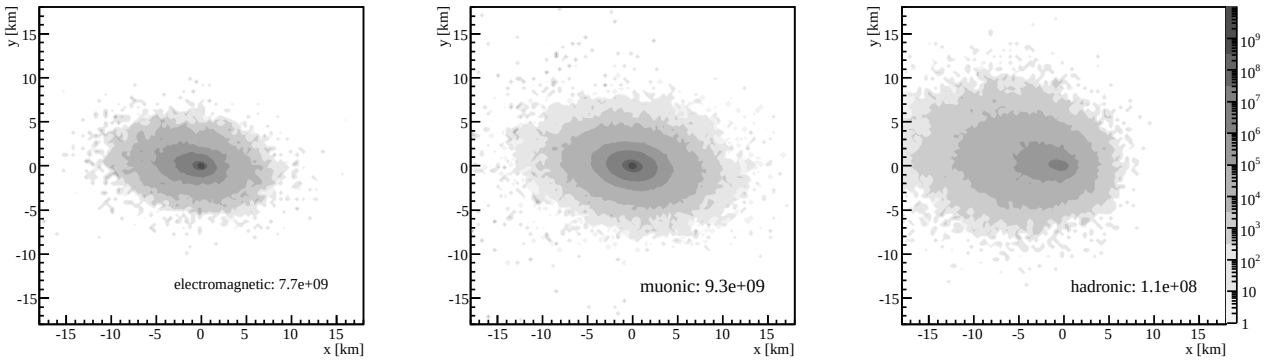


(d) *elongation rate*

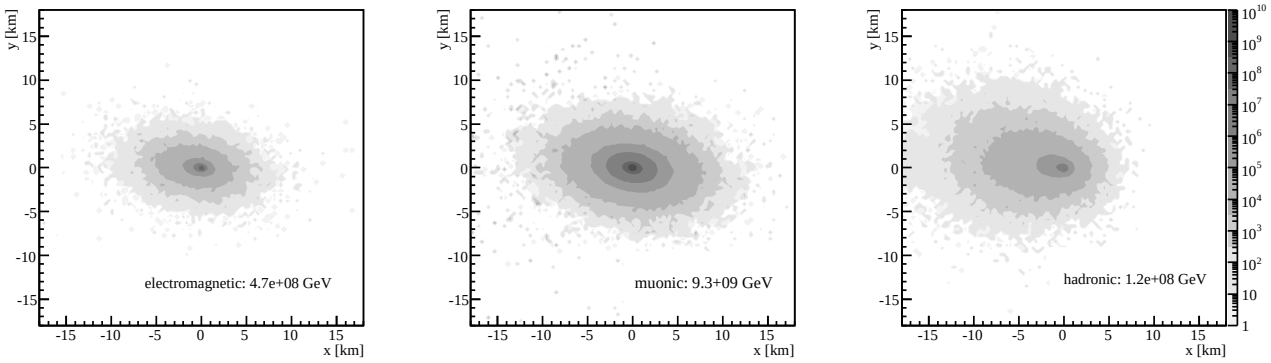
Figure 1.7: (a)-(c) Longitudinal profiles [64]. (d) The mean $X_{\max}$ versus energy [65], data from [31,38,66,67] compared to air shower simulations for iron. proton and photon primaries. For hadronic primary particles, the differing model predictions are given as solid line for QGSJET01c [68,69], dotted line for SIBYLL2.1 [70,71], dashed line for EPOS1.61 [72] and dashed-dotted line for QGSJETII.3 [73].

with the same energy.

The longitudinal profile, i.e. the number of particles as a function of the amount of matter crossed (atmospheric depth) is observed with fluorescence telescopes. Charged particles excite nitrogen molecules that emit light in the ultra violet (UV) wavelength band. The energy deposited is proportional to the total number of particles. Not all the energy is going into fluorescence light, since a part of it is carried away by neutrinos (produced mainly in pion and muon decays) and very low energetic particles. This contribution called the *invisible energy*, was obtained from simulations and is about 5% of the total energy. A mean longitudinal profile is shown in Fig. 1.7(a), 1.7(b) and 1.7(c) compared to the simulated profiles of iron, proton and photon. The Heitler model, even if it is a very crude approach predicts the observed dependency of $X_{\max}$ with energy in data as shown in Fig. 1.7(d). The cosmic rays in the ultra high energy range suggest a general trend from heavy cosmic ray primaries at



(a) *particle content*



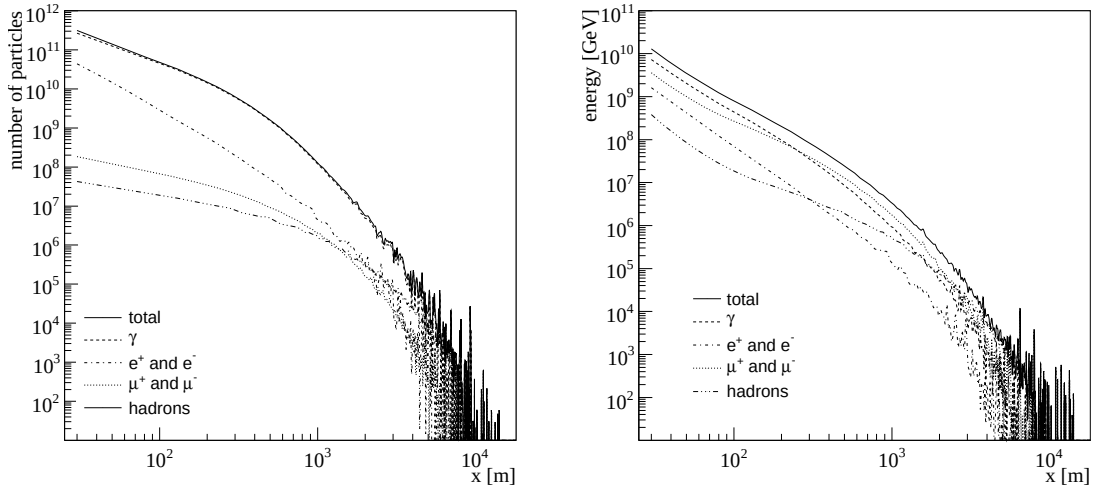
(b) *energy content*

Figure 1.8: The ground particles of an air shower induced by a proton with energy of 140 EeV and a zenith angle of 55°, simulated with CORSIKA [77].

lower energies to light primaries at high energy. The Heitler model prediction coincides very well with the photon simulation, whereas the deviation at highest energies is related to the Landau- Pomeranchuck- Migdal effect [75, 76].

Lateral development The particle spread away from the shower axis is caused by multiple Coulomb scattering and tranverse momenta obtained in interactions with the atmosphere and decays. The density of particles and of the kinetic energy at the mean altitude of the Pierre Auger Observatory for the different components are shown in Fig. 1.8. The simulated shower was initiated by a proton with an energy of 140 EeV and a zenith angle of 55°. The number of muons is equal to the number of electrons as it had developed high in the atmosphere ($X_{\max} = 672\,\mathrm{g/cm^2}$) and the electromagnetic component of the shower has already died out. The hadronic part is mainly composed of neutrons that are not detected. The number of muons relative to the number of electrons reaching the ground is a parameter which is sensitive to the composition of the primary particle.

For high energetic events, with a maximum deep in the atmosphere, the electromagnetic



(a) *particles* (b) *energy*

Figure 1.9: The particle content on ground of a proton induced air shower as a function of the distance to the shower axis. The simulated shower had an energy of 120 EeV and a zenith angle of 10°.

component dominates the region near the shower center and the muonic component is more important in the far region. The density of particles for a vertical shower with energy above 100 EeV is illustrated in Fig. 1.9(a) and their energy density is shown in Fig. 1.9(b). Since all the particles have a very high gamma factor they are concentrated near the core. For the electromagnetic cascade the lateral spread scales with the Moliere radius, which is proportional to the attenuation length and inverse proportional to the critical energy. Within this radius 90% of the air shower energy is contained. Its typical value for the atmosphere is ≈ 100 m.

The surface detector only samples the shower front of the air showers with small detectors. The total number of particles reaching the ground has to be inferred by fitting the individual sampled densities to an assumed lateral distribution function. A transformation from the number of particles reaching the detector to the number of particles at the shower maximum (i.e. energy) is done using air shower simulations. The lateral distribution is subject to fluctuations, from the detector response itself and from shower to shower, therefore the density at a certain distance from the axis that minimizes this effect is used instead. For the Pierre Auger Observatory the optimum distance is at ≈ 1000 m [78, 79].

The lateral distribution of particles obtained from simulations does not describe perfectly the data [80, 81]. In case of the Pierre Auger Observatory a parameterization obtained from data [82, 83] is used in the analysis.

To conversion of the signal at the optimal distance to the cosmic ray energy, independently of simulations, is possible only in the case of Pierre Auger Observatory, using coincident observations of the same air shower with a hybrid technique.

Chapter 2

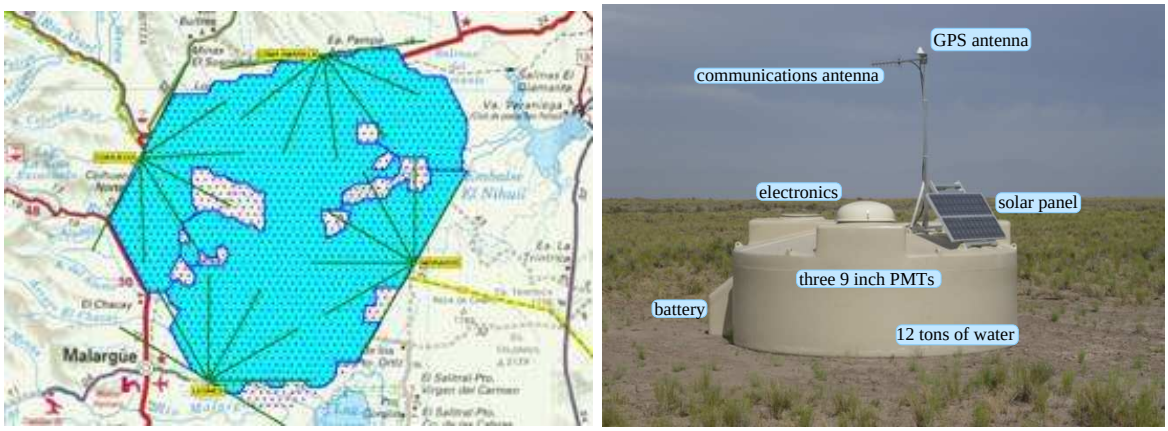
The Pierre Auger Observatory

The Pierre Auger Observatory is designed to have two sites in the northern and southern hemisphere. The northern site is in its planning stage, in southeastern Colorado near the city of Lamar, at an altitude of 1100 m above sea level (asl) and at a latitude of 38° north. The southern observatory, nearing completion, is located near the town of Malargüe in Argentina. Its latitude of 35° South provides a uniform sky coverage when combined with the northern site as well as a good survey of the galactic center. The need of full sky coverage is motivated by anisotropy studies, point source determination and correlations with astrophysical objects, as well as for determining the cosmic ray flux in different regions of the sky, since a different energy spectrum could give insides into different mechanisms of particle acceleration at the source.

The design incorporates two measurement techniques used with success in the past: Detecting the Nitrogen fluorescence in the atmosphere caused by particles of the extensive air shower and measuring the lateral distributions of particles that reach the ground. It consists of an array of water Cherenkov detectors covering 3000 km^2 with 1600 detectors spaced 1.5 km apart in a triangular grid. On the 15th of January 2008 there were 1592 tanks deployed, 1551 with water and 1481 with electronics. The completion will be achieved in March 2008. The mean ground slope of less than 1% and the altitude of 1400 m assures a measurement of the air shower at the same shower age, close to the maximum of the shower development of cosmic rays in the EeV range. On the edges of the surface detector array there are 4 fluorescence detector buildings with 6 telescopes each. The layout of the southern observatory is shown in Fig. 2.1(a).

The large surface detection area allows the collection of a large amount of statistics for analysis in a reasonable amount of time. With both detector sites completed and operational, the cosmic ray flux will be measured within 50 days with more statistics than the total of all experiments in the last 30 years. The stereo observation mode, by two or more fluorescence detectors, allows the understanding and the evaluation of the systematic effects arising from varying atmospheric conditions.

Most importantly the golden hybrid events, air showers recorded and reconstructed with both techniques, are useful for inter-calibration and data consistency checks. They are also valuable by providing two independent estimations of the primary particle nature.



(a) *Pierre Auger Observatory layout* (b) *surface detector unit*

Figure 2.1: (a) Pierre Auger Observatory configuration as of 19th December 2007. Individual tanks are marked with dots. Four telescopes stations with 6 telescope bays each are displayed. (b) Components of one of the SD stations.

2.1 Surface Detector

Water Cherenkov detectors were chosen as the cell unit of the surface detector (SD) of the Pierre Auger Observatory. The same principle of detecting air showers was used successfully in other experiments like Haverah Park [84].

A SD station [10], shown in Fig. 2.1(b), is a cylindrical polyethylene plastic tank with a base of $10\,\mathrm{m}^2$ and a height of 1.55 m. It encloses a Tyvek liner for uniform reflection of the Cherenkov light produced by charged particles that travel with a speed larger than the speed of light in the medium. The liner is filled with purified water (resistivity: $5 - 15\,\mathrm{M\Omega cm}$), up to a level of 1.2 m. Two solar panels combined with batteries provide a power of 10 W for the station.

The signals produced by the Cherenkov light are read out by three large 9" XP1805 Photonis photomultipliers (PMTs). They are equipped with a resistive base having two outputs: Anode and an amplified last dynode. The signal from the last dynode is amplified by a nominal factor of 32. This allows a large dynamic range, totaling 15 bits, extending from a few to about 10^5 photoelectrons. The high voltage (HV) is provided locally. The nominal operating gain of the PMTs is $2 \cdot 10^5$ and can be extended to 10^6. The base, together with the HV module, is protected against humidity by silicone potting.

The signals from anode and dynode are filtered and digitized at 40 MHz using 10 bit Flash Analog Digital Converters (FADCs). A pedestal of 50 channels is added to the signal to observe possible fluctuations of the baseline. The signal recorded by the FADCs is referred to in units of channels and is read by a programmable logic device (PLD) that performs trigger decisions on the signal.

A common time base is established for different detector stations by using the Global Positioning System (GPS) [85]. Each tank is equipped with a commercial GPS receiver (Motorola OnCore UT) providing one pulse per second output and software corrections. This

signal is used to synchronize a 100 MHz clock which serves to time-tag the trigger. Each detector station has a IBM 403 PowerPC micro-controller for local data acquisition, software trigger and detector monitoring, and memory for data storage.

Surface detector calibration

The calibration is designed to establish the signal sizes in each station, to set the high voltages of the PMTs, to give the amplification factor of the anode with respect to the last dynode and to set the trigger levels. The signal produced in a station depends on variable parameters like water quality, liner reflectivity, the coupling of the PMT to the water, the gain of the PMT, the gain of the last dynode, the electronic gain of dynode and anode amplifiers. Thus the calibration has to be independent on the number of PMTs functioning and on differences from station to station.

The basic physics calibration [86] information needed by the SD is the average charge produced by a vertical and centrally through-going muon (VCT). It is named vertical equivalent muon (VEM) and is the signal unit in the water tanks. The conversion to this unit is done to provide a common reference level between tanks. Therefore the calibration goal is to obtain the value of 1 VEM in electronic units for each individual station.

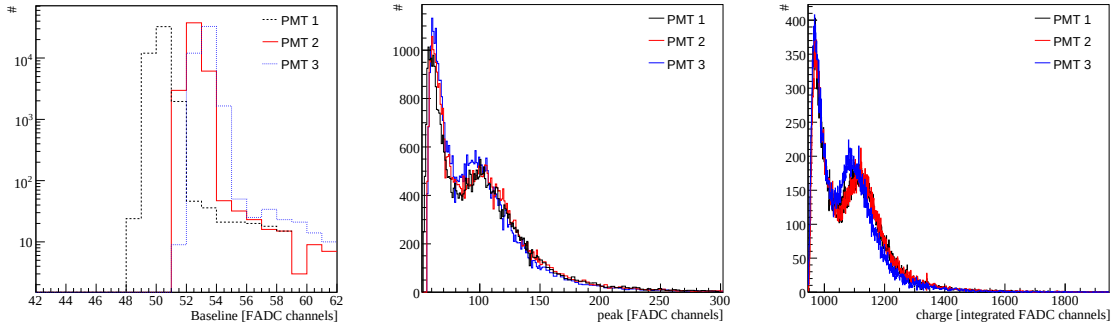
The surface detector in its normal configuration cannot distinguish the VCT from the inclined atmospheric muons passing constantly through. However the distribution of light from atmospheric isotropic muons produces a peak which has to be related to the VEM. The sum of the PMTs is a measure of the total signal deposited in the tank, whereas the individual PMTs are sensitive only to a fraction of the signal, deposited in the proximity. Thus this peak is at ≈ 1.09 VEM for the first case and 1.03 ± 0.02 VEM for individual PMTs.

These values were measured in a reference tank [87], a special setup consisting of a station equipped with two centered scintillators, one on top and the other underneath the station, the trigger requiring coincidences within the two scintillators.

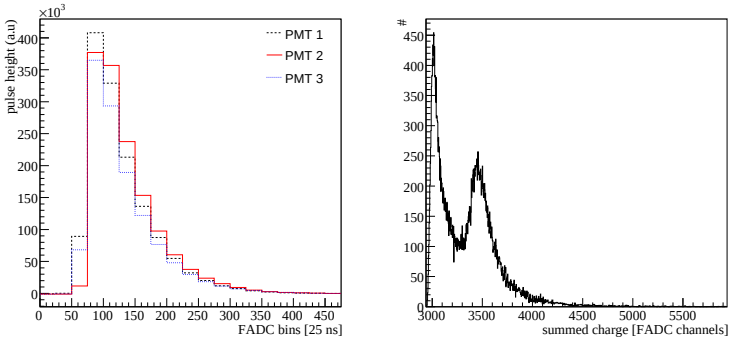
The calibration to VEM units includes three main steps and is done only on the peak produced by the omnidirectional muons: Set up the gains of each of the three PMTs to have the peak in the pulse height histogram of 50 channels; online local calibration to determine the peak and to adust the electronics trigger level, compensating for drifts from the initial value; determination of the peak from the charge histograms. The peak will be afterwards transformed to 1 VEM and will give the conversion from the integrated signal of the PMT to VEM. The baseline and the dynode to anode ratio have also to be determined.

PMT voltage setup The PMTs voltage is set at the deployment time and is determines the dynamic range of each PMT. Each PMT can be re-calibrated on request from the Central Data Aquisition System (CDAS). The initial value is chosen in such a way that a single VCT produces a peak of about 50 FADC counts, meaning a gain of $3.5 \cdot 10^5$ for 94 photoelectrons per VEM per muon.

The gains of the three PMTs are set up by matching a point in the event distribution to a measured rate from the reference tank. The latter is calibrated by obtaining a charge histogram and adjusting the three PMTs until the peak of the histograms agree. The gains of the PMTs are set such that the single rates at 150 channels above baseline are 100 Hz.



(a) *baseline histogram* (b) *pulse height (peak) histogram* (c) *charge histogram*



(d) *average pulse shape* (e) *summed charge histogram*

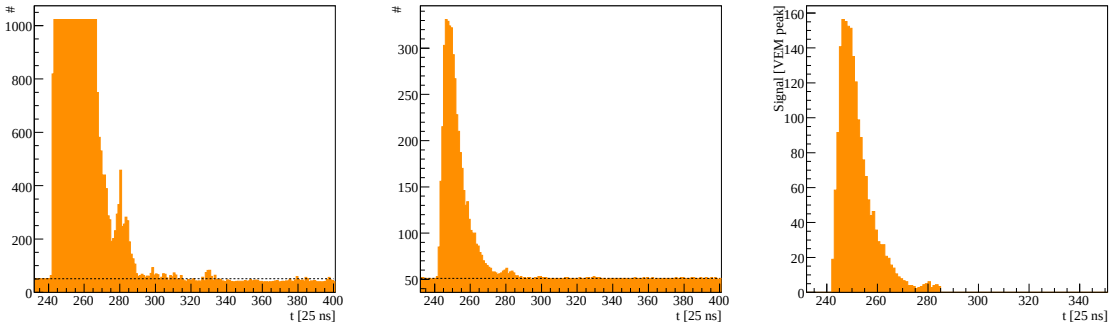
Figure 2.2: Example of the histograms produced by the surface detector unit and transmitted to CDAS for each event triggered by SD.

If one PMT has a worse optical coupling than the others in the tank than it would collect less light and therefore the gain is going to be set at higher values. The gain calibration is also sensitive to the water purity which might slightly differ from tank to tank.

Online calibration To maintain a uniform trigger condition of the array the station must be able to get to a common trigger threshold. Therefore the online calibration has to compensate for possible drifts in the gain of the PMT. This allows a tank independent analysis of the array response. The trigger levels are set on a minute by minute basis and compensate the drifts of the peak for each detector.

For each station of a measured SD event 13 histograms are send to CDAS: Baselines (Fig. 2.2(a)), peak histograms (Fig. 2.2(b)), charge histograms (Fig. 2.2(c)), shape histograms (Fig. 2.2(d)) for each PMT and a charge histogram for the sum of the pulses from each PMT (Fig. 2.2(e)). The shape histograms contain the average pulse of the pulses of charge $1 \pm 0.1 Q_{VEM}$, where Q_{VEM} is the VEM charge.

The goal of the online calibration is to obtain the same pulse height histograms for all the PMTs in the station at a uniform rate over the array. To obtain the value for the estimated VEM peak I_{est}, a so called $\sigma - \delta$ algorithm is applied for each PMT:



(a) *dynode trace* (b) *anode trace* (c) *calibrated trace*

Figure 2.3: Signal traces recorded in a single PMT.

(1) Initialize I_{est} to 50 ch.

(2) In a time t_{cal} (initialized with 5 s) the rate of events satisfying the calibration trigger is measured. If the rate is between 68 Hz and 72 Hz go to step 5 (δ initiated to 1 ch) .

(3) If the rate is between 72 Hz and 90 Hz, I_{est} is increased by $\delta = 1$ ch or if the rate is between 50 Hz and 68 Hz, I_{est} is decreased by $\delta = 1$ ch.

(4) If the rate is more than 90 Hz or less than 50 Hz I_{est} is adjusted by 5 ch in the appropriate direction. Then the procedure is repeated from step 2 with $t_{\text{cal}} = 10$ s

(5) If $t_{\text{cal}} < 60$ s it is increased by 5 s. If $\delta > 0.1$ ch it is decreased by 0.1 ch. The loop is continued at step 2.

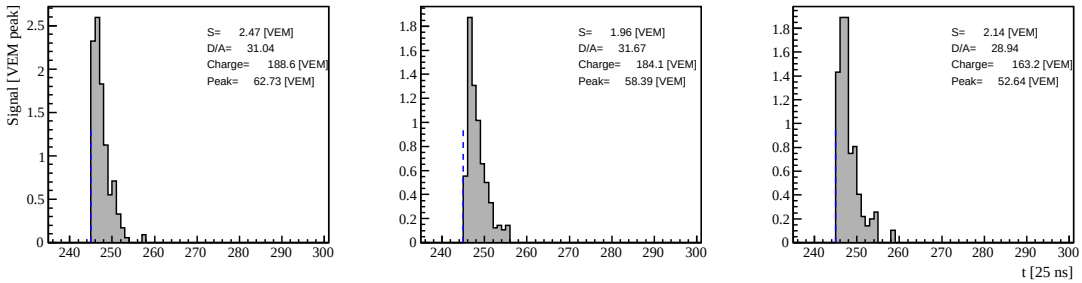
This algorithm has been shown to be very stable, leading to a very fast convergence [87].

Dynode to anode ratio The PMT signals are read from both the anode and the last dynode to extend the dynamic range for the measured signals. The nominal gain between the dynode and the anode is 32. An example of traces from a station where the high gain readout is saturated is given in Fig. 2.3. For a signal which is saturated in more than 50 time bins on the dynode ($\approx$ 950 channels with a 50 channels baseline, maximum 1023 channels) the anode signal will be merely 300 channels above baseline.

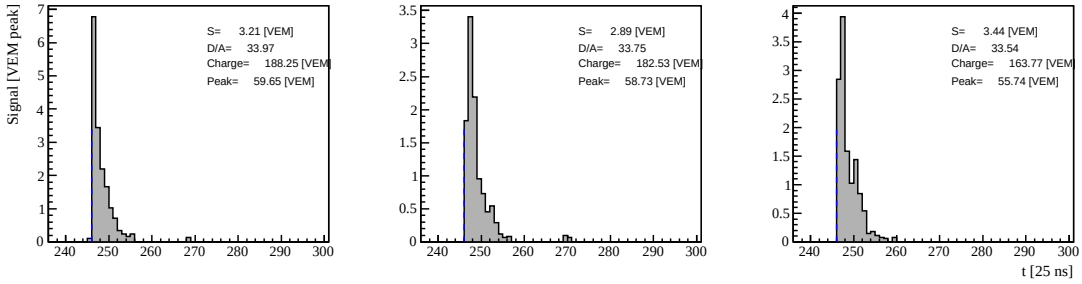
The relation between the dynode signal and the anode signal is

$$A(t) = \frac{1}{R}[(1-\varepsilon)D(t) + \varepsilon D(t+1)] \tag{2.1.1}$$

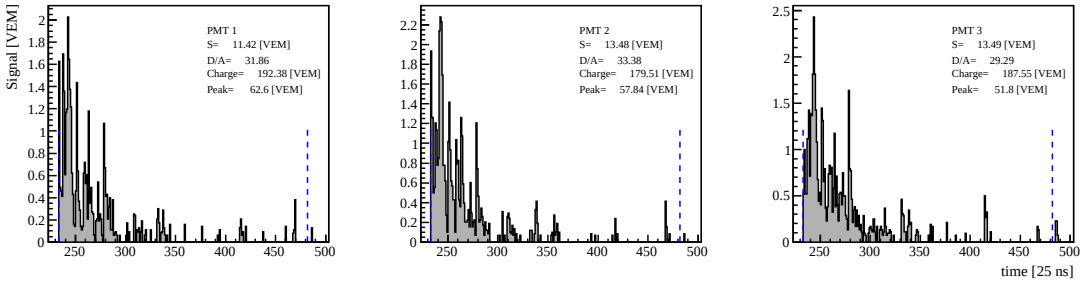
where $A(t)$ is the anode signal in the time bin t, R is the dynode to anode ratio, $D(t)$ is the dynode signal in the time bin t and ε is the fractional bin offset of the dynode. In the ideal case ε would be zero, leaving a linear relation between the two signals. In the case of the Pierre Auger Observatory the amplification stages add a small delay of ≈ 6 ns to the dynode signal. R and ε are determined performing a fit according to Eq. (2.1.1). For actual stations ε is on average 0.23 ± 0.04. Because the delay is only a fourth of the time bin size of the traces (25 ns) it does not have a big influence on the final signal and is thus neglected in the data analysis. The dynode to anode ratio is important for large signals, typically close to the shower center.



(a) *FADC traces for a station satisfying T1 threshold*



(b) *FADC traces for a station satisfying T2 threshold*



(c) *FADC traces for a station satisfying T2 ToT*

Figure 2.4: FADC traces for a tank satisfying the station triggers.

Baseline The baseline is computed from each of the 100 Hz calibration triggers obtained over a 60 second interval as well as its standard deviation from each of the six station traces.

Surface detector triggering system

The surface detector triggering system [88] can be seen as a chain, from the low level trigger received from the single tank (T1, T2) to the array trigger (T3) and to the physical events trigger (T4 and T5). The T2 and T3 conditions allow to detect the cosmic rays in a wide range of energies with an efficiency $> 95\%$ for cosmic rays with energies above 10^{18} eV.

Station trigger The local station trigger (T1) is evaluated by the PLD units and identifies the signals in a tank that could be part of real air showers. The station has to meet the

demands of a threshold trigger, a check for a coincidence between the 3 PMTs crossing the threshold value of 1.75 VEM above the baseline (Fig. 2.4(a)) or a coincidence condition in at least 2 PMTs of more than 12 FADC bins with a signal of more than 0.2 VEM above the baseline in a window of 120 time bins. The calibration procedure previously described ensures a T1 rate of about 100 Hz.

The second station trigger (T2) is processed by the local software and it asks for either a

1. threshold trigger (Thr2), requiring a coincidence of 3 PMTs above 3.2 VEM. This is meant to select fast large signals, produced by the particles from either high energetic EAS very close to the tank or to the muonic component in horizontal showers. The Thr2 rate is ≈ 20 Hz. The traces of a station obeying this condition are show in Figure 2.4(b).

or a

2. time over threshold trigger (ToT) which requires a coincidence of 2 PMTs with traces having at least 13 bins above 0.2 VEM within a sliding window of 3 μs width (120 bins). This selects small signals given by particles far from the EAS core or low energy showers. The ToT rate is 1 Hz. The traces of a station passing this condition are illustrated in Fig. 2.4(c).

When a tank trace satisfies both Thr2 and ToT only the latter is marked. If a tank satisfies the T2 threshold condition then it automatically satisfies the T1 condition. The Thr1 is usually observed for stations that are not part of an event, these stations are triggered by accidental muons, having a total signal of around 1-2 VEM.

Central Data Aquisition System (CDAS) trigger The lowest CDAS trigger (T3) identifies time coincidences between the signals in different tanks that could be associated with a real air shower. It does not guarantee that the data are physics events while a large number of chance coincidences in accidental tanks is expected due to low energy showers and to single cosmic muons. It considers any of the following requests:

(a) a 3-fold condition, which requires a coincidence within a time interval depending on the distance of three tanks passing the ToT condition.

(b) a 4-fold coincidence which requires the coincidence within a time window depending on the tank distance among 4 tanks having passed any T2 condition, with 2 tanks inside 2 hexagonal crowns from a triggered tank and a further one within 4 crowns. A crown is formed by the stations at equal distance from the center one and are numbered depending on this separation.

(c) a 3-fold condition which requires the coincidence of three aligned tanks passing any T2 condition

(d) an external condition generated by the fluorescence detector (FD)

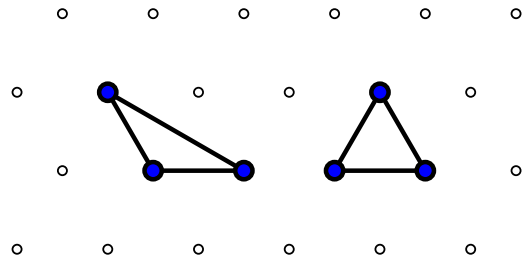


Figure 2.5: Two possible 3TOT compact configurations.

Every time a station has a T2 trigger it sends a signal to the CDAS containing the trigger time. The station trigger times are sorted. One by one all stations within a sliding time window of 50 μs are searched for the above patterns. If a pattern is found, the search stops and a first event trigger is formed.

For every T3 a request to all the stations in the array that had any trigger in coincidence with the station at the center is send and the timing and the FADC traces are stored. Trigger times are considered to be in coincidence if they are within (6+5n) μs from the trigger of the station at the center of the crown, where n indicates the crown number. For every event, a readout of the entire array is done. Some other information, e.g. the station position, id, calibration histograms and an error code, are also stored in the data file. The error code indicates whether the communication with the station was successful and can have the following values:

(1) The station already cleared the buffer for the requested time.

(2) No error, the station did not have a trigger at the requested time.

(3) T3 too early, the requested time was in the future.

(4) The FADC data at the requested time was already sent in a previous event.

(5) The station was not in acquisition mode.

(6) The data received from the station were corrupt.

(7) The tank did not respond.

The T3 trigger does not ensure that the events that are taken are physical events, but the philosophy is that a large set of events is recorded among which all the physical events are contained and the sorting is left for a subsequent analysis.

Physics trigger The first physics trigger (T4) was designed to distinguish air showers from random coincidences of single atmospheric muons and is also the first step to select reconstructible vertical events. It consists of

(a) 3ToT trigger requires at least 3 stations with a ToT trigger in a non-aligned configuration. This simple compact trigger is not effective for events with large zenith angles due to the dominance in this case of the muonic component which gives origin to fast less spread signals. It was shown in [88, 89] that this requirement selects 99% of the events with zenith angle less than 60° (Fig. 2.5).

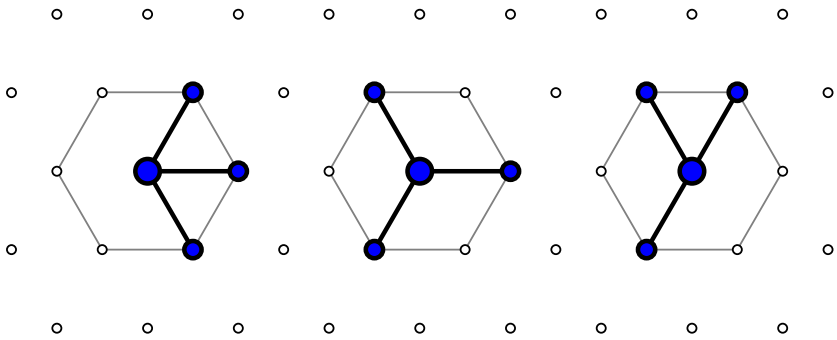


Figure 2.6: The three (minimal) 4C1 configurations.

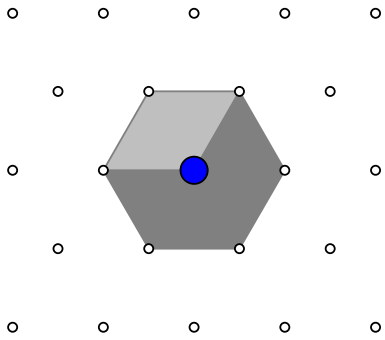
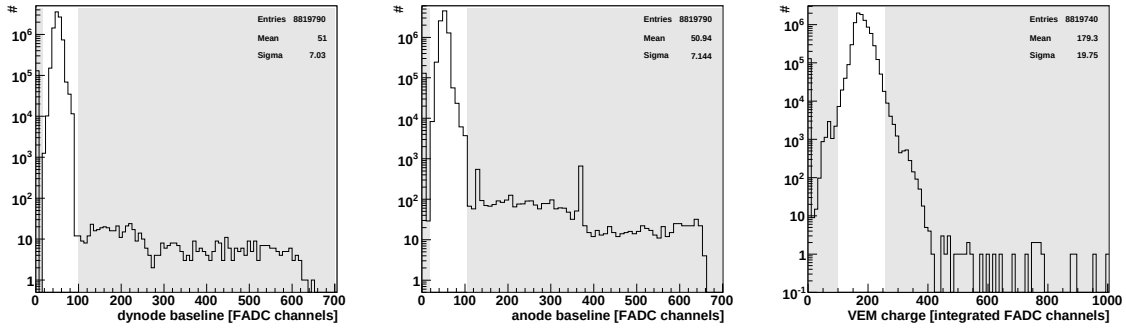


Figure 2.7: The T5 configuration. The central station (blue) with the largest signal is surrounded by 6 functioning stations.

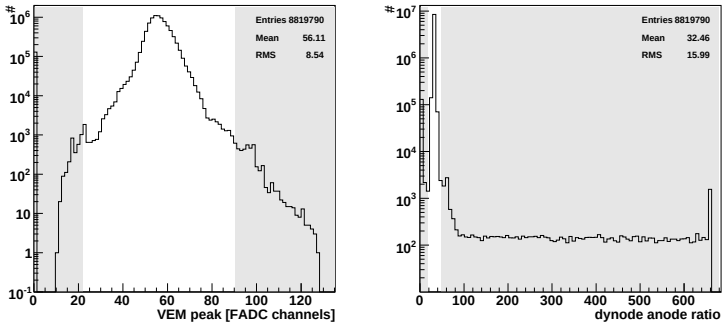
(b) 4C1 trigger is passed only by events that have 4 tanks with a T2 trigger each, and a configuration of one station with 3 close neighbors (Fig. 2.6).

In all cases of the T4 trigger, compatibility in time between stations part of the trigger is required. The difference in their start time has to be lower than the distance between them divided by the speed of light, allowing for a marginal limit of 200 ns.

Quality trigger The quality trigger (T5) [90] is meant to exclude events that fall too close to the edge of the SD array. For this class of events, due to a possible missing signal, the reconstruction of the air shower variables may not be reliable. Another reason is that it is very hard to compute the acceptance which would take into account events that are highly energetic but far away from the array. In such cases the trigger probability for 4 tanks on the edge of the array depends on fluctuations which are very hard to simulate. This quality trigger is based on a criterion related to the core position: For the station with the largest signal it is required to have six nearest neighbors that were present and functioning (but not necessarily triggered) at the time of the shower impact (Fig. 2.7). This assures a good and unbiased reconstruction of the event.



(a) *dynode baseline* (b) *anode baseline* (c) *VEM charge*



(d) *VEM peak* (e) *dynode anode ratio*

Figure 2.8: Allowed ranges for the PMT calibration constants.

2.2 Stability of the Surface Detector

The number of T2 triggers which a given station has sent to the CDAS is stored each second in a special file called T2 file. These files are the basis of the calculation of the instantaneous area of the SD (see Section 7.1) by monitoring the number of active stations second by second. In addition, every 10 minutes the station reports to CDAS the values read by a set of sensors that monitor the station, including the voltage of the batteries, water temperature and trigger levels.

The PMT calibration constants can vary in time, due to water losses in some tanks, electronic problems, PMTs with a bad coupling to the liner, etc. Even though they do not influence the acceptance calculations they can influence the reconstruction quality of the air shower parameters. In order to see the stability of the calibration in a long time period, data from January 2004 to June 2007 have been analyzed.

As illustrated in Fig. 2.8 the distributions of the calibration constants have large non-Gaussian tails, which might influence the final reconstruction. Therefore signals from *mal functioning* PMTs are rejected. The allowed values for the calibration variables are deduced as 5σ deviations from the mean values given in Table 2.1. In Appendix A examples of the influence on the reconstructed energy of some special events are shown. The number of rejected stations from the total number of stations in the data until August 2007 is less than 1%, therfore it

Table 2.1. Mean values for the calibration constants.

	dynode baseline	anode baseline	VEM charge	VEM peak	dynode-anode ratio
mean	51.0	50.9	179.3	56.11	32.46
RMS	7.03	7.11	19.75	8.54	16.0

does not influence the acceptance of the surface detector.

Another occurring problem at the PMT level is the so-called *raining* behavior [91]. These PMTs show a drop of the VEM value of the order of $\leq 30\%$. This variations appear in the scale of tenth of minutes resulting in an increase of the ToT trigger rate and influencing the acceptance of the detector. The impact of the raining PMTs on the measured cosmic rays flux is presented in Section 7.1. It turns out that the problem is caused by a faulty cabling procedure of the PMTs.

2.3 Fluorescence Detector

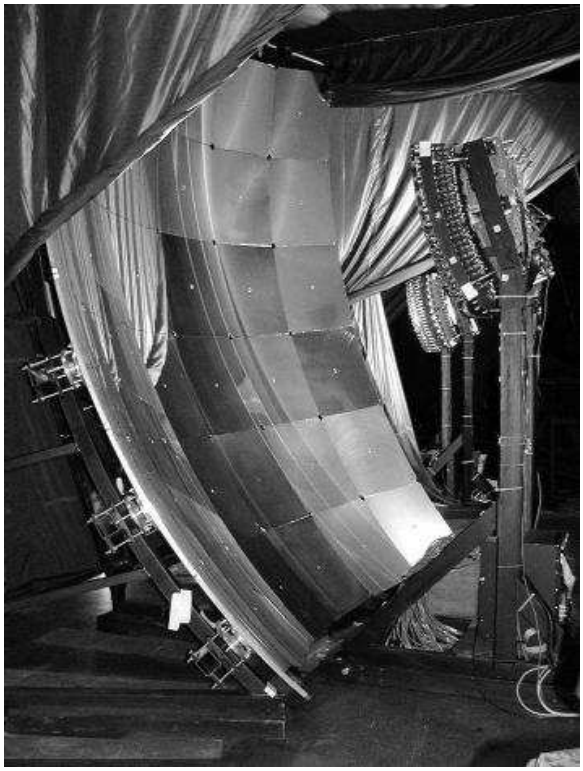
The atmosphere above the SD detector is part of the fluorescence detector (FD), being viewed by 24 Schmidt telescopes placed in 4 buildings at the border of the array. One of the buildings, Los Leones, is shown in Fig. 2.9(a). The field of view for each telescope is $30° \times 30°$.

The optical system consists of a diaphragm of $3.8\,\mathrm{m}^2$ aperture, the mirrors of $12\,\mathrm{m}^2$ with a radius of curvature of 3.4 m. To increase the signal to noise ratio in the fluorescence band a UV transmitting filter, Schott M-UG6 is installed at the aperture. The optical aberrations are reduced with a ring of corrector lenses installed [92] around the diaphragm (Fig. 2.9(c)), increasing the collection area of the telescope, while keeping a maximum angular size of the light spot reflected on the camera to 0.5°. The camera is equipped with 440 hexagonal PMTs (Photonis XP-3062) and is installed at the focal surface of the structure. The PMTs have a diameter of 1.5° and are arranged in a 22×20 matrix. A picture of the optical system can be seen in Fig. 2.9(b). Each PMT is equipped with a head electronic unit [93]. The signal from the PMT is send to the analog board where it is filtered and amplified [94]. Each analog board serves 22 channels, controlling their gain. Programmable potentiometers equalize the gains of the PMTs to guarantee a uniform time response and amplitude. The analog board is connected to the digital front-end board which hosts the trigger boards.

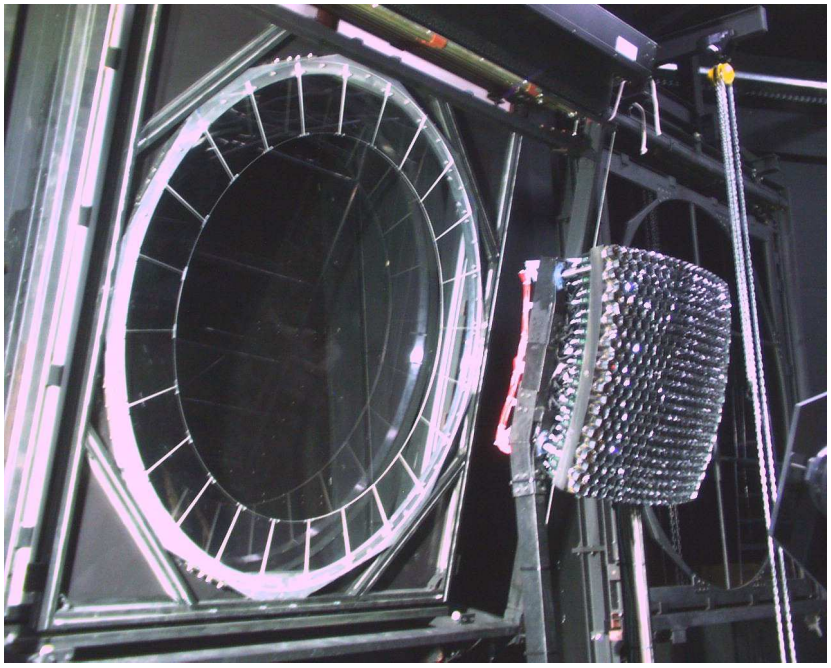
Trigger The fluorescence detector (FD) trigger is a 4 level trigger [95]:

1. The pixel trigger is built by four Field-Programmable Gate Arrays (FPGAs), each controlling 6 channels. The FADC values are integrated over 10 bins improving the signal to noise ratio by a factor of $\sqrt{10}$. The threshold is adjusted continuously to achieve a pixel trigger rate of 100 Hz.

(a) *Los Leones*



(b) *mirror and PMT camera*



(c) *UV filter with the corrector ring and the PMT camera*

Figure 2.9: Setup of the fluorescence detector (FD).

2. The second level trigger is also decided by the FPGA trigger board. It consists of a purely geometric pattern recognition. It searches for 4 or 5 adjacent pixels overlapping in a time window of 1 to 32 μs. The rate is 0.1 Hz per mirror.

3. The last trigger is implemented in software [96]. It checks for the time structure of an event. The average trigger rate is 0.02 Hz per mirror.

4. After the last mirror-level trigger the data are collected by the eye PC, where the events have to pass the eye trigger level T3 which performs rudimentary event reconstruction of the direction and of the time of impact on the ground that is used in CDAS for reading the corresponding part of the array. The slow control system [97] handles the operational data and experimental conditions which are stored in a local database available for the on-line monitoring together with calibration tests and various performance parameters [98].

Calibration and atmosphere monitoring

To correctly determine the size of the shower at a given depth, the number of photons emitted via nitrogen fluorescence for an electron that travels through a certain distance in the atmosphere, the photon yield, must be known. The fluorescence light is attenuated through scattering in the atmosphere, therefore the atmosphere must be monitored during data taking to parameterize attenuation lengths and scattering due to aerosols. Finally, the calibration of the detector must be such that for a given pixel the integrated signal can be converted into an absolute number of photons.

The fluorescence yield has been measured by independent experiments [99–102]. Within the Pierre Auger collaboration the absolute measurement of the fluorescence yield is taken from [99], while the wavelength, temperature and pressure dependence is taken from [100].

Detector calibration The absolute calibration [103] of the FD is an end-to-end calibration in the sense that it accounts for all the components of the system from filters to mirrors to PMTs to the readout electronics. The calibration is done with a diffuse light source that is 2.5 m in diameter and can be placed on the aperture of a telescope. The light intensity and uniformity are measured in the lab. The diffuse light source is able to uniformly illuminate all the pixels in the camera with a known light intensity in order to calculate the conversion from integrated signal to number of photons incident on the pixel. The absolute calibration is done 3 or 4 times a year and there is relative calibration [104] procedure that is run nightly to monitor any changes in the system.

This absolute calibration of the detector has currently an uncertainty of 12%.

Atmosphere monitoring In addition to the absolute calibration of the telescope, there is still a need to correct for attenuation effects in the atmosphere in order to obtain the absolute number of photons emitted at the shower axis.

The amount of aerosols in the air are monitored with backscatter Light Detection and Rangings (LIDARs) which are steerable UV lasers located at each FD eye [105]. Each LIDAR has a PMT that detects the backscattered light from the UV laser pulses. The

timing information gives information about the aerosol content of the air at any given spot along the path of the laser. This system can check various locations and directions in the sky and eventually will be able to measure the aerosol content along the development of an air shower.

In addition to the aerosol content of the atmosphere, it is important to know the atmospheric density and temperature profiles. The fluorescence yield has both a pressure and temperature dependence, which change with altitude. In the past, a parameterized atmosphere was assumed in analyzing data based on the assumption that atmospheric conditions were relatively stable.

For the atmosphere seen by FD meteorological radio soundings have been performed. Radiosondes [106] are launched with helium balloons and data is taken every 20 m in altitude until reaching 25 kmasl. The profiles are recorded and then compared to the parameterization used previously. If the parameterized values are used instead of the measured profiles, the shower depth at maximum values change on average $15\,\mathrm{g/cm^2}$ while the energy changes less than 1%.

A useful tool to cross-check the calibration and atmospheric monitoring is the Central Laser Facility (CLF) [107]. The CLF is used to check the angular reconstruction, the atmospheric conditions, the relative timing between the SD and FD [108], and the calibration of the telescopes. The pulsed laser light is scattered by the air and is detected in the FD providing a test beam to cross-check all important quantities in determining the properties of an extensive air shower.

Chapter 3

Event Reconstruction

The air showers recorded by the Pierre Auger Observatory are characterized by triggered stations on the ground array, i.e. discrete samples of the arriving particles in a plane, or by triggered pixels in the fluorescence detectors, samples of the energy deposit in the atmosphere. From this informations the lateral distribution of the particles on the ground and the lateral development of the shower are reconstructed. The variables important for this work are the S(1000 m), obtained from the SD reconstruction, and the energy of the primary particle, obtained from the hybrid reconstruction. For this reason the way of obtaining these two variables is emphasized next.

The reconstruction is done within the $\overline{\text{Off}}\underline{\text{line}}$ framework [109]. Another compatible SD reconstruction is done in CDAS [110]. The difference in S(1000 m) given by the two algorithms is presented in Appendix A.

3.1 SD event reconstruction

The goal of the reconstruction is to obtain the geometry of the air-shower, the energy estimator S(1000 m) and mass sensitive parameters (curvature and the rise time at 1000 m from the shower core) as it has been described in [111].

After selecting the stations that are belonging to the event, the impact point on the ground and the arrival direction of the air shower are estimated, assuming a simple plane front as first approximation of the shower geometry. Then the shape parameters of the lateral distribution function (LDF), which are dependent on the zenith angle, are initialized. The S(1000 m) initial value estimation is based on the signal in the station closest to a distance of 1000 m to the shower center. In the next step the S(1000 m) and core location fitting is performed with a maximum likelihood method and depending on the number of candidate stations, the LDF parameters are gradually included as variable parameters.

Fixing the core position, the curvature is reconstructed and all the previous steps are repeated iteratively because the axis of the shower might change in this last step and therefore the whole configuration of the shower.

Station selection In order to obtain only signals that belong to the actual air shower a selection at the level of PMTs and station is necessary. The selection algorithms has been described in [112]. The reasons to remove a station from the reconstruction are bad calibrations and/or accidental timing information.

The accidental triggered stations given by the atmospheric muons are removed based on the distance to the first neighbors, requiring at least one other triggered station within 1800 m or more than one within 5000 m, and on timing information, applying the bottom up selection algorithm. This algorithm requires compatibility between the timing in the stations with a planar front propagating with speed of light. An initial guess of the plane front is done using a seed, the three stations that maximize the sum of signals. If these stations are compatible with the analytical solution the event is considered good enough for a proper reconstruction.

For the other stations, i, the allowed range for the time delay to this initial planar shower front time is

$$-1000\,\text{ns} < \Delta t_i < 2000\,\text{ns}, \tag{3.1.1}$$

the asymmetry in the values is caused by to the probability that the signal is delayed in stations with lower signal, due to the curved shower front.

Additionally to the regular distribution, some stations are placed in special configurations. For signal and timing accuracy studies two stations are placed one near each other at a distance of ≈10 m. The one with the higher id in this configuration is removed by default from the event (even if the partner station is not present in the event). This allows the rejection of any bias that might be introduced by the sampling of the LDF at such small distances, and an easy computation of the acceptance. The stations placed outside of the regular grid in order to locally increase the density of the array, *infill* stations, are rejected as well.

Geometry reconstruction The simplest model of the shower front can be approximated by a plane, which is a robust estimator of the shower axis direction.

A shower track (see Fig. 3.1) can be visualized as a point $\widehat{x}(t)$ moving with the speed of light c along the straight line with (normalized) axis $\vec{a}$, and passing the origin at time t_0,

$$c(t_0 - t) = (\widehat{x}(t) - \widehat{b})\vec{a}. \tag{3.1.2}$$

The *origin* $\widehat{b}$ from where all the distances are measured is set as the signal-weighted barycentre of the all the stations. The weighted bary-time is set as time origin, t_0.

The *shower plane* is a plane perpendicular to the shower axis, moving along with the same speed and containing the shower forehead. To infer on the time $t(\widehat{x})$ when the shower plane is passing through some chosen point $\widehat{x}$ on the ground, the point has to be projected to the shower axis,

$$ct(\widehat{x}) = ct_0 - (\widehat{x} - \widehat{b})\vec{a}. \tag{3.1.3}$$

Assuming that the positions of the stations are given with absolute precision and the only deviations can be due to the time uncertainty σ_t of the signal start, the function to minimize is the squares of the time differences between the measured signal start and the model (Eq. (3.1.3)) time prediction. Denoting the components of the shower axis as $\vec{a} = (u, v, w)$,

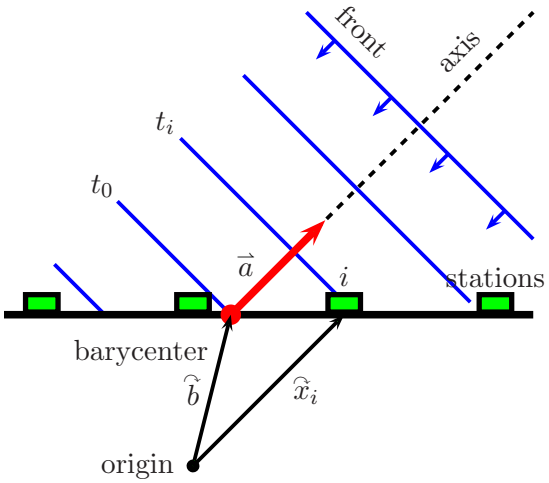


Figure 3.1: Schematic of the plane front arrival.

the station coordinates with $\vec{x}_i = (x_i, y_i, z_i)$, and σ_i the uncertainty of the time multiplied with the speed of light

$$\chi^2 = \sum_i \frac{[ct_i - ct_0 + x_i u + y_i v + z_i w]^2}{\sigma_i^2}. \tag{3.1.4}$$

with a constraint of $u^2 + v^2 + w^2 = 1$ inherited. Due to this constraint, the problem is not linear, but an approximate solution can be obtained in the ansatz that all stations lay close to a plane, $z_i \ll x_i$, y_i and therefore the z-component can be neglected.

The minimization can fail only in one case, when there is a linear dependence of the station positions (as when having three stations in a line). For higher station multiplicity the occurrence of such a situation is highly unlike.

The more realistic shower front model is based on a curved front fit, as illustrated in Fig. 3.2, done by extending the plane fit method with a parabolic term that describes the curvature of the shower front near the impact point $\hat{c}$, i.e. $\rho \ll R_c$. Using $\vec{x} = \hat{x} - \hat{c}$, Eq. (3.1.3) can be extended to get

$$ct(\hat{x}) = ct_0 - \vec{a}\,\vec{x} + \frac{\rho(\vec{x})^2}{2R_c}, \tag{3.1.5}$$

with perpendicular distance $\rho(\vec{x})^2 = (\vec{a} \times \vec{x})^2 = x^2 - (\vec{a}\,\vec{x})^2$. A first approximation to the radius of curvature is obtained from a slightly different model, where the time propagation of the shower front is described as an expanding sphere where the timing information can be decoupled from any information on the impact point. The fit parameters are the radius of curvature, R_c and the shower axis. The shower axis is a derived quantity obtained only after the position of the impact point is known, therefore the curvature fit is done only after the lateral distribution function fit, after the core is known.

The solid angle differences between the plane-fit and curvature-fit axis $\vec{a}$ are of the order of a half degree.

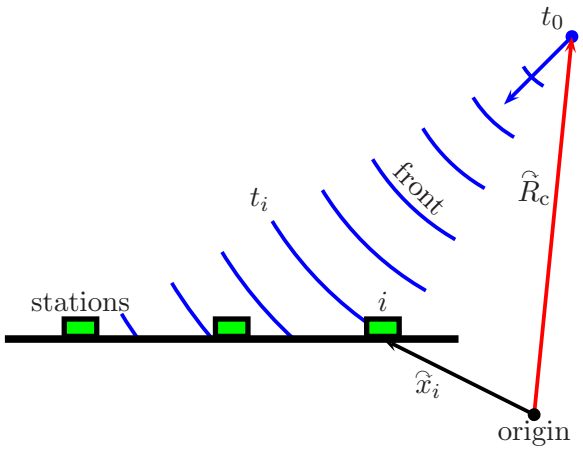


Figure 3.2: Schematic of the spherical shower front development.

An exact three-dimensional minimization of a function

$$\chi^2 = \sum_i \frac{[c(t_i - t_0) - |R_c\vec{a} - \vec{x}_i|]^2}{c^2\sigma_{t_i}^2} \tag{3.1.6}$$

is also attempted with accurate $z_i \neq 0$ treatment. The differences between the approximate estimation of R_c and this one are of the order of few 10 m, while the solid angle difference between the axes is of the order of a few 0.1°.

The intrinsic time variance model from [110,113] is used by default. In this model the number of particles (muons) n in the signal is estimated from the VEM station signal S, corrected for the zenith angle dependence of the average track length $\bar{L}$,

$$n = \frac{S}{\ell(\theta)}, \qquad \ell = \frac{\bar{L}(\theta)}{\bar{L}(0)}, \qquad \bar{L}(\theta) = \frac{\pi}{\pi\cos\theta + 2(h/r)\sin\theta} \tag{3.1.7}$$

where h and r are tank height and radius, respectively.

Lateral distribution function

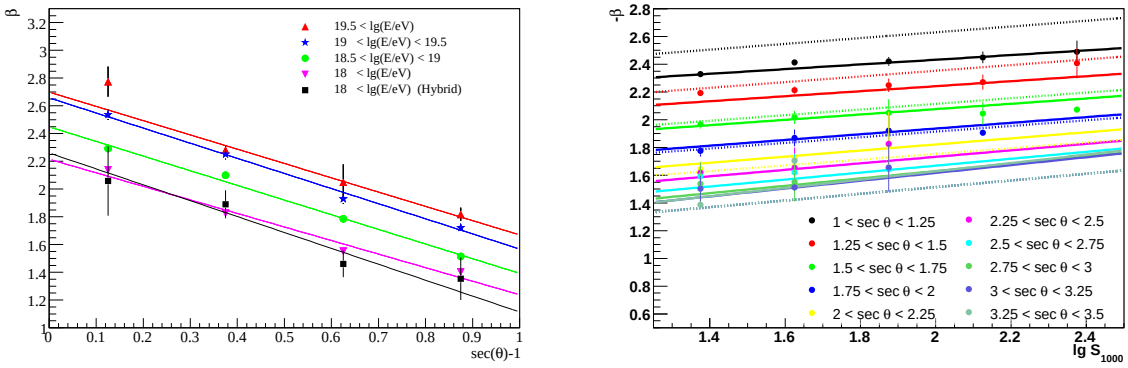
The surface detector samples only a part of the particles arriving at ground therefore a fit of the lateral distribution has to be done. The lateral dependence of the signal measured in tanks is modeled as

$$S(r) = S(1000\,\mathrm{m})\, f_{\mathrm{LDF}}(r), \tag{3.1.8}$$

where $f_{\mathrm{LDF}}(r)$ is a particular shape parameterization normalized to S(1000 m) at 1000 m. The uncertainty of the signal [114–116] is taken as

$$\sigma_S(\theta) = (0.32 + 0.42\,\sec\theta)\sqrt{S}. \tag{3.1.9}$$

The shower impact point must be accurately defined since S(1000 m) is directly related to this quantity. It is defined to lie in the plane tangent to the Earth's reference ellipsoid containing the barycentre of the stations. The core location $\widehat{c}$, obtained by the previous stages in the



(a) *theta dependency* (b) *S(1000 m) dependency*

Figure 3.3: (a) Parameterization of the shape parameter as a function of the zenith angle. The values obtained from the golden hybrid data, by fixing the core in the reconstruction procedure to the one given by the hybrid reconstruction are shown for comparison. (b) The latest parameterization of the LDF parameter β on S(1000 m) [118]. The parameterization used in this thesis is shown by dashed lines.

reconstruction chain is projected along the shower axis $\vec{a}$ towards the plane defined by the barycentre $\widehat{b} = (0,\, 0,\, 0)$ and a normal $\vec{n} = (0,\, 0,\, 1)$

$$\widehat{c}' = \widehat{c} + \frac{\vec{n}(\widehat{b} - \widehat{c})}{\vec{n}\,\vec{a}}\,\vec{a}, \tag{3.1.10}$$

Several functional forms of the LDF have been investigated in [117, 118]. It turned out that the best description is given by a power-law-like function and a modified Nishimura Kamata Greisen (NKG) function [119, 120]. The function used in this analysis is a slightly modified NKG function

$$f_{\mathrm{LDF}}(r) = \left(\frac{r}{r_{1000}}\right)^{\beta} \left(\frac{r + r_{700}}{r_{1000} + r_{700}}\right)^{\beta}, \tag{3.1.11}$$

with $r_{700} = 700\,\mathrm{m}$ and the slope β fixed according to the following parameterization

$$\beta(\theta) = \begin{cases} a + b(\sec\theta - 1), & \sec\theta < 1.55 \quad (\theta \geq 50^\circ), \\ a + b(\sec\theta - 1) + f(\sec\theta - 1.55)^2, & \sec\theta \geq 1.55. \end{cases} \tag{3.1.12}$$

with

$$\begin{aligned} a &= 2.26 + 0.195\,\lg e, \quad b = -0.98, \\ c &= 0.37 - 0.51\,\sec\theta + 0.30\,\sec^2\theta, \quad e = c\,(S_{1000})^d, \\ d &= 1.27 - 0.27\,\sec\theta + 0.08\,\sec^2\theta, \quad f = -0.29. \end{aligned} \tag{3.1.13}$$

A simpler parameterization as deduced with data before June 2005, is shown in Fig. 3.3(a). The parameterization evolved meanwhile adding a quadratic term in Eq. (3.1.12). The increase of event statistics made a more accurate parameterization [118] possible, but it was not implemented in the Offline framework at the time of writing. A comparison to the default values is shown in 3.3(b). The effect of the new parameters on S(1000 m) is at a level of 2%.

If β is estimated to be larger than -1, it is fixed to this value. To notice is that as this parameter is negative the LDF is divergent for $r = 0$. The reconstruction of the LDF is done only with three free parameters: S(1000 m) and the core location (two components, x and y). The uncertainties on S(1000 m) from fixing β are obtained doing two additional reconstructions with $\beta \pm 3\%$.

Maximum likelihood fit of the lateral distribution function The water-Cerenkov tanks provide information about Cerenkov photons, which are released by muons, electrons, or converted photons when passing the tanks. The first step is to define an effective particle number, which can be used to apply a maximum likelihood method, being able to include zero-signal stations, small signals (i.e. small particle densities) by means of Poisson statistics, large signals by a Gaussian approximation and handle the signal of saturated stations.

The total signal measured in a tank has two major contributions: Electromagnetic part and the muonic part

$$S = S_\mu + S_{e/\gamma} \tag{3.1.14}$$

in the assumption that a single converted photon and a single electron equally energetic deposit the same mean signal in a tank. A muon is considered to deposit 1 VEM irrespective of incoming angle, distance, etc.

The signal $S_{e/\gamma}$ is much smaller on average than S_μ and the mean conversion factor for electrons and photons to signal is smaller than 1 VEM. The total number of particles that have produced the signal is then estimated as

$$n = p(r, \theta|\, E, A)\, S, \tag{3.1.15}$$

where $p(r, \theta|E, A)$, the so-called *Poisson factor*, is approximated to 1 for $\sigma_S(\theta) \geq 1$ and $1/\sqrt{\sigma_S(\theta)}$ otherwise, independent of primary energy and mass as well as core distance and zenith angle [114]. The signals recorded in tanks close to the trigger threshold S_{thresh}, specially at larger radii, have a large muon content, therefore p is taken to be 1. Assuming that the transition to half-signal electron deposit content takes place when the signal exceeds the threshold $S^{\text{G}}_{\text{thresh}}$, the final simplified conversion between signal and particles is

$$n(r, \theta|E, A) = n = \begin{cases} S(r, \theta|\, E, A) & ; S < S^{\text{G}}_{\text{thresh}} \\ 2S(r, \theta|\, E, A) & ; S \geq S^{\text{G}}_{\text{thresh}} \end{cases}. \tag{3.1.16}$$

where $S^{\text{G}}_{\text{thresh}} = 15$ VEM corresponding to an estimated number of particles of ≈ 30.

The likelihood function, gathering the sampled information of tank i at distance r_i to be maximized is

$$L = \prod_i f_{\text{P}}(n_i, \mu_i) \prod_i f_{\text{G}}(n_i, \mu_i) \prod_i F_{\text{sat}}(n_i, \mu_i) \prod_i F_{\text{zero}}(n_i, \mu_i) \tag{3.1.17}$$

and thus the log likelihood function gives

$$\ell = \sum_i \ln f_{\text{P}}(n_i, \mu_i) + \sum_i \ln f_{\text{G}}(n_i, \mu_i) + \sum_i \ln F_{\text{sat}}(n_i, \mu_i) + \sum_i \ln F_{\text{zero}}(n_i, \mu_i), \tag{3.1.18}$$

with n_i the effective number of particles detected in the tank and μ_i the corresponding theoretical expectation. The different contributions are from

(a) Small signals, $n < 30$ particles which have a Poisson distribution

$$\ln f_{\mathrm{P}}(n_i, \mu_i) = n_i \ln \mu_i - \mu_i - \sum_{j=1}^{n_i} \ln j. \tag{3.1.19}$$

(b) large signals, $n > 30$ particles, distributed Gaussian

$$\ln f_{\mathrm{G}}(n_i, \mu_i) = -\frac{(n_i - \mu_i)^2}{2\sigma_i^2} - \ln \sigma_i - \ln \sqrt{2\pi}. \tag{3.1.20}$$

(c) stations without signal. The assumed threshold to trigger a tank is $n_{\mathrm{thresh}} \geq 3$, i.e. at least 3 muons hitting the tank. Therefore their contribution is a sum over all Poisson probabilities with a predicted particle number μ_i and actual particle number $n_i \leq n_{\mathrm{thresh}}$

$$\ln F_{\mathrm{zero}}(n_{\mathrm{thresh}}, \mu_i) = -\mu_i + \ln \left(\sum_{n=0}^{n_{\mathrm{thresh}}} \frac{\mu_i^n}{n!} \right) \tag{3.1.21}$$

There might be a slight complication due to different local trigger algorithms. In a first approximation the *threshold trigger* ($S_i \simeq 3\,\mathrm{VEM}$) fulfills the n_{thresh} condition, while a *time over threshold* (ToT) has certainly a lower threshold ($S_i \simeq 1.7\,\mathrm{VEM}$) but is unlikely to be the trigger at larger radii.

(d) saturated signal, for which n_i represents a lower limit on the actual signal. Integrating f_{G} over all possible values larger than n_i, an estimate of the probability detecting a signal larger than n_i is obtained as

$$F_{\mathrm{sat}}(n_i, \mu_i) = \int_{n_i}^{\infty} f_{\mathrm{G}}(n, \mu_i)\, \mathrm{d}n = \frac{1}{2} \mathrm{Erfc} \left(\frac{n_i - \mu_i}{\sqrt{2}\sigma_i} \right), \tag{3.1.22}$$

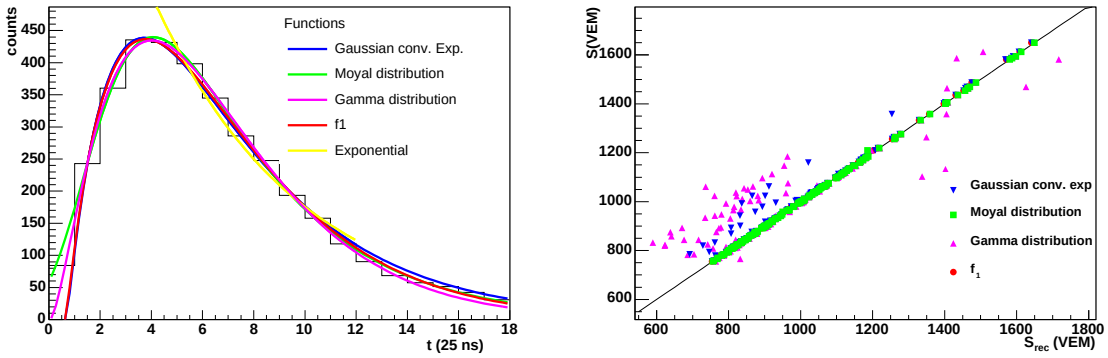
where $\mathrm{Erfc}() = 1 - \mathrm{Erf}()$ is the complementary *error function*.

The missing signal can be recovered as described in section 3.2. The inclusion of the recovered signal in the maximum likelihood function will be described later.

3.2 Saturation recovery

When a station is close to the air shower core, expecially for vertical high energetic showers, saturation of the tank electronics may occur, and therefore the recorded signal is incomplete. One contribution to the missing signal comes from the FADC saturation of the anode recordings ($\approx$ 950 channels with a 50 channels baseline). For high signals the PMTs are deviating from the linear behavior, adding an extra contribution to the inaccuracy of the signal.

To avoid the effect of the overflow of the FADC channels, two studies [121,122] were performed. They are measuring the signal undershoot (change in the baseline, before and after the signal) due to the coupling capacitors on dynode and anode circuits. It was found that the undershoot is nearly proportional to the total charge. As a drawback these methods can not take into



(a) *example of a fitted trace* (b) *reconstructed signal vs true signal*

Figure 3.4: Description of the PMT signal for different trial functions.

account the PMT nonlinearity. Another approach to recover the signal, part of the present work, is based on a parameterization of the shape of the time distribution of the signals [123].

The two different methods are combined in this section using the undershoot method, the measured non-linear response of the PMT and the charge shape to derive the value of the total signal.

Pulse shape In order to deduce the pulse shape the first test was done on the non-saturated traces. Four trial functions that give a good description of the signal distribution in time are considered.

(a) Supanitsky et al. [124] showed that the traces are well described by the function

$$f_1(t) = \frac{a}{\sigma^2} \cdot t \exp\left(-\frac{t}{\sigma}\right). \tag{3.2.1}$$

(b) and by the gamma distribution

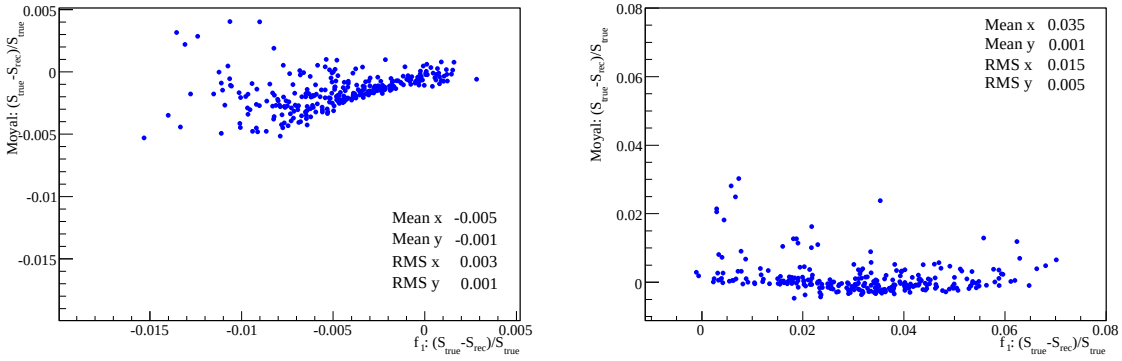
$$f_2(t) = a \cdot (t - t_0)^b \cdot \exp\left(-c \cdot (t - t_0)\right) \cdot c^{b-1} \tag{3.2.2}$$

Furthermore, choosing f_1 for simplicity a parameterization of the σ parameter as a function of the distance to the core in shower plane coordinates, R, was obtained. We use the same parameterization for the width of this function.

(c) Gaussian convoluted with exponential

$$f_3(t) = b \cdot \exp\left(-a(t - t_0) + \frac{1}{2}a^2\sigma^2\right)\left[\mathrm{Erf}\left(\frac{a\sigma}{\sqrt{2}}\right) + \mathrm{Erf}\left(\frac{a\sigma^2 - (t - t_0)}{\sqrt{2}\sigma}\right)\right]. \tag{3.2.3}$$

(d) The Moyal distribution gives a good description of the electromagnetic as well as of the muonic component of simulated traces [125]. Accordingly, the total shape of the traces is reproduced by the sum of two Moyal distributions. If only the part close to the shower



(a) *variable parameters* (b) *fixed parameters*

Figure 3.5: The relative difference between the true signal and the reproduced signal with the function f_1 and the Moyal distribution.

core is of interest, the traces are dominated by the electromagnetic component and it is sufficient to restrict to only one distribution

$$f_4(t) = a \cdot \exp\left(-\frac{t-t_0}{2\sigma} - \frac{1}{2}\exp\left(-\frac{t-t_0}{\sigma}\right)\right). \tag{3.2.4}$$

In Fig. 3.4(a) a trace fitted by the functions given above is shown. All the functions estimate the signal with mean errors smaller than 5%. The main difference is constituted by the underestimation of the true signal in case of gamma distribution and Gaussian convoluted with an exponential (Fig. 3.4(b)).

The Moyal parameters depend on the distance to the core in shower plane coordinates, with an exponential behavior

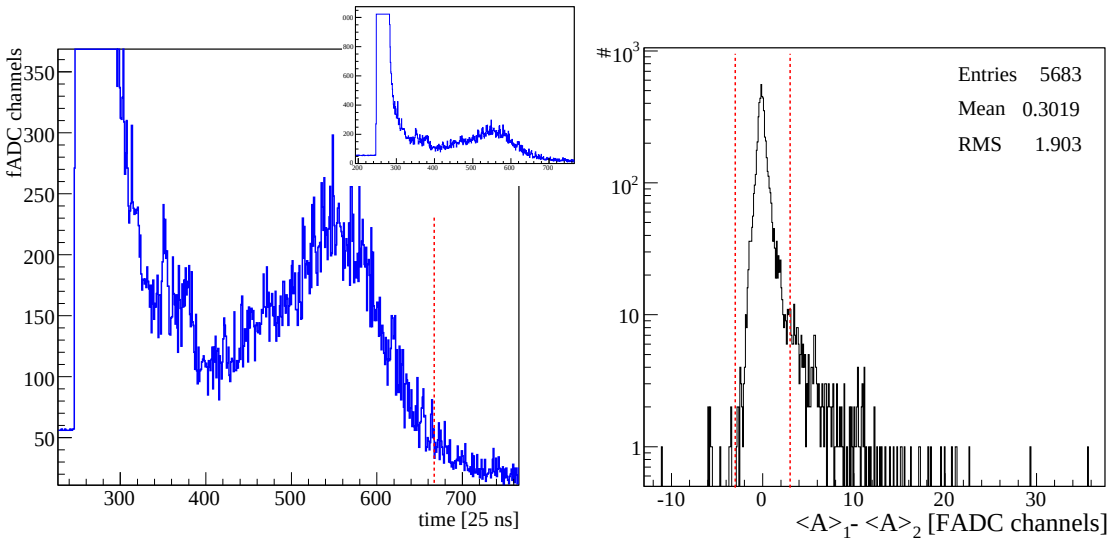
$$\sigma = 0.90 + 0.20 \cdot \exp(0.42 \cdot R)\ 25\,\text{ns} \tag{3.2.5}$$

and

$$t_0 = 1.92 + 0.49 \cdot \exp(0.38 \cdot R)\ 25\,\text{ns}. \tag{3.2.6}$$

The Moyal distribution and the function f_1 give a more stable fit (due to the smaller number of fitting parameters). They reproduce the VEM traces, with a deviation from the true value of the total signal of $-0.40\% \pm 0.3\%$ for the gamma function and $-0.15\% \pm 0.14\%$ for the Moyal distribution (Fig. 3.5(a)). Fixing the width parameter for both distributions, according to Eq. (3.2.5) and the values given in [124] respectively, the mean of residual distributions becomes 0.031 with a spread of 0.015 for f_1 and respective 0.001 with a spread of 0.005 for Moyal in relative units (Fig. 3.5(b)). The Moyal distribution has been tested inducing fake saturation of the signal and it was shown that it can reproduce satisfactory the true signal [123]. Considering the presented results the Moyal distribution is used further on.

Undershoot method The value of the undershoot is determined by the difference in the baseline of the FADC traces before and after the signal, considering the average of the first and last 100 bins of the trace. The readout chain can be simplified with a single RC circuit



(a) *a bad trace* (b) *baseline stability*

Figure 3.6: (a) An example of a trace that is not used in the saturation recovery. In the last 100 bins (after the dotted line) the baseline is still not stable. (b) Histogram of the difference between the average of the first half and the second half of the last 100 bins of the trace, used to evaluate the undershoot. Only traces between the two red dotted lines are accepted.

with the proper time constant. Therefore a linear dependence between the anode signal charge Q_a and its undershoot U_a is expected [121,122]. Since the anode undershoot amounts to very few FADC counts a recovering method based on it alone would have a poor resolution. It is then considered the usage of the dynode undershoot because of its larger amplitudes. The dependence of the dynode charge Q_d, or (better) at once the anode one Q_a, on the dynode undershoot U_d has then to be analyzed and exploited.

Very large values of the dynode undershoot might not be properly measured because of the same order of the baseline level (about 40 FADC counts depending on the PMT and the electronic channel), in these cases the anode undershoot is used.

In some cases, due to big pulses after the signal, the baseline in the last 100 bins is not stable, as exemplified in Fig. 3.6(a). This would lead to a big deviation of the undershoot value from the true one. A check on the stability of the baseline after the signal is done comparing the average of the signal in bins from 668 to 718 and from 719 to 768. If the difference between the two values is greater than 3 channels the undershoot value is not to be trusted. Therefore we reject these PMTs from this analysis.

The histogram of such difference is shown in Fig. 3.6(b). The traces that are rejected (e.g. the undershoot is not computed) comprise $\approx 5\%$ of the total number of saturated PMTs, but only to 1% of the total saturated stations.

The two different works that use the undershoot method to recover the FADC saturation have been compared. We report here the final results and the revised analysis done in fruitful collaboration with the developers of the methods.

The undershoot-based saturation recovery method presented in [121] (further referred to as

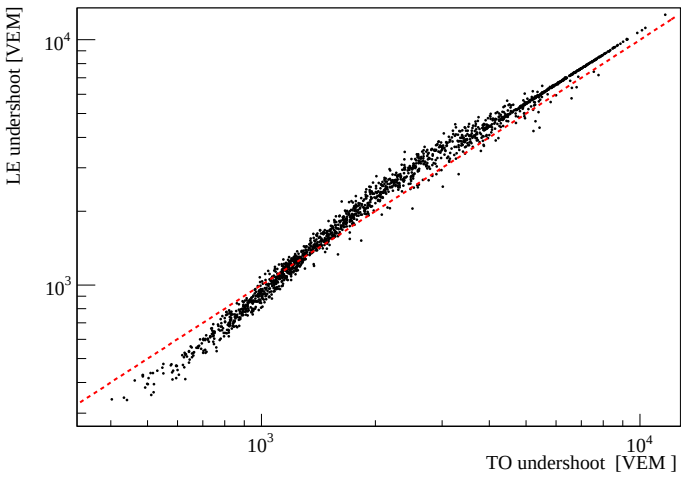


Figure 3.7: Comparison of the undershoot recovery methods.

To, from the city of residence of the developers - Torino) uses the assumption that the anode output circuit is a simple RC circuit. The time constants are deduced from the nominal values of the electric components, which have been measured during the testing of the bases [126]. The parameterization for the anode charge in this case is

$$Q_a = 770 \cdot (U_d - 1.35 \cdot 10^{-4} Q_d - 0.055 K_d) - 222 \quad \text{[adc ch. } \times \text{ 25 ns], for } U_a \leq 1 \tag{3.2.7}$$

and

$$Q_a = U_a \cdot 20000 - 2216 \quad \text{[adc ch.} \times \text{ 25 ns], for } U_a > 1. \tag{3.2.8}$$

where K_d is the number of saturated bins on the dynode trace.

In [122] it has been decided to adopt a heuristic approach, namely to get this dependence from the data sample itself. Since for very large signals, i.e. for a small fraction of traces, the dynode undershoot might be inaccurately measured, the anode undershoot has been used in these cases. The FADC saturation recovered anode charge is given by

$$Q_a = \alpha \cdot U_d^{\beta} \quad \text{[adc ch. } \times \text{ 25 ns]} \tag{3.2.9}$$

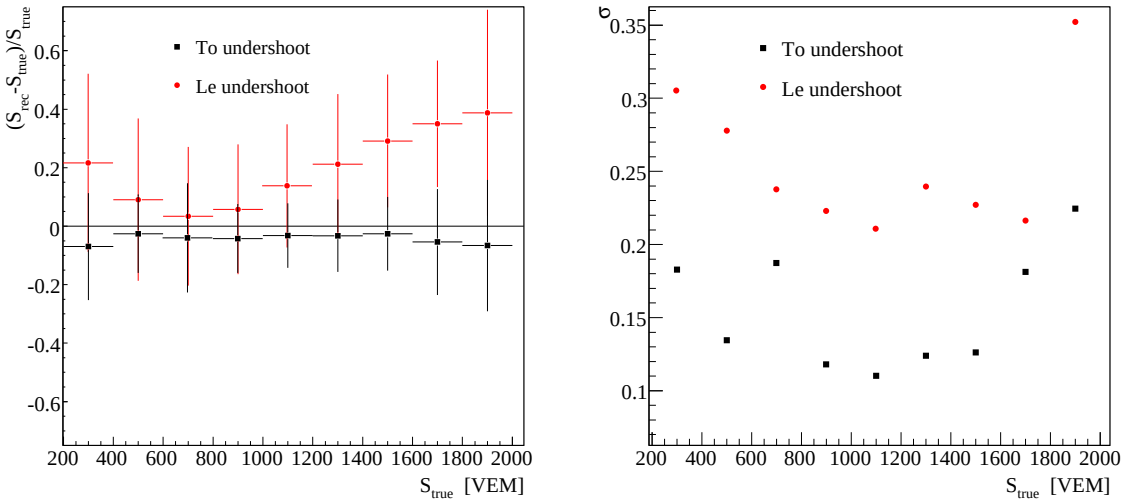
for $U_d \leq 30$, with $\alpha = 144 \pm 40$ and $\beta = 1.53 \pm 0.04$, and

$$Q_a = a \cdot U_a + b \quad \text{[adc ch. } \times \text{ 25 ns]} \tag{3.2.10}$$

for $U_d > 30$, with $a = (21 \pm 1) \cdot 10^3$ and $b = -(1.3 \pm 0.7) \cdot 10^3$. Because the developers of this method are coming from Lecce, it will be further referred as Le.

In Fig. 3.7 a comparison of the signal recovered with the two undershoot methods is shown. The difference in the recovered signal is ≈10% for big signals.

To test the two methods they have been applied to not saturated signals. The mean value of the relative difference between the recovered signal and the *true* (not saturated) one is always



(a) *relative difference* (b) *spread*

Figure 3.8: The uncertainties from the undershoot methods. (a) The mean value of the relative difference between the recovered signal and the *true* one for not saturated traces, for the two undershoot methods. The vertical error bars denote the RMS dispersion around the mean. (b) sigma of distributions of $(S_{\rm rec} - S_{\rm true})/S_{\rm true}$ as a function of $S_{\rm true}$

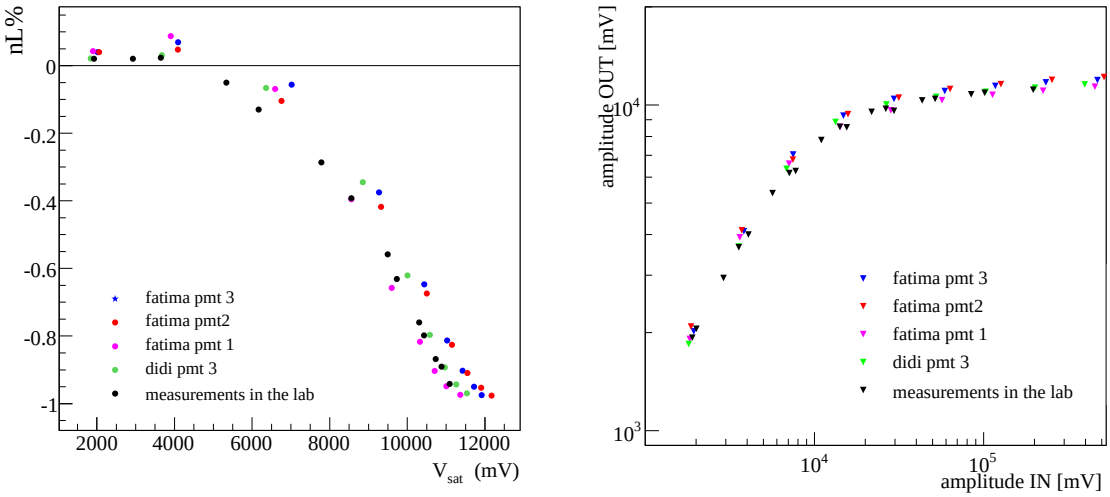
less than 10% for To, and about 25% for Le as represented in Fig. 3.8(a). The dispersions around the mean value are shown in Fig. 3.8(b). The deviations from null are within the statistical uncertainties for both methods.

In the low signal region, the fluctuations of the two algorithms around the right values are almost anti-correlated. For this reason we decided to use as default in our recovery procedure the average between the Le and To values and to use the half-difference as an estimation of the related systematic uncertainty.

PMT non-linearity The PMT saturation has been studied in the past mostly in the *small saturation regime*, i.e. the region around 50 mA. Such measurements have now been extended to the *high saturation regime*. This has required the use of different light-emitting diodes (LEDs) and of a different driving system as the present LEDs operating on the tanks in the field do not reach such levels. The new system has been operated in laboratory and in some tanks in the field. The signal can be injected in the last case through the *emergency* LED opening. Non linearity, nL, is measured with the two LED technique as:

$$\mathrm{nL} = \frac{C - (A + B)}{A + B} \tag{3.2.11}$$

where C denotes the light given by two LEDs A and B switched on together. Since after the first few measurements the pulses of the individual LEDs (A and B) are in the non linear range, the measurement has be done choosing A and B near the sum of the two at the previous step; in such way it is possible for each step to correct the value of A and B for the



(a) *PMT non-linearity* (b) *saturation curve.*

Figure 3.9: (a) Non linearity, Eq. (3.2.11), curve measured in the lab and in the field. (b) Relation between the input (non saturated amplitudes) and the output (saturated amplitudes).

non linearity measured in the step before. The saturation measured in laboratory and in the field for different PMTs is shown in Fig. 3.9(a).

It has been found that the PMT non-linearity depends only on the amplitude of the signal, and not on the total charge. This implies that the saturated signals can be studied through the saturations of the individual amplitudes. The actual curves used in the following, relating the input amplitude (A_{in} - non-saturated amplitudes) and the output (A_{out} - saturated amplitudes) are shown in Fig. 3.9(b).

Theoretical model The model for the recovery of a saturated signal has as input the assumptions that the pulse shape is described by the Moyal function, the PMT behavior on the nonlinear range has been measured and is the same for all PMTs and that the missing signal due to the FADCs overflow is recovered by the dynode or anode undershoot with a good accuracy.

The signal distribution in time is well described by the Moyal function,

$$f_M(t) = A \cdot e^{-1/2\left(\frac{t-t_0}{\sigma}+e^{\frac{-t+t_0}{\sigma}}\right)} = A \cdot f(t). \tag{3.2.12}$$

where t is time, t_0 is time bin of the maximum position, and σ represents the spread of the signal and A is the maximum amplitude of the signal. The trace, A_{true}, is attenuated due to the PMT non-linearity, described by a function g, depending only on the amplitude and leading to the measured signal distribution

$$A_{\text{m}}(t) = g(A_{\text{true}}) = g(f_M(t)) = A \cdot g(f(t)). \tag{3.2.13}$$

The goal is to obtain the true signal distribution, knowing the measured one. To be able to obtain an analytical solution g was approximated with a set of linear functions (Fig. 3.9(b)), defined in six intervals

$$g_i(A_{\text{true}}) = \alpha_i \cdot A_{\text{true}} + \beta_i, \text{ for } A_{\text{true}} \in [A^i_{\text{in}}, A^i_{\text{fi}}) \text{ and } i \in 1,..6 \tag{3.2.14}$$

where g_i is the attenuation curve function, A_{true} is the input amplitude, α_i and β_i are the free parameters, and $[A^i_{\text{in}}, A^i_{\text{fi}})$ is the interval where g_i is valid. The amplitude that would have been measured without FADC saturation, A_{m} is not known, but it can be obtained from the undershoot method. The total signal that would have been measured, I_u is equal to the total calibrated charge,

$$I_u = \int A_{\text{m}}(t)dt = A\sum_{i=0}^{n}\int_{t^i_{\text{in}}}^{t^i_{\text{fi}}} g_i(f(t))dt. \tag{3.2.15}$$

Inserting Eq. (3.2.14) it becomes

$$I_u = A\sum_{i=0}^{n}\int_{t^i_{\text{in}}}^{t^i_{\text{fi}}} (\alpha_i f(t) + \beta_i)dt \tag{3.2.16}$$

where $[t^i_{\text{in}}, t^i_{\text{fi}})$ is the time interval corresponding to the values of the amplitudes in $[A^i_{\text{in}}, A^i_{\text{fi}})$. Separating the integral parts

$$I_u = A\sum_{i=0}^{n}\int_{t^i_{\text{in}}}^{t^i_{\text{fi}}} \alpha_i f(t)dt + \sum_{i=0}^{n}\int_{t^i_{\text{in}}}^{t^i_{\text{fi}}} \beta_i dt \tag{3.2.17}$$

the amplitude parameter of the Moyal function can be extracted

$$A = \frac{I_u + \sum_{i=0}^{n}(t^i_{\text{in}} - t^i_{\text{fi}})\beta_i}{\sum_{i=0}^{n}\int_{t^i_{\text{in}}}^{t^i_{\text{fi}}} \alpha_i f(t)dt} \tag{3.2.18}$$

The integral of the denominator can be computed and gives

$$I_i = \int_{t^i_{\text{in}}}^{t^i_{\text{fi}}} f(t)dt = \sigma\cdot\sqrt{2\pi}\left[\text{Erf}\left(\frac{1}{\sqrt{2}}\exp\left(-\frac{1}{2}\frac{t^i_{\text{in}} - t_0}{\sigma}\right)\right) - \text{Erf}\left(\frac{1}{\sqrt{2}}\exp\left(-\frac{1}{2}\frac{t^i_{\text{fi}} - t_0}{\sigma}\right)\right)\right] \tag{3.2.19}$$

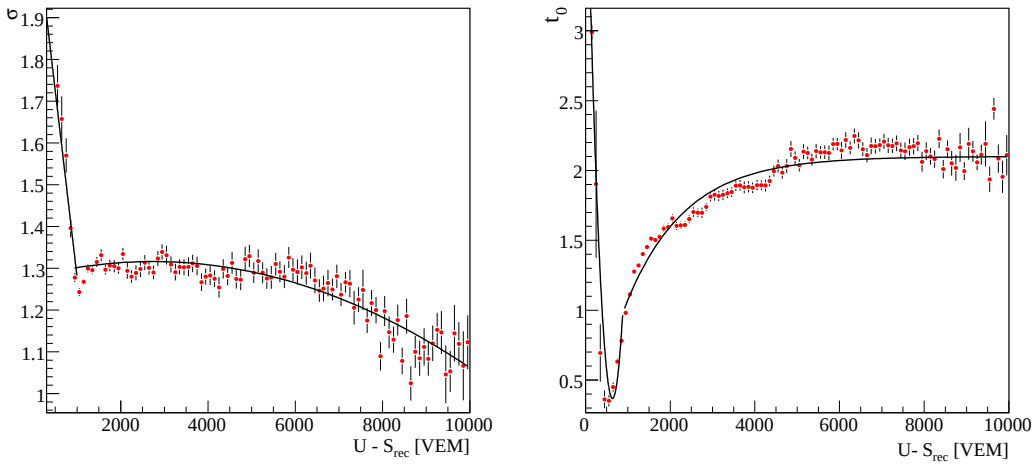
The amplitude parameter is then

$$A = \frac{I_u + \sum_{i=0}^{n}(t^i_{\text{in}} - t^i_{\text{fi}})\beta_i}{\sum_{i=0}^{n}\alpha_i \cdot I_i} \tag{3.2.20}$$

and it is completely determined by the attenuation curve (α_i, β_i), the signal determined by the undershoot value (I_u), and the integration intervals (t_{in}, t_{fi}).

Unfortunately the integration intervals cannot be analytically determined, as the equations

$$f_M(t_{\text{in/fi}}) = A_{\text{in/fi}} \tag{3.2.21}$$



(a) σ (b) t_0

Figure 3.10: The parameterization (Eq. (3.2.23) and (3.2.24)) used for the initial values of σ (a) and t_0 in units of 25 ns. (b). Red dots denote the mean values of the distributions of the parameters.

are transcendental, thus the solution is obtained numerically.

At the end the function that is used to describe the signal shape is

$$f_M(t) = \frac{I_u + \sum_{i=0}^{n}(t^i_{\text{in}} - t^i_{\text{fi}})\beta_i}{\sum_{i=0}^{n}\alpha_i \cdot I_i} \cdot \exp\left(-1/2\left(\frac{t-t_0}{\sigma} + \exp\left(\frac{-t+t_0}{\sigma}\right)\right)\right) \tag{3.2.22}$$

with 2 free parameters t_0 and σ. One should remark that the fitting function should be $g(f_M)$, but in the time interval where the bins are not saturated (e.g. the fitting intervals) $g(f_M) = f_M$, the PMT is in the linear response range.

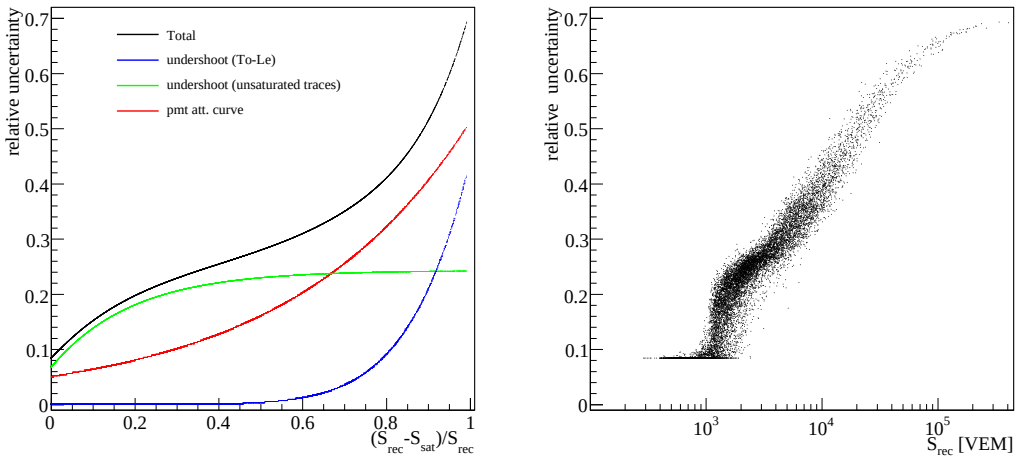
Parameterization for t_0 and σ The parameterization presented in [123], and briefly described earlier, dependends on the shower-plane distance to the core of the station. In order to perform the saturation recovery independently from the shower reconstruction the initial values have been parameterized as a function of the undershoot-recovered signal (Fig. 3.10) .

$$t_0(25\,\text{ns}) = \begin{cases} 4.645 - 0.0134 \cdot I_u + 1.05 \cdot I_u^2 & \text{for } I_u < 878\,\text{VEM} \\ 2.1 \cdot (1 - e^{0.00072 \cdot I_u}) + I_u^{-3.8} & \text{for } I_u \geq 878\,\text{VEM} \end{cases} \tag{3.2.23}$$

and

$$\sigma(25\,\text{ns}) = \begin{cases} 2.17 - 0.00089 \cdot I_u & \text{for } I_u < 974.5\,\text{VEM} \\ 1.28 + 2.63 \cdot 10^{-5} I_u - 4.808 \cdot 10^{-9} \cdot I_u^2 & \text{for } I_u \geq 974.5\,\text{VEM} \end{cases} \tag{3.2.24}$$

For for the first PMT of the saturated tank the values are set according to the parameterization, with a limit of $0.5 \cdot 25$ ns for σ and $3 \cdot 25$ ns for the position of the maximum. If the fit was successful, for the next 2 PMTs the fitting parameters are initialized with the values from the previous fit. The parameterization is shown in Fig. 3.10. Given t_0 and σ and the



(a) *uncertainty versus fraction of recovered signal* (b) *relative uncertainty vs recovered signal*

Figure 3.11: The total uncertainty on the recovered signal.

undershoot-recovered charge, the amplitude is completely determined. Since the integration intervals depend on an initial amplitude an iterative procedure for the amplitude computation was implemented. As a result of all the constraints the convergence of the fit is achieved even for a low number of degrees of freedom.

The recovered signal becomes

$$S = \int f_M(t)dt. \tag{3.2.25}$$

Uncertainties The total uncertainty on the recovered signal can be separated into 3 terms: The uncertainty coming from the fitting procedure itself σ_{fit}, a second term related to the PMT non-linear response, σ_{att}, and one for the uncertainties given by undershoot recovery, σ_{under} methods when they extrapolate to large charges. We have also to consider the intrinsic accuracy of a Auger tank signal, which is nearly Poissonian

$$\sigma^2_{\text{tot}} = \sigma^2_{\text{fit}} + \sigma^2_{\text{att}} + \sigma^2_{\text{under}} + \sigma^2_{\text{intrinsic}}. \tag{3.2.26}$$

The first term has been evaluated applying the recovery procedure to non saturated traces cut at different levels in order to simulate the saturation. Traces have been cut at the level $\epsilon_i A$, where A is the peak of the trace and $\epsilon_i = 0.9, 0.8, ...0.5$. Only traces with maximum higher than 600 FADC counts and more than 4 fitted bins are considered.

The recovery procedure has been performed using both the undershoot methods, and for each trace the relative difference between the real signal (S_{true}) and the reconstructed one (S_{rec}) is computed. The dispersion of the distribution of such differences increases with the fraction

of recovered signal. It can be parameterized as:

$$\sigma_{\rm fit} \equiv (S_{\rm rec} - S_{\rm true})/S_{\rm true} = a \cdot (1 - e^{-b \cdot x}) + c, \tag{3.2.27}$$

where $x = (S_{\rm true} - S_{\rm cut})/S_{\rm true}$. The parameterization gives a= 0.17, b=5.21, c=0.068 for the To undershoot and a=0.18, b=9.04, c= 0.117 for the Le undershoot.

To evaluate the uncertainties due to the PMT non-linearity recovery, for each trace the fitting procedure has been performed twice, using two different PMT non-linearity response curves, corresponding to the extreme measurements available.

The relative error is computed as $\sigma/\sqrt{2}$ of the distribution of $(S_{\rm att1} - S_{\rm att2})/\langle S \rangle$ as a function of $x = (S_{\rm rec} - S_{\rm sat})/S_{\rm rec}$. It is well decribed by an exponential function $\sigma_{\rm att} = a_1 \cdot x^{b_1}$. The obtained parameters with $a_1 = 0.48$, $b_1 = 7.65$ for the To method and $a_1 = 0.44$, $b_1 = 7.12$ for the Le method agree very well. An increase with recovered signal up to 35% is observed as expected, because the attenuation curve has a big weight in the signal recovery for high signals.

Comparing the recovered signal using the two undershoot methods, we can estimate the uncertainty on the recovered signal due to the systematic uncertainty that affects the undershoot method. The relative difference in the recovered signals ($(S_{ULe} - S_{UTo})/S_{average}$) as a function of the same variable x defined before is described by

$$\sigma_{\rm under} = 0.05 \cdot e^{x \cdot 2.325}. \tag{3.2.28}$$

The different terms that contribute to the total uncertainty and their quadratic sum are shown in Fig 3.11(a). The total uncertainty is also shown as a function of the recovered signal in Fig. 3.11(b); it varies from 10% for small signal up to 60% when the recoverd signal is greater than 10^5 VEM.

The recovery method has been implemented in the SaturationRecoveryKLT module of the Offline software. A flag for using the average undershoot value, the Torino value or the Lecce one was implemented.

Validation of the method A good saturation recovery method should give as a result a distribution of the signal from individual stations with a power law behavior. An approximative solution for this is given below. Let us start with the signal spectrum

$$\frac{dN}{dS} = \int {\rm d}E \int {\rm d}\cos(\theta) \int {\rm d}t \int {\rm d}A \int 2\pi r {\rm d}r \frac{{\rm d}N}{{\rm d}E\ {\rm d}\Omega\ {\rm d}t\ {\rm d}A} \frac{{\rm d}P}{{\rm d}S} \rho_{\rm st}, \tag{3.2.29}$$

where E is the energy, θ is the zenith angle, t is the time, r is the radial distance to the core in shower plane coordinates, Ω is the solid angle, A is the effective area and ρ_{st} is the station distribution (includes the detector geometry).

We assume a trivial lateral distribution function

$$S(r) = a \cdot S_{1000} \cdot e^{b \cdot r} \tag{3.2.30}$$

to make the following calculations simple, but other LDF assumptions should give qualitatively similar results.

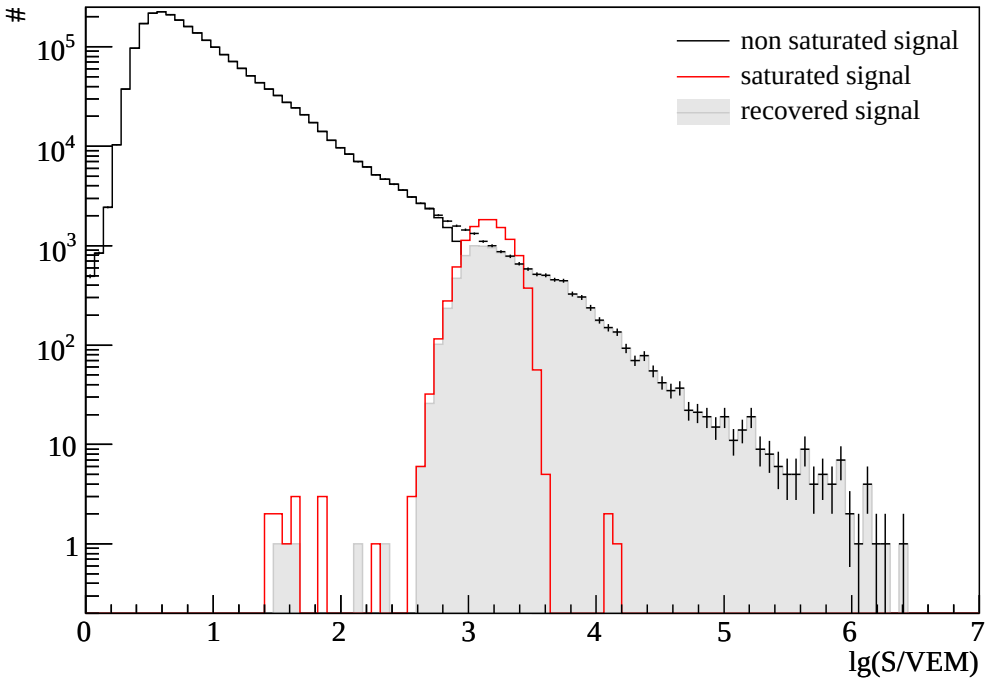


Figure 3.12: Signal spectrum. The black histogram denotes the signal from all the events with signal recovery. The recovered signal is shown as a grey histogram. The red histogram shows the signals from the saturated events. It is interesting to note that almost all the signals above 1000 VEM are saturated.

The energy is given by

$$E(S_{1000}, \theta) = A \cdot \left(\frac{S_{1000}}{f(\theta)} \right)^B \tag{3.2.31}$$

where $f(\theta)$ is the correction factor applied to S(1000 m) for obtaining a zenith angle independent variable. The energy probability distribution is

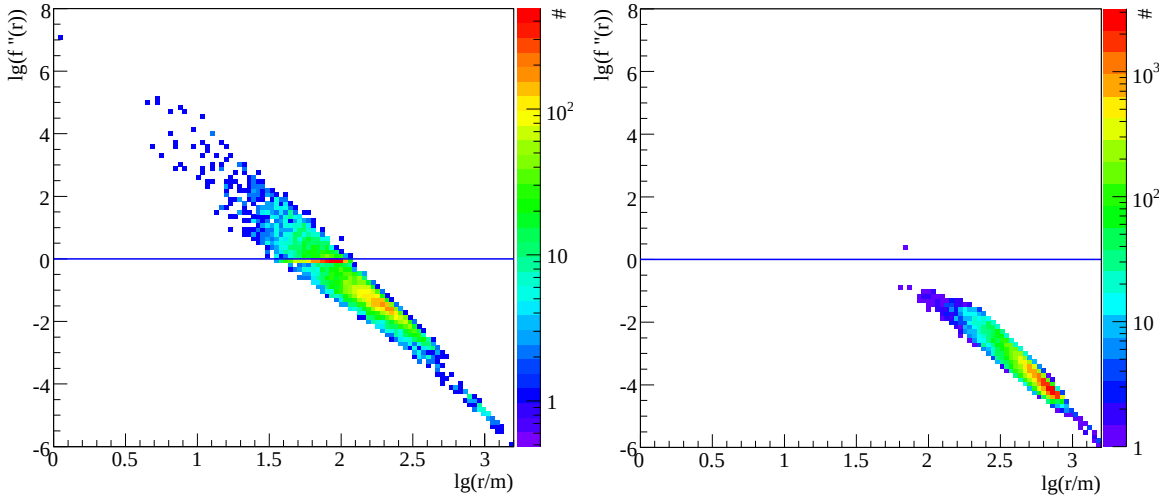
$$\frac{\mathrm{d}P}{\mathrm{d}E} = d \cdot E^{-\alpha}. \tag{3.2.32}$$

The computation was simplified assuming no trigger threshold, thus we do not model the signal distribution at small VEM values. Also the zenith angle influence on the shape of the signal spectrum was considered to be negligible in first order. With these assumptions the signal probability distribution becomes:

$$\frac{\mathrm{d}P(S)}{\mathrm{d}S} = \frac{d}{b \cdot (1-\alpha)} \cdot S^{B(1-\alpha)-1}. \tag{3.2.33}$$

The ratio $^{dN}/_{dS}$, depends directly on S

$$\frac{\mathrm{d}N}{\mathrm{d}S} \propto S^{c(1-\alpha)-1}, \tag{3.2.34}$$



(a) *saturated stations* (b) *unsaturated stations*

Figure 3.13: LDF second derivative as a function of the distance to the core in shower plane coordinates. Stations above the blue line are used as lower limits in the maximum likelihood fit. The applied cut does not influence the non-saturated events.

therefore the signal spectrum should obey a power law. If the deviation is considerable, the method in use is strongly biased.

In Fig. 3.12 we show the signal spectrum for the default signal recovery. The obtained value for a fit in the region $\lg(S_{\rm rec}/{\rm VEM}) \in [2,3.5]$ gives a slope of -1.88 ± 0.01. A small underestimation of the recovered signal is observed in the region of $\lg(S_{\rm rec}/{\rm VEM}) \in [3.5, 4]$, but this is within the given uncertainties.

The dynamic range is extended from $\approx 10^3$ VEM to 10^6 VEM obeying the expected power law.

Impact on the air-shower reconstruction The reconstruction of the LDF of the particles on ground in the $\overline{\rm Off}\underline{\rm line}$ framework is based on a maximum likelihood method. It considered previously the saturated signals as lower limits in the fit. The LDF is not well known in the region close to the core and it is even divergent. This would lead to a signal of more than 10^{10} VEM in a distance to the core of less than 50 m. Almost all the signals at this distance are saturated and a good parameterization of the LDF is not feasible.

In such cases the recovered signal should be used as a lower limit until better knowledge of the LDF is achieved. The best criteria for using recovered signals as lower limits turned out to be a cut on the second derivative of the normalized LDF at 1000 m with respect to the distance to the core in shower plane. In Fig. 3.13(a) the second derivative as a function of the shower plane distance for all the saturated stations is illustrated. All signals from the stations above the blue line, $f(r) = 1$, are used as a lower limit in the reconstruction. This cut does not affect the non-saturated events as shown in Fig. 3.13(b).

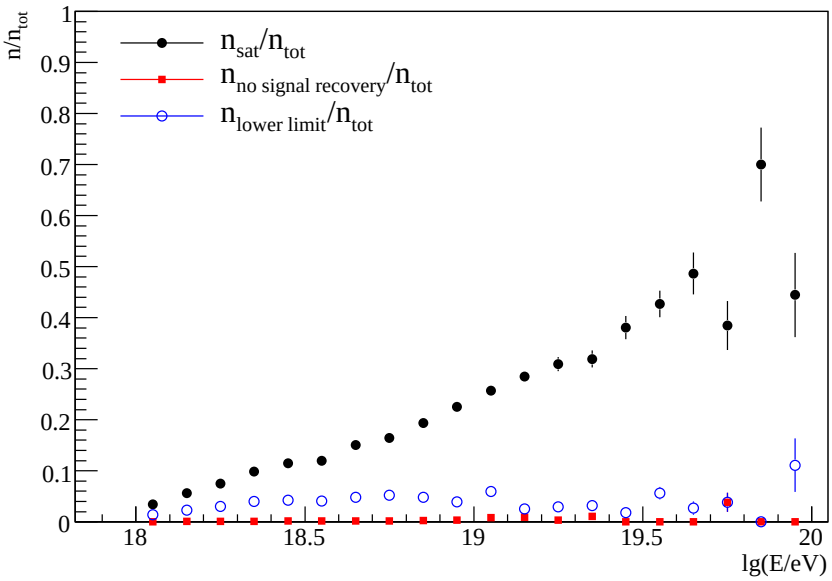
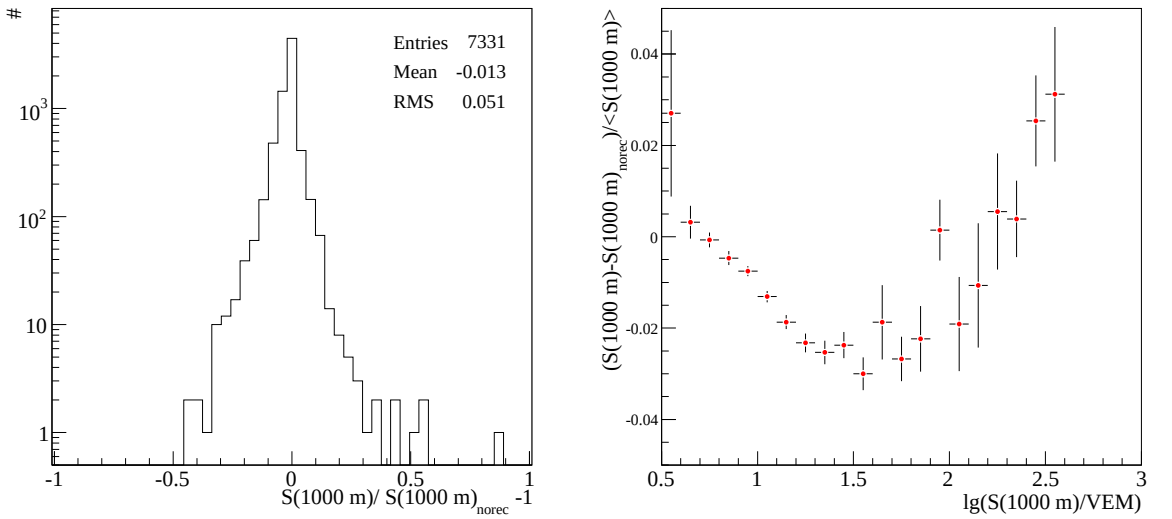


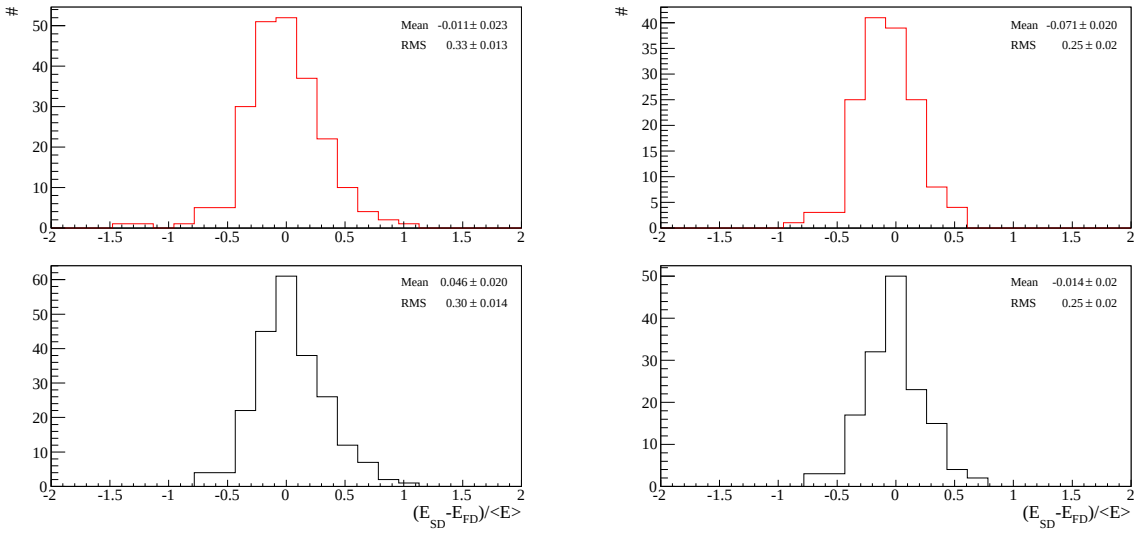
Figure 3.14: The number of saturated events relative to the total number of events (black dots). At high energies about half of the total number of events have at least a saturated station. The blue open circles denote the number of events which do not fulfill the LDF criteria cut, relative to the total number of events. The fraction of events with failed saturation recovery is shown with red squares.



(a) *S(1000 m) relative difference*

(b) *S(1000 m) relative difference vs S(1000 m)*

Figure 3.15: Impact of the signal recovery on the reconstructed S(1000 m).



(a) *all energies* (b) *above E= 2 EeV*

Figure 3.16: Validation of the recovery method with golden hybrid data. Energy resolutions are shown for the case when the saturated signal is used as a lower limit (upper plots) and when signal recovery is used (lower plots).

The number of events where the recovered signal is used as lower limit is a fraction of $\approx 20\%$ of the total number of saturated events. In Fig. 3.14 the relative number of saturated events to the total number of events is illustrated. At 3 EeV 10% of our data are saturated, raising to 60% at 30 EeV. The signal is not recovered at all due to the variation of the baseline in the FADC traces, described in the previous section, for less than 1% of the saturated events.

The obtained change in S(1000 m), of about 2% is shown in Fig. 3.15(a). In the case of low multiplicity events S(1000 m) is underestimated when there is no saturation recovery performed. At high S(1000 m) (Fig. 3.15(b)) the difference becomes negative of about -2%.

Another test that was done using the golden hybrid data set recorded until August 2007. The events were not required to have the vertical aerosol optical depth (VAOD) parameters measured by the CLF. In the whole energy calibration set there are 222 saturated events with a decent FD reconstruction, 149 above 2 EeV.

Using the energy calibration (see Section 5) from Eq. (5.2.1) the mean of the distribution of the $E_{\rm FD} - E_{\rm SD}$ over the average energy is changed from a negative value of $-1.2(\pm 2.2)\%$ to $4.6(\pm 2.0)\%$, as illustrated in Fig. 3.16(a). A positive mean is expected as in the low energy region only the air showers with upper fluctuations are triggering the surface detector. The RMS is improved from 0.33 ± 0.013 to 0.30 ± 0.014. Taking only events with energies greater than 2 EeV, to reduce the influence of the upward fluctuations, an improvement in the mean is observed from $-7.1(\pm 2)\%$ to $-1.4(\pm 2)\%$.

3.3 FD event reconstruction

One of the aims of the FD reconstruction is to estimate the energy needed for the energy calibration of the surface detector. First the signal pulses from individual PMTs are determined, then the geometry of the shower is deduced with the help of the timing from a SD station. Finally the energy is obtained from the total energy deposit in the atmosphere [127].

Pulse finder and geometry reconstruction The data taking software of the fluorescence detectors writes out the ADC-counts for a period of 100 μs. In order to obtain the pulse related to the detection of fluorescence light from an air shower with this trace, a signal over noise (S/N) maximization algorithm is used. Starting from the first triggered time bin, the boundaries of the suspected pulse are shifted to find the maximal S/N-ratio.

$$S/N = \frac{S(\Delta t)}{\sqrt{\Delta t} \times RMS}. \tag{3.3.1}$$

$S(\Delta t)$ is the total signal within the tested pulse length and RMS is the fluctuation of the trace baseline. In order to reject accidental noise from pixels triggered by background light a minimum requirement on this ratio is set and usually only pixels with pulses having a S/N greater than 5 are kept for the subsequent analysis. In the reconstruction the variables used are the pulse time, t_i^{meas}, for each pixel i, defined as the centroid of the found pulse, and the total integrated signal, w_i. To determine all pixels connected to the shower image and further suppress noise pixels a pattern recognition algorithm is used. The geometrical pointing direction, $\vec{p}_i$, together with the signals determine the plane containing both the air shower and the track on the camera called shower detector plane (SDP)

$$Q^2 = \sum_i w_i [\vec{p}_i \cdot \vec{n}] \tag{3.3.2}$$

where $\vec{n}$ denotes the searched vector normal to the SDP. The Q^2 should be 0 in the ideal case. An illustration of the SDP is given in Fig. 3.17.

The next step is to determine the position of the shower within the SDP: The impact parameter R_p (i.e. the position of the shower core, see Fig. 3.17), the corresponding time T_0 and the angle between the shower axis and the ground plane χ_0. The pointing directions of the pixels can be translated into an elevation angle $\tilde{\chi}_i$ and minimization of the following χ^2 leads to the determination of the position of the shower axis inside the SDP (for details see [128])

$$\chi^2 = \sum_i \left(\frac{t_i - t_i^{\mathrm{meas}}}{\sigma_{t_i}} \right) \tag{3.3.3}$$

where t_i is the theoretical expectation of the pulse time

$$t_i = T_0 + \frac{R_p}{c} \cdot \tan\left(\frac{1}{2} \cdot (\chi_0 - \tilde{\chi}_i) \right). \tag{3.3.4}$$

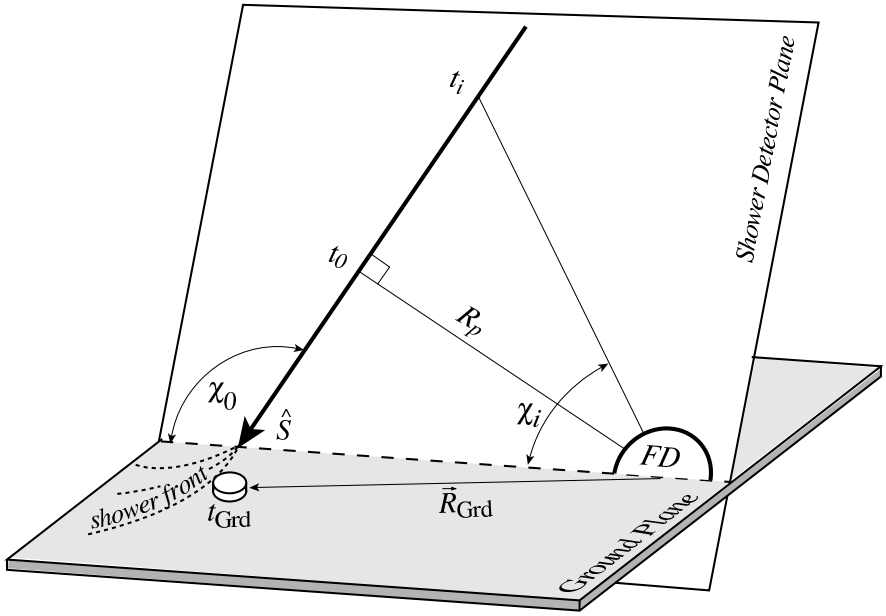


Figure 3.17: Illustration of the hybrid shower detector plane.

given by geometrical considerations. The correlation between the variables is large, a slight deviation in the elevation angle χ_0 induces a big change of the other two parameters. This degeneracy can be broken by adding timing information from a single SD station. Assuming a plane shower front the expected trigger time of the station is given by:

$$t_i^{\mathrm{meas}} = T_0 + \frac{1}{c}\vec{R} \cdot \vec{n} \tag{3.3.5}$$

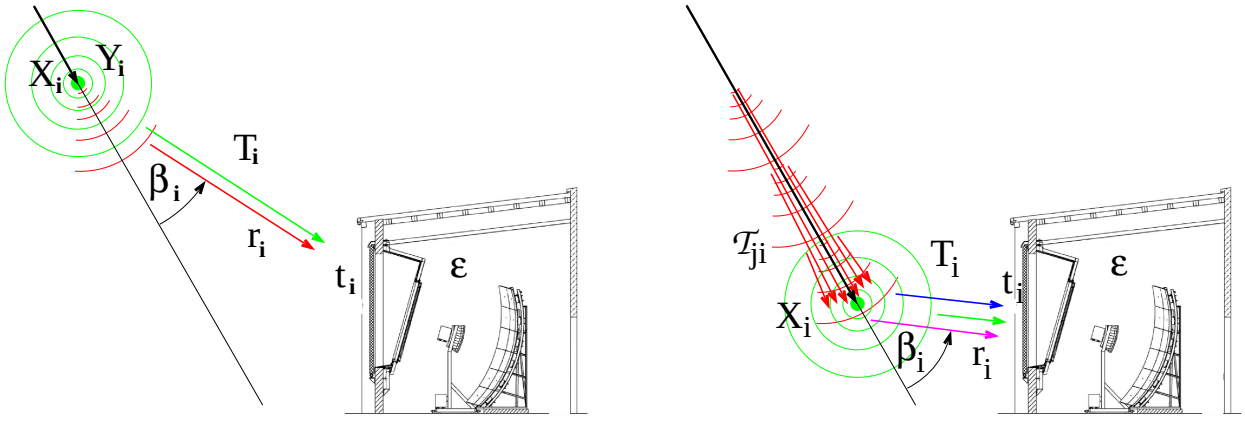
where $\vec{R}$ is the position of the station with respect to the eye and $\vec{n}$ is the shower direction. Of all the stations present in the event, the station is chosen from those that are within 2 km from the intersection of the axis and the ground. In most cases the station with the highest signal satisfies these requirements.

The analysis of laser shots [129] suggests that the directional uncertainties of the SDP reconstruction are of an order of 0.1 degree.

Energy reconstruction

The two major contributions to the light at the FD aperture are the fluorescence light from nitrogen molecules and Cherenkov radiation photons [130–134]. Both contributions are affected by scattering and absorption in the atmosphere. The important scattering processes are Rayleigh scattering, when photons are scattered by particles much smaller than their wavelength, and Mie scattering, from particles larger than about tenfold of the light wavelength.

The amount of fluorescence light is directly proportional to the energy deposited by the air shower along its path in the atmosphere. Given the fluorescence yield Y_i^{f} [99–101] at a point



(a) *direct light contribution* (b) *scattered light contribution*

Figure 3.18: Illustration of the light flux received at the FD. Green: Isotropic fluorescence light, red: forward direct Cherenkov light, blue: Rayleigh-scattered Cherenkov light, and magenta: Mie-scattered Cherenkov light. [127]

in the atmosphere, the number of photons produced by the shower is

$$N_\gamma^{\mathrm{f}}(X_i) = Y_i^{\mathrm{f}}\,\mathrm{d}E/\mathrm{d}X_i\,, \tag{3.3.6}$$

where $\mathrm{d}E/\mathrm{d}X_i$ denotes the energy deposited at slant depth X_i.

Due to Rayleigh and Mie attenuation only a fraction of the photons, T_i, can be detected at the aperture, as sketched in Fig. 3.18. The direct fluorescence light emitted at this slant depth is measured at the detector at time t_i. Given the light detection efficiency of ε and aperture A, the fluorescence light flux y_i^{f} measured at the FD is

$$y_i^{\mathrm{f}} = \frac{A\,\varepsilon\,T_i}{4\,\pi\,r_i^2}\,Y_i^{\mathrm{f}}\,\mathrm{d}E/\mathrm{d}X_i\,, \tag{3.3.7}$$

The number of photons emitted through Cherenkov radiation is proportional to the number of charged particles above a certain energy cutoff, which is in a good approximation just the number of electrons and positrons,

$$N_\gamma^{\mathrm{C}}(X_i) = Y_i^{\mathrm{C}}\,N_{\mathrm{e}}(X_i). \tag{3.3.8}$$

The threshold energy with the height is included in the Cherenkov yield Y_i^{C} [135–138]. Given the fraction $f_{\mathrm{C}}(\beta_i)$ of photons emitted at an angle β_i with respect to the shower axis, the light flux at the FD aperture originated from direct Cherenkov light is

$$y_i^{\mathrm{Cd}} = \frac{A\,\varepsilon\,T_i}{4\,\pi\,r_i^2}\,f_{\mathrm{C}}(\beta_i)\,Y_i^{\mathrm{C}}\,N_{\mathrm{e}}(X_i). \tag{3.3.9}$$

Although the Cherenkov photons are emitted in a narrow cone along the particle direction, they cover a considerable angular range with respect to the shower axis, because the charged particles are deviated from the primary particle direction due to multiple scattering. Due to the forward peaked nature of the Cherenkov light production, an intense Cherenkov light

beam can build up along the shower as it traverses the atmosphere (cf. Fig. 3.18(b)). If a fraction $f_{\rm s}(\beta_i)$ of the beam is scattered towards the detector it can contribute significantly to the total light received. In a simple one-dimensional model the number of photons in the beam at depth X_i is just the sum of Cherenkov light produced at all previous depths X_j attenuated on the way from X_j to X_i by $\mathcal{T}_{ji}$:

$$N_\gamma^{\rm beam}(X_i) = \sum_{j=0}^{i} \mathcal{T}_{ji}\, Y_j^{\rm C}\, N_{\rm e}(X_j). \tag{3.3.10}$$

The total light received at the detector at the time t_i is obtained by adding the scattered and direct light contributions:

$$y_i = y_i^{\rm Cs} + y_i^{\rm Cd} + y_i^{\rm f}. \tag{3.3.11}$$

To obtain the shower energy from the light at the aperture the energy deposited in the atmosphere has to be determined. The total energy deposit is just the sum of the energy loss of electrons in the atmosphere, $\mathrm{d}E/\mathrm{d}X_i$, which is related to the number of electrons $N_{\rm e}(X_i)$ via

$$\mathrm{d}E/\mathrm{d}X_i = N_{\rm e}(X_i) \int_0^\infty f_{\rm e}(E, X_i)\, \mathrm{d}E/\mathrm{d}X_{\rm e}(E, X_i)\, \mathrm{d}E, \tag{3.3.12}$$

where $f_{\rm e}(E, X_i)$ denotes the normalized electron energy distribution and $\mathrm{d}E/\mathrm{d}X_{\rm e}(E, X_i)$ is the energy loss of a single electron with energy E. The electron energy spectrum $f_{\rm e}(E, X_i)$ is universal in shower age [135, 137, 138], i.e. it does not depend on the primary mass or energy and since the electron energy loss depends only weakly on the local density, Eq. (3.3.12) is simplified to

$$\mathrm{d}E/\mathrm{d}X_i = N_{\rm e}(X_i)\, \alpha_i. \tag{3.3.13}$$

Here α_i is the average energy deposit per electron at shower age $s_i = 3/(1+2X_{\rm max}/X_i)$, where $X_{\rm max}$ denotes the shower maximum. It is parameterized from simulations as given in [138].

In general the FD will not be able to observe the full profile because of its limited field of view. Since for the calculation of the Cherenkov beam and the shower energy the full profile is required, the extrapolation to depths outside the field of view is done with a Gaisser-Hillas function [139] which gives a good description of existing fluorescence data [140]

$$f_{\rm GH}(X) = \mathrm{d}E/\mathrm{d}X_{\rm max} \cdot \left(\frac{X - X_0}{X_{\rm max} - X_0}\right)^{(X_{\rm max}-X_0)/\lambda} e^{(X_{\rm max}-X)/\lambda}. \tag{3.3.14}$$

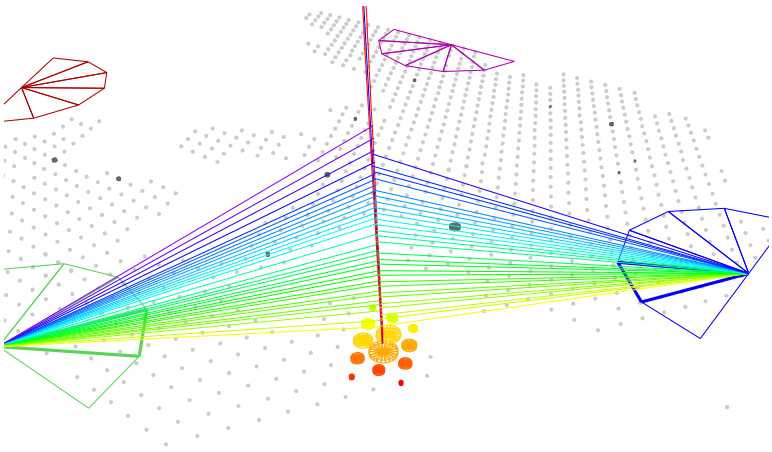
where $X_{\rm max}$ is the depth where the shower reaches its maximum energy deposit $\mathrm{d}E/\mathrm{d}X_{\rm max}$.

The electromagnetic energy is given by the integral over the energy deposit profile

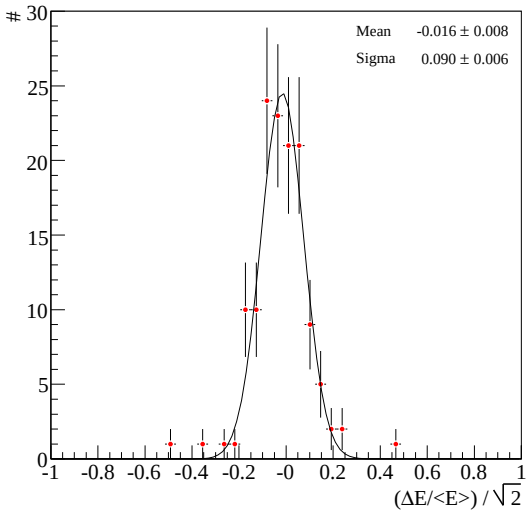
$$E_{\rm em} = \int_0^\infty f_{\rm GH}(X)\, \mathrm{d}X\ . \tag{3.3.15}$$

Not all of the energy of a primary cosmic ray particle ends up in the electromagnetic part of an air shower. Neutrinos escape undetected and muons need long path lengths to release their energy. This is usually accounted for by multiplying the electromagnetic energy (3.3.15) with a correction factor $f_{\rm inv}$ determined from shower simulations to obtain the total primary energy

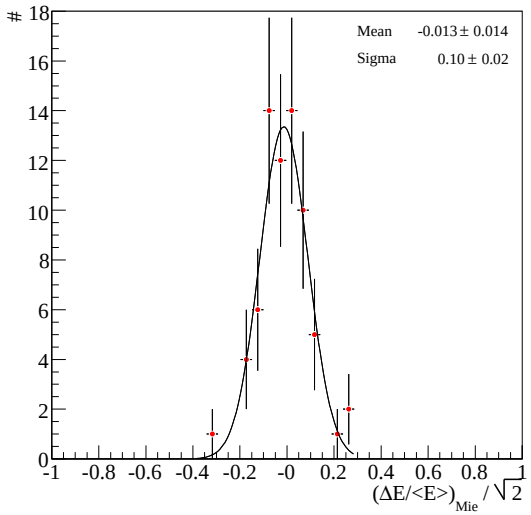
$$E_{\rm tot} = f_{\rm inv}\, E_{\rm em}. \tag{3.3.16}$$



(a) *Example of a golden stereo hybrid event (id 200626602265)*



(b) *stereo - hybrid*



(c) *stereo hybrid with Mie database*

Figure 3.19: FD energy resolution from stereo hybrid events.

Due to the energy dependence of the meson decay probabilities in the atmosphere, and thus the neutrino and muon production probabilities, the correction depends on the energy [141] and is also subject to shower-to-shower fluctuations [142].

Statistical uncertainties of the total energy The uncertainties of $\mathrm{d}E/\mathrm{d}X_{\max}$, $X_{\max}$, X_0 and λ obtained after the fit of the profile with Eq. (3.3.14), reflect only the statistical uncertainty of the light flux, $\sigma_{\mathrm{flux}}(E_{\mathrm{em}})$. Additional uncertainties arise from the uncertainties on the core location and shower direction (σ_{geo}) and the correction of invisible energy (σ_{inv}).

The statistical variance of the total energy is

$$\sigma_{\rm stat}(E_{\rm tot})^2 = E_{\rm tot}^2\sigma(f_{\rm inv})^2 + \left(\frac{{\rm d}f_{\rm inv}}{{\rm d}E_{\rm em}}E_{\rm em} + f_{\rm inv}\right)^2 \left(\sigma_{\rm geom}(E_{\rm em})^2 + \sigma_{\rm flux}(E_{\rm em})^2\right). \quad (3.3.17)$$

Air showers detected with more than one eye (Fig. 3.19(a)) offer the opportunity to compare two independent energy estimations of the same event. The energy resolution deduced from these so-called stereo events is about 9-10%, as shown in Fig. 3.19 independent of the availability of the Mie calibration constants. It was shown that the energy resolution from simulated data is of about 8% [127].

3.4 Typical events

A typical golden hybrid event recorded at the beginning of December 2005 is shown in Fig. 3.20(a). [1] This air shower has been *seen* in one of the telescopes of the Los Morados building and it also left a nice footprint on the SD. The color coding in this figure is according to the signal timing, early (blue) to late (red).

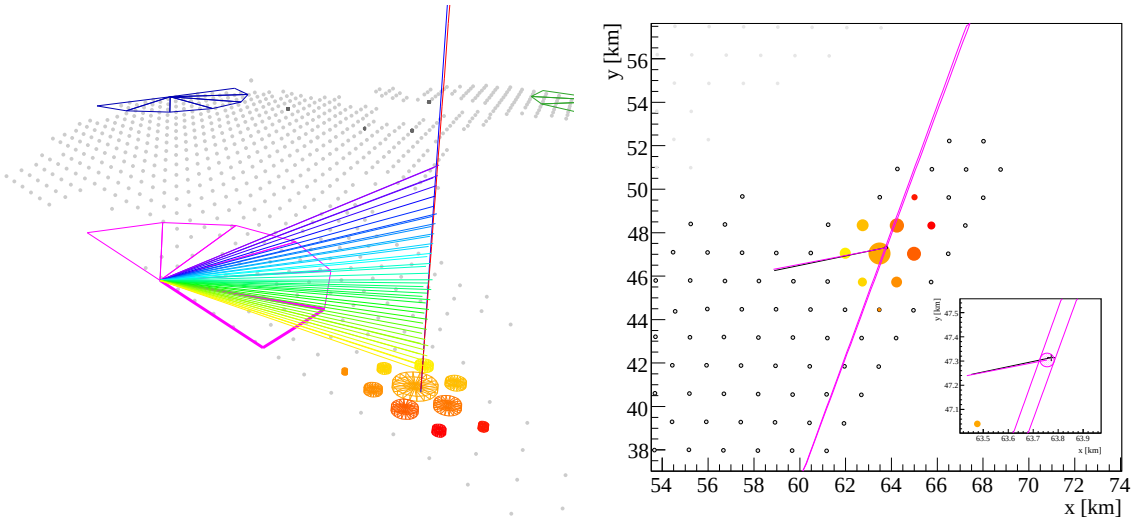
The core position has been independently reconstructed from the fluorescence detector measurement and from the surface detector information. In the infill of the Fig. 3.20(b) the good agreement between the two geometry reconstructions is shown. This air shower had an arrival direction of $(\theta, \phi) = (26.7 \pm 0.3, 191.4 \pm 0.6)_{\rm FD}/(26.5 \pm 0.2, 192.2 \pm 0.5)_{\rm SD}$.

For this event 10 stations have a signal recorded. The calibrated traces from one of the PMTs are shown in Fig. 3.20(c). Close to the core the signal time distribution is compact ($50\times$ 25 ns), and reaches amplitudes of 600 VEM. Far away from the shower center the signal is twice as spread in time, and is dominated by single muon peaks, with amplitudes up to 4 VEM. The signal from the station closest to the core is saturated in both gains, and a saturation recovery has been applied. The second closest to the core has recorded a signal saturated only in the high gain regime, and the low gain trace has been used for the calibrated trace.

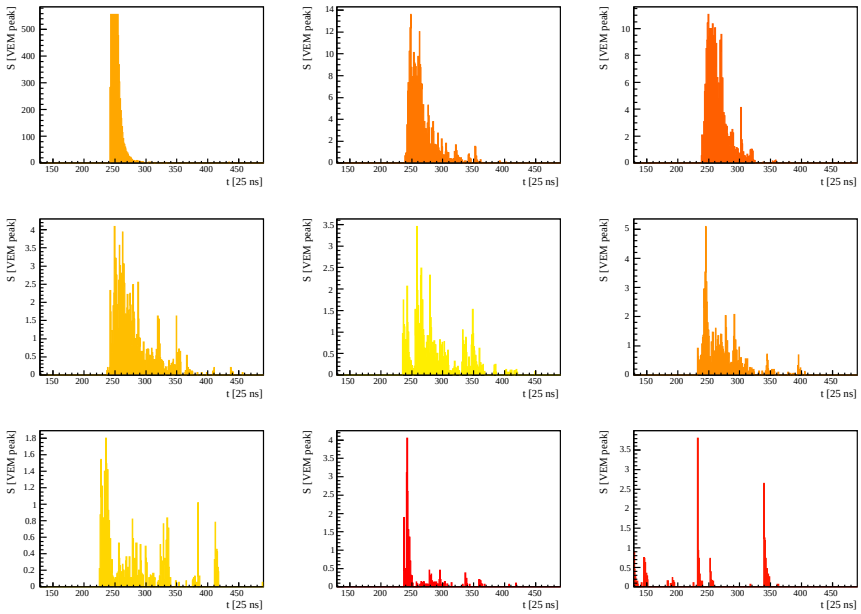
The optical image on the camera is illustrated in Fig. 3.21(a). The number of pixels is almost 10 times larger than the number of the stations, therefore the geometry of the shower, reconstructed from the timing information (Fig. 3.21(c)) is more reliable than the one from the surface detector. The typical recording time for the observed longitudinal development is of the order of 220×100 ns, while the shower front reaches the ground in ≈ 2000 ns (Fig. 3.21(d)).

After separating the different contributions to the light at aperture, reproduced in Fig. 3.21(b), the profile reconstruction is performed. For this event the scattered Cherenkov photons dominate over the direct Cherenkov light because the air shower direction is pointing away from the telescope. The energy deposit as a function of the slant depth is illustrated in Fig. 3.21(e). The energy of the extensive air shower is computed as the integral over the whole slant depth of the Gaiser Hillas function which describes the profile with a reduced χ^2 of 0.77. The shower maximum is at $775 \pm 6\,{\rm g/cm^2}$.

[1] Plots are made within the Advanced Data Summary Tree (ADST) software for data storage and visualization, part of which was developed during this thesis.



(a) *event pattern* (b) *footprint on array*



(c) *calibrated FADC traces*

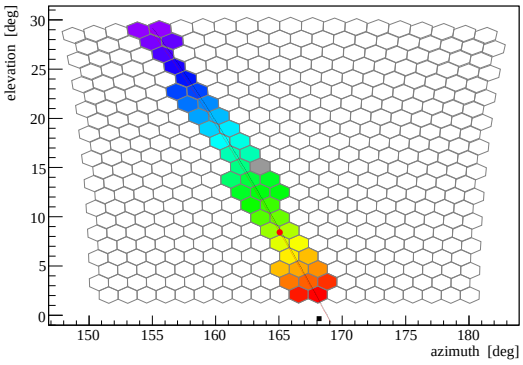
Figure 3.20: A typical golden hybrid event (id 200534001935) recorded at the Pierre Auger Observatory.

In Fig. 3.21(f) the signal as a function of the distance to the shower core in the shower plane coordinates is illustrated. The fit of the lateral distribution of particles has been done with the modified NKG function. The recovered signal of the station closest to the core is shown with a blue dot, and the recorded signal for this station is displayed as a red circle. The low

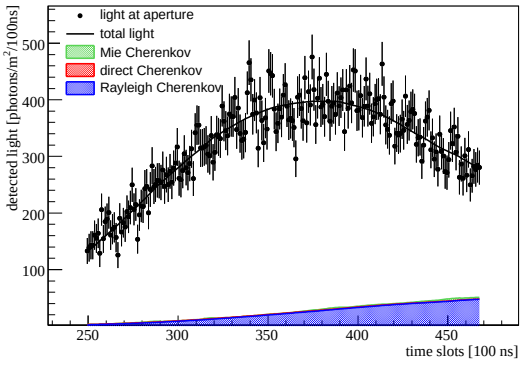
triangles represent the stations with no signal, used as upper limits in the fit. The signal at 1000 m, S(1000 m), marked with a red cross, is $142.1 \pm 5.5(\text{stat}) \pm 2.2(\beta)$ VEM.

The cosmic ray that entered the atmosphere had an energy of 26.1 ± 2.3 EeV reconstructed from the longitudinal profile, in agreement with the energy given by the surface detector of 27 ± 1 EeV.

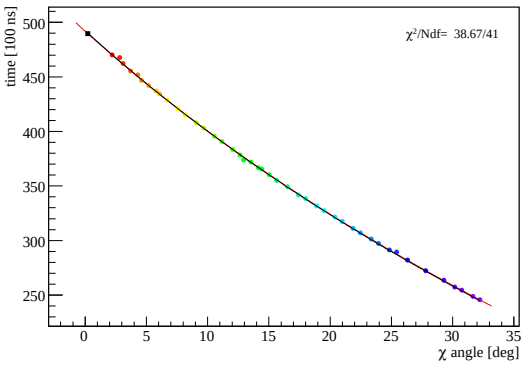
Pierre Auger Observatory is sensitive also to inclined events. A typical event, with zenith angle larger than 60° is shown in Fig. 3.22(a). Another independent data set is provided by the hybrid events (Fig. 3.22(b)), that triggered the telescope and only one, or two stations of the array have recorded a signal. A SD reconstruction is not feasible for these class of events. The timing information from one station is used in the geometry reconstruction.



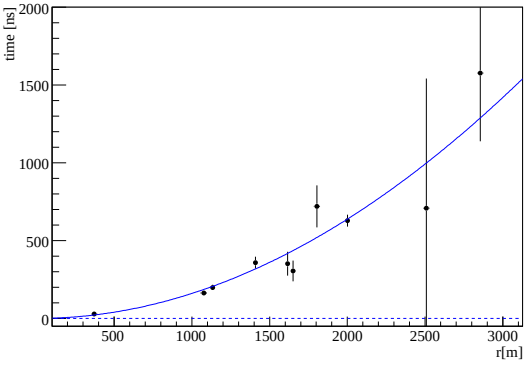
(a) *camera view*



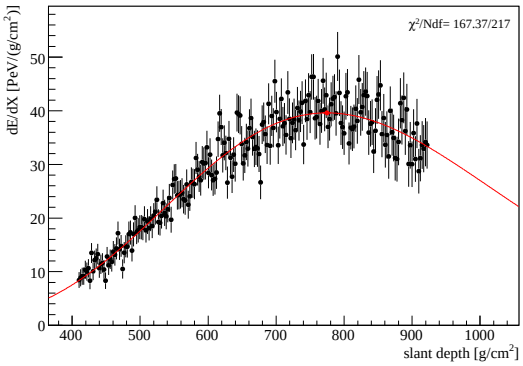
(b) *light at aperture*



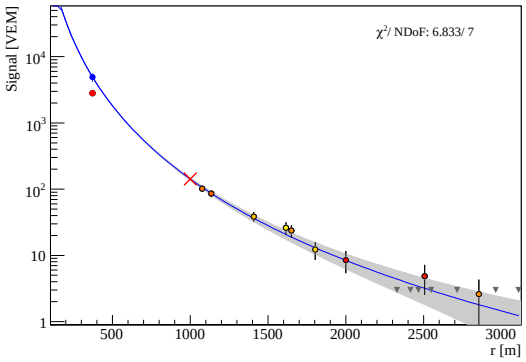
(c) *FD timing information*



(d) *SD timing information*

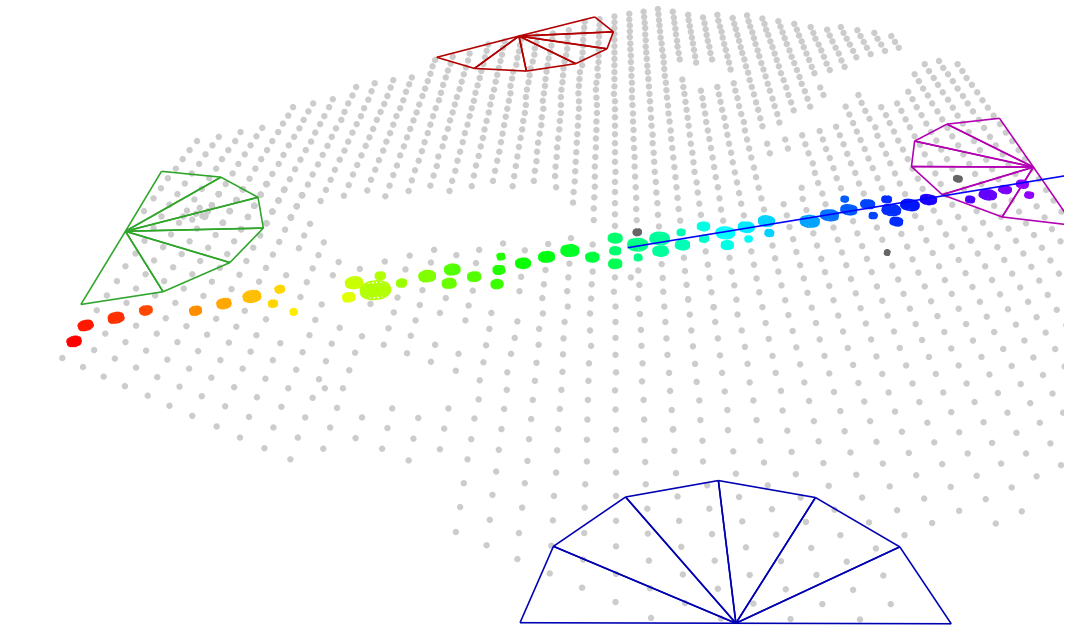


(e) *longitudinal profile*

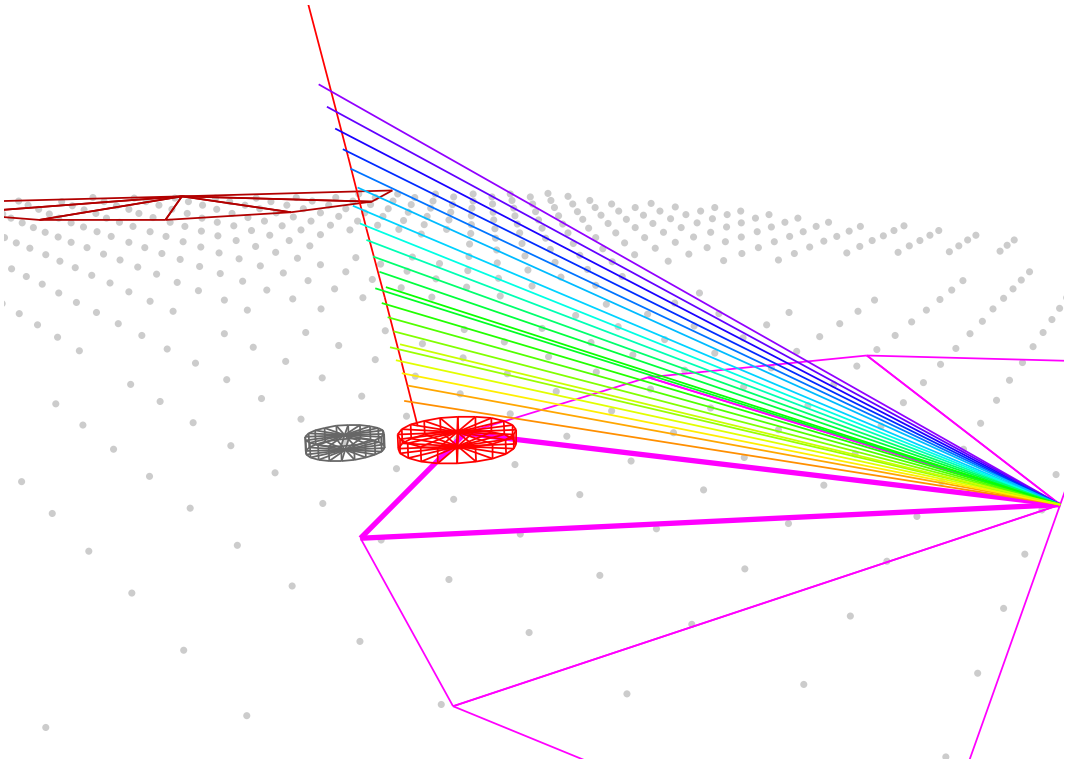


(f) *lateral distribution function*

Figure 3.21: Event reconstruction.



(a) *inclined event (id 200717502656)*



(b) *hybrid event (id 200701102325)*

Figure 3.22: Typical inclined and hybrid events.

Chapter 4

Constant Intensity Cut

The standard variable that relates the SD measurement to the energy of the primary air shower particle is the shower size at a distance of 1000 m from the core, $S(1000\,\mathrm{m})$. The stage of the development of an air shower depends on the traversed atmospheric depth. For the same energy of the primary particle the surface detector will measure different particle densities, depending on zenith angle. In order to correct for geometrical as well as for attenuation effects the hypothesis is taken that the flux, or the intensity (integrated flux above a certain energy) is isotropic for all energy ranges. This method has a long history going back to the work of the MIT group at sea-level and mountain altitudes [143].

4.1 Deducing the attenuation curve

Even though it was shown that the flux J is anisotropic in equatorial coordinates at high energies [55], the hypothesis of constant flux still holds at low energies and in local coordinates:

$$\frac{d}{d(cos\theta)}J = \frac{d}{d\theta}\left(\frac{1}{A_{\mathrm{eff}}}\frac{d^3N}{dt\,dE\,d\Omega}\right) = 0 \text{ and } \frac{d}{d\varphi}J = 0, \tag{4.1.1}$$

where θ is the zenith angle. Since the detector measurement is symmetric in azimuth, no dependence on φ of $\mathrm{S}(1000\,\mathrm{m})$ is expected. The effective area, A_{eff} is the area seen by the air shower, the projection of the geometrical flat area of the detector on the shower plane $A_{\mathrm{eff}} = A\cos\theta$. The other zenith angle contribution comes from the solid angle, $d\Omega = d\cos\theta d\varphi$.

Neglecting all other zenith angle independent variables (time t and the geometrical area), Eq. (4.1.1) reads as

$$\frac{dN}{dE\,d(\cos^2\theta)} = \text{const.}, \tag{4.1.2}$$

or integrating above a certain energy E_0, the intensity I, should be constant

$$\frac{dI}{d(\cos^2\theta)} = \left.\frac{dN}{d(\cos^2\theta)}\right|_{E>E_0} = \text{const.} \tag{4.1.3}$$

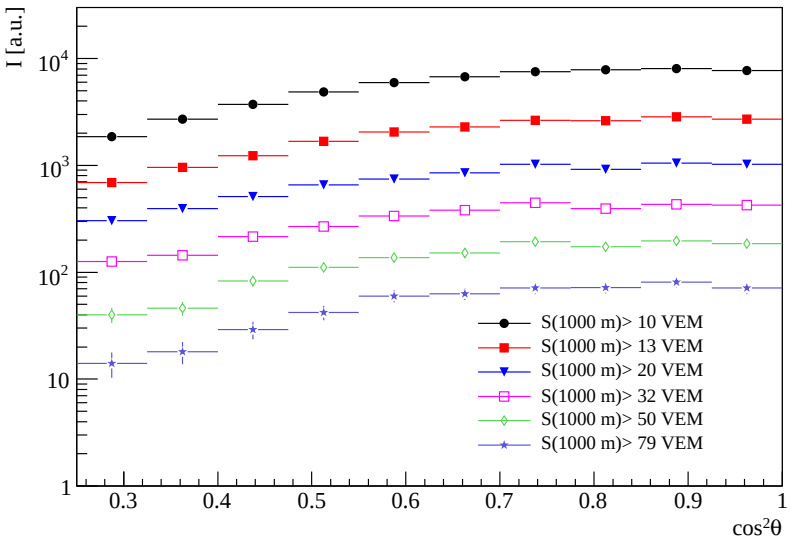


Figure 4.1: Intensity, number of events above a certain S(1000 m) versus zenith angle.

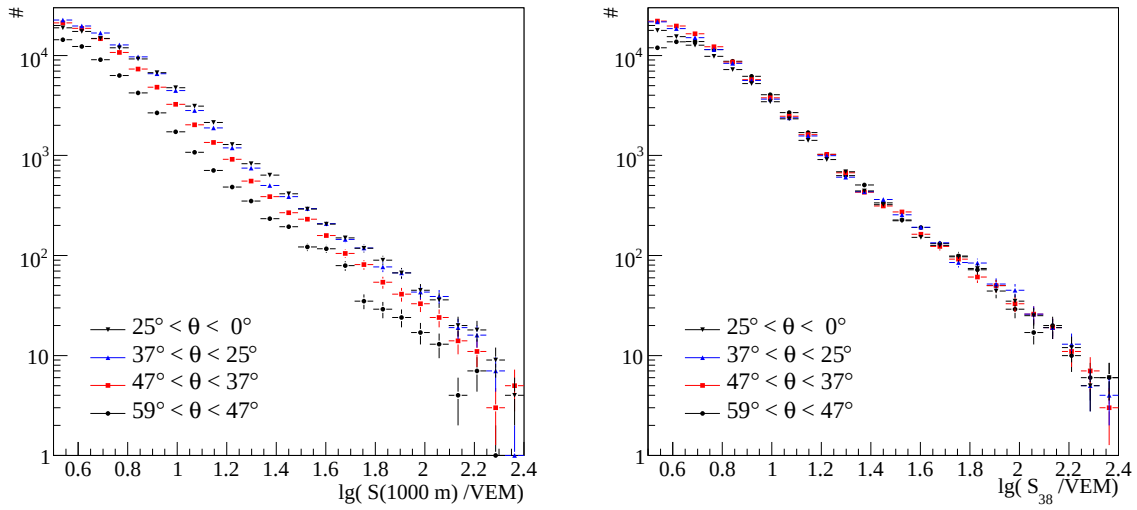
The zenith angle independent variable related to energy is S_{38}, the S(1000 m) that would be measured if the air shower comes from a zenith angle of 38°. Since our data have a zenith angle distribution peaked at 38° the impact of the correction function is minimized. Assuming that $E = f(S_{38})$, the relation between S(1000 m) and S_{38} is deduced from the assumption that the number of events above a certain S_{38}^0 is constant with zenith angle,

$$\left.\frac{dI}{d(\cos^2\theta)}\right|_{S_{38}>S_{38}^0} = \text{const.} \tag{4.1.4}$$

The attenuation curve might depend on mass composition or observation level of the shower development (i.e. energy). In the last part of this chapter the evolution of the attenuation curve with these quantities is analyzed. In case of the standard analysis of Auger the attenuation curve is assumed to be universal in energy.

The usage of the constant intensity cut method is to derive the attenuation curve, the relation between S(1000 m) and S_{38}. In Fig. 4.1 the number of events above a certain S(1000 m) is shown as a function of $\cos^2\theta$. The intensity uncertainties for different S(1000 m) cuts at a given zenith angle are correlated. In order to obtain the dependence of S(1000 m) versus zenith angle at a certain intensity two almost equivalent analyses are performed: Interpolation and power law fits.

Explicit error propagation Data are binned in 0.1 lg(S(1000 m)/VEM) intervals from 1 to 1.8 and 10 intervals in $\cos^2\theta$, from 0.2 to 1 (see Fig. 4.1). In case of the interpolation method, for each zenith angle bin the intensity versus S(1000 m) is obtained (Fig. 4.2(a)). The logarithm of two neighboring intensity bins, I_0 and I_1 in this histogram, can be interpolated



(a) *S(1000 m) distribution* (b) S_{38} *distribution*

Figure 4.2: Number of events as a function for equal $\cos^2\theta$ intervals.

by a line (corresponding to a local power law):

$$\ln S_0 = \alpha - \beta \cdot \ln I_0 \tag{4.1.5}$$

$$\ln S_1 = \alpha - \beta \cdot \ln I_1 \tag{4.1.6}$$

With the knowledge $I_0 = n_0$, $I_1 = n_0 + n_1$ where n_0 and n_1 are independent one obtains

$$\beta = \ln \frac{S_0}{S_1} \cdot \left(\ln \frac{n_0 + n_1}{n_0} \right)^{-1} \tag{4.1.7}$$

$$\alpha = \ln S_0 + \ln \frac{S_0}{S_1} \cdot \left(\ln \frac{n_0 + n_1}{n_0} \right)^{-1} \cdot \ln n_0. \tag{4.1.8}$$

The uncertainties from

$$S = \exp(\alpha - \beta \cdot \ln I) \tag{4.1.9}$$

are propagated by constructing the covariance matrix

$$U_S = \begin{bmatrix} \left(\frac{\partial\alpha}{\partial n_0}\right)^2 n_0 + \left(\frac{\partial\alpha}{\partial n_1}\right)^2 \cdot n_1 & \left(\frac{\partial\alpha}{\partial n_0}\right)\left(\frac{\partial\beta}{\partial n_0}\right) n_0 + \left(\frac{\partial\alpha}{\partial n_1}\right)\left(\frac{\partial\beta}{\partial n_1}\right) n_1 \\ \left(\frac{\partial\alpha}{\partial n_0}\right)\left(\frac{\partial\beta}{\partial n_0}\right) n_0 + \left(\frac{\partial\alpha}{\partial n_1}\right)\left(\frac{\partial\beta}{\partial n_1}\right) n_1 & \left(\frac{\partial\beta}{\partial n_0}\right)^2 n_0 + \left(\frac{\partial\beta}{\partial n_1}\right)^2 n_1 \end{bmatrix}. \tag{4.1.10}$$

where n_0 and n_1 have only Poisson fluctuations of $\sigma_n^2 = n$.

In the case of a power law fit of the intensity versus S(1000 m), data are correlated in the same way. For each zenith angle bin above the threshold shower size S_i the intensity I_i is given by

$$I_i = \sum_{j=0}^{i} n_j, \tag{4.1.11}$$

where n_i are the number of events in a specific $\cos^2\theta$ and lg(S(1000 m)) interval. $I_0 = n_0$ corresponds to the intensity above the highest signal bin. In matrix notations Eq.(4.1.11) reads as

$$\begin{bmatrix} I_0 \\ I_1 \\ I_2 \\ ... \\ I_N \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & ... & 0 \\ 1 & 1 & 0 & ... & 0 \\ 1 & 1 & 1 & ... & 0 \\ ... & & & & \\ 1 & 1 & 1 & ... & 1 \end{bmatrix} \begin{bmatrix} n_0 \\ n_1 \\ n_2 \\ ... \\ n_N \end{bmatrix} \tag{4.1.12}$$

in short

$$\vec{I} = \mathbf{B} \cdot \vec{n}. \tag{4.1.13}$$

The covariance matrix of $\vec{n}$ is diagonal with $(\mathbf{V_n})_{ii} = n_i$. The statistical uncertainties of $\mathbf{I}$ are obtained from standard error propagations as

$$\mathbf{V}_I = \mathbf{B}^{-1} \cdot \mathbf{V}_n \cdot (\mathbf{B}^{-1})^{\mathrm{T}}. \tag{4.1.14}$$

Since $(\mathbf{B}^{-1})^{\mathrm{T}} = (\mathbf{B}^{\mathrm{T}})^{-1}$, the inverse of $\mathbf{V}_I$ is given by

$$\mathbf{V}_I^{-1} = \mathbf{B}^{\mathrm{T}} \cdot \mathbf{V}_n^{-1} \cdot \mathbf{B} \tag{4.1.15}$$

where $(\mathbf{V_n})_{ii}^{-1} = 1/n_i$.

This leads to the inverse of the covariance matrix of the intensities, which even if the elements of $\vec{n}$ are independent has non-diagonal elements

$$\mathbf{V}_I^{-1} = \begin{bmatrix} n_0^{-1} & n_0^{-1} & n_0^{-1} & ... & n_0^{-1} \\ n_0^{-1} & n_0^{-1}+n_1^{-1} & n_0^{-1}+n_1^{-1} & ... & n_0^{-1}+n_1^{-1} \\ n_0^{-1} & n_0^{-1}+n_1^{-1} & n_0^{-1}+n_1^{-1}+n_2^{-1} & .... & n_0^{-1}+n_1^{-1}+n_2^{-1} \\ ... & & & & \\ n_0^{-1} & n_0^{-1}+n_1^{-1} & n_0^{-1}+n_1^{-1}+n_2^{-1} & .... & \sum_0^N n_i^{-1} \end{bmatrix}. \tag{4.1.16}$$

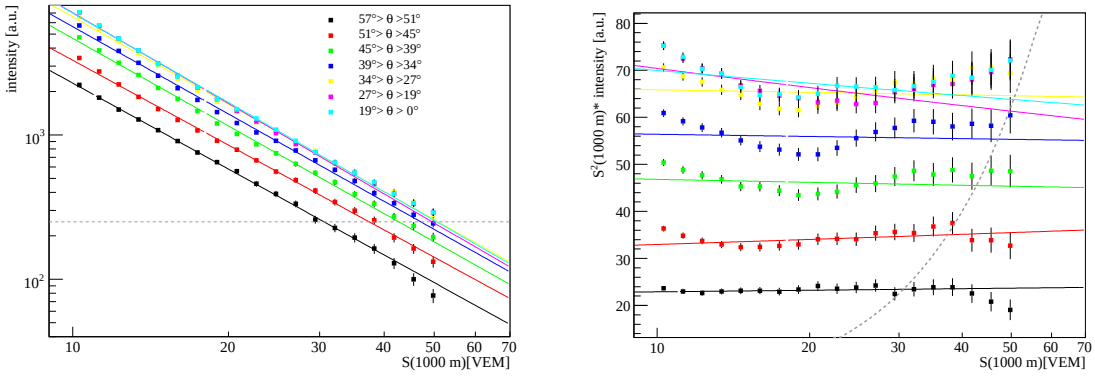
For a given measured intensity $\vec{I}$ a minimization function can be constructed

$$\chi^2 = [\vec{I} - f(\vec{S})]^{\mathrm{T}} \mathbf{V}_I^{-1} [\vec{I} - f(\vec{S})] \tag{4.1.17}$$

where $f(\vec{S})$ is the expectation value of the intensity at a certain signal. It can be written as

$$\chi^2 = \sum_{i=0}^{N} \left(\sum_{j=0}^{i} \frac{1}{n_j} \right) \left((I_i - f(S_i))^2 + 2 \cdot \sum_{k=i+1}^{N} (I_i - f(S_i)) \cdot (I_k - f(S_k)) \right). \tag{4.1.18}$$

The power law fit might introduce biases from a changing spectral index with energy, therefore the interpolation will be used as a default method. Even if they reduce the statistical uncertainties on the attenuation curve, the power law fit will be utilized only in a small study of the S(1000 m) integrated spectra.



(a) *power law fit* (b) *power law fit, scaled*

Figure 4.3: Intensity,i.e. number of events above a certain S(1000 m), versus S(1000 m) for equal $\cos^2\theta$ intervals, assuming a simple power law behavior (Eq. (4.2.1)). The data points under the dashed line are not used in the fit.

4.2 S(1000 m) integrated spectra

To observe any change with energy of the zenith angle dependency of S(1000 m) it is instructive to look at the integrated spectra for different zenith angles. All changes of the spectral indices in the cosmic ray differential flux are reflected also in S(1000 m) distributions. A first estimation of spectral features can be made.

In Fig. 4.3(a) the intensity as a function of S(1000 m) for different zenith angles is shown. The expected dependency of the energy on the shower size is a simple power law. In the hypothesis that there are no spectral features expected in the energy spectrum, the data should be well described by

$$f_1(S_i, k) = \zeta(k) S_i^{\gamma_1(k)} \tag{4.2.1}$$

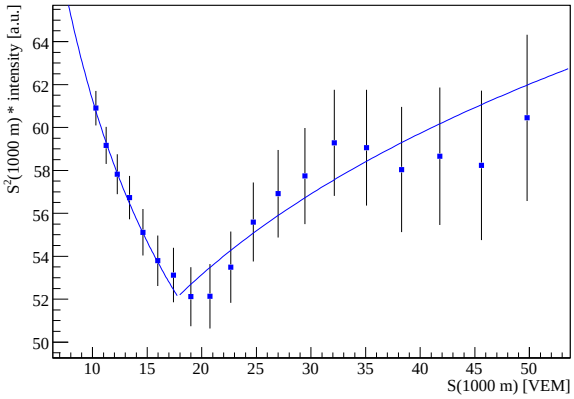
where k defines equal-$\cos^2\theta$ intervals. Moreover, assuming that the attenuation curve is constant with energy then $\gamma_1(k)$ should be the same for all zenith angle intervals.

For each zenith angle interval k, $\zeta(k)$ and $\gamma_1(k)$ were obtained. The intensity, scaled with S(1000 m)^2 to enhance eventual spectral features is shown in Fig. 4.3(b). As expected there are 2 deviations from the power law, first in the region of 10− 20 VEM, and the second at 40− 60 VEM, which are the equivalent of the ankle and the flux suppression in the cosmic ray flux.

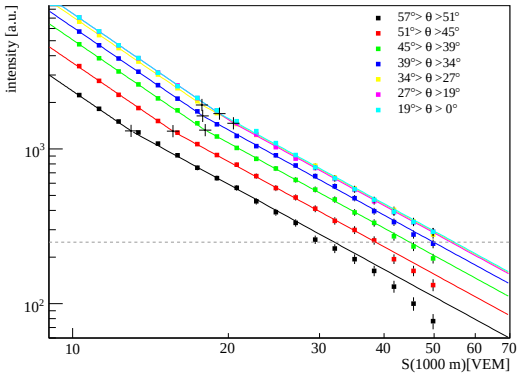
A broken power law is more appropriate to describe the S(1000 m) spectra

$$f_2(S_i, k, I_0(k)) = \begin{cases} \zeta(k)\ S_i^{\gamma_1(k)} & \text{for } S_i < S(I^0(k)) \\ \zeta(k)^{\frac{\gamma_2(k)}{\gamma_1(k)}} I_0(k)^{1-\frac{\gamma_2(k)}{\gamma_1(k)}}\ S_i^{\gamma_2(k)} & \text{for } S_i > S(I_0(k)), \end{cases} \tag{4.2.2}$$

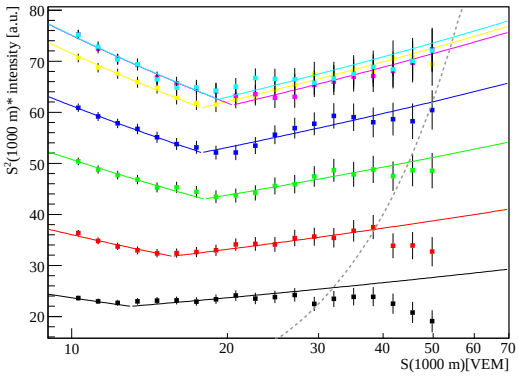
where $I_0(k)$ is the break position. To avoid the last change of the spectral index, due to the flux suppression, data points having an intensity lower than 250 events are rejected in the fit.



(a) *broken power law fit* ($34° < \theta < 39°$)



(b) *broken power law fit*



(c) *broken power law fit, scaled*

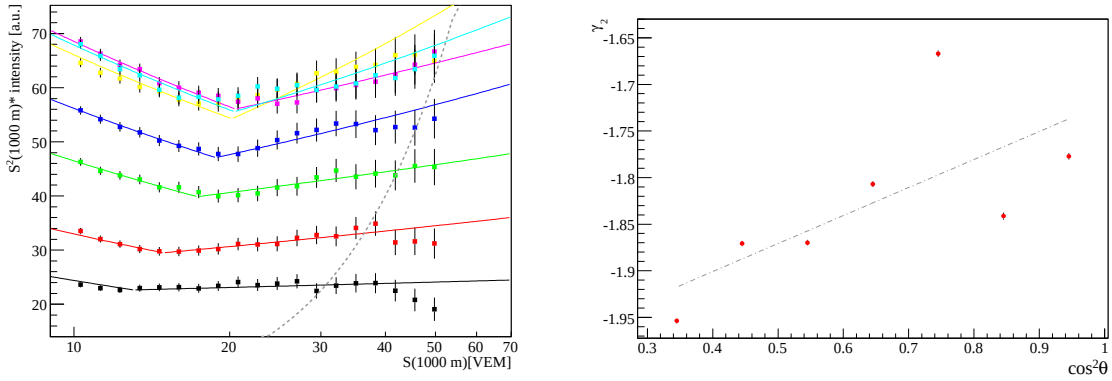
Figure 4.4: Intensity above a certain S(1000 m) versus S(1000 m) for equal $\cos^2\theta$ intervals.

Any change of the indices of the power laws with the zenith angle, or of the break intensity will result in an energy dependent attenuation curve.

The first power law index varies mainly because of the trigger threshold behavior with zenith angle, described in Section 7.1. The zenith angle bin where the threshold saturation is achieved first is for $\theta \approx 35°$. The scaled intensity versus S(1000 m) of this zenith angle bin is shown in Fig. 4.4(a). The spectrum for this zenith angle will be used as default one for the next analysis.

The first step is to fix the slopes $\gamma_1(k)$ and $\gamma_2(k)$ to standard values and let the break position $I_0(k)$ vary as a parameter. The result is shown in Fig. 4.4(b). The intensity (marked with black crosses) is dependent on zenith angle. This indicates that there is a slope difference at large S(1000 m) values for the different equal $\cos^2\theta$ spectra.

The next step is to fix I_0 and only γ_2, the spectral index at high S(1000 m), is a free parameter in the minimization. γ_1 is fixed to avoid threshold effects (Fig. 4.5(a)). The second slope at large S(1000 m) as a function of zenith angle is shown in Fig. 4.5(b). The



(a) *power law fit, scaled* (b) γ_2

Figure 4.5: (a) Intensity versus S(1000 m) for equal $\cos^2\theta$ intervals. The fit was done assuming the same intensity for the change of the power law index. Also, in order to get rid of the threshold effect the slope of the first power law was fixed to the same value. The data points on the right hand side of the dashed line are not used in the fit. (b) γ_2 as a function of $cos^2\theta$.

linear correlation coefficient is $r_0 = 0.68$, giving a probability of 92% (1.7σ) that the variables are not independent. This study gives an indication that there is a change of the S(1000 m) attenuation at highest energy, but it is not statistically significant.

Several functional forms have been used within the Auger collaboration to describe the attenuation behavior: Gaussian, third degree polynomial, different exponential functions and second degree polynomials. They are illustrated in Fig. 4.6. There is a difference of up to 30% depending on threshold intensity and on different parameterizations. To note is that there is a small intensity dependence in all the studies (attenuation curves deduced at different intensities are shown in the same figures with dotted lines).

Two functions are used to describe the attenuation and are compared: A second degree polynomial in $\cos^2\theta$

$$S(1000\,\mathrm{m}) = S_{38}(1 + a \cdot (\cos^2\theta - \cos^2\theta_{\mathrm{ref}}) + b \cdot (\cos^2\theta - \cos^2\theta_{\mathrm{ref}})^2) \tag{4.2.3}$$

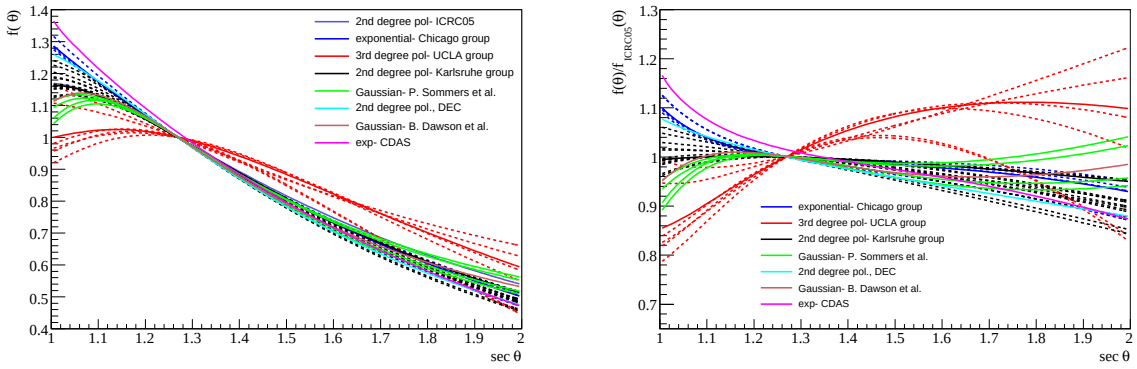
with two free parameters a and b and an exponential function

$$S(1000\,\mathrm{m}) = S_{38} \exp\left(X_0/\lambda \cdot (\sec\theta - \sec\theta_{\mathrm{ref}})\right), \tag{4.2.4}$$

where X_0 =879 g/cm^2 and the free parameter is λ.

In Fig. 4.7 the reduced χ^2 for the fit as a function of a lower limit on S_{38} is shown. The interpolation method is used to derive the corresponding S(1000 m) at a certain zenith angle. A polynomial describes the data better, the reduced χ^2 being lower than for exponential, independent of the intensity, i.e. S_{38}.

The attenuation curve parameters as a function of the corresponding S_{38} are shown in Fig. 4.8. Red circles represent the obtained parameters using a simple interpolation, i.e. assuming a local power law behavior. An energy dependence of the parameters with energy is observed. It might be due to either a change in the mass composition of the cosmic rays, or a different



(a) *attenuation curves* (b) *attenuation curve ratios*

Figure 4.6: (a) Different attenuation curves used within the Auger collaboration. (b) The same attenuation curves relative to the attenuation curve used in [144]. Dashed lines represent the same functions as the corresponding colors, but different cut intensities.

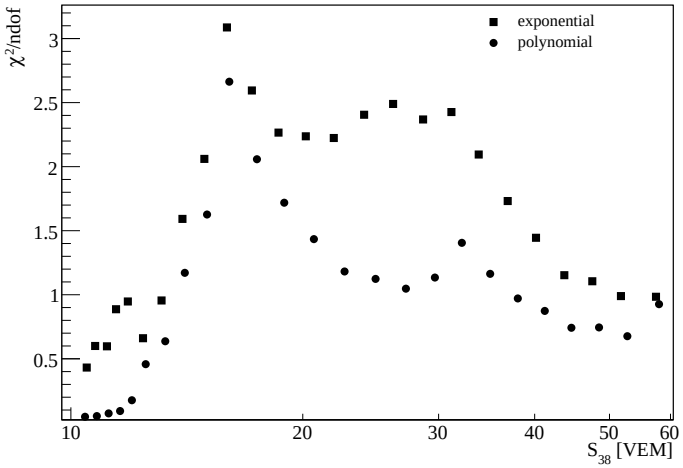


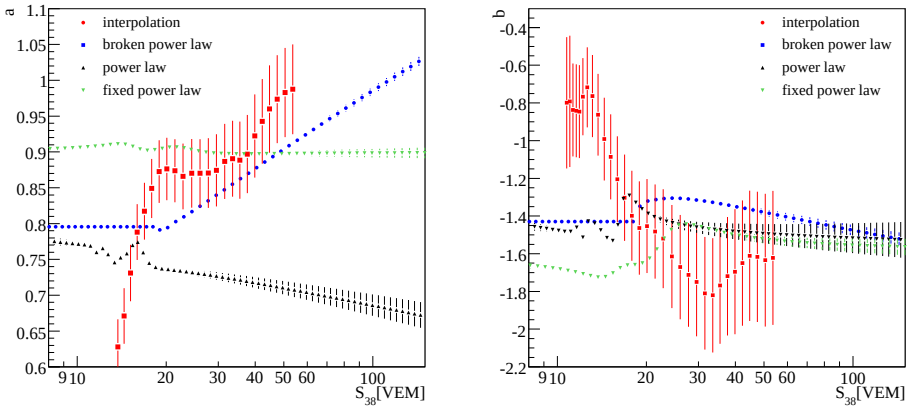
Figure 4.7: The reduced χ^2 for the attenuation curve described by Eq. (4.2.4) and (4.2.3).

level of observation. The data points are correlated and the deviations are not statistically significant. The first part of the evolution, up to a signal of S_{38}=15-20 VEM is caused by detector threshold effects.

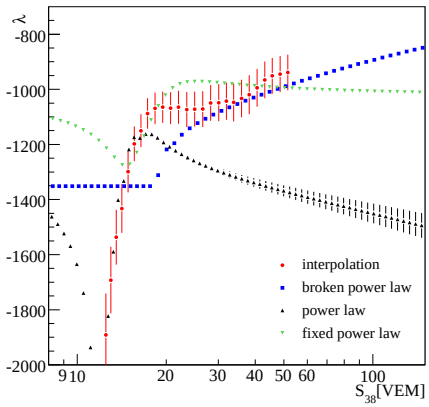
A simple power law fit, represented by black upper triangles, is not appropriate. The results have large deviations from the interpolation parameters and are dominated by low S(1000 m), in the threshold region where they intersect the values obtained by interpolation.

Neglecting threshold events, and assuming the same power index for all zenith angles and all S(1000 m) intervals (shown in the same figures with green markers) the parameters agree with the interpolation method in a S_{38} interval of [20 VEM, 45 VEM].

The same behavior as in the interpolation method is observed (blue squares) if the second



(a) *parameter a* (b) *parameter b*



(c) *parameter λ*

Figure 4.8: The parameters of the attenuation curve as a function of S_{38}. (a) and (b) parameters of the polynomial fit as described by Eq. (4.2.3); (c) exponential fit (Eq. (4.2.4))

slope of Eq. (4.2.2) is a free parameter in the minimization.

The results with the power law fit are not conclusive. The assumption that the attenuation curve is universal and that the observed behavior is just a statistical fluctuation is taken. The default intensity is corresponding to a S_{38}= 47.4±0.8 VEM in the region of the measured energy spectrum (see Section 7.1). The attenuation curve obtained with this default intensity is shown in Fig. 4.9, with the fits using the two functional forms. The obtained parameters and their correlation coefficient are:

$$\begin{aligned} a &= \ 0.919 \pm 0.055\ (\text{stat})\ ^{+0.02}_{-0.09}\ (\text{sys}) \\ b &= -1.13 \pm 0.26\ (\text{stat})\ ^{-0.04}_{+0.19}\ (\text{sys}) \\ \rho_{ab} &= 0.50\ (\text{stat}) \end{aligned} \tag{4.2.5}$$

and

$$\lambda = -986 \pm 52\ (\text{stat})\ ^{+94}_{-81}\ (\text{sys})\ \text{g/cm}^2. \tag{4.2.6}$$

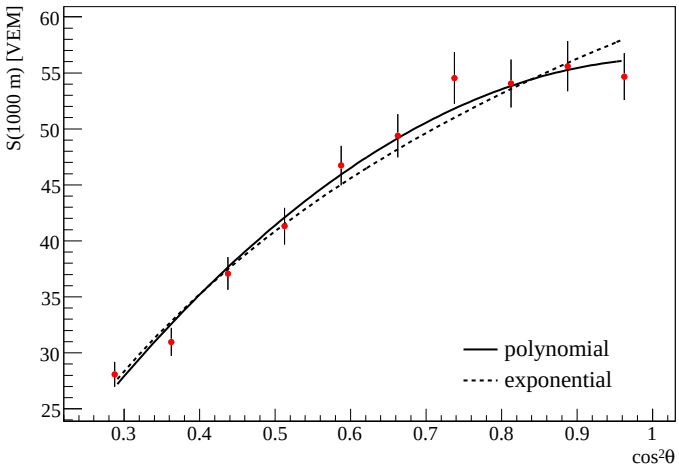
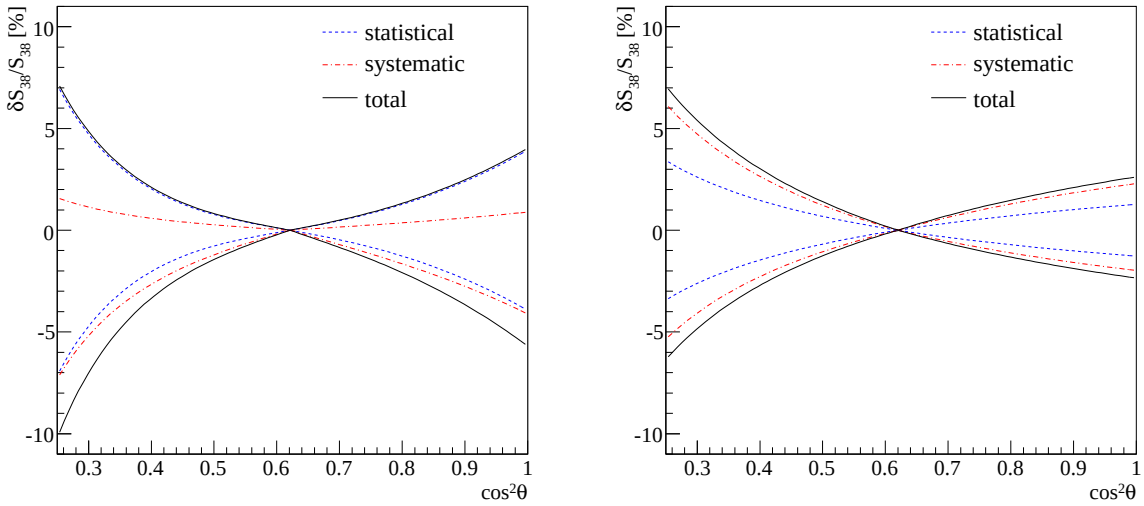


Figure 4.9: The default attenuation curve described by 4.2.3 with the parameters from Eq. (4.2.5). (Eq. (4.2.4) is shown for comparison with dashed line).



(a) *polynomial* (b) *exponential*

Figure 4.10: The contribution of the attenuation curve uncertainties to the S_{38} uncertainties.

The systematic uncertainties for the parameters were deduced as the maximum differences of the other attenuation parameterizations to the default one in the S_{38} region above 10 VEM. The statistical and systematic contribution to S_{38} uncertainties as a function of the zenith angle are shown in Fig. 4.10. The total uncertainty is smaller than 10% for all zenith angles.

To investigate if the dependency is caused by mass composition of the primary particles, the same analysis has been performed using simulated air showers (Chapter 6).

Chapter 5

Energy Calibration

The energy calibration of the SD measurements is based on the derivation of a relation between S_{38} and the energy given by the hybrid reconstruction.

A new method to calibrate directly the energy without the constant intensity assumption is used to cross check the attenuation curve and the energy conversion.

5.1 Event selection

The data sample used to calibrate the energy contains SD events of same quality as in the energy spectrum to avoid possible biases. The only selection applied are T5 condition and requirement of a zenith angle smaller than 60°.

Quality cuts Good quality FD events are a need and thus the rejection criteria, similar to the ones used in the mean $X_{\max}$ analysis [67], are applied.

(0) Events that are in time periods with no data for the absolute calibration of the FD available are rejected.

(1) The shower maximum should be within the detected slant depth (Fig.5.1(a)). A precise fit of the longitudinal shower development requires an observation of a significant fraction of it. Especially if only the rising or falling edge of a shower profile is observed no reliable estimate of the energy can be given.

(2) As the profile reconstruction algorithm propagates both, light flux and geometrical uncertainties, the estimated uncertainties of $X_{\max}$ and energy are good variables to reject poorly reconstructed showers. Therefore the uncertainty on $X_{\max}$ is required to be smaller than $\sigma(X_{\max}) < 40\,\mathrm{g/cm^2}$ (Fig.5.1(b)) and

(3) the relative uncertainty of the total energy smaller than 20% (Fig.5.1(c)),

(4) Atmospheric conditions such as clouds and fog can distort the profiles and these events can be identified by their deviation from the assumed Gaiser-Hillas shape (Fig.5.1(d)). Thus the reduced χ^2 of the Gaiser Hillas fit should be smaller than 2.5.

Table 5.1. Table head: Cut number; first line: The number of events that are rejected only by the corresponding cut even if all the other cuts are applied; second line: The number of events left after applying consecutively the cuts; third line: The step-by-step efficiency on the remaining events selected with the previous cuts.

cut	1	2	3	4	5	6	7	8	9
(n-1)	31	66	3	65	84	7	1042	25	2293
events left	10928	7643	7620	7377	7173	7113	5228	5178	2885
efficiency	68%	69%	99%	96%	97%	99%	73%	99%	56%

(5) In order to avoid uncertainties from the modeling of Cherenkov light, events with Cherenkov light contributions of more than 50% are rejected (Fig.5.1(e)).

(6) By requiring the distance in the shower plane to the station used in the hybrid geometry fit to be smaller than 1000 m, mono events with random SD triggers are suppressed. This value was chosen also to account for the fact that in the FD geometry reconstruction no curvature for the shower front is assumed.

(7) Faint showers at low energies can sometimes be well fitted by a straight line (Fig. 5.1(f)), therefore a requirement that the χ^2 obtained with a linear fit and a Gaiser-Hillas fit is at least greater than 4 (in analogy to one parameter least square fit where $\chi^2_{\rm min} + 4$ corresponds to a 2σ deviation).

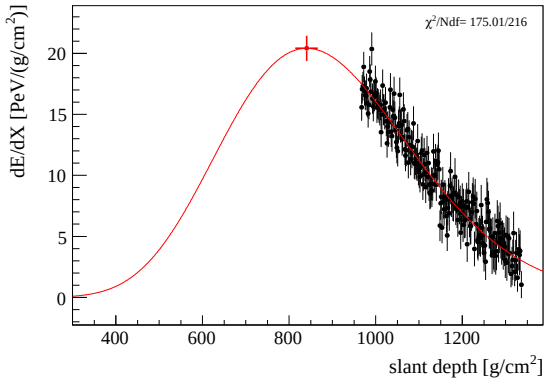
(8) Due to a flaw in the current light collection algorithm, the light at the boundaries of the cameras can be lost for events triggering more than one telescope. Therefore profiles with gaps greater than 20% of the observed length are not considered.

(9) Events which are reconstructed using VAOD measurements by the CLF have a better energy resolution. This cut reduces to half the event statistics so it will be applied only for systematic studies.

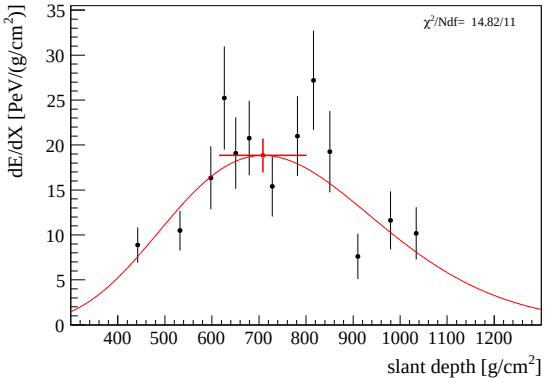
The total number of golden data events that fall in the time periods with good calibration is 15983. The number of events that are at the end of the selection available is 2885. The step-by-step efficiency as well as the number of events rejected by each cuts are shown in Table. 5.1. The most stringent ones are the requirement of X_{max} in field of view and the profile reconstruction cuts which select high quality data. The number of events is highly reduced by the requirement on the difference between the χ^2 obtained with the Gaiser-Hillas fit and a linear fit, but the rejected events are mostly in the low energy region. After applying all the quality cuts (1-8) the data are reduced to half by the requirement of the availability of the Mie attenuation parameters measured by the CLF and used in reconstruction.

Anti-bias cuts

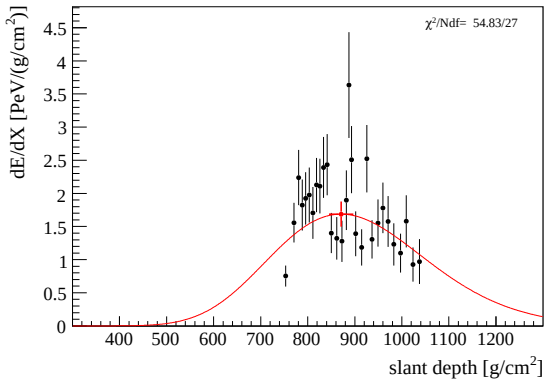
The air showers always trigger the surface detector and have a reasonable reconstruction if their energy is greater than 3 EeV. Below this energy the detector is more sensitive to upward



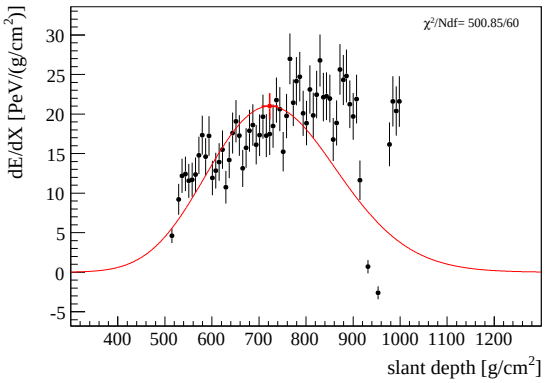
(a) $X_{\max}$ *not in field of view*



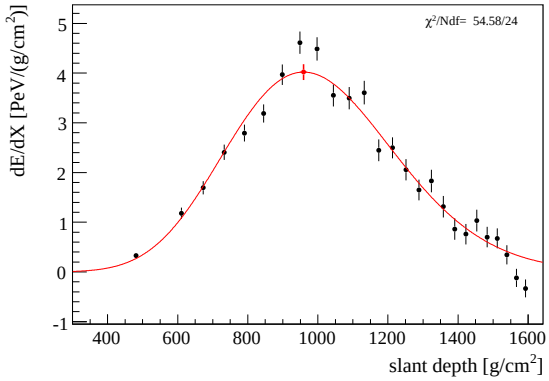
(b) $X_{\max}$ *uncertainty* > 40 *g/cm*^{2}



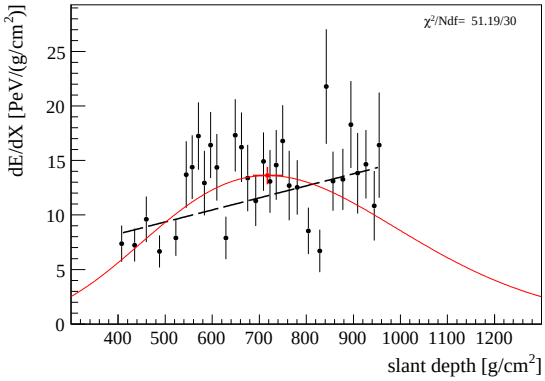
(c) *relative energy uncert.* $> 20\%$



(d) *reduced* $\chi^2 > 2.5$



(e) *Cherenkov fraction* $> 50\%$



(f) *straight line - Gaiser Hillas* χ^2 *difference* < 4

Figure 5.1: Some of the profiles from the events rejected by quality cuts.

fluctuations of S(1000 m). The rejection of the events under this energy might introduce important biases which are reduced with an anti bias cut based on a toy Monte Carlo.

The energy versus S_{38} is shown in Fig. 5.2. As can be seen at low energies ($E < 10^{18}$ eV) the correlation deviates from a linear behavior, which can be attributed to a trigger threshold effect (only upward fluctuations of S_{38} can form a low energy 3ToT). Previous analysis have

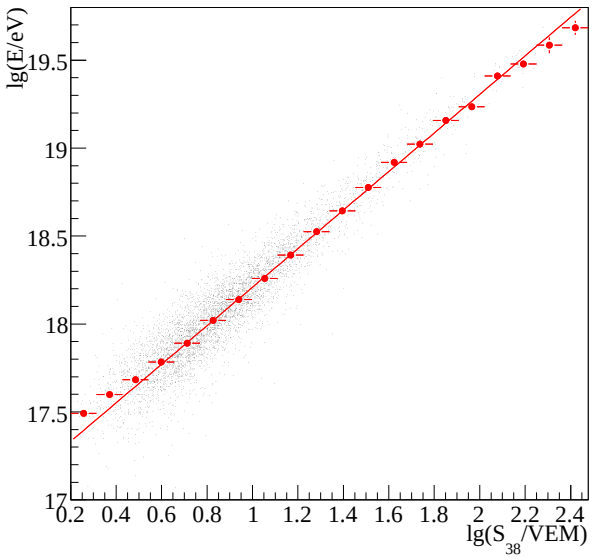
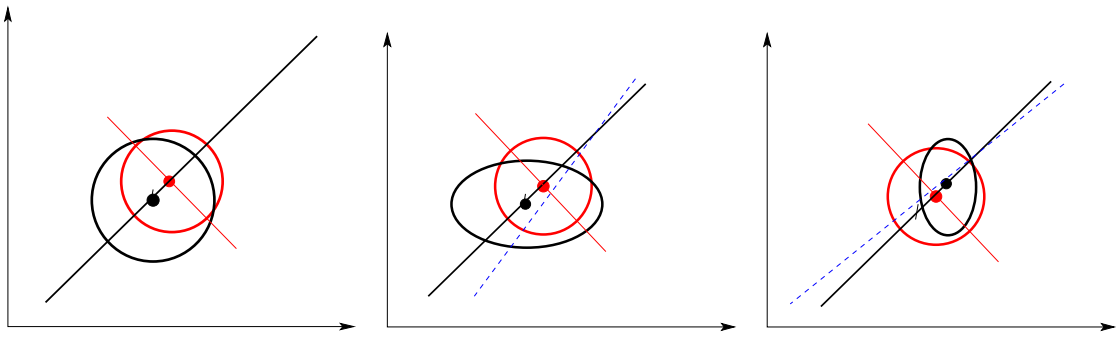


Figure 5.2: The energy versus S_{38}.



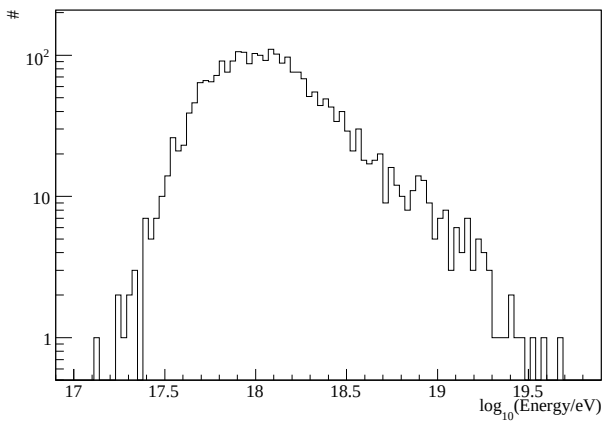
(a) *constant* σ_x/σ_y (b) *variable* σ_x/σ_y (c) *variable* σ_x/σ_y

Figure 5.3: (a) Constant uncertainties; Energy dependent uncertainties: (b) downward contributions from rejected events (c) upward contributions from the nearby accepted events.

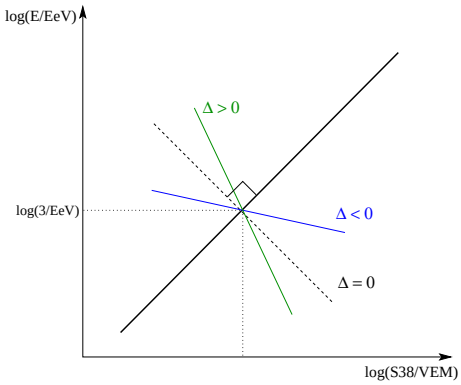
discarded events in the low energy region if the events were falling under a line perpendicular to the fit line. But due to different relative uncertainties of the energy and of S_{38}, this simple rejection criteria introduces a bias and can influence the cosmic ray flux up to 20%, especially in the low energy range.

To reduce a bias in a two dimensional fit, the influence of up- or downward fluctuations from both variables has to be understood. This corresponds to minimizing the influence of the upward fluctuations of S_{38} caused by the trigger efficiency of the surface detector.

When the uncertainties of the two variables, x and y, are constant, an optimal line for the



(a) *golden hybrid events*



(b) Δ *parameter*

Figure 5.4: (a) Energy distribution of the golden hybrid events. (b) Illustration of the Δ parameter.

cut is described by

$$B = -\frac{\sigma_x^2}{\sigma_y^2}\frac{1}{B_0} \tag{5.1.1}$$

where σ_x and σ_y are the relative uncertainties, B is the slope of the cut line and B_0 is the fit line. Thus, if the FD energy, y, was measured with infinite precision, a simple cut on energy would not introduce any bias. If the uncertainties of the two variables, x and y, are equal and normal distributed this condition is fulfilled when the events below a line perpendicular to the fit function are rejected.

Eq. (5.1.1) does not hold when the ratio $\sigma_x/\sigma_y = \sigma_{S_{38}}/\sigma_E$ is energy dependent. This is explained in Fig. 5.3. For each point from the energy calibration there is a transformation to a coordinate system where the uncertainties in x and y are equal, e.g. the 1σ contour plot is a circle. If $\sigma_x/\sigma_y = \text{const.}$, then for any point from the energy calibration the 1σ contour plot becomes a circle, therefore one does not introduce any bias by rejecting the events under a perpendicular line to the true one (Fig. 5.3(a)). This is not the case for the golden data set, where the relative energy uncertainties are constant and the relative uncertainties of S_{38} decrease with energy. In the same representation the 1σ contour plots will be ellipses for the neighboring events. The neighboring point with lower energy would introduce downward fluctuations (Fig. 5.3(b)), while the upper energetic events would contribute more to the upward fluctuations (Fig. 5.3(c)). Schematic biased slopes are shown in the figures with dotted lines. The contribution to the bias from higher energy neighbors is of less importance for our data sample because the number of events decreases rapidly with energy.

To solve this problem we need to deduce an *anti-bias* cut based on Monte Carlo studies.

Determining an unbiased cut by toy Monte Carlo The input for the toy Monte Carlo (MC) are two functions for the conversion of S_{38} to energy

$$\log(E/\text{EeV}) = A + B \cdot \log(S_{38}/\text{VEM}), \tag{5.1.2}$$

with input values $A = -2.02$ and $B = 1.1$ and

$$E = A \cdot S_{38}^{B} \ \ [\mathrm{EeV}], \tag{5.1.3}$$

with $A = 0.134\,\mathrm{EeV}$ and $B = 1.1$. B is referred from now on as *slope*.

This is motivated by the need of investigating the difference between fitting a linear function in log-log, equivalent to assuming log-normal uncertainties for E and S_{38}, and a power law function, assuming Gaussian uncertainties on these variables. The data are not conclusive in this respect.

The events are drawn randomly from the real data energy histogram, shown in Fig. 5.4(a). This enables a toy MC free of any spectral index assumption or any T4 trigger parameterization. Also a spectral index of -2.7 was investigated to see if there is a strong dependency on the assumed power law.

To deduce the optimal cut line, the deviation of the obtained slope from the true one is our benchmark. The cut-line is defined by a slope of

$$B = -\left(\frac{1}{B_0} + \Delta\right), \tag{5.1.4}$$

with Δ taken between -1 and 1 in steps of 0.1; B_0 is the resulting energy calibration slope. For $\Delta = 0$, the cut and reconstructed lines are perpendicular. Their intersection is set at 3 EeV, as sketched in Figure. 5.4(b). Negative values of Δ determine a cut closer to horizontal, appropriate when $\sigma_E/\sigma_{S38} < 1$ and positive values correspond to $\sigma_E/\sigma_{S38} > 1$.

Three MC samples are generated with respect to the uncertainties on S_{38} and hybrid energy

(a) **Equal uncertainties:** The relative uncertainties of both variables were assumed to be normal distributed with a mean of 10% and a spread (i.e. 'error of the error') of 1%. In this situation rejecting events under a perpendicular cut should not introduce any bias.

(b) **Different, but constant uncertainties:** To test Eq. (5.1.1), the relative uncertainties were generated from Gaussian distributions with $\sigma_E = 5 \pm 1\%$ and $\sigma_{S_{38}} = 10 \pm 1\%$.

(c) **Realistic uncertainties:** The golden hybrid data uncertainties (Fig. 5.5) were used as input for the MC.

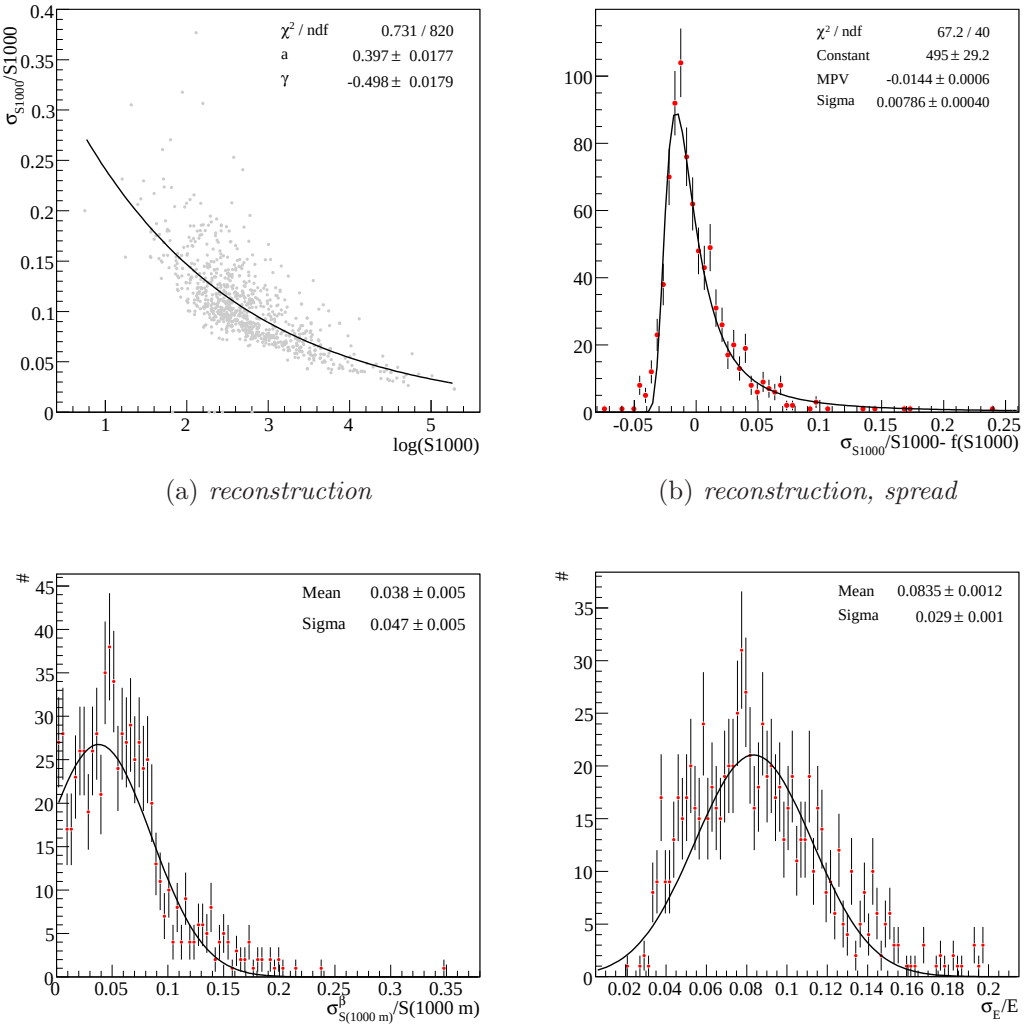
The relative hybrid energy uncertainty was considered to be $8 \pm 3\%$ and constant (Fig. 5.5(d)). The contributions to the S_{38} uncertainties come from reconstruction, from LDF parameterization, and from shower to shower fluctuations. The uncertainties are obtained from data by parameterizing the estimated uncertainties from the minimizations. In this way the toy MC is based on uncertainties similar to data.

The reconstruction uncertainties of S(1000 m) decrease with energy

$$\sigma^{\mathrm{rec}}_{S(1000\,\mathrm{m})} = (0.40 \pm 0.02) \cdot S(1000\,\mathrm{m})^{(-0.50 \pm 0.01)}. \tag{5.1.5}$$

The spread is described by a Landau function with a mean probable value of -0.0144 ± 0.0006 and $\sigma = 0.0072 \pm 0.0004$. The uncertainties of $S(1000\,\mathrm{m})$ from LDF parameterization are assumed constant and distributed Gaussian

$$\sigma^{\mathrm{LDF}}_{S(1000\,\mathrm{m})} = 0.04 \pm 0.05. \tag{5.1.6}$$



(a) *reconstruction* (b) *reconstruction, spread*

(c) *LDF parameterization, spread* (d) *energy*

Figure 5.5: Uncertainties of golden data used as input for the toy MC. (b) the spread around the fit function shown in (a).

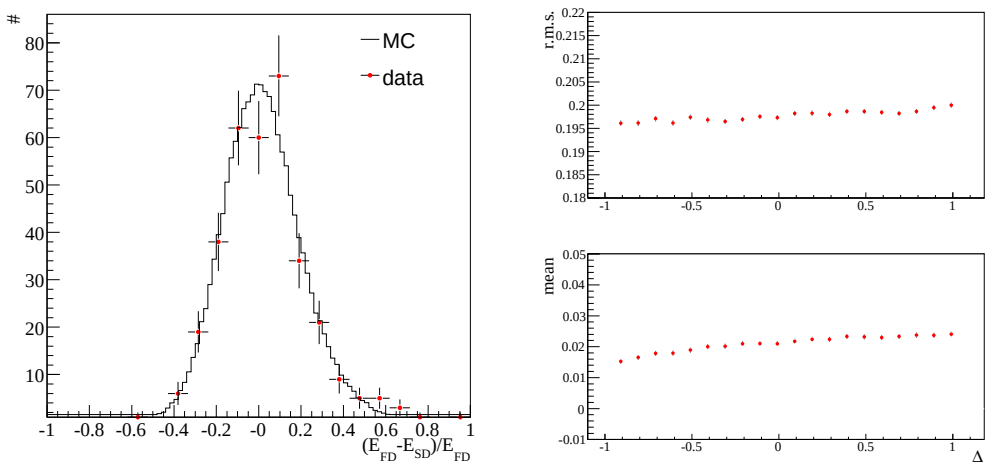
The shower to shower fluctuations are taken as

$$\sigma^{\mathrm{s2s}}_{S(1000\,\mathrm{m})} = 0.1, \tag{5.1.7}$$

independent of energy and zenith angle, as approximation of the parameterization deduced in [145].

Zenith angle uncertainties are neglected, since their contribution is at the level of 0.1%. Thus

$$\sigma_{S_{38}} = \frac{\delta S_{38}}{S_{38}} = \frac{\delta S(1000\,\mathrm{m})}{S(1000\,\mathrm{m})} = \sigma_{S(1000\,\mathrm{m})} \tag{5.1.8}$$



(a) *data and MC comparison* (b) *mean and RMS for MC*

Figure 5.6: Check for the consistency between data and the toy MC; the $(E_{FD} - E_{SD})/E_{FD}$ distribution with a mean of $\approx 0.02 \pm 0.01$ and RMS $\approx 0.19 \pm 0.1$ is well reproduced.

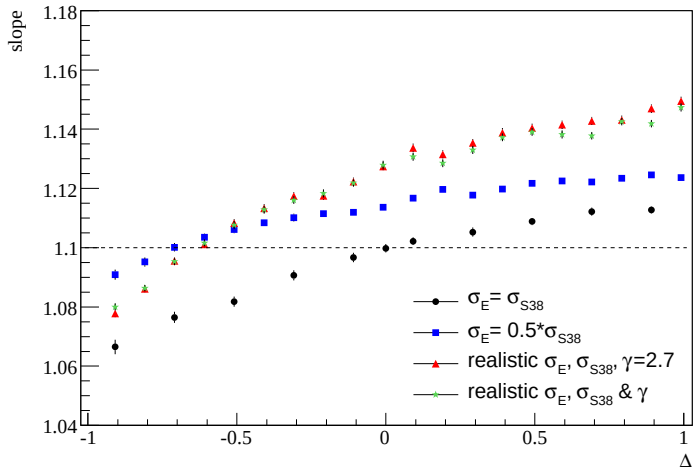
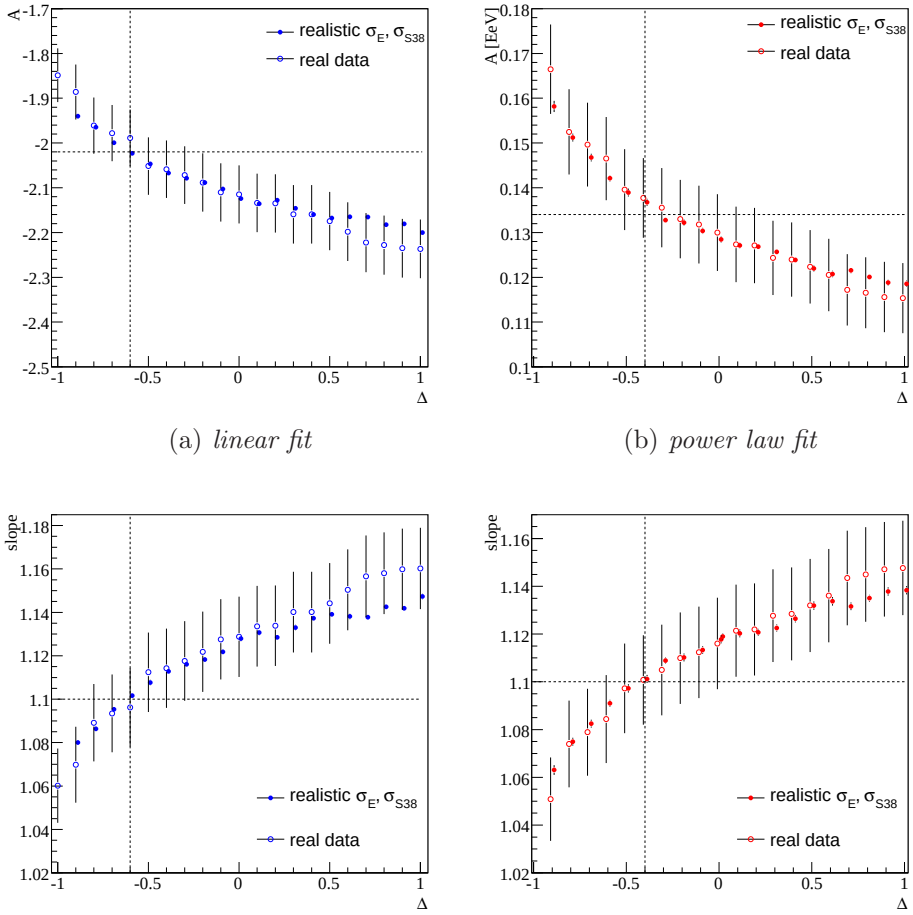


Figure 5.7: 'Measured' slope using different diagonal cuts as a function of Δ. The true slope in simulations was set to 1.1.

Accuracy of the toy MC and results The resolution of the energy calibration in real data is between 0.19 and 0.20 with a mean of 0.02. As shown in Fig. 5.6(a) the toy MC reproduces this values very well and is constant for all the values of Δ (Fig. 5.6(b)).

For event samples in the first two cases, (a) and (b), the spectral index of -2.7 and the calibration defined by Eq. (5.1.2) are taken. For each case, Δ and energy distribution, 200 ensembles of MC experiments with ≈ 350 events above 3 EeV are generated.

The reconstructed slope as a function of the Δ is illustrated in Fig. 5.7. In case of equal



(a) *linear fit* (b) *power law fit*

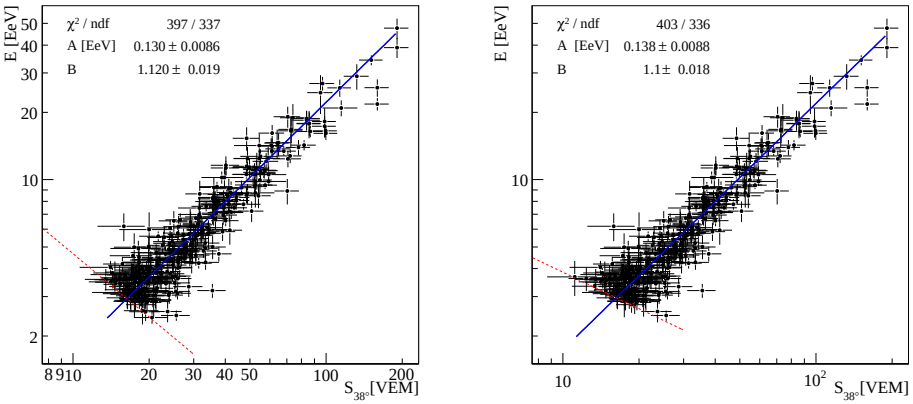
(c) *linear fit* (d) *power law fit*

Figure 5.8: The evolution of the fit parameters with Δ. The MC input parameters are denoted by horizontal lines. In case of a linear fit the most unbiased curve is given by $\Delta = -0.6$, while for a power law fit $\Delta = -0.4$.

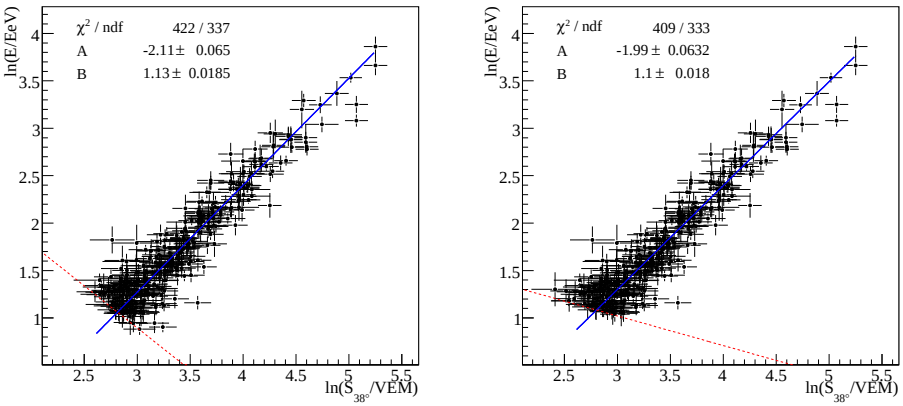
relative uncertainties (black circles) a perpendicular cut-line gives no bias (shown by the coincidence of the obtained slope with the real slope at $\Delta = 0$). When constant uncertainties (blue squares) are assumed the minimum bias is given when applying $\Delta = -0.70$ which is very close to $\Delta = -0.68$, the value given by Eq. (5.1.1). There is no difference between a realistic energy distribution, or a spectral index of -2.7.

The ratio of relative uncertainties of energy and S_{38} in the neighborhood of the cut region is 0.506, which would result according to Eq. (5.1.1) to $\Delta = -0.67$. The optimal cut deduced with the toy MC and realistic uncertainties(c) is -0.60. This shows that the energy dependency of the S(1000 m) uncertainties has a small contribution to the bias. Nevertheless, they start to be important if a Δ dependence of the deviation of the energy calibration as in data is needed.

The two fitting functions, described by Eq. (5.1.2), referred further as linear fit, and



(a) *perpendicular cut, power law fit* (b) *anti-bias cut, power law fit*

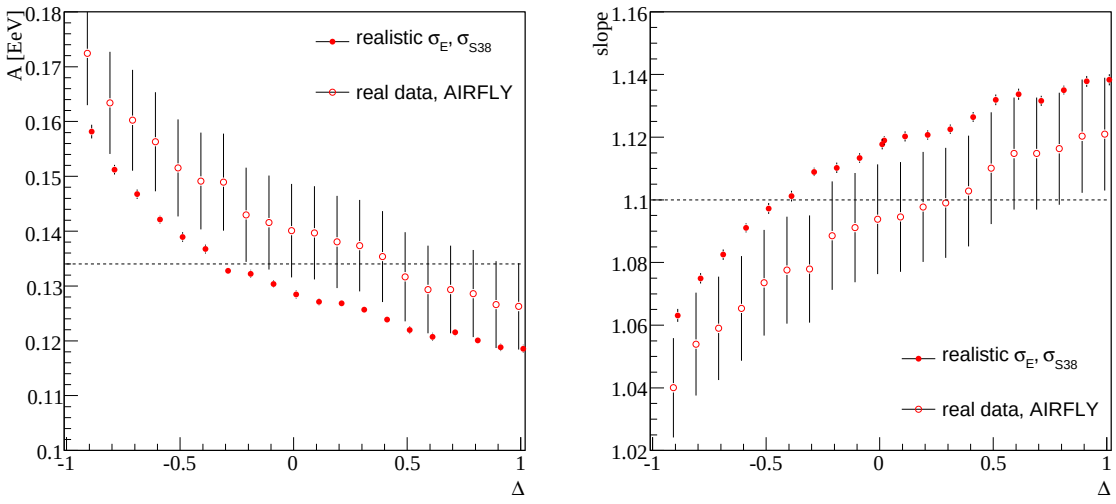


(c) *perpendicular cut, linear fit* (d) *anti-bias cut, linear fit*

Figure 5.9: The energy calibration for different lower cuts and for a power law (a, b) and a linear fit (c, d). In the case of an unbiased cut the values from the two fits agree.

Eq. (5.1.3), referred as power law fit, were used to obtain the energy calibration. For this study the fluorescence detector events are reconstructed assuming the fluorescence yield measurement from [99]. The behavior of the deviation of the energy calibration is very well reproduced by our toy MC, as shown in Fig. 5.8, for both fits.

The optimal cut is not described by the same Δ in the two cases where different uncertainty distributions are assumed: log normal or Gaussian. The obtained unbiased energy calibration is: $A = 0.138 \pm 0.009$ EeV and $B = 1.10 \pm 0.02$ (Fig. 5.9(b)) for the power law fit in agreement with $A = -1.99 \pm 0.06$ and $B = 1.10 \pm 0.02$ (Fig. 5.9(d)) for the linear fit. The perpendicular cut results are also shown in Fig. 5.9(a) and 5.9(c).



(a) *power law fit* (b) *power law fit*

Figure 5.10: AIRFLY fluorescence yield: The evolution of the fit parameters with Δ.

5.2 Results and cross-checks

In the toy MC study the hybrid reconstruction was done using the relative as well as the normalization of the fluorescence yield measured by [99]. A new measurement of the pressure dependency of the fluorescence yield has been performed recently by the AIRFLY collaboration [100]. This values of the relative fluorescence yield are utilized to deduce the final energy calibration.

Even though the final shape of the observed line for the reconstruction has a power index of ≈ 1.08 the evolution of the slope with Δ is well reproduced (Fig. 5.10(a) and 5.10(b)), showing that there is no dependency on the final reconstructed slope of the anti-bias cut.

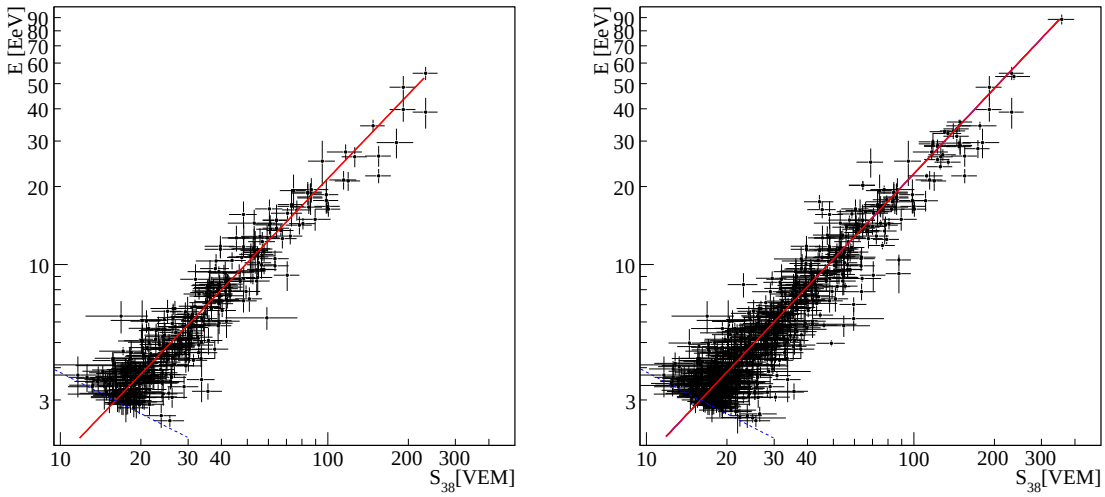
In order to increase the event statistics, especially at high energies, the events without the requirement of the measured Mie scattering parameters are used. The events without measurements were reconstructed using average aerosol profiles.

The total of events after selection is 685 and the energy calibration parameters obtained with this data set is

$$\begin{aligned} A &= 0.144 \pm 0.006(^{+0.005}_{-0.008}) \text{ [EeV]} \\ B &= 1.096 \pm 0.011(^{+0.008}_{-0.014}) \\ \rho_{AB} &= -0.91. \end{aligned} \tag{5.2.1}$$

Within the statistical uncertainties, the parameters are in good agreement with the ones stated above (Eq. (5.2.2)), therefore, the average aerosol profiles do not introduce a systematic bias.

The energy calibration (Fig. 5.11(a)) obtained only from events with measured VAOD



(a) *measured Mie parameters* (b) *all events*

Figure 5.11: AIRFLY fluorescence yield: The energy calibrations obtained with the anti-bias cut of $\Delta = -0.4$ (a) Only events with Mie parameters measured by CLF (b) all events.

parameters

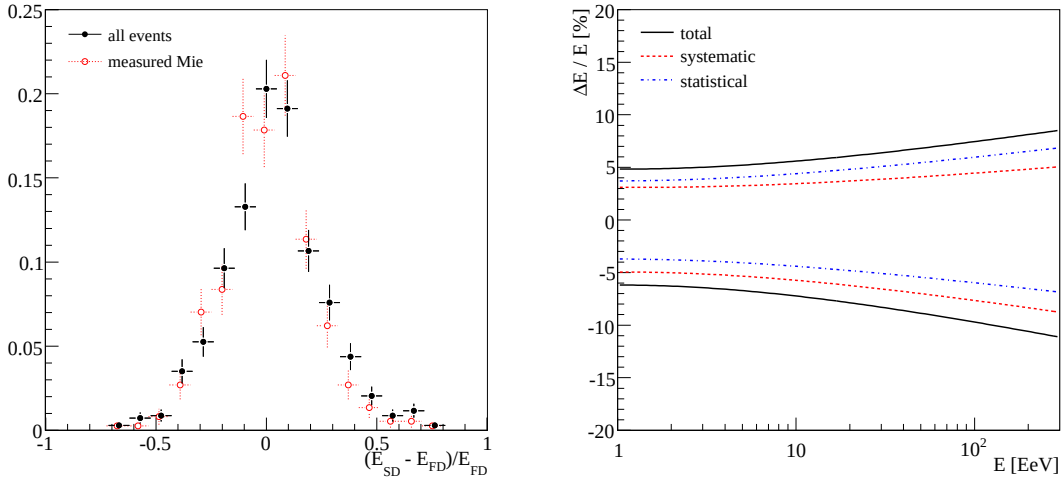
$$\begin{aligned} A &= 0.148 \pm 0.008 \text{ [EeV]} \\ B &= 1.080 \pm 0.016 \\ \rho_{AB} &= -0.99 \end{aligned} \tag{5.2.2}$$

was deduced with 370 events.

The normalized distributions of $(E_{\rm SD} - E_{\rm FD})/E_{\rm FD}$ obtained with Eq. (5.2.2) and Eq. (5.2.1) are shown in Fig. 5.12(a) with red open circles and, respectively, black dots. For the full data set the RMS is 0.22 ± 0.01 and the mean of 0.03 ± 0.01 of the last distribution is larger than the ones obtained with the reduced data set of RMS=0.20 ± 0.01 and mean=0.004 ± 0.01. The two distributions agree within the total uncertainties.

A contribution to the systematic uncertainties on the energy is estimated letting the Δ parameter vary from -0.60 to -0.20. The interval is determined by the accuracy of the toy Monte Carlo, i.e by the assumed uncertainties in the golden hybrid data energy and S(1000 m). The systematic uncertainty on the SD energy, taking into account the correlation between the parameters is shown in Fig. 5.12(b) as a function of energy.

The applied quality cuts on the hybrid data could induce systematics uncertainties. The relative difference between the measured energy, E_{FD}, and the reconstructed energy as a function of the variables used to define the rejection criteria are illustrated in Fig. 5.15 and Fig. 5.16. In the left panels all the events above 2 EeV were considered without applying any quality cuts. The arrows indicate which part of the distributions are not considered in energy calibration. In right panels the same distributions are shown after applying all the quality cuts, for events with energies larger than 2 EeV.



(a) *energy calibration resolution* (b) *energy calibration uncertainties*

Figure 5.12: (a) The energy resolution. (b) The contribution to the energy systematics uncertainties of the energy calibration.

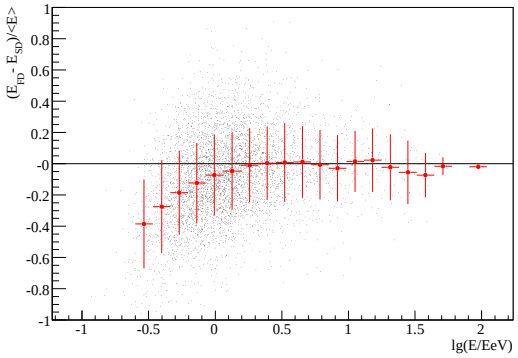
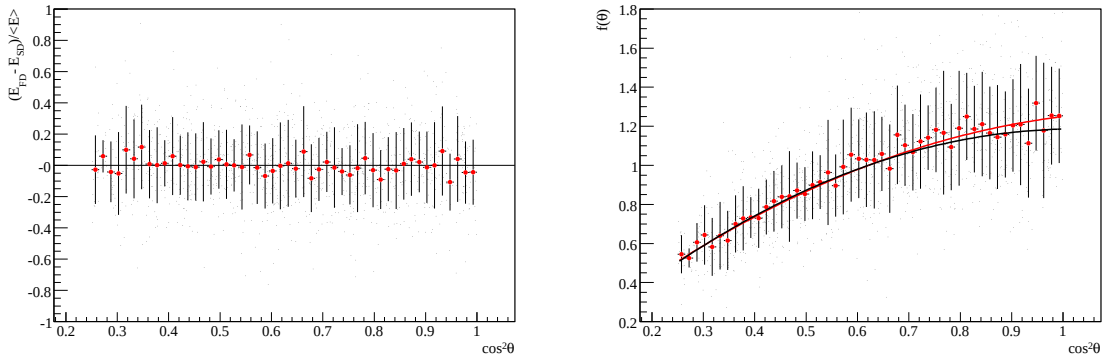


Figure 5.13: Energy dependency of the energy estimation. Under 1.6 EeV the upward fluctuations caused by the trigger efficiency of the SD is obvious.

No dependency is observed on the variables, even in the neighborhood of the rejection value. Releasing the criteria or making it harder induces only statistical fluctuations and no systematic effects on energy.

Another important cut is the minimum energy given by the trigger probability for the surface detector. The energy above which we assume that the trigger efficiency of SD is 1 is ≈ 3 EeV. The relative difference between estimated and reconstructed energies as a function of energy is illustrated in Fig. 5.13. The upward fluctuations from the surface detector appear below $\lg(E/\mathrm{EeV}) = 0.2$ (1.6 EeV), which is 40% lower than our chosen energy.

The relative energy difference is constant with zenith angle as shown in Fig. 5.14(a) and the S(1000 m) attenuation factor $f(\theta)$ obtained by inverting the energy conversion function is



(a) *zenith angle dependency* (b) *attenuation curve*

Figure 5.14: (a) Energy dependence on the zenith angle. (b) The attenuation curve from golden hybrid data (red line). Black line shows the function obtained from the SD data only (Eq. (4.2.5)).

shown in Fig 5.14(b). Superimposed in black is the attenuation curve deduced with the SD data only, Eq. (4.2.5).

A fit of the profile of the attenuation function described by Eq. (4.2.3) (red line) gives $a = 0.99 \pm 0.12$ and $b = -0.88 \pm 0.40$ in agreement within the statistical uncertainties of the attenuation curve deduced with the constant intensity method (interrupted line). The small difference for vertical showers arises from slightly different zenith angle distribution of golden hybrid and SD data.

5.3 Direct energy calibration

The energy dependence on S(1000 m) and zenith angle can be deduced directly from golden hybrid events without any constant intensity assumption. The knowledge of the functional form of the dependency of S(1000 m) with the zenith angle and of the relation between energy and S_{38} enables the construction of a simple χ^2. The primary particle energy is related as a power law to S_{38}.

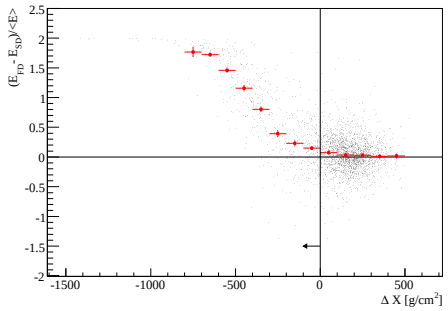
Within this framework the uncertainties in all variables are straight forward to be included: The reconstruction and intrinsic uncertainties of the S(1000 m), the statistical uncertainties of the zenith angle and of the energy. The attenuation curve is described by

$$S_{\theta_{\mathrm{ref}}} = \frac{S(1000\,\mathrm{m})}{f(\theta)} \tag{5.3.1}$$

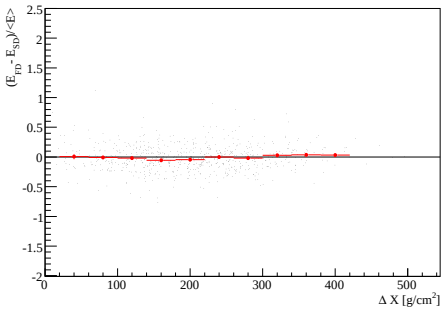
$$E = A \cdot S_{\theta_{\mathrm{ref}}}^{B} = A \cdot \left(\frac{S(1000\,\mathrm{m})}{f(\theta)}\right)^{B} = E_{\mathrm{SD}}. \tag{5.3.2}$$

The function to be minimized is

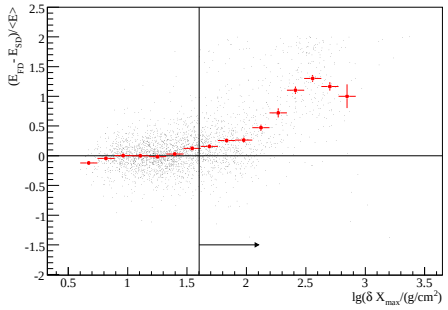
$$\chi^2 = \sum_i \frac{(E^i_{\mathrm{FD}} - E^i_{\mathrm{SD}}(S^i(1000\,\mathrm{m}), \theta^i))^2}{\sigma^2_{E^i_{\mathrm{FD}}} + \sigma^2_{E^i_{\mathrm{SD}}}} \tag{5.3.3}$$



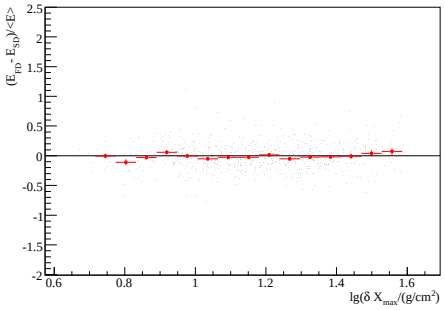
(a) *the smallest difference between* $X_{\max}$ *and the borders of the profile*



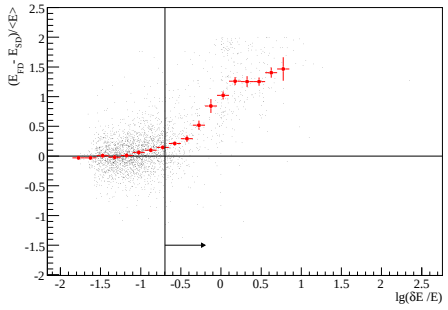
(b) *the smallest difference between* $X_{\max}$ *and the borders of the profile*



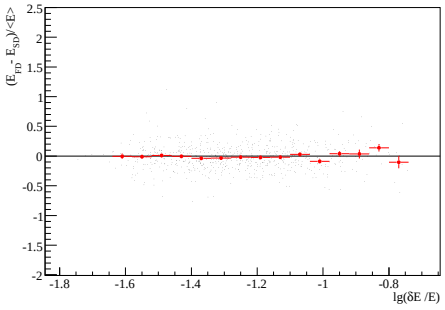
(c) *uncertainty of* $X_{\max}$



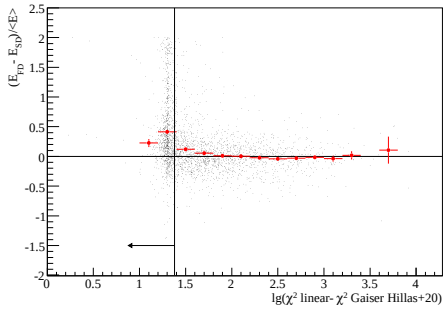
(d) *uncertainty of* $X_{\max}$



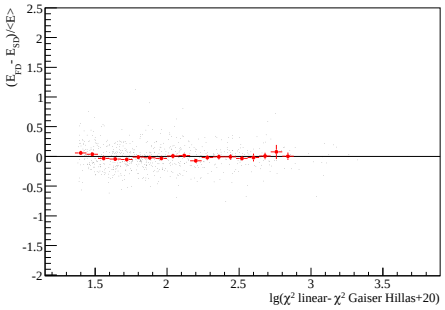
(e) *energy uncertainty*



(f) *energy uncertainty*



(g) χ^2 *diff. between linear and GH fits*



(h) χ^2 *diff. between linear and GH fits*

Figure 5.15: Test for the most stringent hybrid quality cuts. (left) before applying any cuts. The vertical line and the arrow indicate the cut value and the direction. (right) Cleaned data.

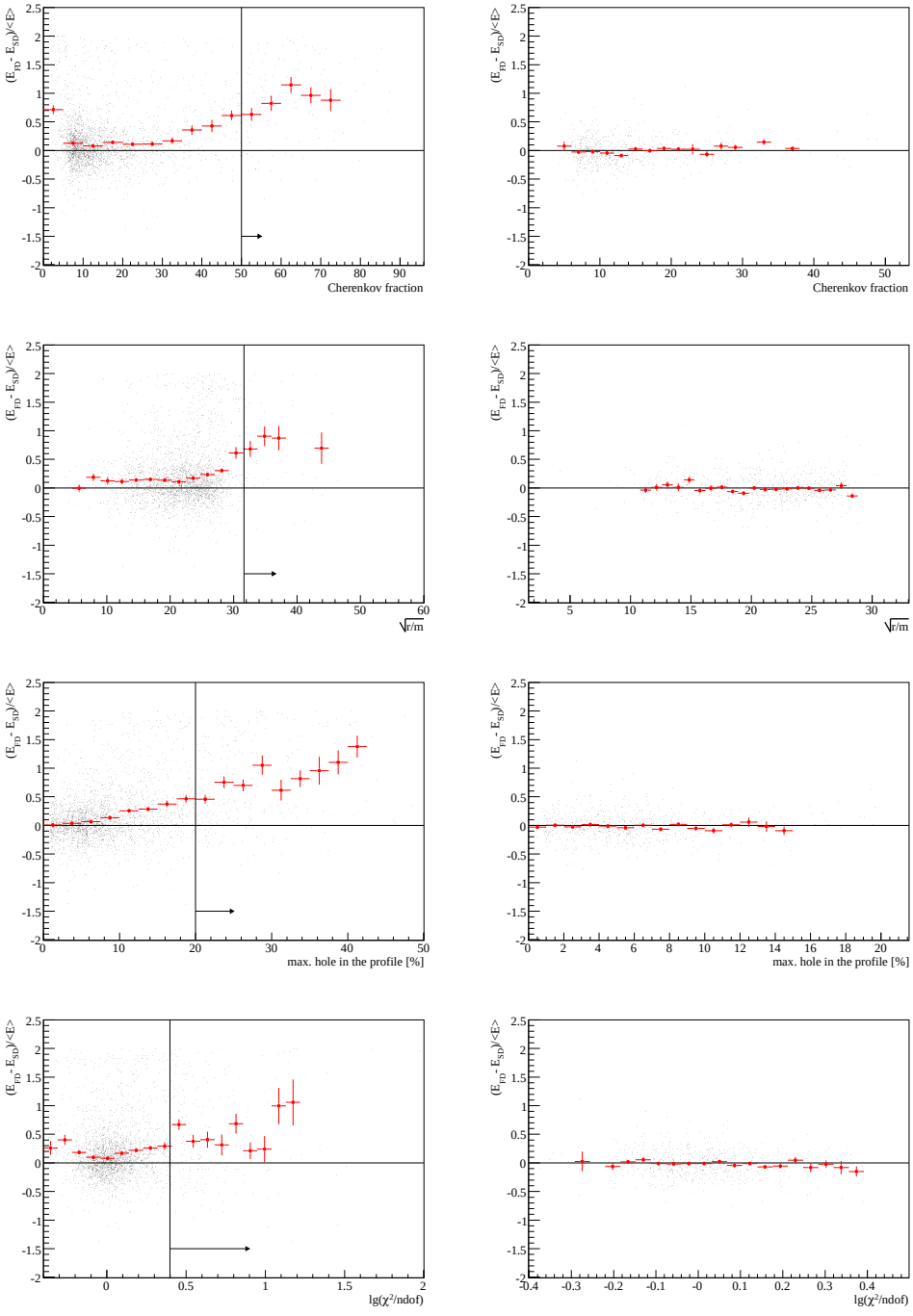
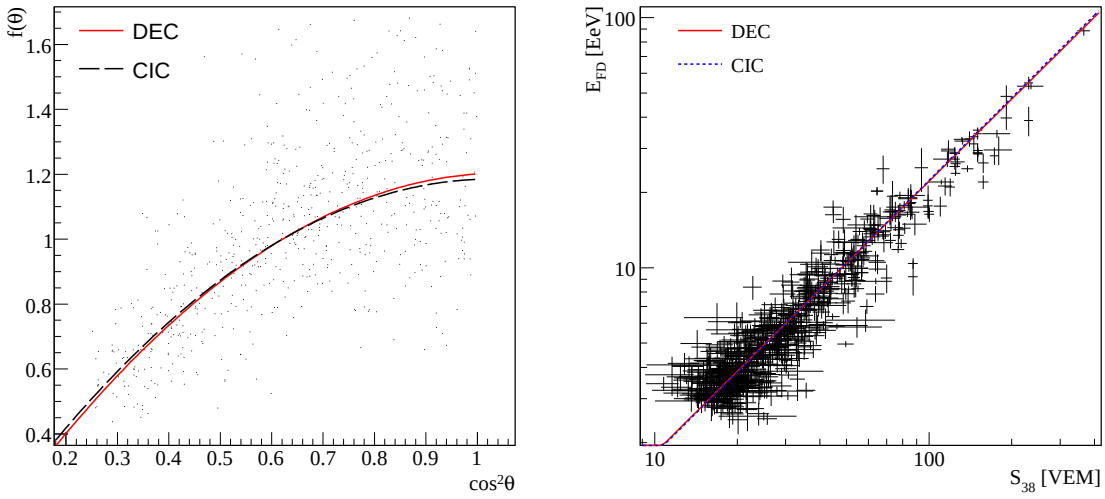


Figure 5.16: Test for the less important hybrid quality cuts. (left) before applying any cuts. The vertical line and the arrow indicate the cut value and the direction. (right) Cleaned data.



(a) *attenuation curve* (b) *energy-S_{38} dependency*

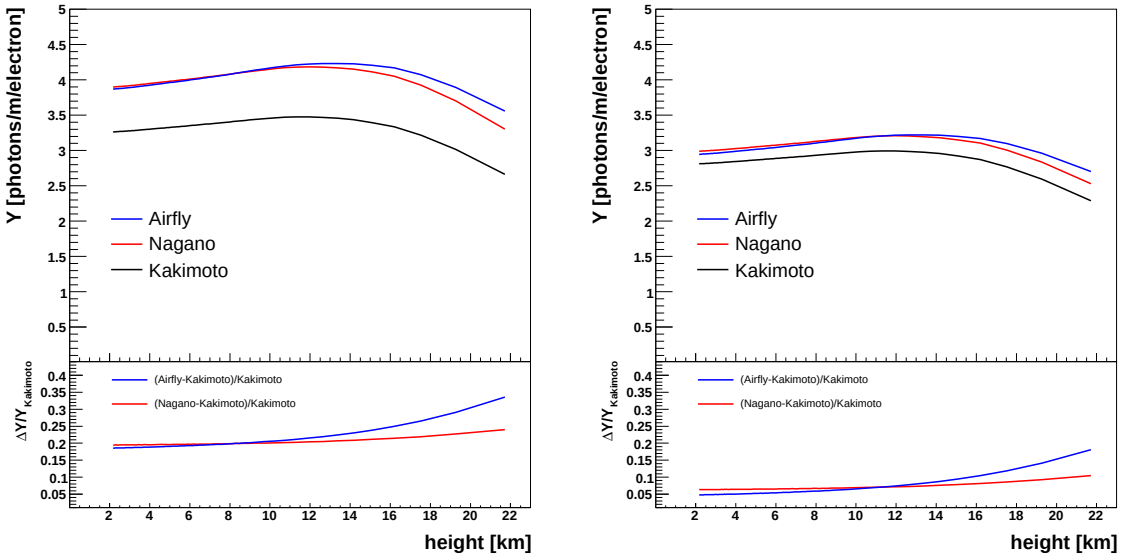
Figure 5.17: Direct energy calibration (a) The attenuation curve deduced without the constant intensity assumption. The one obtained with the constant intensity method is also shown for comparison. (Eq. (4.2.5)). (b) The energy-S_{38} dependency.

to deduce the parameters from the attenuation curve and the energy calibration. The functional form of the attenuation is described, similar to the constant intensity method, by a second degree polynomial in $\cos^2\theta$ (Eq. (4.2.3)) and is shown in Fig. 5.17(a). There is a very good agreement to the attenuation deduced by the constant intensity method (shown in black). The parameters are $a = 0.96 \pm 0.03$, $b = -1.104 - 0.137$, $A = 0.148 \pm 0.002$ and $B = 1.088 \pm 0.011$. The corresponding energy calibration is shown in Fig. 5.17(b) and the spread of the distribution $(E_{\rm SD} - E_{\rm FD})/E_{\rm FD}$ is 0.207 ± 0.008.

Systematic uncertainties of the estimated energy

The energy calibration of the surface detector inherits all systematic uncertainties of fluorescence detector energy measurement [147]. The absolute energy scale of FD is determined by the absolute fluorescence yield measured by [99]. The difference to the relative calibration of fluorescence yield measured in [101] is shown in Fig. 5.18. Even though there is a 20% difference between the measurements, it translates in a 5% difference in energy. The wavelength dependence of the FD response happens to filter the spectral lines with largest disagreement. The quoted uncertainties for the absolute measurement is 14%.

The pressure dependency of the fluorescence yield has a systematic uncertainty of 1%. It depends on the amount of water vapor present in the atmosphere, estimated at the observatory site to 5%. There are no measurements published concerning temperature dependence of the collisional cross section. Assuming a dependency of $\sqrt{T}$, in the hypothesis that the collisional cross section is independent of temperature, an uncertainty of 5% is assigned. The uncertainties due to the profile reconstruction are estimated by comparing two reconstruction algorithms of light collection at the camera [148, 149] contributing to the total systematic



(a) *fluorescence yield* (b) *detector response folded*

Figure 5.18: (a) The fluorescence yield as a function of the height in the atmosphere. (b) The fluorescence yield after the efficiency of the detector has been folded [146].

Table 5.2. The energy systematic uncertainties

source	$\Delta E/E[\%]$
Fluorescence yield: absolute calibration	14
Fluorescence yield: pressure dependence	1
Fluorescence yield: humidity dependence	1
Fluorescence yield: temperature dependence	5
FD calibration: absolute calibration	11
FD calibration: wavelength dependence	3
atmosphere: Rayleigh	1
atmosphere: wavelength dependence of aerosol scattering	1
atmosphere: aerosol phase function	1
FD reconstruction: method	10
FD reconstruction: invisible energy	4

uncertainties with 10%. The systematic uncertainties related to FD calibration, together with fluorescence yield uncertainties are summarized in Table 5.2. The total systematic uncertainty is obtained by adding the individual contributions in quadrature, which leads to $\sigma_{sys}^{E_{\rm FD}} = 22\%$.

Chapter 6

Monte Carlo Studies

As explained in the last chapter, due to the hybrid approach of the Pierre Auger Observatory the energy spectrum of ultra high energy cosmic rays can be measured without the need of air shower simulations. However, it is interesting to compare the predictions of these simulations to data in order to test the validity of models of high energy hadronic interactions.

Moreover, with the help of a detailed simulation of the detector response it is possible to deduce the energy resolution of the apparatus, which is needed to correct the measurement for detector effects.

6.1 Air shower and detector simulation

Air shower simulations The Monte Carlo framework used for in this work is COsmic Ray SImulations for KAscade (CORSIKA) [77]. It is a four-dimensional particle transport code that handles interactions and decays of particles in the atmosphere. As primary particles proton and iron were considered, generated with isotropic arrival directions within a zenith angle range from 0 to 65° and an energy distribution following a power law with a spectral index of -3 below 3 EeV and -2.5 above this energy.

The number of particles in an extensive air shower grows proportionally to the energy and so does the computational time. In order to speed up the simulation a procedure called thinning is applied: Similar energetic particles are merged and only one is tracked further with a weight proportional to the represented number of particles. This process induces artificial fluctuations by removing information from the air shower development. The maximum weight allowed for the produced simulations was 20000. The chosen thinning of 10^{-6} is a good compromise between computational time and air showers simulation quality [150].

The output of the CORSIKA program is a two dimensional distribution of the weighted particle positions and their momenta at a certain height. For the case of the used library of simulations was set at the mean Pierre Auger Observatory altitude of 1452 m (a.s.l). The reversed process of obtaining particles from the weighted one within a certain volume, called *un-thinning* is limited by the assumed spread area and by the imperfect knowledge of the energy distribution of the particles in the bunch [151–153].

The description of the hadronic interactions is divided in two parts, low and high energetic hadron interactions with a transition energy of 200 GeV. All high energy hadronic interactions are simulated with the QGSJetII model [73,154] and at low energies the FLUKA model [155] is used. The electromagnetic component is described by EGS4 [156] implemented in CORSIKA. The interaction of particles at the highest energies is the largest source of uncertainty in the description of the air showers [157], mainly due to the poor knowledge of the corresponding hadronic cross section, diffractive interactions, elasticity (the fraction of energy carried away by the leading particle) or secondary particle multiplicity. Even though the chosen model cannot describe precisely some of the above mentioned quantities, it can give a good description of the existing air shower data [35, 44, 158–162].

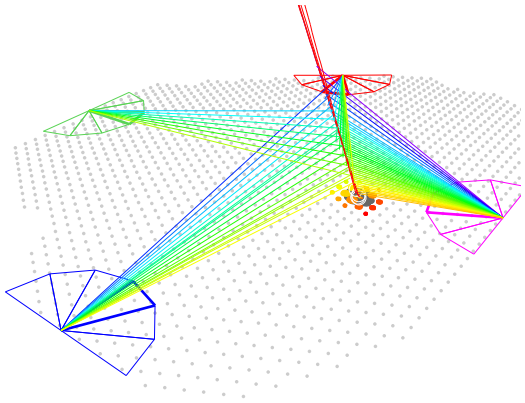
Detector simulation The detector simulation is included in the same framework, $\overline{\text{Off}}\underline{\text{line}}$, as the data reconstruction algorithm. This makes it possible to reconstruct of simulated and measured events in the same way. In the FD simulation [163, 164] the first step is to generate the light produced by air shower particles and to propagate it through the atmosphere to obtain the light received at the diaphragm.

The detector response is simulated in detail, from ray-tracing of generated photons on diaphragm through the telescope optics to the camera response. Finally the FD trigger is reproduced. The background light received permanently by the telescopes from the atmosphere, the moon, the milky way, etc [165] is calculated from the recorded variance of the ADC traces after subtracting the electronic noise. The fluorescence detector simulation does not rely on GEANT4 code, allowing the production of large sets of events, especially if CONEX [166] is used for the simulation of air showers (one dimensional simulation code).

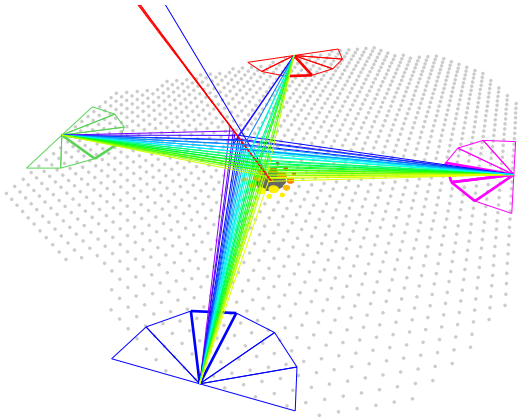
This is not the case for SD simulations, where the lateral distribution of particles is a need. First the CORSIKA program is time consuming. Secondly the tank behavior to passage of muons, electrons and photons from the interaction of particles in water, optical transmission of the materials involved and the PMT response has to be understood and implemented. This is possible with the GEANT4 [167, 168] software package, which provides an abundant set of physics processes to handle interactions of particles with matter across a wide energy range.

There are a lot of studies on the simulation quality of the tank response [169–176] in case of the Auger surface detector. The tank response to vertical and omni-directional muons can be compared with data obtained from the calibration [177]. The distribution of charges of atmospheric muons has a peak that is well correlated with the VEM charge and is the basis for the SD calibration. The ratio of the peak to the VEM charge is the benchmark for the tank simulations. Using a realistic spectrum of multi-directional electrons, muons and photons from [178] the position of the peak is found to be at 1.09 in a very good agreement with the data. Good agreement is also found for inclined muons and electrons from muon decays and for different water levels in the tank.

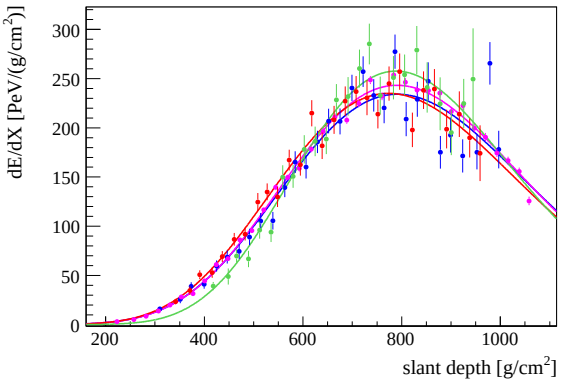
Simulation examples In Fig. 6.1 and 6.2 the output after the detector simulation for showers initiated by a proton with energy 160 EeV and by an iron nucleus with 163 EeV is shown. The showers have triggered all the four eyes of the FD, the longitudinal profiles from individual telescopes being in a very good agreement. The reconstructed energy is also compatible with the generated energy.



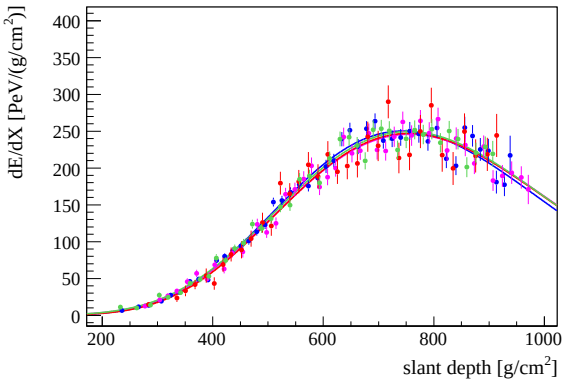
(a) *proton*



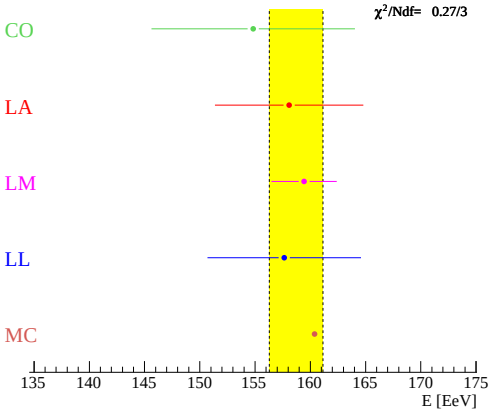
(b) *iron*



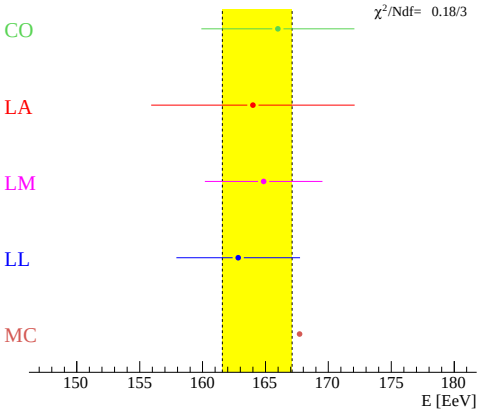
(c) *profiles proton*



(d) *profiles iron*

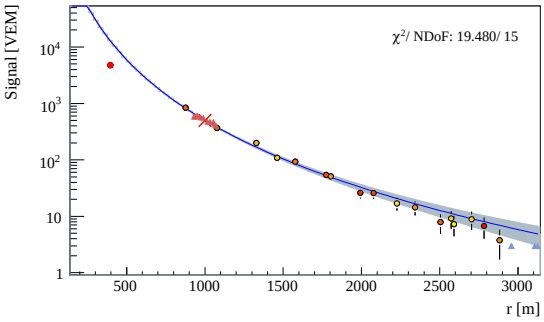


(e) *energy reconstruction proton*

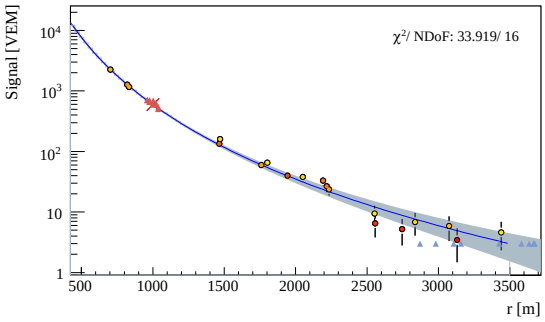


(f) *energy reconstruction iron*

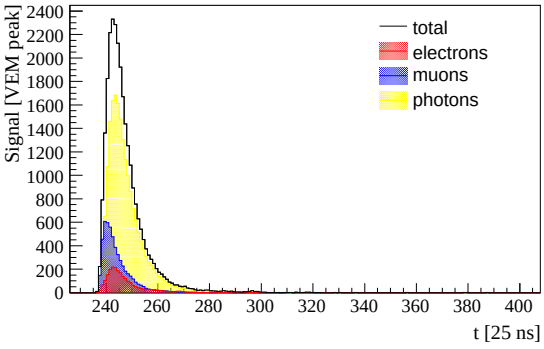
Figure 6.1: Event simulation: FD output.



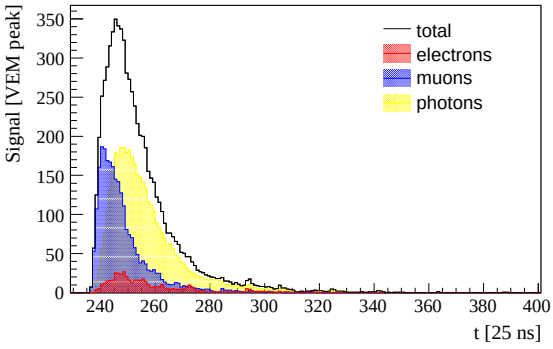
(a) *LDF proton*



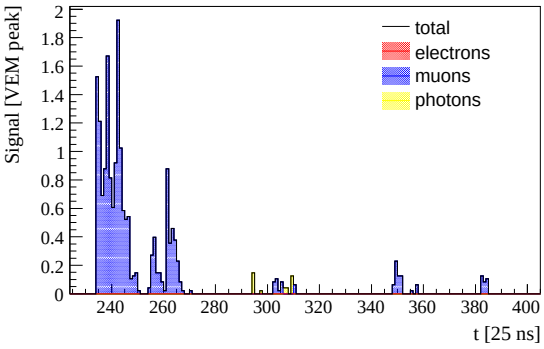
(b) *LDF iron*



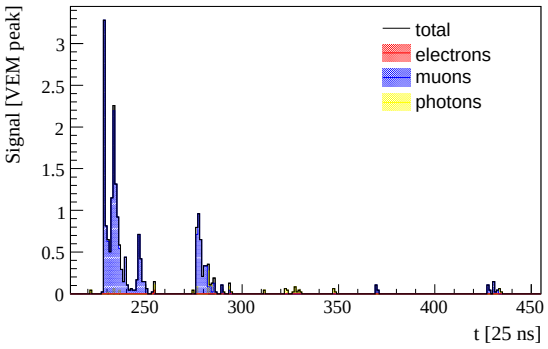
(c) *close trace proton*



(d) *close trace iron*



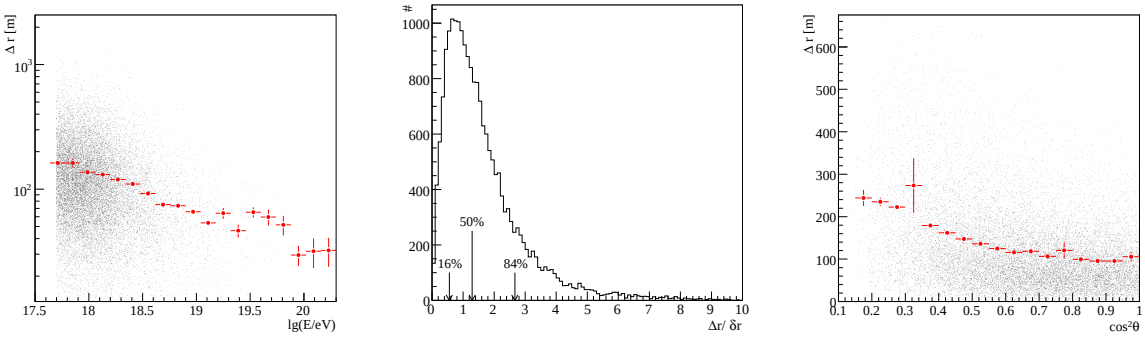
(e) *far trace proton*



(f) *far trace iron*

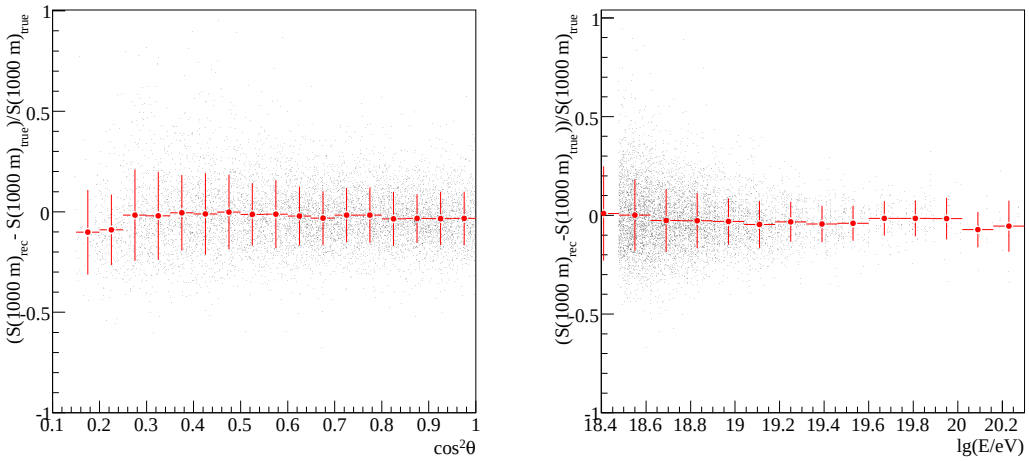
Figure 6.2: Event simulation: SD output.

Lateral distributions are shown in Fig. 6.2(a) and 6.2(b). A ring of stations (appearing in the same figures with red triangles) has been simulated at a distance of 1000 m from the shower center to obtain a good estimation of S(1000 m). They were not used in the LDF reconstruction. The signal from the closest station to the shower core has a large electromagnetic contribution compared to the muonic part for the proton induced shower (Fig. 6.2(c)) and at a distance 150 m further from the core iron has almost equal contributions 6.2(d). At large distances the signal is produced mostly by muons (Fig. 6.2(e)



(a) Δr *versus energy* (b) *pull distribution* (c) Δr *versus zenith*

Figure 6.3: (a) Distance between the true core and reconstructed core as a function of energy. (b) The same distance divided by the estimated uncertainty. The quantiles of the distribution at 16%, 50% and 84% are indicated with arrows. (d) the same distance as a function of the zenith angle.



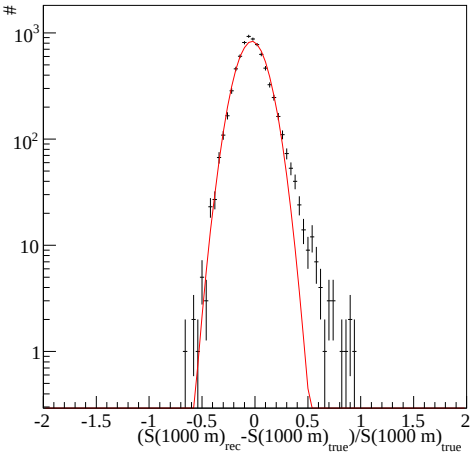
(a) *relative difference. vs. zenith* (b) *relative difference. vs. energy*

Figure 6.4: (a) Relative difference in S(1000 m) as a function of zenith angle. (b) Relative difference in S(1000 m) as a function of the simulated primary energy.

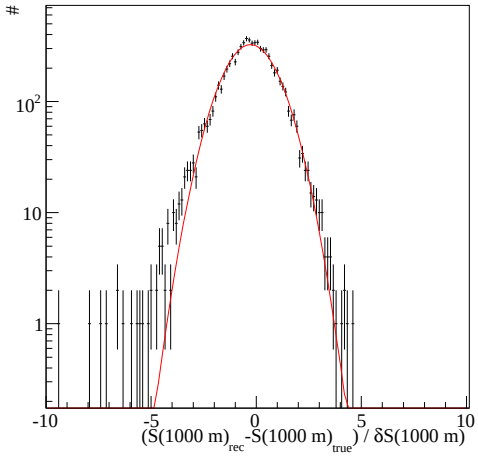
and 6.2(f)).

6.2 Accuracy of the reconstruction

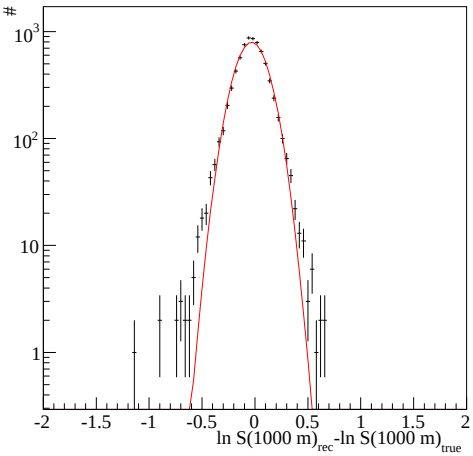
In the reconstruction of the lateral distribution function (LDF) the free minimization parameters are the position of the shower core on ground and S(1000 m). A wrong reconstruction of the core location would lead to an uncertain estimation of S(1000 m). The shower asymmetries introduce a bias on the reconstructed core [179]. The core is pushed towards the upstream part of the shower, i.e. the early part that has higher signal density.



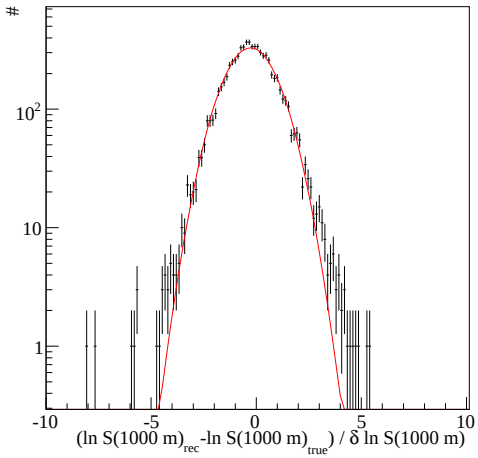
(a) $\chi^2/\mathrm{ndof} = 5.56$,
$\mu = -0.029 \pm 0.002$,
$\sigma = 0.136 \pm 0.001$



(b) $\chi^2/\mathrm{ndof} = 1.90$,
$\mu = -0.26 \pm 0.01$,
$\sigma = 1.17 \pm 0.012$



(c) $\chi^2/\mathrm{ndof} = 4.01$,
$\mu = -0.031 \pm 0.002$,
$\sigma = 0.143 \pm 0.002$



(d) $\chi^2/\mathrm{ndof} = 1.72$,
$\mu = -0.26 \pm 0.01$,
$\sigma = 1.16 \pm 0.012$

Figure 6.5: (a) S(1000 m) resolution assuming normal distributed uncertainties. (b),(d) pull distributions. (c) S(1000 m) resolution assuming log-normal distributed uncertainties. The spread of 1.16 obtained from a Gaussian fit demonstrates that the uncertainties of S(1000 m) are well estimated.

This is due to cylindrical symmetry assumed in the reconstruction. The induced uncertainty on S(1000 m) due to the inaccuracy of the shower core is less than 4%.

The distance between the true core location and the reconstructed one for the simulated showers as a function of simulated energy is shown in Fig 6.3(a). In the energy range $\lg(E/\mathrm{eV}) > 18.4$ the mean distance is less than 100 m. It evolves with zenith angle (Fig 6.3(c)), the core location for vertical shower (which have less asymmetries) is better reconstructed

than for inclined showers.

In Fig 6.3(b) the same distance divided by the estimated error from the minimization is shown. The 50% quantile is 1.29, the median of the distribution is at 0.94, with the mean at 1.69. The 16% and 84% quantiles are 0.55 and 2.67. The core location uncertainties are well estimated, so their contribution to S(1000 m) are well taken into account.

The statistical uncertainties in S(1000 m) have two contributions, one given by the fit of LDF and the other arising from fixing the slope parameter, β, in the reconstruction. In the following S(1000 m) uncertainties are taken as the quadratic sum of them.

For each simulated shower the $S(1000\,\mathrm{m})_{\mathrm{true}}$ is estimated from the ring of stations that were simulated at 1000 m from the shower core in shower plane coordinates. It is obtained as the mean of the signals from 4 stations that have the azimuth angle closest to 90°, where the asymmetries of the shower are minimal.

The relative difference between the reconstructed S(1000 m) and the true one as a function of zenith angle is shown in Fig. 6.4(a). A small deviation of about 5% is observed for inclined showers which might be also caused by a biased estimation of the true S(1000 m).

In the next plots, data are restricted to simulated events with $\cos^2\theta > 0.35$ (54°). There is no clear dependency of the bias in S(1000 m) as a function of energy. It is compensated therefore by the energy calibration. The spread of S(1000 m) is decreasing with increasing energy from $\approx 16\%$ to $\approx 10\%$.

To decide if the distribution of the uncertainty of S(1000 m) is normal or log-normal, we compare the results shown in Fig. 6.5(a) and 6.5(c). The distribution of the normal spread shows a non-Gaussian tail in the region of large values. Reduced χ^2 values of a Gaussian fit for the log-normal distribution of 4.01 and 5.56 for the normal distribution favor the log-normal distributed uncertainties.

The pull distributions are shown in Fig. 6.5(b) and 6.5(d) containing the uncertainties of the reconstructed S(1000 m) as well as of the estimated true S(1000 m). The uncertainties of S(1000 m) are well estimated, proved by the spread of the pull distribution of 1.16 ± 0.012.

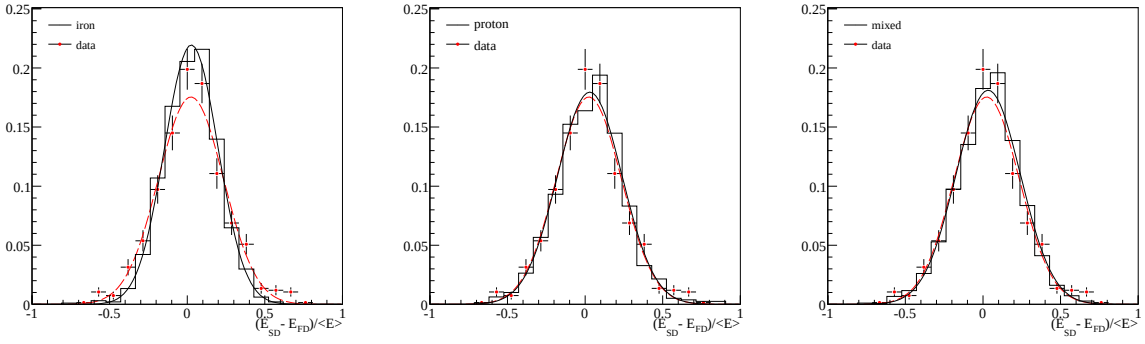
6.3 Energy calibration and attenuation curve

Direct Energy calibration In the performed detector simulation the atmosphere was assumed constant in the horizontal direction and also the temperature variation of the atmosphere was neglected. The standard parameterization for the US atmosphere was used.

Reconstructing showers with different sets of aerosol measurements it is possible to estimate the effect on energy of horizontal non-uniformity of the atmosphere. It was shown in [180] that this effect contributes 7% to energy fluctuations.

The set of events used in the energy calibration consists of events with and without the VAOD parameters measurement. The use of average aerosol profiles introduces another 9% fluctuations on the hybrid energy [181] and are considered for the correspondent fraction of events (0.46).

Temperature changes induce variations of the S(1000 m), due to the dependence of the lateral distribution function on the density of the atmosphere [182]. Looking at the difference in the



cm

(a) *iron* (b) *proton* (c) *mixed*

Figure 6.6: Energy resolution from simulations.

event rate as a function of temperature in data [183], the induced fluctuations are 6%. This contribution has been added to S(1000 m).

The energy calibration for simulations is obtained applying the direct energy calibration (Section 5) from which the attenuation curve determination is not limited by the number of simulated events as in the case of a constant intensity cut method. Proton, iron and mixed composition (defined as an equal part combination of iron and proton) are analyzed.

The normalized distributions of $(E_{\rm SD} - E_{\rm FD})/\langle E\rangle$ are shown in Fig. 6.6. The proton and mixed composition spreads of 20.8% are in a very good agreement with data and the spread of the iron distribution is lower, 17.2%.

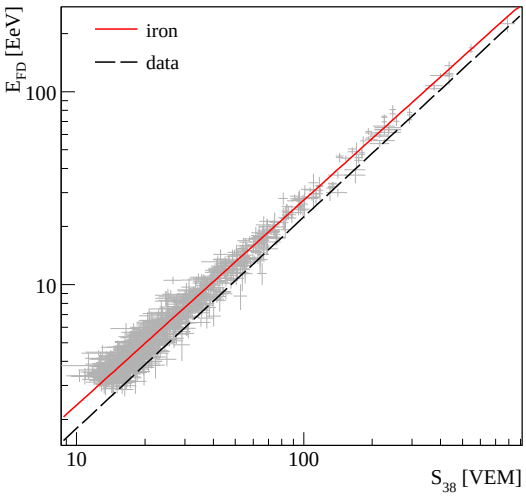
A difference in energy assignment (Fig. 6.7) between Monte Carlo and data, at an order of 50% in S_{38} for protons and 20% for iron is seen.

The energy conversion for proton is different from that for iron, the ratio of the energies is $E_{\rm iron}/E_{\rm proton} = 0.85$. The correction factors to obtain the same energy for the same S_{38} as in data are 0.645 for proton and 0.745 for iron induced showers. These factors are used to obtain the surface detector trigger efficiency uncertainty, described later on (Section 7.3).

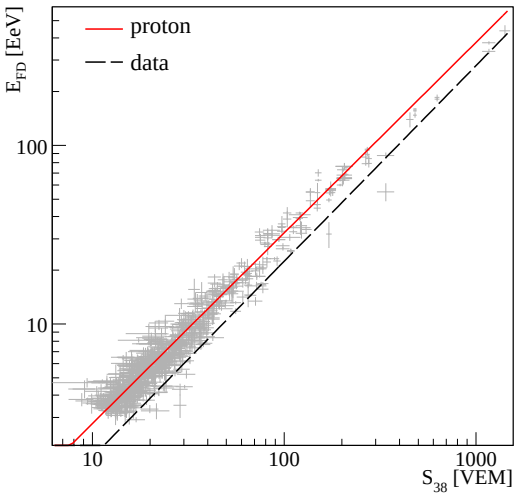
The attenuation curves, shown in Fig. 6.8 normalized at 38°, deviate from the data especially at high zenith angles where the muonic content is higher. The obtained values are given in Table. 6.1 together with the parameters for converting S_{38} to energy.

Attenuation curve The generated spectrum distribution and the flatness in $\cos^2\theta$ allows to apply the constant intensity cut method to obtain the attenuation curve. Previous studies have shown that there is a difference in the number of muons that reach the ground between simulations and data [184–186]. From Auger data it was inferred [161] that the number of muons measured in data is about 1.5 times larger than that predicted by QGSJET II for proton showers and 1.39 for iron.

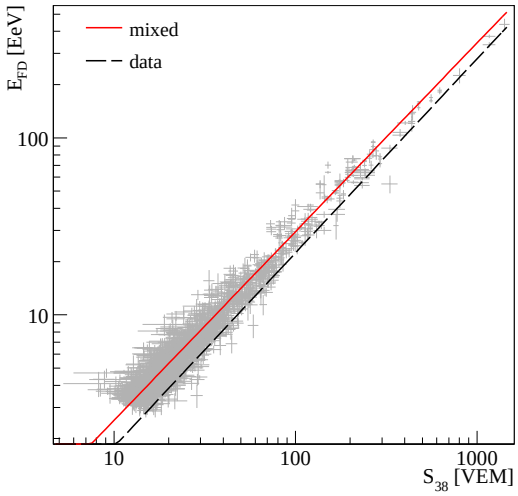
As no correction for the missing number of muons is applied different attenuation curves are expected. The ratio of muonic to electromagnetic component on the particles that reach the ground is higher for iron, as it develops higher in the atmosphere. The deviations of the



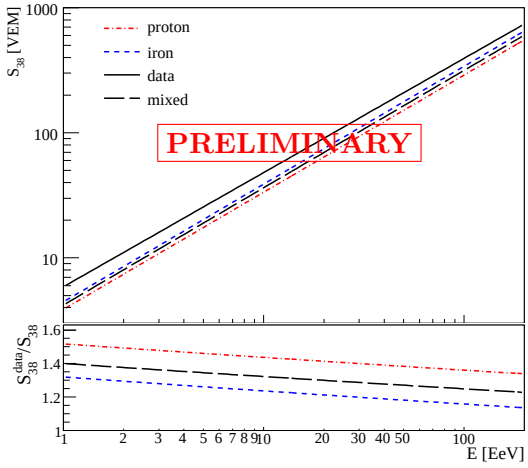
(a) *iron*



(b) *proton*



(c) *mixed*



(d) *difference to real data*

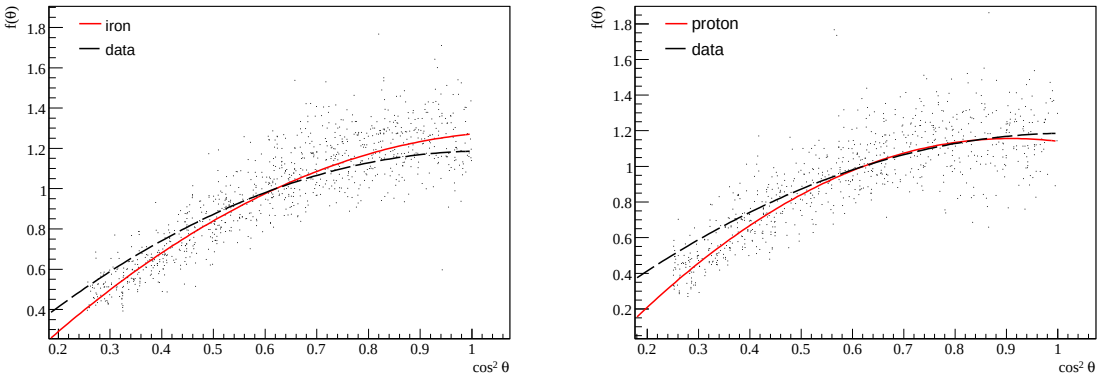
Figure 6.7: Energy calibration for simulations with different primary composition assumptions: iron(a), proton(b) and mixed(50-50%)(c).(d) Comparison with real data. Lower panel the ratio between S_{38} deduced from data and the one deduced from simulations.

shape of the attenuation curve from data should be higher than for proton. This is shown in Fig. 6.9. The parameters are deduced assuming no change of the attenuation curve with energy, i.e. making a simultaneous fit at all zenith angles for the S(1000 m) spectra (for details see Section 4).

If the observed difference between proton and iron is not an artefact of the hadronic interaction models then the observed energy dependence of the attenuation curve (although not statistically significant) for the data can be assumed to be change in the composition of

Table 6.1. The energy calibration and attenuation parameters for simulations obtained with the direct energy calibration and as defined by Eq. (5.3.2). The spread of the distribution $(E_{SD} - E_{FD})/\langle E \rangle$, before ($\sigma$) and after($\sigma$ smeared) introducing the missing fluctuations in the simulations are given.

primary	iron	proton	mixed	data
A [EeV]	0.206 ± 0.002	0.237 ± 0.002	0.217 ± 0.001	0.144 ± 0.006
B	1.067 ± 0.007	1.062 ± 0.007	1.066 ± 0.005	1.096 ± 0.011
a	1.17 ± 0.02	1.09 ± 0.03	1.17 ± 0.02	0.92 ± 0.05
b	-1.21 ± 0.10	-1.88 ± 0.10	-1.60 ± 0.07	-1.13 ± 0.26
E/E_{data}	0.745	0.645	-	-
σ [%]	13.4 ± 0.3	18.8 ± 0.4	18.7 ± 0.3	-
σ smeared [%]	17.2 ± 0.4	20.8 ± 0.6	20.8 ± 0.4	20.9 ± 0.4



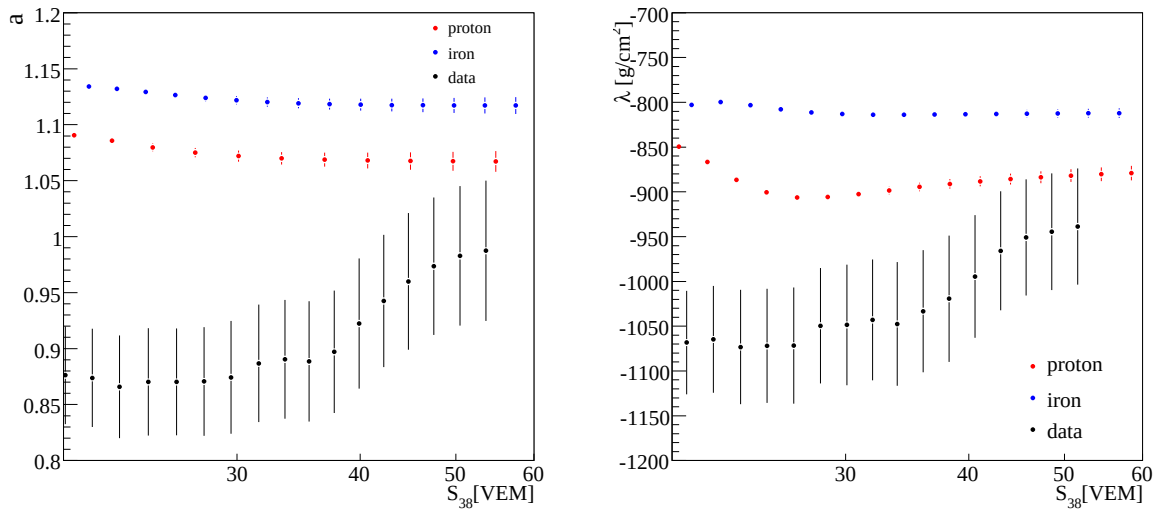
(a) *iron* (b) *proton*

Figure 6.8: S(1000 m) attenuation obtained with the direct energy calibration from simulations.

the cosmic ray flux between S_{38} in 40 VEM and 60 VEM corresponding to an energy of about $8 - 12$ EeV.

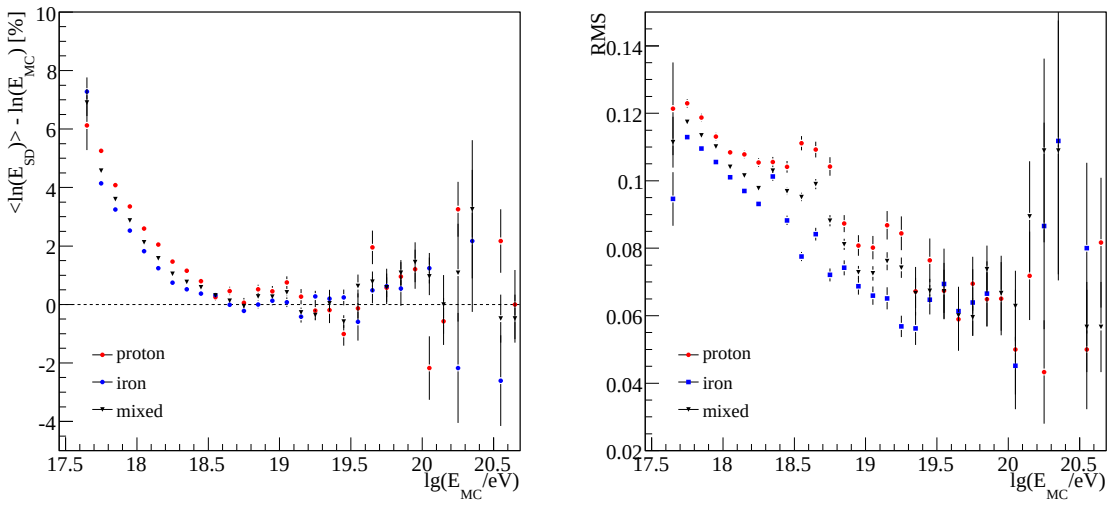
6.4 Migration matrix

To determine the detector resolution effects on the cosmic ray flux, the detector response for SD has to be known. Applying exactly the same data analysis for simulations as for data the migration matrix is built. The observed difference between simulations and data, introduce inaccuracies in the estimation, but the majority (e.g. attenuation curve differences) are compensated by the energy calibration. This can be seen in the good agreement of the energy calibration residuals (Fig. 6.6). The results from simulations will be used only as a brief estimation of the true flux.



(a) *parameter a* (b) *parameter λ*

Figure 6.9: The attenuation curve parameters (see Eq. (4.2.3) and (4.2.4)) as a function of S_{38} for data and simulations obtained with the constant intensity method.

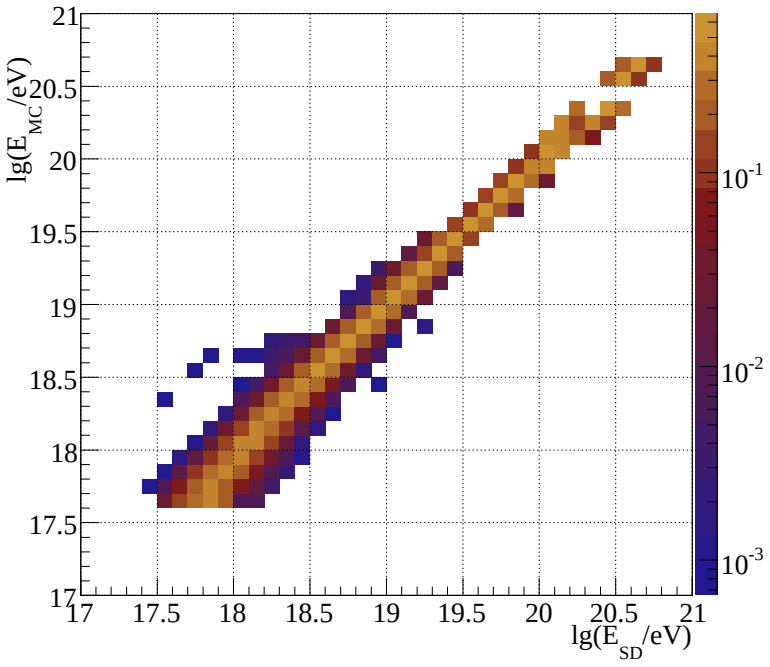


(a) *bias* (b) *RMS*

Figure 6.10: RMS and bias for SD simulations.

The SD energy resolution as a function of the generated energy is illustrated in Fig. 6.10(b). For $\lg(E_{\rm MC}) < 18.5$ the spread is larger that 0.1, the bin width in the energy spectrum. Therefore migrations of events from bin to bin are expected to be important.

The mean of $\Delta E/E$ is compatible with zero value (Fig. 6.10(a)) for lg(E) greater than 18.5, and in the lower region it has positive values reflecting the enhanced sensitivity of the SD



(a) *bias*

Figure 6.11: SD migration matrix for mixed composition.

trigger to the upward fluctuations of the signal. The fluctuations for iron are $\approx 5\%$ smaller than for proton and they vary from 12% at low energies to 7% at the highest energies.

The matrix elements are built as the probability for measuring an energy E_{SD} with the detector, given a true energy E_{MC} for all events with a T5 trigger (i.e. the trigger probability is *not* part of the matrix)

$$R_{ij} = P(E_i^{\mathrm{SD}}|E_j^{\mathrm{MC}}) = \frac{n_i(E_{\mathrm{SD}})}{n_j(E_{\mathrm{MC}})} \tag{6.4.1}$$

where $n(E_{\mathrm{SD}})$ is the number of events with a measured energy E_{SD} and $n(E_{\mathrm{MC}})$ is the total number of simulated events with the energy E_{MC}. A representation of the matrix can be seen in Fig. 6.11. The largest probabilities are on the diagonal elements for high energies and at low energies the probability is deviating due to the SD trigger properties.

The matrix is used in Section 7.2 to estimate the influence of the event migration on the measured spectrum.

Chapter 7

Energy Spectrum

7.1 SD energy spectrum

The standard SD data set used in this analysis is composed of events with a zenith angle smaller than 60°. The T5 trigger is applied to the data to guarantee that no crucial information is missed for the shower reconstruction, rejecting events located close to the array boundary or close to a temporarily missing tank. The choice of the T5 trigger allows us to simply exploit the regularity of the array. The aperture of any array configuration (see Fig. 7.1) is obtained as a multiple of the elementary hexagon aperture [187].

At full efficiency the detection area per internal tank is $D^2\sqrt{3}/2 \approx 1.95\,\mathrm{km}^2$, where D=1.5 km is the array spacing. The corresponding aperture for showers up to 60° is then a_{cell}= $4.59\,\mathrm{km}^2$ sr. To calculate the integrated exposure over a given period of time, only the number of cell-seconds have to be counted. The changes of the array configuration are monitored second by second [188]. For each configuration the number of elementary hexagons, N_{cell} is computed and the exposure incremented by $N_{\mathrm{cell}}\, a_{\mathrm{cell}}\, \Delta t$, where Δt is the time when the configuration has been stable.

The use of the T2 information is not sufficient for the evaluation of a correct exposure, as they give single station trigger rates, but do not take into account the hidden dead times of the central trigger and CDAS. For example if the central trigger and CDAS are not functioning, but the T2 informations are available there might be an overestimation of the exposure, or an opposite effect if T2 files are not available, but CDAS is still receiving data. The estimation of the dead times is based on the study of the arrival times of T5 events, under the hypothesis that the rate is constant for the T5 daily rate per hexagon [189, 190].

Acceptance from simulations The threshold energy where the array is fully efficient has to be deduced from simulations. At high energy, air showers have a large enough size on the ground to trigger many stations independently of where they exactly fall on the array. At lower energies the granularity of the array becomes important and the trigger probability depends on the position of the shower core with respect to the closest tanks. The particle density in the shower is a steeply decreasing function of the distance to the shower axis and so is the probability for a given shower to trigger a tank.

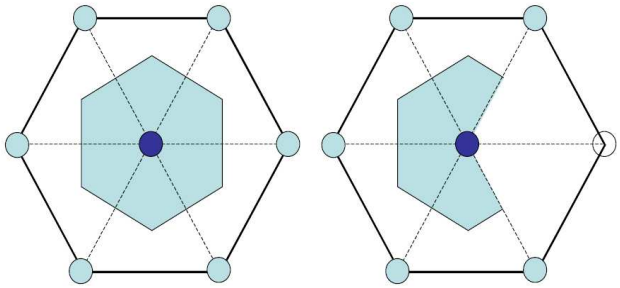
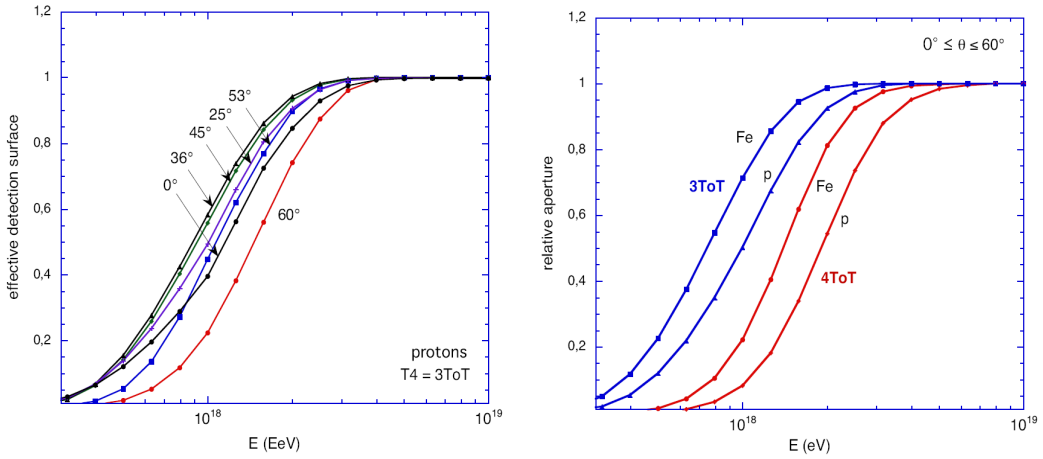


Figure 7.1: Schematic view of the area (shaded region) where the shower core must be located inside an elementary hexagonal cell in order to pass the T5 trigger (left: full hexagon; right: hexagon with a missing vertex [187]).



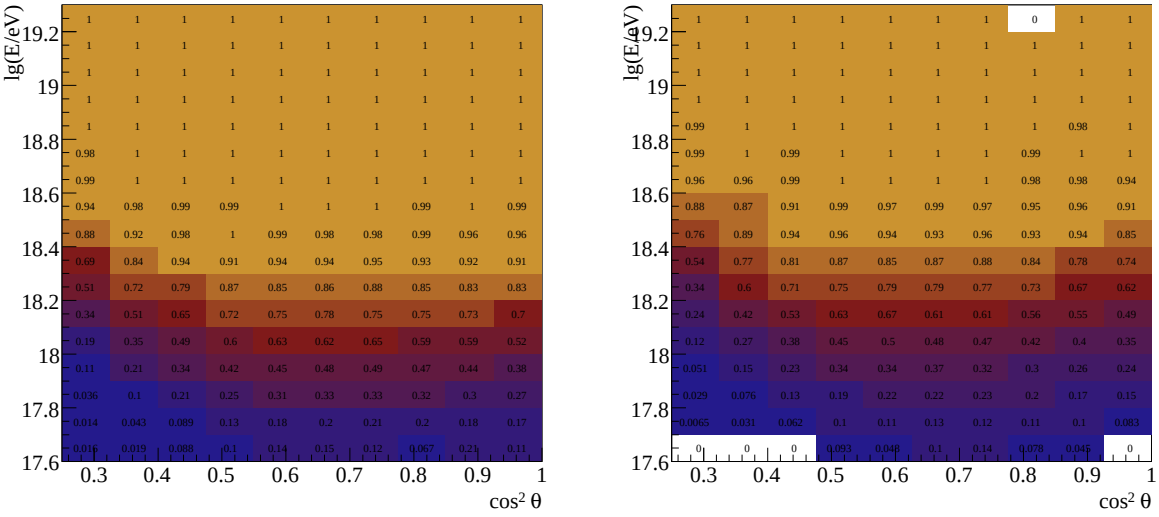
(a) *zenith angle dependence* (b) *trigger dependence*

Figure 7.2: (a) Effective detection surface as a function of energy for proton showers at different zenith angles. (b) Aperture saturation curves for proton or iron induced showers [187].

The threshold energy was deduced from the so called Lateral Trigger Probability (LTP) function, giving the individual tank ToT trigger probability [187, 191] as a function of the distance of the station to the shower axis for different energy, zenith angle, and azimuth of showers induced by different primary particles. From the LTPs the trigger probability is

$$P_{T4} = 1 - P(0)\left(1 + \sum_i \frac{P_i}{Q_i} + \sum_{j>i} \frac{P_i}{Q_i}\frac{P_j}{Q_j} + \sum_{\text{aligned}} \frac{P_i}{Q_i}\frac{P_j}{Q_j}\frac{P_k}{Q_k}\right) \tag{7.1.1}$$

where i, j and $k \in [1, 7]$ (only in aligned configuration for the last term), P_i is the probability that tank i triggers, $Q_i = 1 - P_i$ (probability that it does not trigger) and $P(0) = Q_1 \cdot Q_2 \cdot ... \cdot Q_7$ (probability that not a single tank triggers). The relative aperture to the total one for proton and iron induced showers as a function of energy, without any scaling is shown in Fig. 7.2. From this study the SD detection efficiency is 1 above 3 EeV and 7 EeV for the 3ToT and



(a) *iron* (b) *proton*

Figure 7.3: Full simulation chain. Trigger probability for the surface detector as a function of zenith angle and energy.

4ToT triggers respectively.

Another independent analysis on the efficiency of the surface detector has been done using a complete chain simulation as described in Section 6. The results are similar to the one obtained with the LTPs. The energy above which the SD is fully efficient is at 2.5 EeV for iron induced showers and at 3.5 EeV for showers generated by proton primaries as shown in Fig. 7.3. The trigger efficiency is zenith angle dependent and even at high energies SD is not fully efficient, having a trigger probability of 80% at 60° and energy of 4 EeV for protons. The zenith angle dependency of the flux will be investigated in a later subsection.

Acceptance from hybrid events A complementary method to derive the T5 trigger probability is using hybrid data [192], and does not rely at all on simulations. The probability for a shower to trigger the FD as well as the SD, i.e. golden hybrid data, is given by

$$p(\mathrm{SD}|\mathrm{FD}) = \frac{p(\mathrm{SD\&FD})}{p(\mathrm{FD})} \tag{7.1.2}$$

where $p(\mathrm{SD})$, the probability to trigger the surface detector is supposed to be independent of the probability to trigger the fluorescence detector $p(\mathrm{FD})$. The left hand side is directly observable and can be measured as a function of energy and zenith angle. In Fig. 7.4 the T5 trigger probability as measured from the hybrid data for different bins in FD energy and zenith angle is illustrated.

This method has been tested with full Monte Carlo simulations, assuming a mixed composition and applying the energy correction factors. The trigger efficiency, deduced from golden hybrid simulations, as function of the simulated shower energy and zenith angle is shown

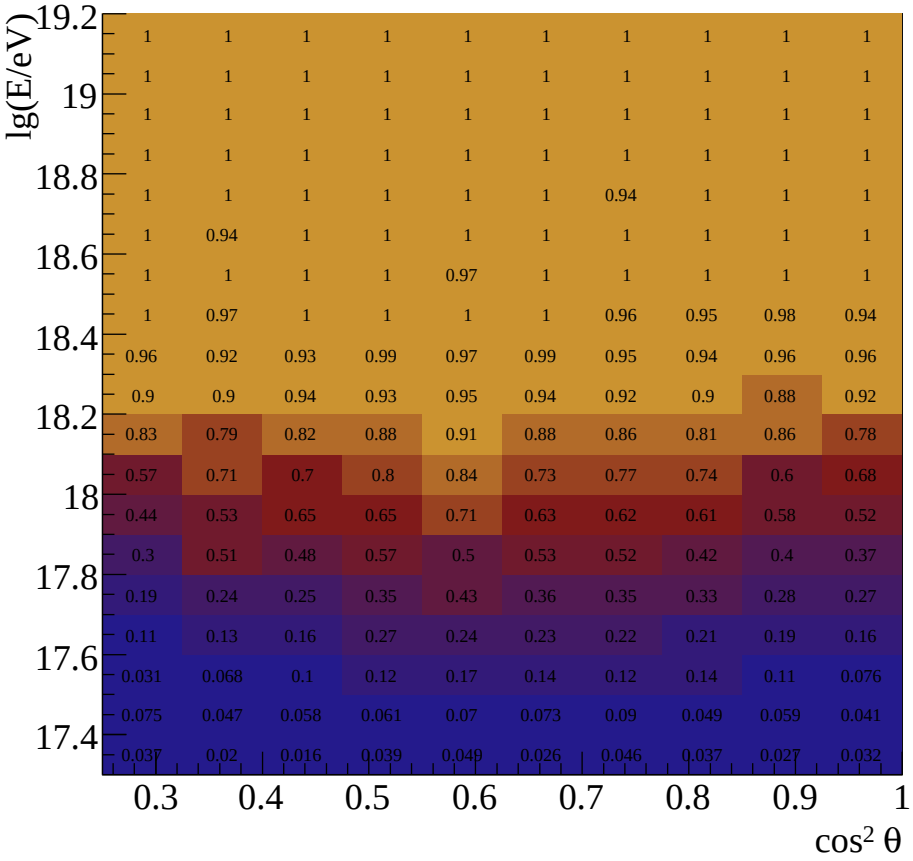


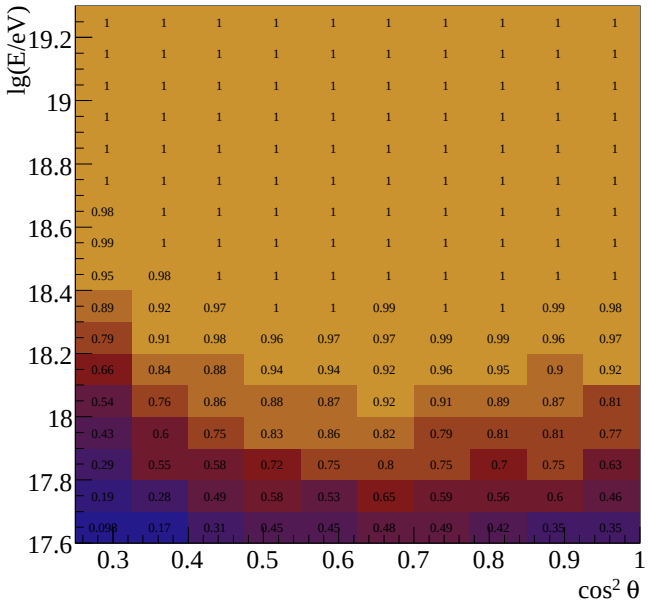
Figure 7.4: T5 efficiency as deduced from the hybrid and golden hybrid data.

in Fig. 7.5(a). Compared to the results obtained using the SD data only (see Fig. 7.5(b)) the efficiency is overestimated. This, illustrated in Fig. 7.6(a), is mainly due to the correlation between the SD trigger probability and FD trigger probability. Showers that develop high in the atmosphere do not have the maximum number of particle depth in the field of view of the FD and also they cannot trigger the SD, therefore the trigger probability is overestimated.

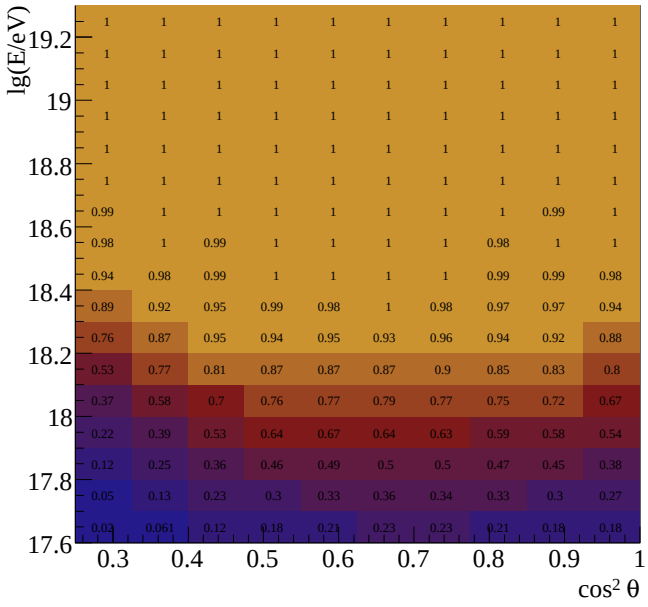
The difference between the trigger efficiency deduced based on SD simulations only and on the FD data, is represented in Fig. 7.6(b). It is smaller than in the previous case, but the statistical uncertainties from the data are larger. For this work the numbers obtained from the simulations only will be used as a systematic uncertainty. In Table B.1 from Appendix the T5 trigger efficiency is given explicitly.

Cosmic ray energy flux from the surface detector data The cosmic ray energy spectrum obtained from the surface detector data can be extended below the full efficiency threshold energy due to the knowledge of the trigger probability and its uncertainty.

The flux multiplied by the energy ($dJ/d\ln E$) above 1.4 EeV is shown in Fig 7.7. The number

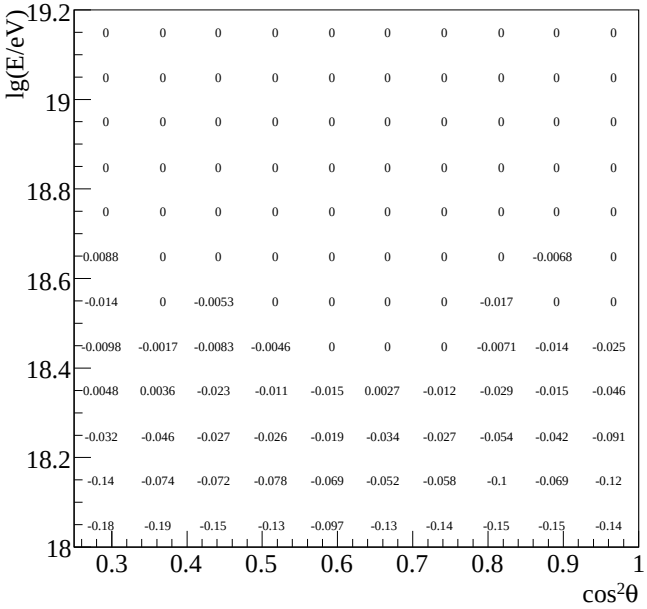


(a) *golden hybrid simulation (energy scaled)*

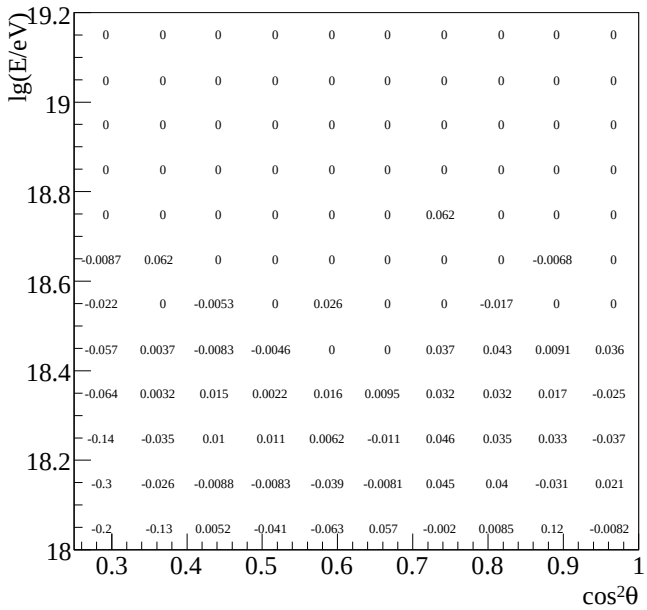


(b) *SD simulation (energy scaled)*

Figure 7.5: Trigger efficiency after scaling the MC energy for mixed composition. (a) The trigger efficiency obtained from golden hybrid simulations in the same way as with the golden hybrid data (b) Trigger efficiency as deduced from SD simulations only.



(a) *P(SD) - P(golden hybrid) from simulations*



(b) *P(SD) simulations - P(golden hybrid) data*

Figure 7.6: (a) The difference between the trigger efficiency as deduced from energy scaled simulations with both methods and (b) the difference between the trigger efficiency deduced from golden hybrid data and from the scaled SD simulations.

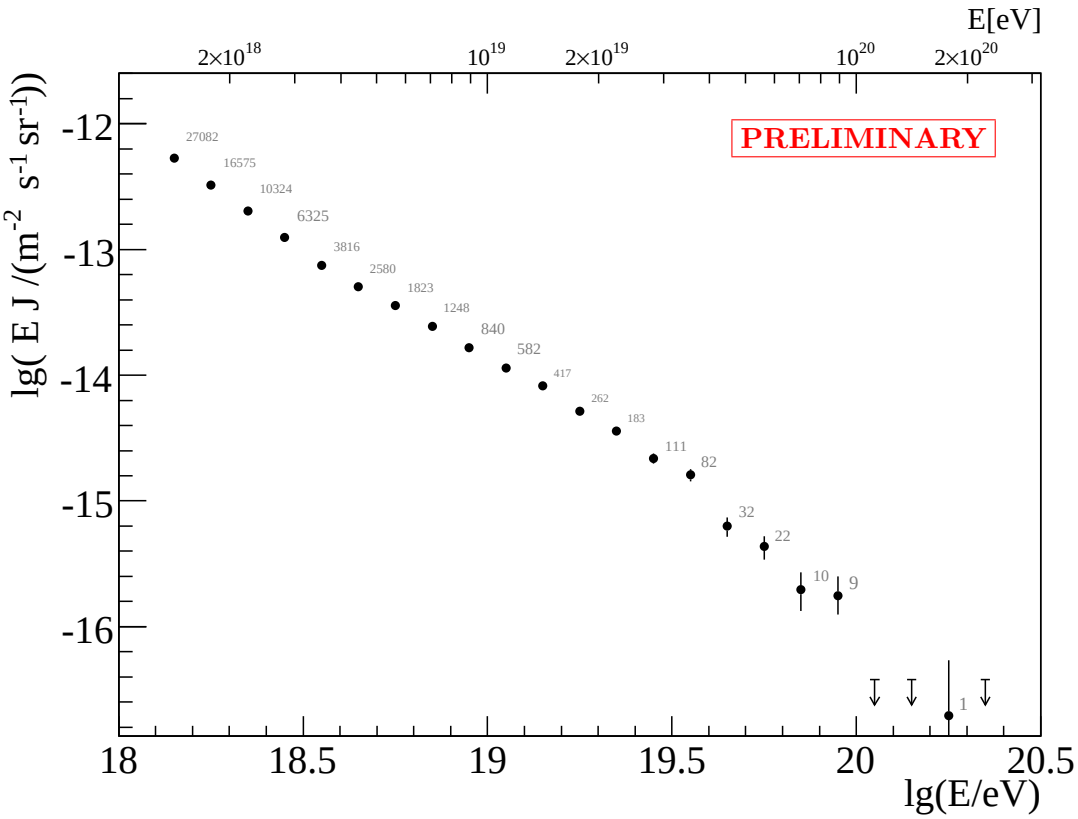


Figure 7.7: The comic rays flux as a function of energy.

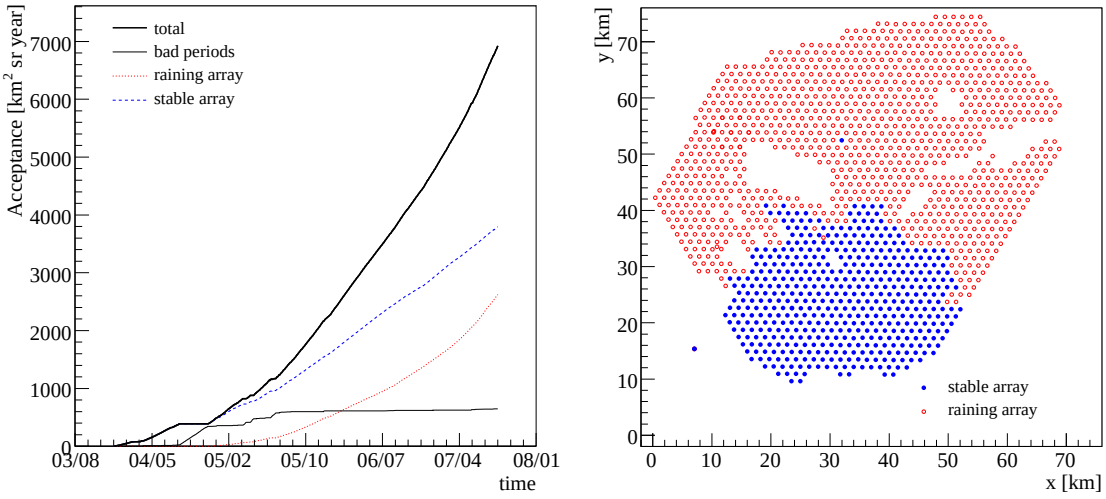
of events in each bin is given, and only the statistical uncertainties [193] are shown. The total acceptance for the data in use, from January 2004 until the end of August 2007, of 6923 km^2 sr year is 4 times that achieved by AGASA [9] and comparable to that of the monocular HiRes detectors [8]. The number of events is larger than the sum of all events recorded with previous efforts.

One possible inaccuracy in the spectrum might be given by the contamination from photon primaries as the values of S(1000 m) can be a factor 2-3 reduced in photon showers compared to nuclear primaries. No indication for photon primaries were found over the entire energy range and limits to the photon flux were derived [194–196].

Flux uncertainties and threshold energy The flux uncertainty has two major contributions: From the energy scale uncertainties and from the normalization (i.e. acceptance).

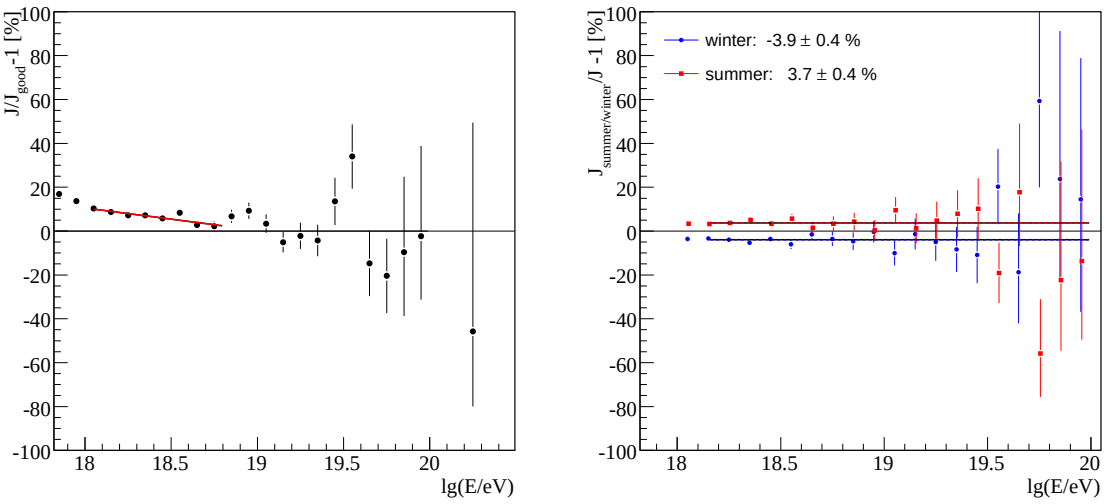
The energy uncertainties can be propagated to the cosmic ray flux only after establishing the flux distribution with energy. The uncertainties resulting from the determination of S_{38} and energy calibration are energy dependent. Therefore we propagate them completing the spectra by varying the energy within $\pm 1\sigma$. The same method has been applied on the contribution from the 22% systematic uncertainties of the FD energy.

The acceptance uncertainties have contributions from the trigger efficiency, and from the



(a) *acceptance* (b) *array*

Figure 7.8: The acceptance as a function of time (a) for different parts of the array, as defined in (b).



(a) *different arrays* (b) *array*

Figure 7.9: (a) The difference between the measured flux calculated with the whole array and the one obtained only with the stable array. The red line indicates the region where the correction factors will be applied. (b) Temperature effects. The relative difference between the summer/winter and the whole period.

detector response. The contribution from the non regularity of the array, the spacing between stations may vary up to 10 m, obtained taking into account the real sizes for each hexagon, is $\approx 0.03\%$ [197]. The unstable period rejection has an effect of 0.5% on the exposure [190].

Table 7.1. The correction factors obtained from the raining PMTs.

lg(E/eV)	correction ± stat	$\frac{\Delta J}{J}$ [%] (sys)
18.05	0.906 ± 0.005	5
18.15	0.920 ± 0.005	4
18.25	0.933 ± 0.007	3
18.35	0.933 ± 0.009	3
18.45	0.946 ± 0.012	2
18.55	0.921 ± 0.016	2
18.65	0.973 ± 0.020	1

The uncertainty, including the calculation of the bad periods rejection sums up to 1%.

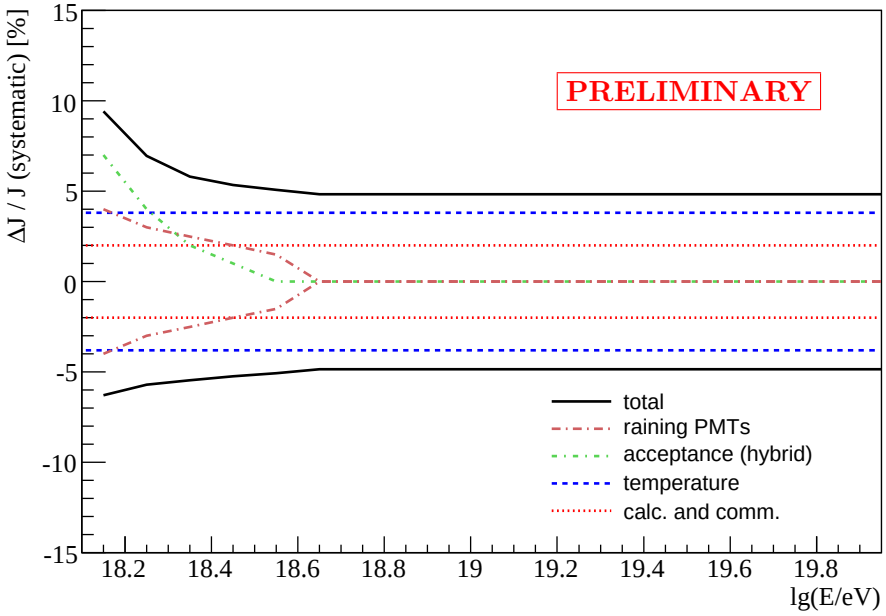
Above efficiency saturation, the unstable ToT characteristic due to the raining PMTs should not create any problem. Below the threshold energy limit they induce an increase of the trigger efficiency. To estimate this effect the array is divided in two sub-arrays, based on the stability of the PMTs, as shown in Fig. 7.8(b). In figure Fig. 7.9(a) the relative difference between the flux measured with the whole array and the one measured with the stable array is shown. The acceptance until August 2007 for the obtained array is $3797\,\mathrm{km}^2$ sr year comparable with the one obtained for the sub-array containing the PMTs with a strange behavior $2622\,\mathrm{km}^2$ sr year. The effect is 3% at $\lg(E/\mathrm{eV})= 18.55$ raising to 10% at $\lg(E/\mathrm{eV})= 18$. The correction factors as well as their uncertainties are given in Table B.1 for each energy bin. The systematic uncertainty was taken as half of the correction.

The acceptance depends on the temperature, in summer there are more events recorded than in winter. The data were divided in winter time from April until September and summer time as the rest of the year. Until August 2007 the exposure is $3572\,\mathrm{km}^2$ sr year for winter and $3351\,\mathrm{km}^2$ sr year for summer. The spectra difference is shown in Fig. 7.9(b). The systematic uncertainty due to temperature variation will be taken as 3.8%.

An effect that has not yet been investigated is the communication problems of the data acquisition. They have a small effect on the total acceptance, but nevertheless they should be at the level of 1-2%. Together with the uncertainty from the bad period computation it will be assumed to be 2%.

In Fig. 7.10(a) we summarize the systematic uncertainties of cosmic ray flux related to the acceptance as a function of energy. Over the whole energy range the effect is smaller than 10% and it is at the level of 5% in the full acceptance region.

The cosmic ray flux multiplied with $E^{2.65}$ is illustrated in Fig. 7.10(b). The total systematic uncertainty is shown with dashed lines. The largest contribution is from the energy systematics of the FD of 22%. In the future the contributions from the fluorescence yield measurements and from the reconstruction and detector calibration will be reduced to 10%. The SD contributions are shown with filled histograms, the systematics of the energy calibration dominates at high energies and will be reduced with an increased statistics of golden hybrid events.



(a) *acceptance contribution*

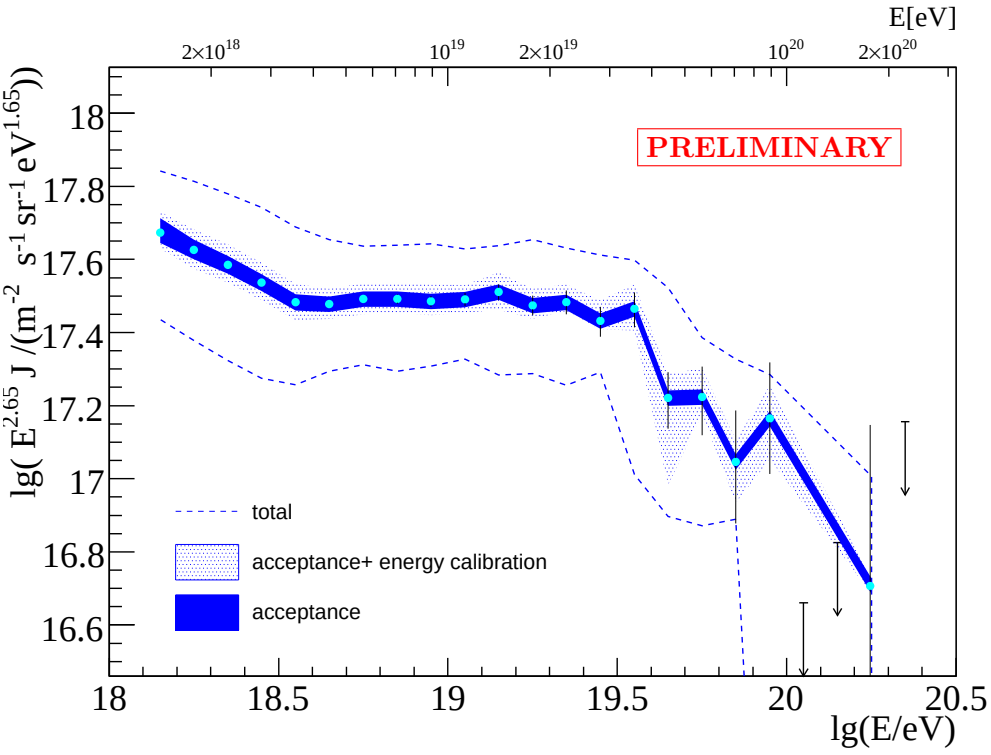


Figure 7.10: (a) The systematic uncertainties on the flux coming from the acceptance. (b) The flux multiplied with $E^{2.65}$. The different contributions to the flux uncertainties are shown.

7.2 Forward folding and the ankle

For a deconvolution of the detector effects on the energy spectrum a proper unfolding [198–204] can be done only after detailed studies of the air shower and detector simulations and after an increase of Monte Carlo statistics.

To determine the possible detector resolution effects only a forward folding [198] of an assumed true flux is done in the present work.

The energy spectrum might be interpreted as a sum of two components: a galactic origin component, which dominates until a certain energy called *ankle* and a second extragalactic component with another spectral index dominating above this energy. Therefore the true energy spectrum can be described by

$$J(E) = A_1 \cdot E^{\gamma_1} \left(1 + \exp\left(\frac{\lg E - \lg E_c^g}{W_c^g}\right)\right)^{-1} + A_2 \cdot E^{\gamma_2} \left(1 + \exp\left(\frac{\lg E - \lg E_c}{W_c}\right)\right)^{-1} \tag{7.2.1}$$

where E_c^g, E_c are the critical energies, at which the galactic and extragalactic fluxes are suppressed by 50% with respect to the pure power law and W_c^g, W_c determine the sharpness of the cut, γ_1 and γ_2 are the power law indices of the components and A_1, A_2 the normalizations. The measured flux extends only down to 1.25 EeV, therefore the data are sensitive only to the falloff of the galactic contribution. Thus in the used function for the fit

$$J(E) = A_1 \cdot E^{\gamma_1} + A_2 \cdot E^{\gamma_2} \left(1 + \exp\left(\frac{\lg E - \lg E_c}{W_c}\right)\right)^{-1} \tag{7.2.2}$$

the exponential behavior is neglected and γ_1 becomes a power law index of the galactic flux suppression.

The flux, corrected by the trigger efficiency is

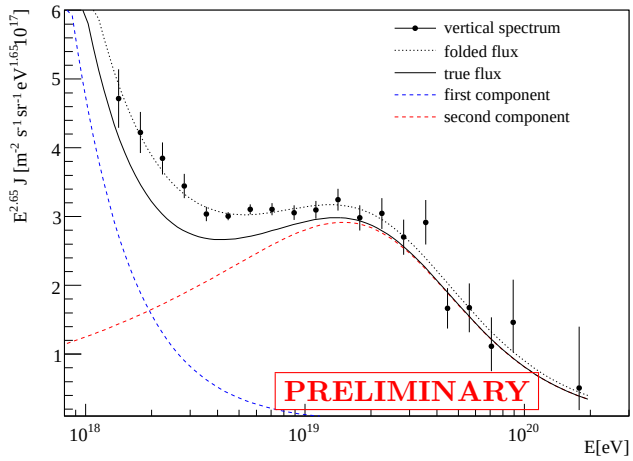
$$J_{\text{meas}} = \mathbf{P}^{-1} \cdot \mathbf{R} \cdot \mathbf{P} J_{\text{true}} \tag{7.2.3}$$

where $\mathbf{R}$ is the migration matrix deduced in the previous chapter from full detector simulations, $\mathbf{P}$ is the trigger efficiency diagonal matrix and J_{true} is the true flux given by Eq. (7.2.2).

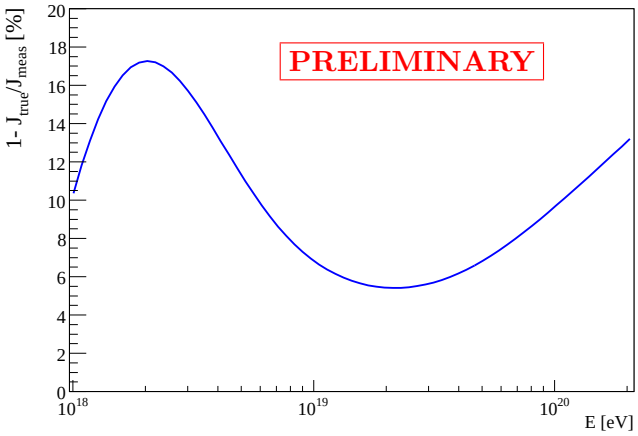
In order to obtain a realistic flux model, Eq. (7.2.3) is fitted to the data with A_1, A_2, γ_1, γ_2, W_c and E_c as free parameters. As can be seen in Fig. 7.11(a), the ansatz (7.2.2) describes the data very well (referred as folded flux in the figure). The obtained parameters are given in Table. 7.2.

The relative difference between the assessed flux and the true flux is illustrated in Fig. 7.11(b). In the low energy part the detector resolution starts to be less important as the trigger efficiency allows only for a smaller migration from the lower energy bins to the next one, therefore the effect on the flux is reduced.

The largest correction is at energies of 2 EeV at the order of 17%, because this is the region of a transition from a steeper falling spectrum to a flatter one. The threshold effects are not anymore important. In the 10 EeV range the effect is minimal (5%). The tail of the evaluated flux at high energies is overestimated with 10%, thus the forward folding increases the flux suppression significance.



(a) *energy spectrum*

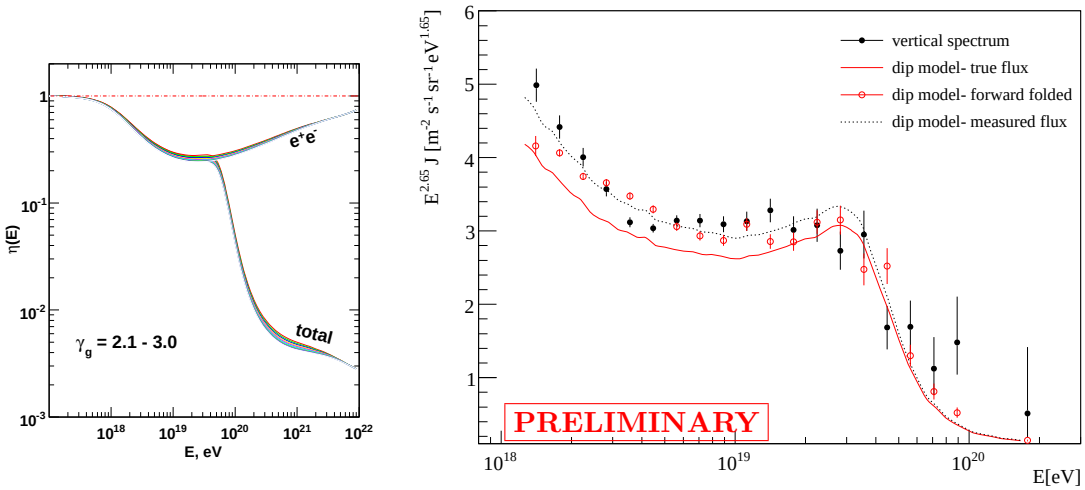


(b) *flux relative difference*

Figure 7.11: (a)The energy spectrum is well described by Eq. (7.2.2). The 2 components are shown with dashed lines and with dotted black line a fit to the measured flux is illustrated. (b) Detector effects on the measured flux. The flux is overestimated over the whole energy range.

Table 7.2. The parameters of the fit, (7.2.2) first line and (7.2.3) second line.

	γ_1	γ_2	$\lg(E_c/eV)$	W_c	χ^2/ndof
flux$_{meas}$	-3.9 ± 0.7	-2.2 ± 0.4	19.46 ± 0.29	0.24 ± 0.04	0.66
flux$_{true}$	-4.2 ± 0.7	-2.2 ± 0.4	19.46 ± 0.30	0.24 ± 0.06	1.2



(a) *modification factor* (b) *fit of the cosmic ray flux*

Figure 7.12: The dip model applied to the vertical spectrum. (a) the modification factor taken from [47] (b) The fit of the dip model applying the forward folding to obtain the true flux. The folded flux is denoted by red dots, having the statistical uncertainties given by simulations, as well as a fit to the measured flux (dotted line).

The dip model One of the models to explain the spectral feature of the ankle is the so-called dip model [47]. Within this model, the flattening of the spectrum between 1-10 EeV is due to energy loss of primary protons by e^+e^- production in interaction with the cosmic microwave background (CMB). If protons are accelerated at the source with a spectrum following a power law $J(E)_{\text{source}} = E^{\gamma}$ then they will arrive at Earth with a modified spectrum

$$\eta(E) = J_{\text{obs}}(E)/J(E)_{\text{source}} \tag{7.2.4}$$

where the modification factor, η, takes into account the GZK effect as well as the energy loss due to pair production. Moreover it is universal, i.e. does not depend on the initial conditions at the acceleration site. If only adiabatic losses, due to the expansion of the Universe, are included, $\eta = 1$, the energy spectrum at the top of the atmosphere would have the same shape as the one at the source.

The protons loose first energy through e^+e^- production, the energy threshold for the interaction with the CMB photons being 0.5 EeV. This process produces a dip in the spectrum. Then pion photoproduction becomes important at 40 EeV, forming the GZK feature. The observable part of the dip extends from 1 EeV up to 40 EeV, and it flattens twice, at 10 EeV and 1 EeV. The first flattening automatically produces the ankle. The cosmic ray flux contains no galactic component above 1 EeV, thus the ankle is not seen as a transition region.

To test if the cosmic ray flux is described by this model, two fits are performed: On the measured vertical spectrum, J_{meas} and applying a forward folding of J_{true}. The χ^2 is

constrained with an additional factor

$$\chi^2 = \sum_i \frac{(A \cdot \eta(\lambda E) \cdot E^\gamma - J_{\mathrm{meas/true}})^2}{\sigma^2_{J_{\mathrm{meas/true}}}} + \frac{(1-\lambda)^2}{\sigma^2_\lambda} \tag{7.2.5}$$

where σ_λ is the systematic uncertainty of the energy scale,λ, taken as 25%. The values for the modification factor are taken from Fig. 7.12(a) and A is the normalization.

The results are shown in Fig. 7.12(b). The reduced χ^2 is 4.54 when the forward folding is applied. The reduced χ^2 in the case of a fit to the measured spectrum it 2.98. The dip model is not describing well the obtained cosmic ray flux. This is mainly due to the sharp transition between the spectral indices in the ankle region. The detector resolution softens this transition (see Fig. 7.11(b)). The folded flux, after applying the migration matrix on the flux is shown with open circles.

The energy scaling factor obtained applying the forward folding is $\lambda = 1.50 \pm 0.001$ and thus two times the estimated energy scale uncertainty of the Pierre Auger Observatory. The injection spectrum has an index of $\gamma = -2.49 \pm 0.02$.

The pure proton model of [47] is thus disfavored by the shape of the measured UHECR flux. It is interesting to note that also the $\langle X_{\mathrm{max}} \rangle$ analysis suggests a mixed composition [67].

7.3 Flux suppression

To check the shape of the spectrum at the highest energies a power law function is fitted between 4×10^{18} eV and 4×10^{19} eV using a binned likelihood method. The spectral index obtained is $\gamma = -2.67 \pm 0.02$. The number of events expected if this power law holds above 4×10^{19} eV are 167 ± 3 and above 4×10^{20} eV 35 ± 1 whereas in the assessed flux there are 64 events and 1 event. The spectral index changes above this energy to -4.46 ± 0.45. A diminishing number of events in this region is clear.

To obtain the significance of the deviation from a pure power law the first recipe given in [205] is applied. The null hypothesis that the maximum energy event is drawn from a power law index γ_i, where γ_i is deduced from different energy intervals is tested.

The probability that the maximum energy event is higher than the maximum of a sample drawn from a power law with an exponent, γ_i, is given by

$$P(z \geq Z_{max}) = (1 - z^{1-\gamma})^N, \tag{7.3.1}$$

where $z = E_{max}/E_{min}$ and N is the number of events observed in the interval $[E_{min}, E_{max}]$. The maximum energy was taken as the highest energy bin with at least one event.

This analysis has been applied on the vertical spectrum, and the systematic uncertainties were computed using the 1σ systematic uncertainties on the flux. The lg(P) as a function of the lower energy bin is shown in Fig. 7.13. The error bars correspond to the statistical uncertainties of the spectral index, γ. The systematics uncertainties from the energy conversion are shown with dotted red line, and the ones coming from the attenuation curve with blue dashed line.

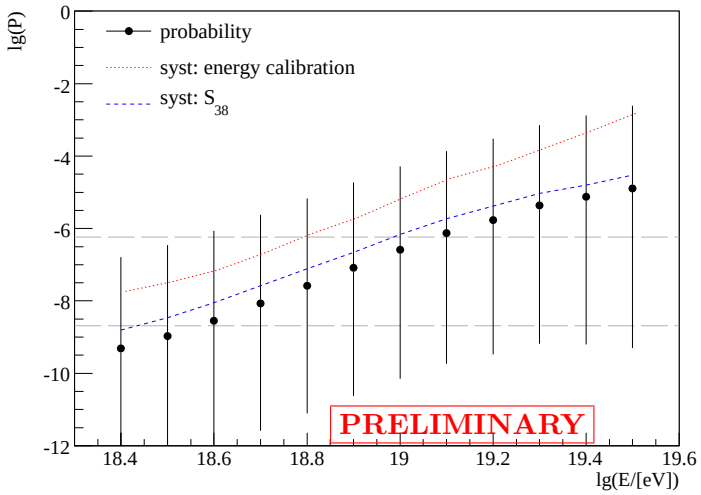


Figure 7.13: Probability that the highest energy event of a power law distributed spectrum will be less then or equal to the highest event obtained in the spectrum. The 5σ and 6σ probabilities are shown with interrupted lines.

The probability, using the lower energy bin lg(E/eV)= 18.4 is

$$P = 4.8 \cdot 10^{-10} + 1.62 \cdot 10^{-7}(\text{stat}) + 1.39 \cdot 10^{-8}(\text{sys}), \tag{7.3.2}$$

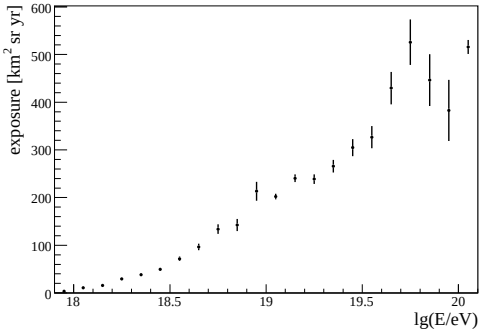
therefore we can reject the hypothesis that the last event is drawn from the same power law with a significance of 5σ within the total uncertainties.

The unfolding correction estimated in the previous section was not applied, but as the unfolding would make the suppression feature *steeper*, the above probability is a conservative estimate.

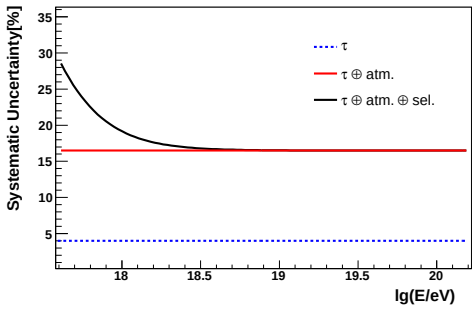
7.4 Combining the vertical, horizontal and hybrid spectra

The energy spectrum can be measured with three independent data sets: hybrid FD data, vertical ($\theta < 60°$) and horizontal($60° < \theta < 80°$) SD data. The energy spectrum derived from the hybrid data only covers the energy range above $\lg(E/eV) = 18$, letting us cross check the correction factors that we apply for the first bins in the vertical spectrum. The total number of events that triggered the FD is about half a million, but after all the quality cuts, similar to the ones applied in the energy calibration analysis, the total number of events is smaller than 5000.

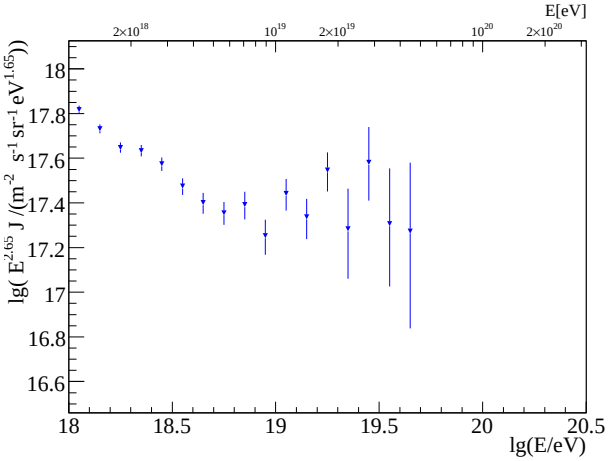
The hybrid exposure (Fig. 7.14(a)) calculation relies on the simulation of the FD and SD response. To reproduce the exact working conditions of the experiment and the entire sequence of given configurations, the fast grow of the array, the seasonal and the instrumental effects, a large sample of Monte Carlo simulations are performed. The simulations are done with the CONEX [62] simulation code, which does not provide the distribution of particles



(a) *hybrid exposure*



(b) *hybrid exposure uncertainties*



(c) *hybrid energy spectrum*

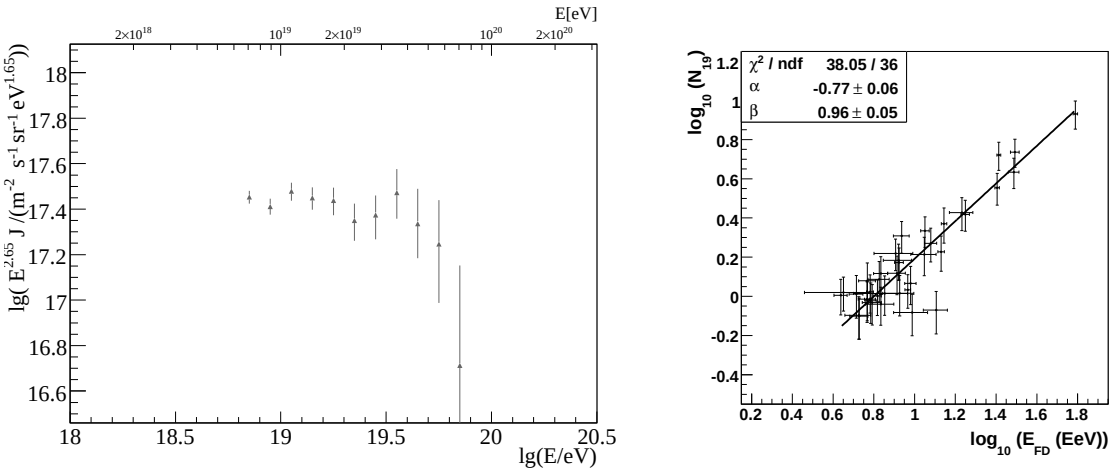
Figure 7.14: FD energy spectrum. Data from [206] as an update of [207].

on ground. Therefore the tank trigger simulation is based on the LTPs, deduced from full CORSIKA simulations.

The systematic uncertainties and their individual contributions are shown in Fig. 7.14(b). The uncertainty is increasing at low energies as a consequence of the aperture calculation at reconstruction level.

The horizontal spectrum [208] is shown in Fig. 7.15(a). It is obtained with inclined events, between 60° and 80°. Vertical and horizontal events are in contrast due to the different reconstruction algorithms.

In the case of the horizontal surface detector data the procedure to derive the energy is equivalent to the vertical one, but instead of using S_{38} the shower size is determined using the relative distributions of the two-dimensional muon number densities at ground level, obtained from simulations. The energy calibration is shown in Fig. 7.15(b). The normalization factor, N_{19}, of the muon map is the estimator to be related to the hybrid energy. It gives the total number of muons, relative to a shower initiated by a proton with an energy of 10^{19} eV. The acceptance calculation is geometrical and the threshold energy where the trigger efficiency is



(a) *horizontal energy spectrum* (b) *horizontal energy calibration*

Figure 7.15: Horizontal energy spectrum. Data and plot (a) are taken from [208].

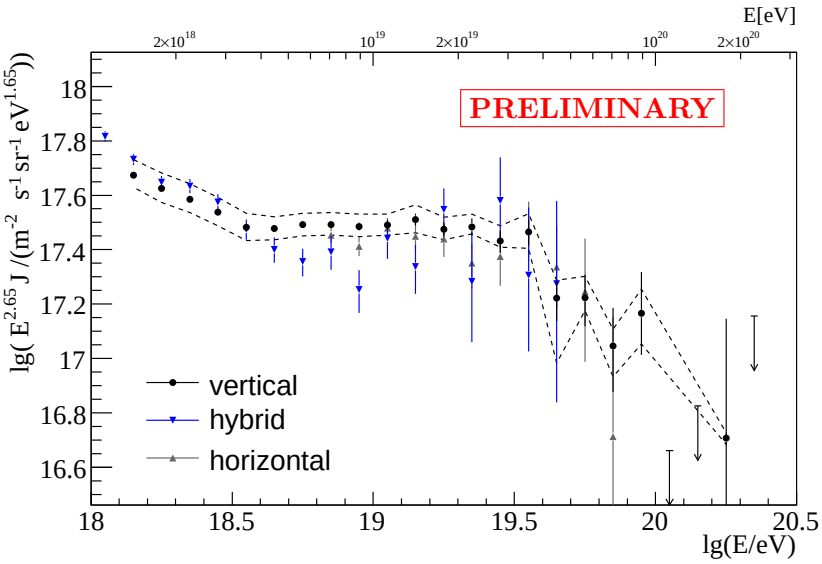


Figure 7.16: The three energy spectra agree within the systematic uncertainties (FD energy systematic is not included) of the vertical spectrum, shown with dashed line.

more than 0.98 is $\approx 6.3\,\mathrm{EeV}$. Above this energy the integrated exposure until end of February 2007 is $1510\,\mathrm{km}^2$ sr year, 29% of the equivalent acceptance of the vertical events.

The vertical, horizontal and hybrid spectra are shown in Fig. 7.16. The uncertainties on the vertical spectrum that are independent of the hybrid uncertainties are shown with black dotted lines. The horizontal spectrum inherits the same uncertainties as the vertical spectrum, due to the equivalent procedure to obtain the integrated exposure and the energy. There is a good agreement within the uncertainties.

Table 7.3. Uncertainties of the exposure for hybrid, vertical and horizontal spectra.

lg(E/eV)	hybrid [%]	vertical [%]	horizontal[%]
18.05	19	-	-
18.15	18	9.5	-
18.25	17	7.0	-
18.35	16	5.8	-
18.45	16	5.3	-
18.55	16	5.1	-
>18.65	16	4.8	4.8

Method to combine the spectra In order to combine the three spectra measured at the Pierre Auger Observatory a maximum likelihood method is used. It has the advantage over a χ^2 method by being able to deal with Poissonian data and ratio distributions.

The notations for the variables in use are

- i: index for lg(E) bins
- k: index for the spectra, $k \in \{1(\text{hybrid}), 2(\text{horizontal}), 3(\text{vertical})\}$
- ψ_{ik}: cosmic ray flux multiplied with energy.
- g_k: the relative normalization with the systematic uncertainty σ_k described below
- $f(\psi_{ik})$: likelihood of ψ_{ik}
- A_{ik}: exposure
- n_{ik}: number of events
- ζ_{ik}: correction factors due to the acceptance as described previously for $(k = 3, i \leq (lg(E/eV) < 18.55))$ and $\zeta_{ik} = 1$, for the rest of the bins
- corresponding measured values are denoted with a hat.

The systematic uncertainties σ_k in the exposure calculations were taken according to Table. 7.3.

We construct the likelihood

$$L = \left(\prod_i f_{\text{hybrid}}\right)\left(\prod_i f_{\text{horizontal}}\right)\left(\prod_i \zeta_i \cdot f_{\text{vertical}}\right)\left(\prod_k G(g_k)\right)\left(\prod_{i=\overline{1,3};k=sd} h(\zeta_{ik})\right) = \left(\prod_k \prod_i f(\psi_{ik})\right)\left(\prod_{i=\overline{1,3};k} g(g_{ik})\right). \tag{7.4.1}$$

For each bin the measured value is

$$\hat{\psi}_{ik} = \hat{n}_{ik}/\hat{A}_{ik} \tag{7.4.2}$$

which is compared to the trial value

$$\psi_{ik} = (1 - g_{ik}) \cdot n_{ik}/A_{ik}. \tag{7.4.3}$$

The distribution of g_k was assumed Gaussian with mean 0 and the spread given by σ_k

$$G(g_k) = \frac{1}{\sqrt{2\pi} \cdot \sigma_k} \exp\left(-\frac{g_k^2}{2\sigma_k^2}\right) \tag{7.4.4}$$

constraining the normalization factors g_k within their uncertainties. Because ζ_{ik} has a Gaussian distribution, $h(\zeta_{ik})$, this contribution can be summed with the g_k giving a normal distribution $g(g_{ik})$ binned in energy. The correlation between the correction factors was neglected.

The likelihood terms are treated in three ways:

If $n_i > 20$ events then A_{ik} as well as n_{ik} are normal distributed

$$f(n_{ik}) = \frac{1}{\sqrt{2\pi} \cdot \sigma_{n_{ik}}} \exp\left(-\frac{(n_{ik} - \hat{n}_{ik})^2}{2 \cdot \sigma_{n_{ik}}^2}\right) \tag{7.4.5}$$

where $\sigma_{n_{ik}} = \sqrt{n_{ik}}$ and

$$f(A_{ik}) = \frac{1}{\sqrt{2\pi} \cdot \sigma_{A_{ik}}} \exp\left(-\frac{(A_{ik} - \hat{A}_{ik})^2}{2 \cdot \sigma_{A_{ik}}^2}\right) \tag{7.4.6}$$

with $\sigma_{A_{ik}}$ being the uncertainty of A_{ik}.

This gives for ψ_{ik} the ratio-distribution [209]

$$f(\psi_{ik}) = \frac{b(\psi_{ik}) \cdot c(\psi_{ik})}{a^3(\psi_{ik})} \frac{1}{\sqrt{2\pi}\sigma_{n_{ik}}\sigma_{A_{ik}}} \left[2 \cdot \Phi\left(\frac{b(\psi_{ik})}{a(\psi_{ik})}\right) - 1\right] + \frac{1}{a^2(\psi_{ik})\pi\sigma_{n_{ik}}\sigma_{A_{ik}}} \exp\left[-\frac{1}{2}\left(\hat{n}_{ik} + \frac{\hat{A}_{ik}^2}{\sigma_{A_{ik}}^2}\right)\right] \tag{7.4.7}$$

$$a(\psi_{ik}) = \sqrt{\frac{1}{\hat{n}_{ik}}\psi_{ik}^2 + \frac{1}{\sigma_{A_{ik}}^2}}$$

$$b(\psi_{ik}) = \psi_{ik} + \frac{\hat{A}_{ik}}{\sigma_{A_{ik}}^2}$$

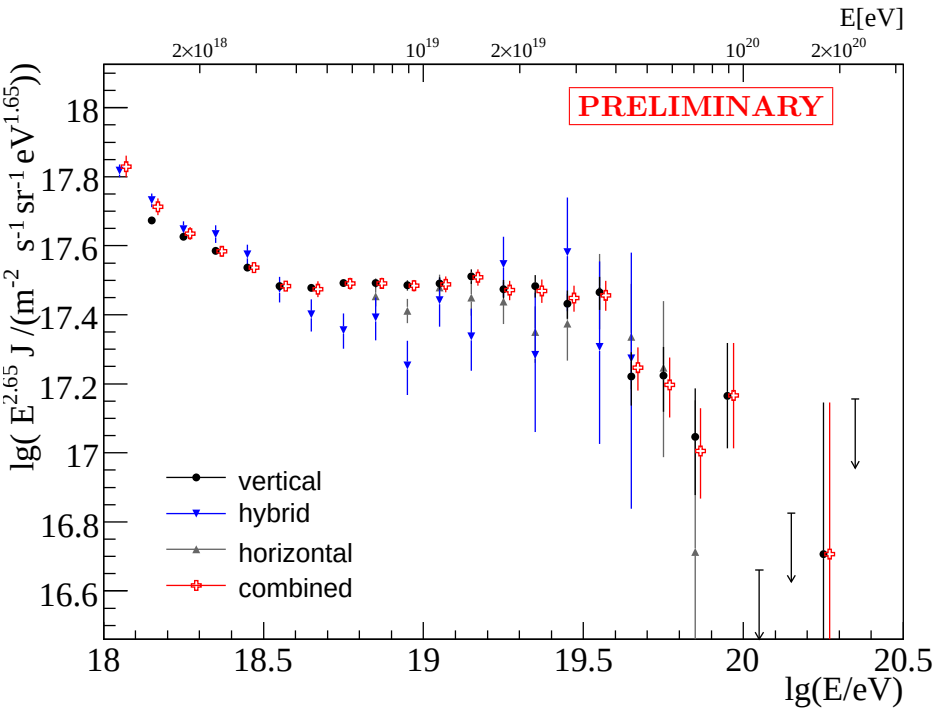


Figure 7.17: The combined energy spectrum. Only statistical uncertainties are shown.

$$c(\psi_{ik}) = \exp\left[\frac{1}{2}\frac{b^2(\psi_{ik})}{a^2(\psi_{ik})} - \frac{1}{2}\left(\hat{n}_{ik} + \frac{\hat{A}_{ik}^2}{\sigma_{A_{ik}}^2}\right)\right]$$

and

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

a. If $n_i > 20$ and the statistical uncertainty of the exposure is ≈ 0, $f(\psi_{ik})$ is treated as Gaussian

$$f(\psi_{ik}) = \frac{\hat{A}_{ik}}{\sqrt{2\pi}\cdot\sigma_{n_{ik}}} \exp\left(-\frac{\hat{A}_{ik}^2(\psi_{ni} - \hat{\psi}_{ni})^2}{2\cdot\sigma_{n_{ik}}^2}\right) \tag{7.4.8}$$

b. If $n_i < 20$ events, n_i has a Poissonian distribution. In this case due to large statistical uncertainties from the data, the ones from the exposure become negligible and $A_{ik} \cong \hat{A}_{ik}$.

$$f(n_{ik}) = \frac{n_{ik}^{\hat{n}_{ik}} \cdot \exp(-n_{ik})}{\hat{n}_{ik}!} \tag{7.4.9}$$

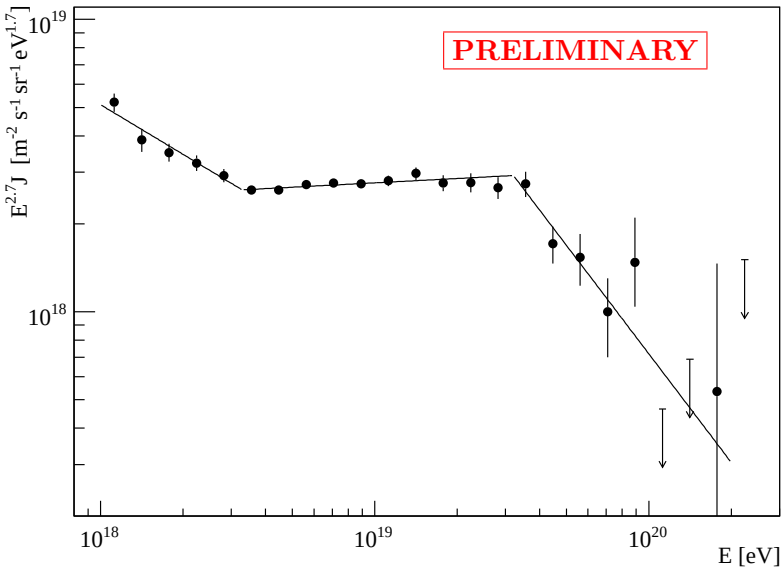


Figure 7.18: The combined energy spectrum. Simple power law assumptions over the whole range.

which in terms of ψ_{ik} is

$$f(\psi_{ik}) = \frac{1}{\hat{A}_{ik}} \frac{(\psi_{ik} \cdot \hat{A}_{ik})^{\hat{n}_{ik}} \cdot \exp(-\psi_{ik} \cdot \hat{A}_{ik})}{\hat{n}_{ik}!}. \tag{7.4.10}$$

At the end

$$\ell = -\ln L = -\sum_{k=1}^{3} \sum_{i=1}^{N} [\ln f(\psi_{ik}) + \ln g(g_{ik})] \tag{7.4.11}$$

is minimized.

The obtained normalization factors for the three spectra are:

$$g_{\text{vertical}} = 0.01 \pm 1.06; \% \tag{7.4.12}$$

$$g_{\text{hybrid}} = 4.06 \pm 3.78\ \% \tag{7.4.13}$$

$$g_{\text{horizontal}} = -0.36 \pm 4.6; \% \tag{7.4.14}$$

well within the uncertainties of each measurement.

The combined spectrum is shown in Fig. 7.17. It is dominated by the vertical spectrum in the whole energy range due to the smaller statistical and systematic uncertainties. The agreement between the three spectra, given by the required normalization factors that have to

be applied is at the level of 3%, much smaller that the independent systematic uncertainties of each measurement.

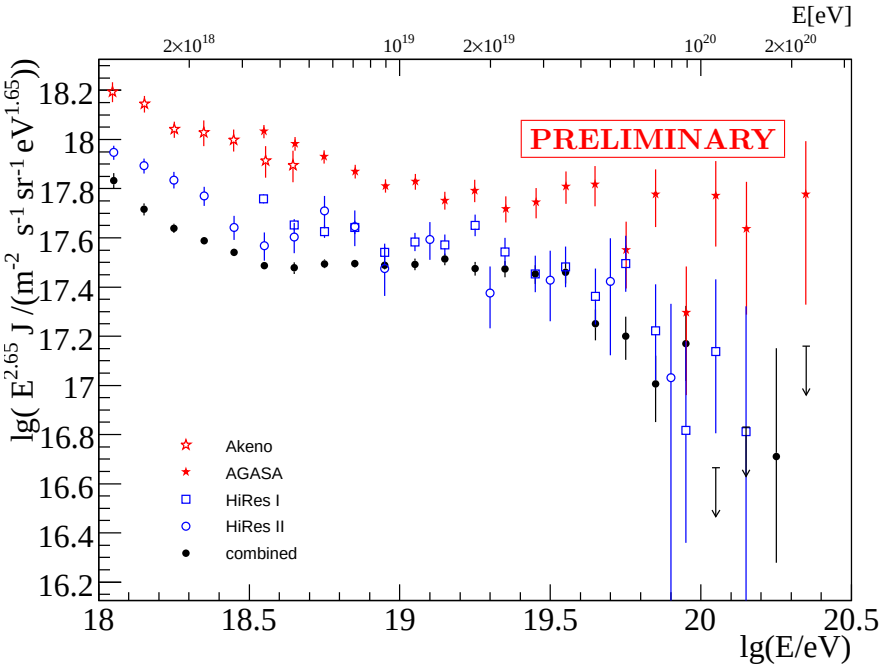
In order to obtain the spectral indices three power laws were fitted to the combined spectrum as shown in Fig. 7.18. The spectral indices change at $\lg(E/\mathrm{eV}) = 18.52 \pm 0.03$ and at $\lg(E/\mathrm{eV}) = 19.51 \pm 0.04$. The first measured spectral index is $\gamma_1 = 3.26 + -0.03$, changing to $\gamma_2 = 2.65 \pm 0.02$ and at the end it is $\gamma_3 = 3.9 \pm 0.3$.

Comparison with other experiments and the energy scale The combined spectrum is illustrated in Fig. 7.19(a) together with all the available data in the energy range on interest.

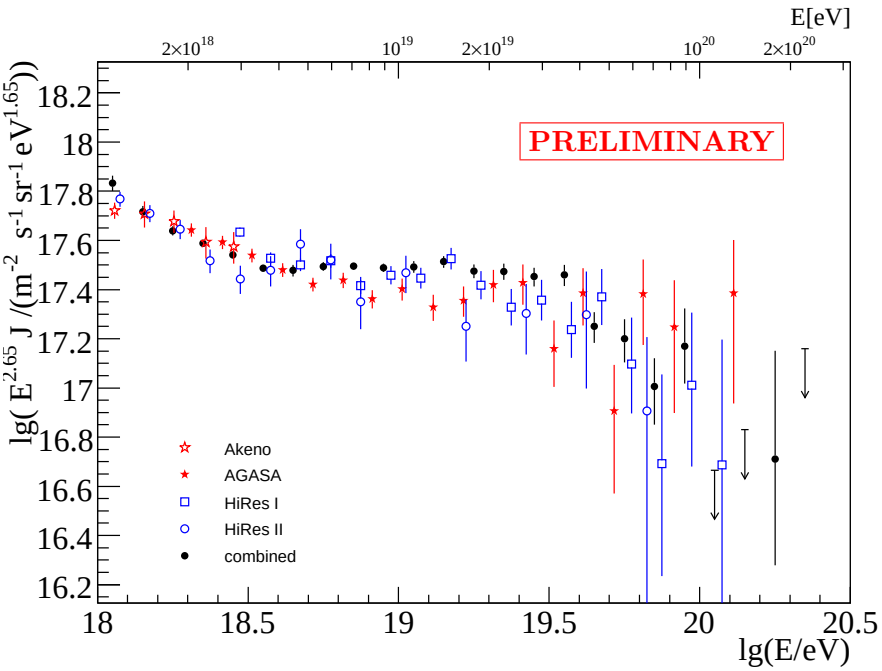
The disagreement between the measurements is due to the different calibration of energy with different inherited systematic uncertainties. The Akeno and AGASA experiments are calibrating their measurement with simulations. The energy determined from MC simulations might be ill defined as it was shown in Chapter 6, in case of Auger data. A scaling correction on energy is well motivated. In order to get agreement with the combined spectrum a lowering with 42% for AGASA and 36% for Akeno is necessary. The HiRes measurement is based on a different fluorescence yield measurement, and also the acceptance is calculated from simulations. A shift of 16% in energy is within the systematic uncertainties of the HiRes experiment (17%).

After the re-calibration all the measurements agree with each other as shown in Fig. 7.19(b), except perhaps at $E \gtrsim 10^{20}\,\mathrm{eV}$ for AGASA.

The total uncertainties (from the energy calibration, acceptance and FD energy scale) are shown in Fig. 7.20 for the vertical spectrum. The detector resolution effect on the measured spectrum is given as well. The uncertainties are dominated by the systematics of FD energy scale.



(a) *different cosmic ray measurements*



(b) *energy rescaled for AGASA(42%), Akeno(36%) and HiRes (17%)*

Figure 7.19: Cosmic ray flux measurements from AGASA, Akeno and HiRes compared to the combined spectrum.

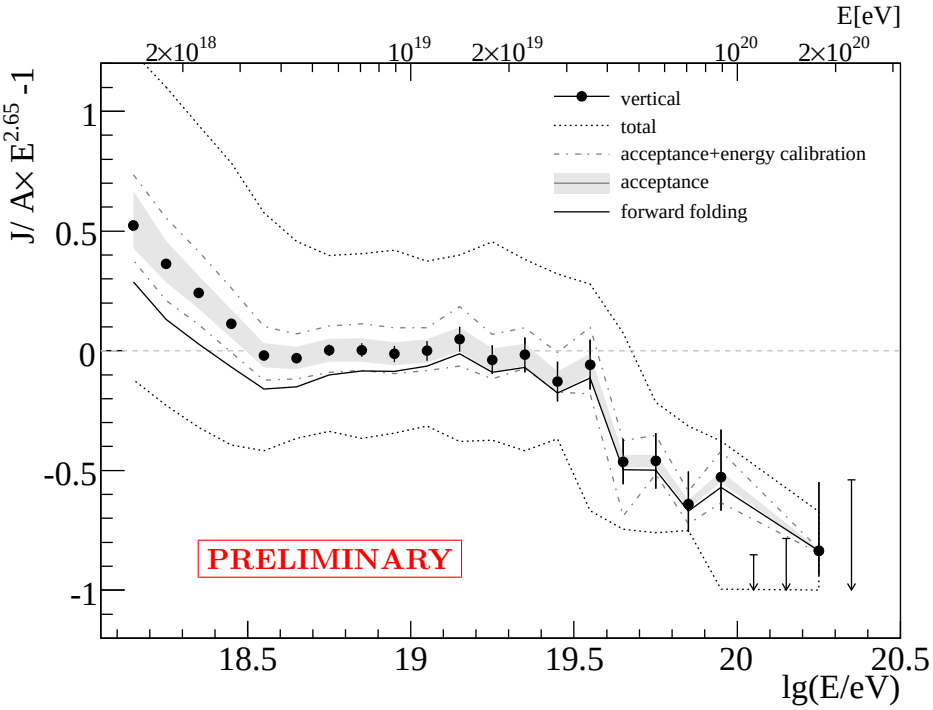


Figure 7.20: The vertical energy spectrum above 1.25 EeV as deduced in this thesis with its systematic uncertainties.

Summary and Conclusions

The aim of this work has been the measurement of the cosmic ray flux above 1 EeV using data from the Pierre Auger Observatory. The high statistics of the surface detector allows a good estimation of the cosmic ray flux above an energy of 3 EeV, with 100% trigger efficiency. This bound could be lowered to $E = 1.25\,\mathrm{EeV}$ here by measuring the low energy trigger efficiency with hybrid data.

For the purpose of measuring the cosmic ray flux, a number of improvements have been introduced during this thesis.

- The estimation of the signal in the tanks is done after rejecting PMTs with unreasonable calibration constants. The rejection is not affecting the acceptance, but only individual events.
- A signal recovery procedure has been developed. It considers the pulse shape of the signal, the FADC overflow and the PMT non-linearity. The dynamic range of the signal recorded in the stations is increased with 3 orders of magnitude, improving the quality of the SD reconstruction. The mean of the relative difference in the energy of hybrid and SD events is improved from $-7.1(\pm2)\%$ to $-1.4(\pm2)\%$. The uncertainties for values above 10^5 VEM are large, 70%. Only a *hardware* solution could give a better resolution [210].
- An initial parameterization of the lateral distribution function was obtained from data. An improved parameterization [111] is used in this work.
- The shape of the attenuation curve, deduced with the constant intensity cut method, has been tested with a polynomial in $x = \cos^2\theta$ and an exponential in $\sec\theta$. The data, above an intensity of 0.18 events/(km^2 sr year) are best described by

$$S_{38} = S(1000\,\mathrm{m}) / \left(1 + \left(0.919 \pm 0.055\,\left(^{+0.02}_{-0.09}\right)\right)\,x - \left(1.13 \pm 0.26\,\left(^{+0.04}_{-0.19}\right)\right)\,x^2\right).$$

- The upward fluctuations in the energy calibration caused by the SD trigger efficiency are rejected with an anti-bias cut deduced from a toy Monte Carlo. The properties of the golden hybrid data are used as input, and the generated energy resolution is in agreement with the data.
- It was checked that the selection of good quality hybrid events do not introduce any systematic uncertainties in the final energy conversion. The events reconstructed with the average aerosol profiles, giving similar energy resolution, are included in the data set used for the energy calibration.

- The conversion from S_{38} to energy is a power law with an index of $1.096 \pm 0.011^{+0.008}_{-0.014}$ and a normalization of $0.144 \pm 0.006^{+0.008}_{-0.006}$ EeV.
- A direct energy calibration was obtained from the golden hybrid events without the constant intensity assumption. It was used as cross-check for the attenuation curve and for the energy conversion. The results are in very good agreement with the standard algorithm. The energy resolution, computed as $(E_{SD} - E_{FD})/\langle E \rangle$ is 20.9%.
- A possible evolution with energy of the S(1000 m) attenuation curve is observed, but it is not statistically significant. The evolution might correspond to a change in the cosmic ray composition from a light to heavy component above 10 EeV. More data are needed in the high energy range for a reasonable conclusion.

Simulations of the surface and fluorescence detector have been made according to a realistic energy spectrum. The energy calibration, applying exactly the same methods as for data, has been deduced for proton and iron compositions.

- The air shower asymmetries have a small effect, i.e. less than 300 m in the core location. The uncertainties of S(1000 m) approximately follow a log-normal distribution. The S_{38}, for the same energy is lower for simulations than for data. The corrections to be applied on S_{38} are 1.1-1.3 for iron and 1.4-1.5 for proton depending on energy. The attenuation curve differs mostly at inclined zenith angles. The results from [161], that the energy difference is mainly due to muons are in agreement with the findings in this thesis.
- The energy resolution of 20% of the observed air showers corresponds to a proton or a mixed composition, the iron energy resolution being 17%.
- A migration matrix has been constructed as a first step to estimate the detector resolution effect of the measured spectrum.

The acceptance of the detector has been evaluated from the hybrid data and its systematic uncertainties were deduced from simulations.

- The threshold energy where the SD trigger is full efficient for all zenith angles is 3 EeV, confirming previous results. The acceptance in the 100% trigger efficiency is more than 6900 km^2 sr year for a period of time extending from the 1st of January 2004 to 31st August 2007.
- The trigger probability has been computed down to energies of 0.5 EeV, allowing an extension of the vertical spectrum in lower energies.
- The effects of raining PMTs and of temperature on the measured spectrum are estimated and corrected for.

The systematics in the energy that are affecting the measurement of the vertical energy spectrum will be improved once the absolute scale of the fluorescence yield and its dependence on the atmosphere parameters will be known. The temperature induced uncertainty can be

reduced by applying the temperature corrections on S_{38} [183]. The contributions of the energy calibration and attenuation curve will be minimized drastically with more golden hybrid events available. The systematic uncertainties of the trigger of 10% to 5% in the region lower than the 3 EeV will vanish by using the data of the infill array which will allow the extension of the range of full trigger efficiency region. The behavior of raining PMTs will be stored in online data files. Their rejection in further reconstructions will be possible [211].

The vertical, hybrid and horizontal spectrum are combined with a novel maximum likelihood method taking into account the systematic uncertainties of each measurement. Two spectral features are observed: The ankle at $E = 3.3 \pm 1\,\mathrm{EeV}$ and a flux suppression starting at $E = 32.4 \pm 1.1\,\mathrm{EeV}$. The measured spectral index changes from $\gamma_1 = 3.26 + -0.03$, to $\gamma_2 = 2.65 \pm 0.02$ and at highest energies to $\gamma_3 = 3.9 \pm 0.3$. As a crosscheck the position of the ankle was inferred also from S(1000 m) integrated spectra and its value, converted to S_{38}, of $17 \pm 1\,\mathrm{VEM}$ corresponds to an energy value of $E = 3.1 \pm 1\,\mathrm{EeV}$.

A forward folding is applied to the vertical spectrum with the ansatz of a simple model of a transition from galactic to the extragalactic component. The function, which describes well the measured flux has the form of simple power laws and a flux suppression described by a Fermi function. The true flux is estimated to be lowered by more than 5% in the whole energy range due to detector resolution effects.

The currently en vogue dip model predicts a source spectrum with an index of -2.49 ± 0.02 when applied to the vertical spectrum with the detector resolution taken into account. The data are not described satisfactorily by this model, the reduced χ^2 being 4.54 for the cosmic ray flux fit.

The continuation of the spectrum in the form of a power law is rejected with a $5.2\,\sigma$ significance. The number of events expected if a power law holds, above 40 EeV are 167 ± 3 and 35 ± 1 above 100 EeV whereas in the measured flux there are 64 events and 1 event.

The energy where the flux is suppressed by 50% matches the energy threshold at which the correlation is maximized in the anisotropy found by the Pierre Auger Collaboration [55]. The mentioned energy allows, if attributed to the GZK effect, a visible horizon of $\approx 100\,\mathrm{Mpc}$. This number is almost the same as the maximum distance found for the correlation with the AGN catalog, suggesting that the flux suppression at high energies is indeed a GZK-effect.

The presented cosmic ray flux, due to its small uncertainties, allows already to distinguish between different astrophysical models. The shape of the GZK-region carries information about the distribution and the injection spectrum of the sources [212], but up to now it is dominated by statistical uncertainties. To answer this question is one of the tasks of the planned northern part of the Pierre Auger Observatory.

Bibliography

[1] J. Linsley, L. Scarsi, and B. Rossi, *Phys. Rev. Lett.* **6**, 485 (1961).

[2] A. A. Penzias and R. W. Wilson, *Astrophys. J.* **142**, 419 (1965).

[3] D. Fixsen et al., *Astrophys. J.* **420**, 445 (1994).

[4] D. N. Spergel et al., *Astrophys. J. Suppl.* **148**, 175 (2003), astro-ph/0302209.

[5] K. Greisen, *Phys. Rev. Lett.* **16**, 748 (1966).

[6] G. Zatsepin and V. Kuzmin, *JETP Lett.* **4**, 78 (1966).

[7] T. Abu-Zayyad et al., *Astropart. Phys.* **23**, 157 (2005), astro-ph/0208301.

[8] R. Abbasi et al., *submitted to Phys. Rev. Lett.* (2007), astro-ph/0703099.

[9] M. Takeda et al., *Phys. Rev. Lett.* **81**, 1163 (1998), astro-ph/9807193.

[10] J. Abraham et al., *Nucl. Instrum. Meth.* **A523**, 50 (2004).

[11] S. L. Dubovsky and P. G. Tinyakov, *Astropart. Phys.* **18**, 89 (2002), astro-ph/0106472.

[12] T. Kifune, *Astrophys. J.* **518**, L21 (1999), astro-ph/9904164.

[13] G. Amelino-Camelia and T. Piran, *Phys. Rev.* **D64**, 036005 (2001), astro-ph/0008107.

[14] S. Coleman and S. Glashow, *Phys. Rev.* **D59**, 116008 (1999), hep-ph/9812418.

[15] V. Berezinsky, M. Kachelriess, and A. Vilenkin, *Phys. Rev. Lett.* **79**, 4302 (1997).

[16] I. F. M. Albuquerque, G. R. Farrar, and E. W. Kolb, *Phys. Rev.* **D59**, 015021 (1999), hep-ph/9805288.

[17] P. Bhattacharjee, *Phys. Rev. Lett.* **81**, 260 (1998), hep-ph/9803223.

[18] P. Bhattacharjee and G. Sigl, *Phys. Rev.* **D51**, 4079 (1995), astro-ph/9412053.

[19] V. Hess, *Z. Phys.* **13**, 1084 (1912).

[20] B. Rossi, *Phys. Rev.* **45**, 212 (1934).

[21] P. Auger, R. Maze, and T. Grivet-Mayer, *Compt. Rend. Acad. Sci. (Ser. II)* **206**, 1721 (1938).

[22] P. Auger, P. Ehrenfest, R. Maze, J. Daudin, and A. Robley, *Rev. Mod. Phys.* **11**, 288 (1939).

[23] W. Kolhörster et al., *Naturw.* **26**, 567 (1938).

[24] A. Hillas, *Conf. on Cosmology, Galaxy Formation and Astro-Particle Physics on the Pathway to the SKA, Oxford, England* (2006), astro-ph/0607109.

[25] C. D. Anderson, *Phys. Rev.* **43**, 491 (1933).

[26] J. C. Street and E. C. Stevenson, *Phys. Rev.* **52**, 1003 (1937).

[27] C. M. G. Lattes, G. P. S. Occhialini, and C. F. Powell, *Nature* **160**, 486 (1947).

[28] C. M. G. Lattes, G. P. S. Occhialini, and C. F. Powell, *Nature* **160**, 453 (1947).

[29] C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini, and C. F. Powell, *Nature* **159**, 694 (1947).

[30] E. Fermi, *Phys. Rev.* **75**, 1169 (1949).

[31] T. Abu-Zayyad et al., *Astrophys. J.* **557**, 686 (2001), astro-ph/0010652.

[32] N. Grigorov, V. Nesterov, I. Rapoport, I. Savenko, and G. Skuridin, *Yad. Fiz.* **11**, 1058 (1970).

[33] D. Bird et al., *Astrophys. J.* **424**, 491 (1994).

[34] J.P. Wefel, 2007, private communication.

[35] T. Antoni et al., *Astropart. Phys.* **24**, 1 (2005), astro-ph/0505413.

[36] V. Derbina et al., *Astrophys. J.* **628**, L41 (2005).

[37] G. Kulikov and G. Khristiansen, *Sov. Phys. JETP* **35**, 441 (1959).

[38] D. Bird et al., *Phys. Rev. Lett.* **71**, 3401 (1993).

[39] A. D. Erlykin and A. W. Wolfendale, *J. Phys.* **G23**, 979 (1997).

[40] A. D. Erlykin and A. W. Wolfendale, *Astropart. Phys.* **18**, 151 (2002).

[41] R. Nam, S. Nikolsky, V. Pavluchenko, A. Chubenko, and V. Yakovlev, *Proc. 14th Int. Cosmic Ray Conf, Munich* , 2258 (1975).

[42] V. Ptuskin and V. Zirakashvili, *A&A* , 755 (2005).

[43] J. Candia, E. Roulet, and L. N. Epele, *JHEP* **12**, 033 (2002).

[44] H. Ulrich et al., *Nucl. Phys. (Proc. Suppl.)* **122**, 218 (2003).

[45] V. P. Egorova et al., *Nucl. Phys. Proc. Suppl.* **136**, 3 (2004).

[46] T. K. Gaisser, *Proc. 28th Int. Cosmic Ray Conf., Tsukuba, Japan* .

[47] V. Berezinsky, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007), arXiv:0710.2750 [astro-ph].

[48] T. Stanev, R. Engel, A. Mücke, R. Protheroe, and J. Rachen, *Phys. Rev.* **D62**, 093005 (2000), astro-ph/0003484.

[49] D. De Marco and T. Stanev, *Phys. Rev.* **D72**, 081301 (2005).

[50] T. Wibig and A. W. Wolfendale, *J. Phys.* **G31**, 255 (2005).

[51] D. Allard, E. Parizot, and A. Olinto, *Astropart. Phys.* **27**, 61 (2007), astro-ph/0512345.

[52] D. Allard, A. Olinto, and E. Parizot, *submitted to Astron. Astrophys.* (2007), astro-ph/0703633.

[53] V. Berezinsky, S. Grigoreva, and B. Hnatyk, *Nucl. Phys. (Proc. Suppl.)* **151**, 497 (2006).

[54] M. Roth for the Pierre Auger Collaboration, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007), arXiv:0706.2096 [astro-ph].

[55] J. Abraham et al., *Science* **318**, 939 (2007), arXiv:0711.2256 [astro-ph].

[56] V. Ginzburg and S.I.Syrovatskii, *Pergamon Press, Russian edition* (1963).

[57] P. L. Biermann and P. A. Strittmatter, *Astrophys. J.* **322**, 643 (1987).

[58] J. Rachen and P. Biermann, *Astron. Astrophys.* **272**, 161 (1993), astro-ph/9301010.

[59] R. Protheroe and A. Szabo, *Phys. Rev. Lett.* **69**, 2885 (1992).

[60] G. Benford and R. Protheroe, (2007), arXiv:0706.4419 [astro-ph].

[61] T. Pierog, private communication.

[62] T. Bergmann et al., *Astropart. Phys.* **26**, 420 (2007), astro-ph/0606564.

[63] J. Matthews, *Astropart. Phys.* **22**, 387 (2005).

[64] R. Engel, private communication.

[65] R. Ulrich, private communication.

[66] R. Abbasi et al., *Astrophys. J.* **622**, 910 (2005), astro-ph/0407622.

[67] M. Unger for the Pierre Auger Collaboration, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007).

[68] N. Kalmykov, S. Ostapchenko, and A. Pavlov, *Nucl. Phys. (Proc. Suppl.)* **B52**, 17 (1997).

[69] S. Ostapchenko, *Phys. Rev.* **D74**, 014026 (2006), hep-ph/0505259.

[70] R. Engel, T. Gaisser, T. Stanev, and P. Lipari, *Proc. 26th Int. Cosmic Ray Conf., Salt Lake City, Utah* **1**, 415 (1999).

[71] R. Fletcher, T. Gaisser, P. Lipari, and T. Stanev, *Phys. Rev.* **D50**, 5710 (1994).

[72] K. Werner and T. Pierog, *AIP Conf. Proc.* **928**, 111 (2007), arXiv:0707.3330 [astro-ph].

[73] S. Ostapchenko, *Nucl. Phys. Proc. Suppl.* **151**, 143 (2006), hep-ph/0412332.

[74] J. Alvarez-Muniz, R. Engel, T. Gaisser, J. Ortiz, and T. Stanev, *Phys. Rev.* **D69**, 103003 (2004), astro-ph/0402092.

[75] L. Landau and I. Pomeranchuk, *Dokl. Akad. Nauk Ser. Fiz.* **92**, 535 (1953).

[76] A. Migdal, *Phys. Rev.* **103**, 1811 (1956).

[77] D. Heck, J. Knapp, J. Capdevielle, G. Schatz, and T. Thouw, *FZKA-6019* (1998).

[78] D.Newton, J.Knapp, and A.A.Watson, The Optimum Ground Parameter, S(r_{opt}), GAP-2005-013.

[79] D.Newton, J.Knapp, and A.A.Watson, The Optimum Distance at which to Determine the Size of a Giant Air Shower, GAP-2006-045.

[80] W. D. Apel et al., *Astropart. Phys.* **24**, 467 (2006).

[81] T. Antoni et al., *Astropart. Phys.* **14**, 245 (2001).

[82] D. Barnhill for the Pierre Auger Collaboration, *Proc. 29th Int. Cosmic Ray Conf., Pune, India* (2005), astro-ph/0507590.

[83] S. Klepser, F. Kislat, H. Kolanoski, P. Niessen, and A. Van Overloop, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007).

[84] M. A. Lawrence, R. J. O. Reid, and A. A. Watson, *J. Phys.* **G17**, 733 (1991).

[85] C. L. Pryke and J. Lloyd-Evans, *Nucl. Instrum. Meth.* **A354**, 560 (1995).

[86] X. Bertou et al., *NIM* **A568**, 839 (2006).

[87] M. Aglietta for the Pierre Auger Collaboration, *Proc. 29th Int. Cosmic Ray Conf., Pune, India* (2005).

[88] P. Ghia, A compact 3ToT as SD physics trigger for vertical ($\theta < 60°$) showers, GAP-2004-018.

[89] P. L. Ghia and G. Navarra, Analysis of $E_0 \approx 10^{18}$ eV events with the AUGER-EA surface detector, GAP-2003-007.

[90] E. Parizot, I. Lhenry, D. Allard, P. Ghia, and G. Navarra, First steps towards the definition of a "quality trigger" (T5) for the SD acceptance calculations, GAP-2004-023.

[91] W.Fulgione, About raining PMTs, GAP-2007-081.

[92] R. Sato and C. O. Escobar, *Proc. 29th Int. Cosmic Ray Conf., Pune, India* (2005).

[93] N. Budnev et al., Fabrication of the first 150 head electronics units-results of the acceptance tests, GAP-1999-043.

[94] S. Argiro et al., *IEEE Trans. Nucl. Sci.* **48**, 444 (2001).

[95] H. Gemmeke et al., *IEEE Trans. Nucl. Sci.* **47**, 371 (2000).

[96] A.Schmidt, T.Asch, M.Kleifges, H.-J.Mathes, and H.Gemmeke, New Third Level Trigger for the Fluorescence Telescopes, GAP-2007-118.

[97] N. Barenthien et al., *Proc. 28th Int. Cosmic Ray Conf., Tsukuba, Japan* **2**, 895 (2003).

[98] J. Rautenberg for the Pierre Auger Collaboration, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007).

[99] M. Nagano et al., *Astropart. Phys.* **22**, 235 (2004).

[100] M. Ave et al., *Astropart. Phys.* **28**, 41 (2007).

[101] F. Kakimoto et al., *Nucl. Instrum. Meth.* **A372**, 527 (1996).

[102] T. Waldenmaier, J. Blümer, and H. Klages, (2007), astro-ph/0709.1494.

[103] P. Bauleo et al., (2005), astro-ph/0507347.

[104] C. Aramo et al., (2005), astro-ph/0507577.

[105] S. BenZvi for the Pierre Auger Collaboration, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007), astro-ph/0706.3236.

[106] B. Keilhauer for the Pierre Auger Collaboration, *Proc. 29th Int. Cosmic Ray Conf., Pune, India* (2005), astro-ph/0507275.

[107] F. Arqueros et al., (2005), astro-ph/0507334.

[108] P. Allison for the Pierre Auger Collaboration, *Proc. 29th Int. Cosmic Ray Conf., Pune, India* (2005).

[109] S. Argiro for the Pierre Auger Collaboration, *Proc. 29th Int. Cosmic Ray Conf., Pune, India* (2005).

[110] CDAS software group, Event reconstruction, release v4r4.

[111] D. Veberic and M. Roth, SD Reconstruction; Offline Reference Manual, GAP-2005-035.

[112] Event selection group, The Official Event Selection as of March 2005, GAP-2005-023.

[113] C. Bonifazi and A. Letessier-Selvon, De-biasing the Station Start Time, GAP-2006-050.

[114] M. Ave et al., *Nucl. Instrum. Meth.* **A578**, 180 (2007).

[115] M. Ave, P. Bauleo, and T.Yamamoto, Signal Fluctuation in the Auger Surface Detector Array, GAP-2003-030.

[116] P. Bauleo, A. Castellina, R. Knapik, G. Navarra, and J. Harton, Auger Surface Detector Signal Accuracy – Results from production tanks data, GAP-2004-047.

[117] D. Barnhill for the Pierre Auger Collaboration, *Proc. 29th Int. Cosmic Ray Conf., Pune, India* (2005), astro-ph/0507590.

[118] T. Schmidt, I. C. Mariş, and M. Roth, Fine Tuning of the LDF parameterisation and the Influence on S1000, GAP-2007-106.

[119] K. Kamata and J. Nishimura, *Prog. Theoret. Phys. Suppl.* **6**, 93 (1958).

[120] K. Greisen, *Progress in Cosmic Ray Physics*, volume 3, North-Holland, Amsterdam, 1956.

[121] F. Arneodo, S. Maldera, and G.Navarra, A procedure to derive the anode charges for saturated signals from the dynode and anode undershoots, GAP-2005-025.

[122] I. D. Mitri, S.Maglio, and D. Martello, A method for SD traces saturation recovery, GAP-2006-075.

[123] I. C. Mariş, M. Roth, and T. Schmidt, A Phenomenological Method to Recover the Signal from Saturated Stations, GAP-2006-012.

[124] D. Supanitsky, A. Watson, and M. Zha, Fluctuations of the pulse shape, GAP-2005-015.

[125] J. L. G. Maurin, New parameter for iron, proton and photon discrimination usig FADC traces, GAP-2004-071.

[126] M.Aglietta, A.Castellina, W.Fulgione, P.L.Ghia, F.Gomez, and C.Morello, PAO:SD test bench and testing procedures for PMT XP1805/D1 base production, GAP-2003-034.

[127] M. Unger, B. Dawson, R. Engel, F. Schüssler, and R. Ulrich, *Reconstruction of Longitudinal Profiles of Ultra-High Energy Cosmic Ray Showers from Fluorescence and Cherenkov Light Measurements, submitted to NIM* (2007).

[128] R. Baltrusaitis et al., *Nucl. Instr. Meth* **240**, 420 (1985).

[129] M. Mostafa, Checking the alignment of the FD cameras, GAP-2005-102.

[130] R. Baltrusaitis et al., *Nucl. Instrum. Meth.* **A240**, 410 (1985).

[131] B. Dawson, Fluorescence Detector Techniques, GAP-1996-017.

[132] B. Dawson et al., Present and Possible Future Implementations of Fluorescence Yield in FD Analysis, GAP-1996-017.

[133] B. Dawson, Suggested Improvements for the Treatment of FD Light in the Offline, GAP-2004-055.

[134] P. Sokolsky, *"Introduction to Ultrahigh Energy Cosmic Ray Physics"*, Addison-Wesley, 1989.

[135] A. M. Hillas, *J. Phys. G* **8**, 1461 (1982).

[136] J. W. Elbert et al., *Proc. 18th ICRC* (1983).

[137] M. Giller, G. Wieczorek, A. Kacperczyk, H. Stojek, and W. Tkaczyk, *J. Phys.* **G30**, 97 (2004).

[138] F. Nerling et al., Description of Cherenkov light production in high-energy air showers, 2005, submitted to Astro. Part. Phys., astro-ph/0506729.

[139] T.K. Gaisser and A.M. Hillas, *Proc. 15th Int. Cosmic Ray Conf., Plovdiv, Bulgaria,* **8**, *353, 1977* .

[140] Z. Cao et al., *Proc. 28th Int. Cosmic Ray Conf. 2003* .

[141] H. Barbosa et al., *Astropart. Phys.* **22**, 159 (2004).

[142] T. Pierog et al., *Proc. 29th Int. Cosmic Ray Conf., Pune, India* (2005).

[143] J. Hersil, I. Escobar, D. Scott, G. Clark, and S. Olbert, *Phys. Rev. Lett.* **6**, 22 (1961).

[144] P. Sommers for the Pierre Auger Collaboration, *Proc. 29th Int. Cosmic Ray Conf., Pune, India* (2005).

[145] M. Ave, N. Busca, L. Cazon, and F. Schmidt, Monte Carlo and Hybrid Studies of Shower-to-Shower Fluctuations , GAP-2007-052.

[146] M. Unger, private communication.

[147] B. R. Dawson for the Pierre Auger Collaboration, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007).

[148] M. Unger, Shower Profile Reconstruction from Fluorescence and Čerenkov light, GAP-2006-010.

[149] D. Allard et al., A New Method for the Longitudinal Profile Reconstruction of the AUGER Fluorescence Detector Events, GAP-2006-026.

[150] M. Kobal, *Astropart. Phys.* **15**, 259 (2001).

[151] P. Billoir, VIHKOS Corsika School, 2005.

[152] P. Billoir, Natural and Artificial Fluctuations in the Auger Surface Detector, GAP-2001-005.

[153] P. Billoir, Does the resampling procedure induce distortions in the FADC traces of the Surface Detector?, GAP-2005-109.

[154] S. Ostapchenko, *Phys. Lett.* **B636**, 40 (2006), hep-ph/0602139.

[155] G. Battistoni et al., *AIP Conf. Proc.* **896**, 31 (2007).

[156] W. Nelson, H. Hirayama, and D. Rogers, *SLAC-0265* (1985).

[157] J. Knapp, D. Heck, S. Sciutto, M. Dova, and M. Risse, *Astropart. Phys.* **19**, 77 (2003), astro-ph/0206414.

[158] H. Mielke, M. Foeller, J. Engler, and J. Knapp, *J. Phys.* **G20**, 637 (1994).

[159] T. Antoni et al., *Astropart. Phys.* **16**, 245 (2002), astro-ph/0102443.

[160] T. Antoni et al., *J. Phys.* **G27**, 1785 (2001), astro-ph/0106494.

[161] R. Engel for the Pierre Auger Collaboration, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007), arXiv:0706.1921 [astro-ph].

[162] W. Apel et al., *J. Phys.* **G34**, 2581 (2007).

[163] L. Prado et al., *Nucl. Instrum. Meth.* **A545**, 632 (2005).

[164] R. Ulrich, Measurement of the proton-air cross section using hybrid data of the Pierre Auger Observatory , PhD thesis, Universität Karlsruhe, 2007.

[165] R. Caruso for the Pierre Auger Collaboration, *Proc. 29th Int. Cosmic Ray Conf., Pune, India* (2005).

[166] T. Bergmann et al., *Astropart. Phys.* **26**, 420 (2007).

[167] S. Agostinelli et al., *Nucl. Instrum. Meth.* **A506**, 250 (2003).

[168] J. Allison et al., *IEEE Trans. Nucl. Sci.* **53**, 270 (2006).

[169] C.Pryke, Auger Project Technical Note - Geometrical Design Studies for Water Cerenkov Detectors via Simulation, GAP-1996-008.

[170] C.Pryke, Performance Simulations of a 10 m**2 Water Cerenkov Detector and Comparison with Experiment, GAP-1997-004.

[171] T.McCauley and T.Paul, GEANT4 Simulation of the Surface Detectors, GAP-2000-055.

[172] W. E. Slater, A. Tripathi, and K. Arisaka, A GEANT3 Simulation of Pierre Auger Surface Detector Response to Muons, GAP-2002-063.

[173] S.Ranchon and M.Urban, You thought you understood water tanks!, GAP-2003-047.

[174] T.Ohnuki, G.Rodriguez-Fernandez, D.Barnhill, A.Tripathi, T.McCauley, and T. K.Arisaka, A Systematic Comparison of Three Tank Simulators (G4/SDSim/FastSim) in the DPA Framework, GAP-2004-043.

[175] M.Zha and J.Knapp, Simulation of Surface Detector using Geant4, GAP-2005-031.

[176] A.Creusot and D.Veberic, Towards more realistic tank simulation, GAP-2007-073.

[177] P. L. Ghia for the Pierre Auger Collaboration, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007).

[178] D. Reyna, A simple parameterization of the cosmic-ray muon momentum spectra at the surface as a function of zenith angle, hep-ph/0604145.

[179] M. T. Dova, F. G. Monticelli, and H. Wahlberg, Influence of the reconstructed core position on S(1000) for inclined showers , GAP-2006-035.

[180] M. Prouza for the Pierre Auger Collaboration, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007), arXiv:0706.1719 [astro-ph].

[181] M.I.Micheletti and D.J.Melo, Effect of the Aerosol Profiles on Shower Reconstruction: A general and seasonal study, GAP-2007-104.

[182] B. Wilczynska et al., *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007).

[183] C. Bleve for the Pierre Auger Collaboration, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007), arXiv:0706.1491 [astro-ph].

[184] T. Abu-Zayyad et al., *Phys. Rev. Lett.* **84**, 4276 (2000).

[185] M. T. Dova et al., *Astropart. Phys.* **21**, 597 (2004).

[186] A. A. Watson, *Nucl. Phys. Proc. Suppl.* **136**, 290 (2004), astro-ph/0408110.

[187] D. Allard for the Pierre Auger Collaboration, *Proc. 29th Int. Cosmic Ray Conf., Pune, India* (2005).

[188] X. B. C. Lachaud, Local Stations Trigger Monitoring Files, GAP-2003-046.

[189] C.Bonifazi and A. Letessier-Selvon, Event selection using the T5 time distribution, GAP-2006-042.

[190] C.Bonifazi and P.Ghia, Selection of data periods and calculation of th SD geometrical acceptance, GAP-2006-101.

[191] M. Settimo, L. Perrone, G. Cataldi, I. D. Mitri, and D. Martello, Parameterisation of the "Lateral Trigger Probability" functions at low energies, GAP-2007-069.

[192] N. G. Busca, The Ultra High Energy Cosmic Ray Flux from the Southern Pierre Auger Observatory Data, GAP-2006-108.

[193] G. J. Feldman and R. D. Cousins, *Phys. Rev.* **D57**, 3873 (1998).

[194] J. Abraham et al., *Astropart. Phys.* **27**, 155 (2007), astro-ph/0606619.

[195] M. Healy for the Pierre Auger Collaboration, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007), arXiv:0710.0025 [astro-ph].

[196] M. Risse and P. Homola, *Mod. Phys. Lett.* **A22**, 749 (2007), astro-ph/0702632.

[197] D. Allard for the Pierre Auger Collaboration, *Proc. 29th Int. Cosmic Ray Conf., Pune, India* (2005).

[198] G. Cowan, *"Statistical Data Analysis"*, Oxford University Press, 1998.

[199] G. Zech, Comparing Statistical Data to Monte carlo Simulation - Parameter Fitting and Unfolding, 1995, DESY 95-113.

[200] A. Höcker and V. Kartvelishvili, *Nucl. Instr. and Meth.* **A372**, 469 (1996).

[201] V. Blobel, Unfolding methods in high-energy physics experiments, 1985, CERN 85-02.

[202] A. N. Tikhonov, *Sov. Math.* **5**, 1035 (1963).

[203] L. Lindemann and G. Zech, *Nucl. Instr. and Meth.* **A354**, 516 (1995).

[204] G. D'Agostini, *Nucl. Instr. and Meth.* **A362**, 487 (1995).

[205] J. D. Hague, B. R. Becker, M. S. Gold, J. A. J. Matthews, and J. Urbar, (2007), arXiv:0710.3600 [astro-ph].

[206] F. Schüssler, R. Engel, M. Unger, and R. Ulrich, *Measurement of the UHECR spectrum based on Auger Hybrid data, Auger Collaboration Meeting, Malargüe* (2007).

[207] L. Perrone for the Pierre Auger Collaboration, Measurement of the UHECR energy spectrum from hybrid data of the Pierre Auger Observatory, 2007, arXiv:0706.2643 [astro-ph].

[208] P. Facal San Luis for the Pierre Auger Collaboration, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007), arXiv:0706.4322 [astro-ph].

[209] R. Geary, *J. R. Statist. Soc.* **93**, 442.

[210] M. Aglietta, S. Maldera, G. Navarra, and F. Suarez, SD signals saturation: Summary of the "hardware solution" studies, GAP-2007-134.

[211] P. L. Ghia and I. Lhenry-Yvon, A method to identify 'raining' PMTs in SD, GAP-2007-132.

[212] T. Yamamoto for the Pierre Auger Collaboration, *Proc. 30th Int. Cosmic Ray Conf., Merida, Mexico* (2007), arXiv:0707.2638 [astro-ph].

[213] I.C.Mariş, F.Schuessler, R.Ulrich, and M.Unger, Data Summary Trees and Shower Visualization for Reconstructed Auger Events, GAP-2006-081.

[214] *ROOT, the object-oriented data analysis framework* , http://root.cern.ch.

[215] *The Auger Observer* , http://observer.fzk.de.

Acronyms

SD	surface detector
FD	fluorescence detector
FADC	Flash Analog Digital Converter
PMT	photomultiplier
HV	high voltage
PLD	programmable logic device
GPS	Global Positioning System
VEM	vertical equivalent muon
VCT	vertical and centrally through-going muon
asl	above sea level
CDAS	Central Data Aquisition System
ToT	time over threshold trigger
T1	local station trigger
T2	second station trigger
Thr2	second level threshold trigger
T3	second level lowest CDAS trigger
T4	physics trigger
4C1	configuration of one station with 3 close neighbors
T5	quality trigger
NIST	National Institute of Standards and Technology
LIDAR	Light Detection and Ranging
CLF	Central Laser Facility
LDF	lateral distribution function
NKG	Nishimura Kamata Greisen lateral distribution function
LED	light-emitting diode
EAS	extensive air showers
SDP	shower detector plane
LTP	Lateral Trigger Probability

ADST Advanced Data Summary Tree

GZK Greisen-Zatsepin-Kuzmin

AGN active galactic nuclei

UHECR ultra high energy cosmic ray

UV ultra violet

VAOD vertical aerosol optical depth

CORSIKA COsmic Ray SImulations for KAscade

KASCADE KArlsruhe Shower Core and Array DEtector

CMB cosmic microwave background

Appendix A

Reconstruction and Calibration

Data Summary Trees and Event Browser

Developping together with colleagues from Karlsurhe the data storage format for data and simulations was also part of this thesis [213]. The ADST format (Fig. A.1) and the shower visualization program, based only on ROOT libraries [214], are used within the whole collaboration. This format allows an easy exchange of analysis codes and results between different groups. The reconstructed events are provided on daily basis [215]. The ADST format was used for the main analysis in this thesis. The powerful shower visualization program enabled an easy debugging and understanding of the reconstruction and simulations output of the Offline software. The main tabs are shown in the next figures.

Figure A.1: The class structure of the ADST.

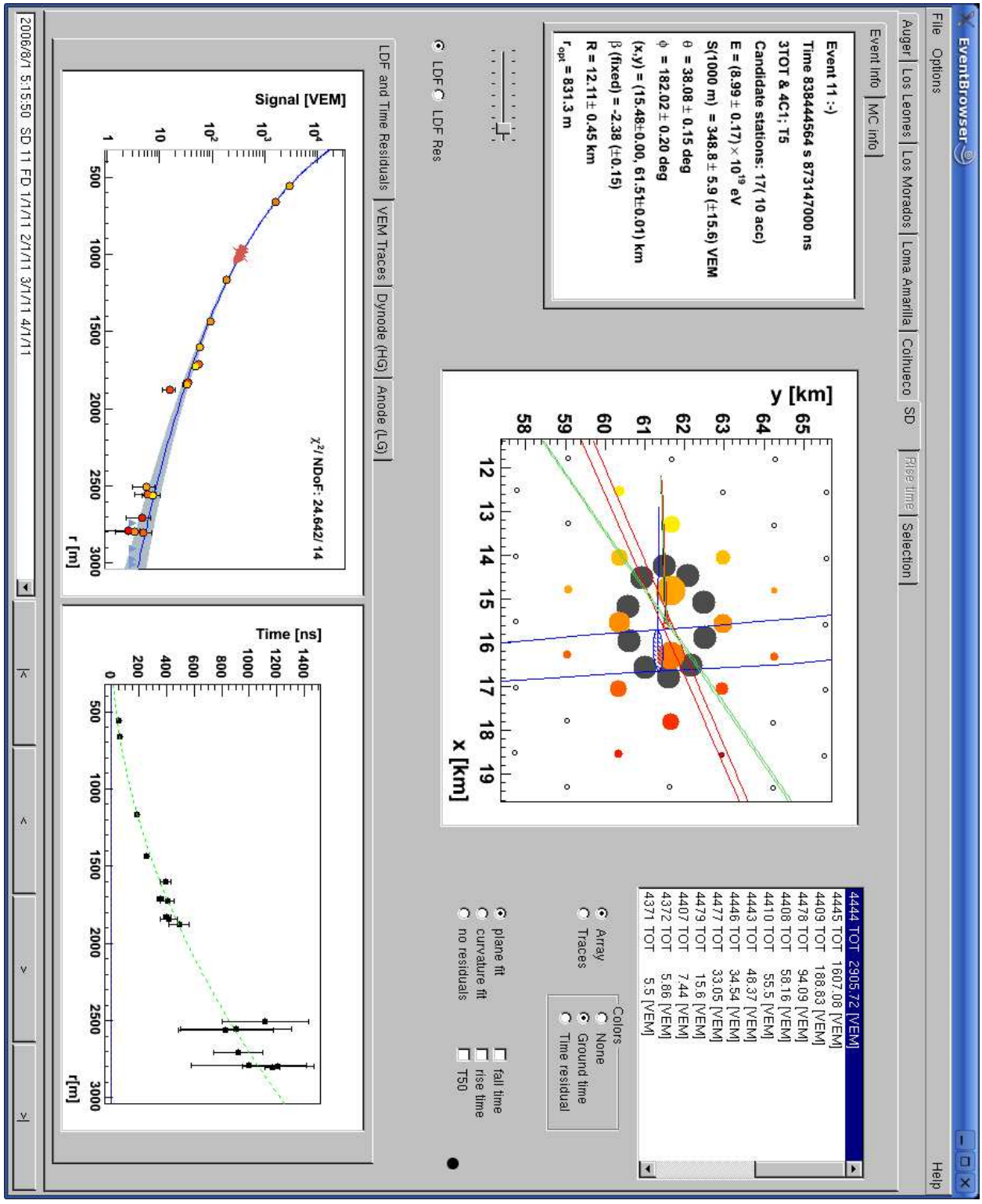


Figure A.2: The SD view. Upper left plot: Event information (Data and MC). Middle plot: Top view on array with all the stations. Upper left: List of stations in the array. Lower tabs: LDF, time residuals and traces.

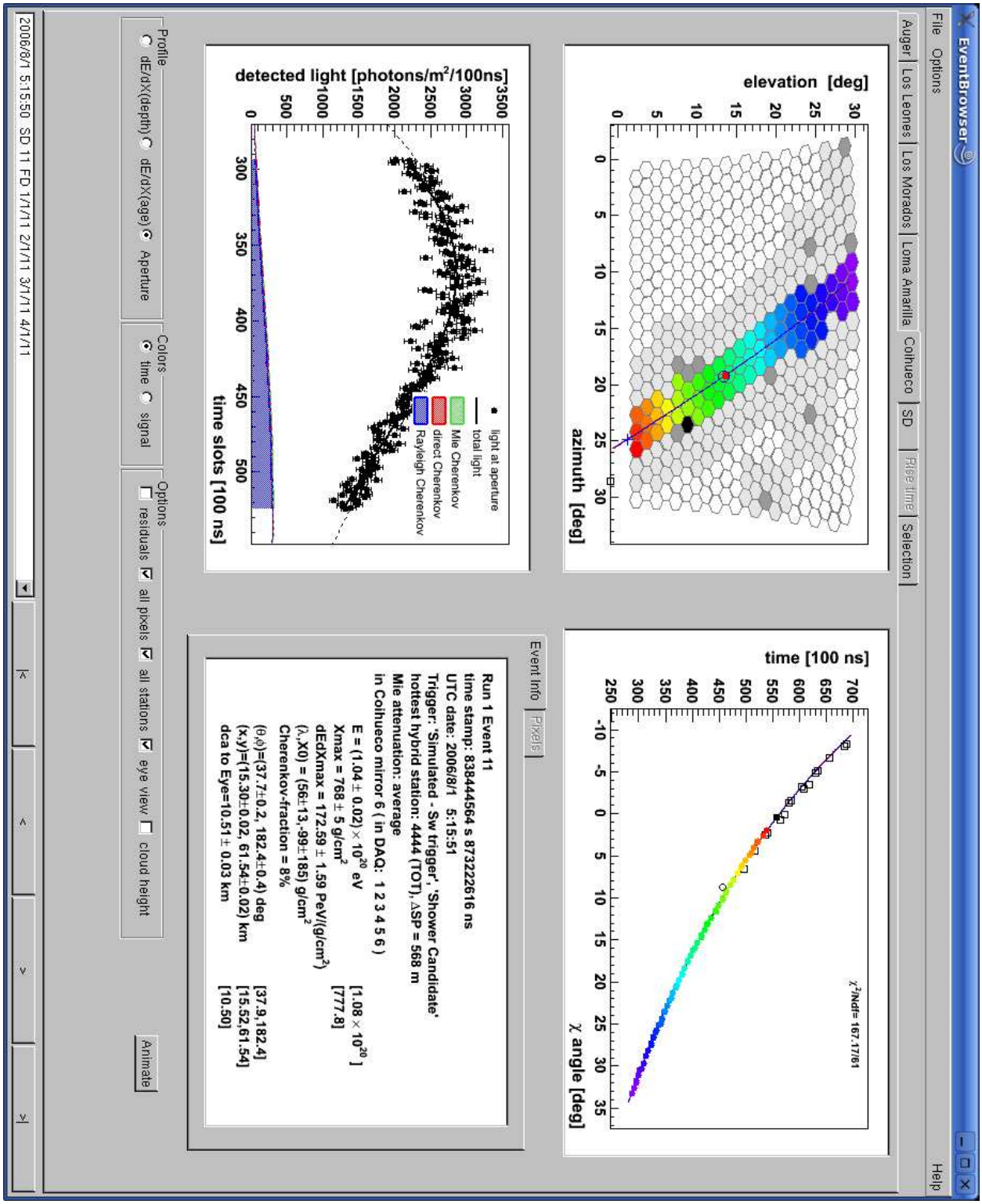


Figure A.3: The FD view. Upper right plot: Camera view. Upper right plot: Time fit. Lower left plot: Light at aperture and energy deposit. Lower right plot: Event informations and pixel ADC trace tabs.

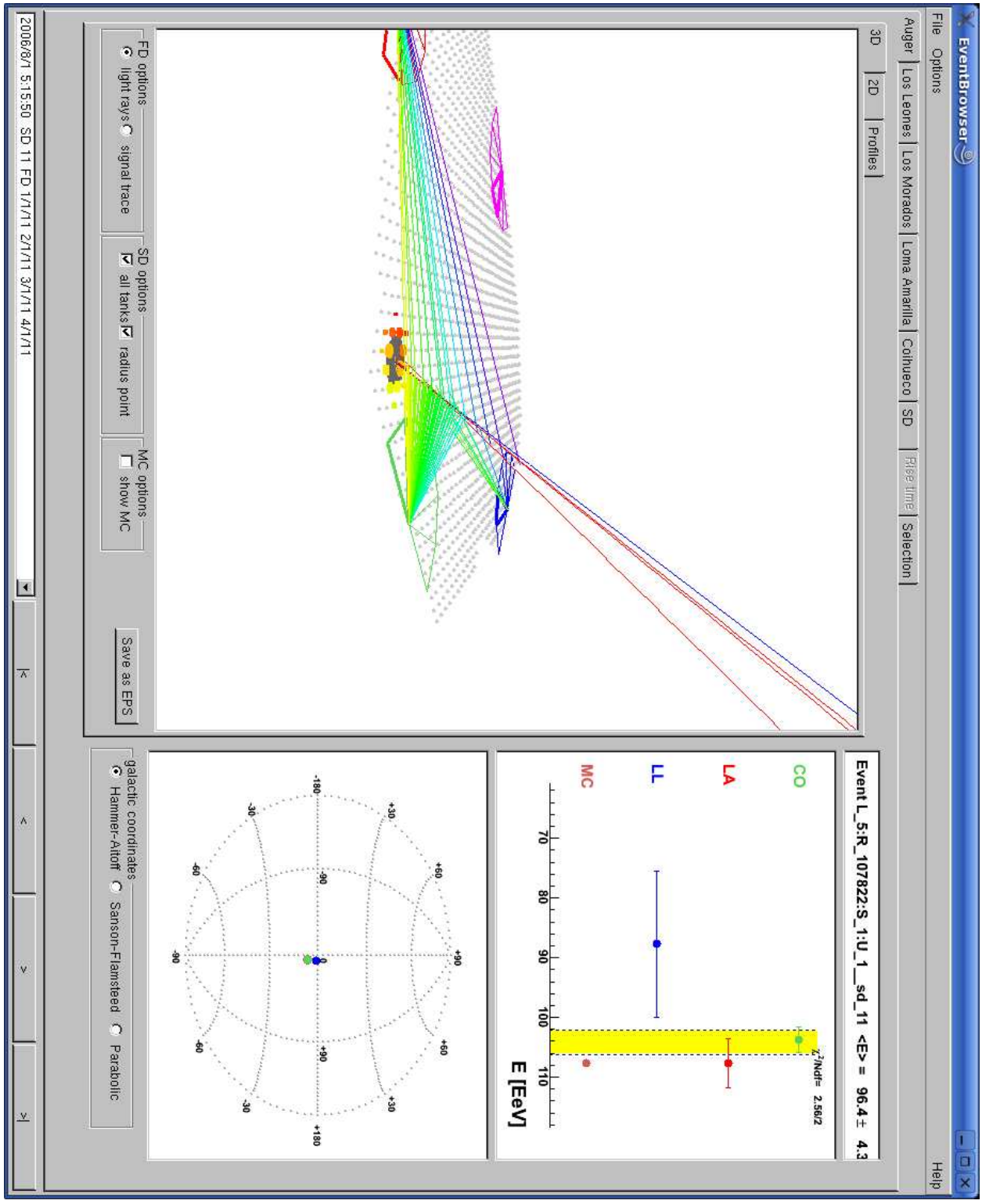
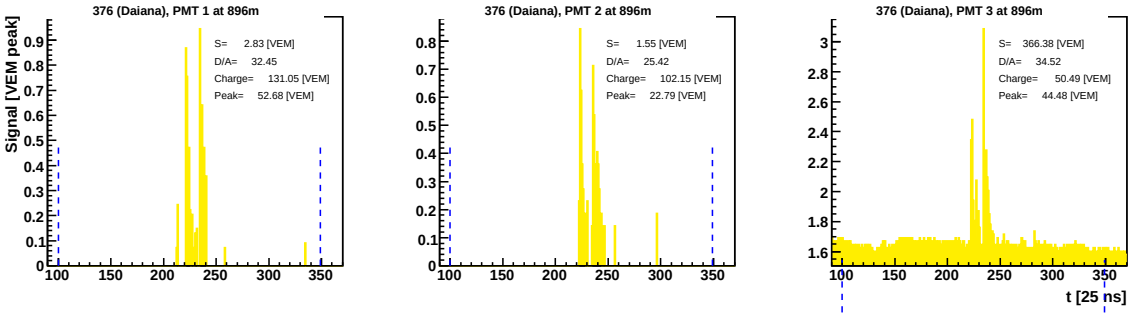
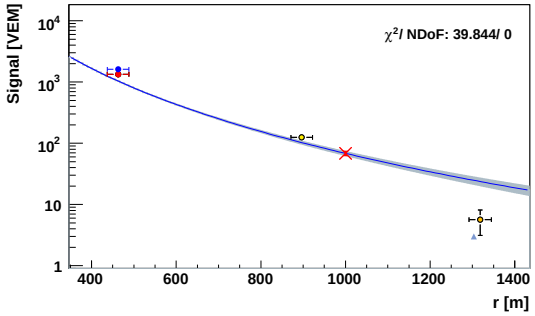


Figure A.4: The Auger view. Left plot: 3D view of the air shower. Upper right plot: FD/MC energy comparison. Lower right panel: comparison of the directions in galactic coordinates.

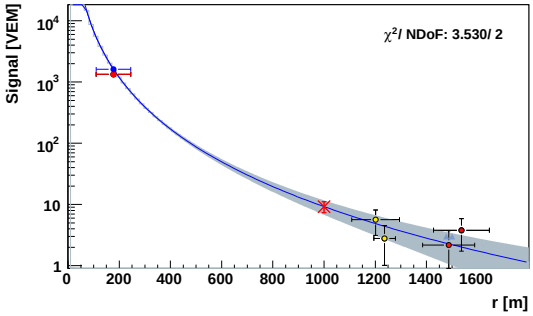
Calibration effects on the reconstruction



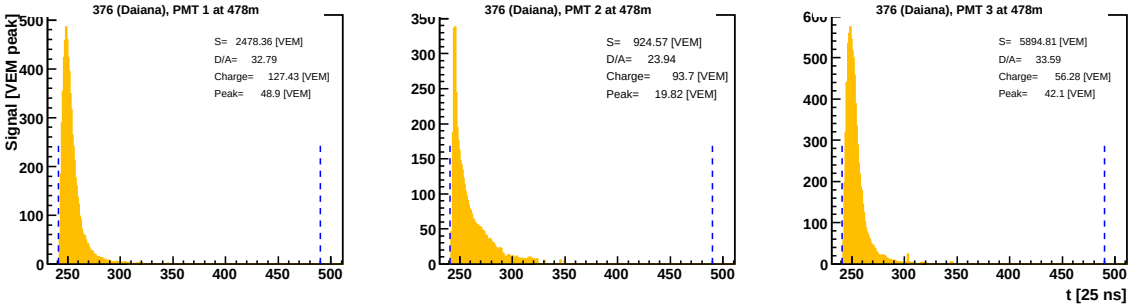
(a) *mis reconstructed baseline*



(b) *before cleaning*

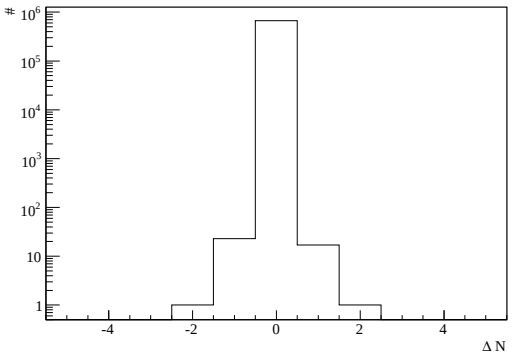


(c) *after cleaning*

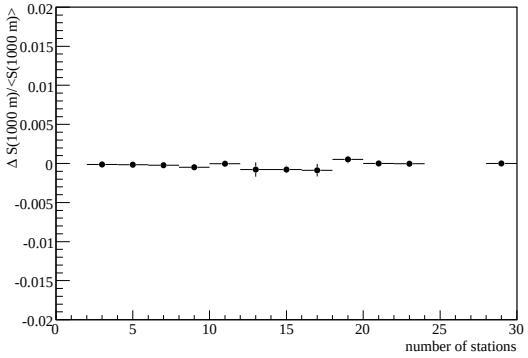


(d) *bad charge and peak values*

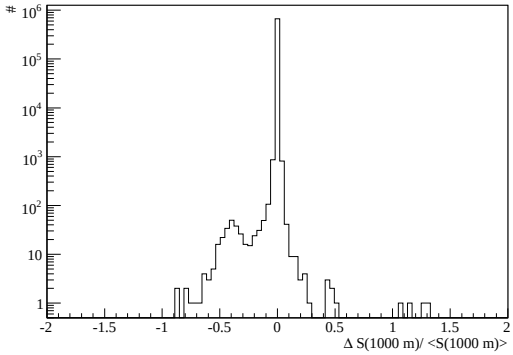
Figure A.5: Examples of bad calibration constants. (a) The third PMT has a base line not well calibrates giving a signal much larger than in the other PMTs. Due to this, some stations are not used in the LDF fit (b). Rejecting this PMT and using only the other two signals the reconstruction is improved (c). The energy of the event changes from 13 EeV to 1.3 EeV. (d) Example of bad peak and charge calibration. Rejecting in this case the 2 PMTs the energy changes from 44 EeV to 41 EeV even if the event is in a stable configuration.



(a) *difference in number of stations/event*



(b) *S(1000 m) difference vs nr. of stations*

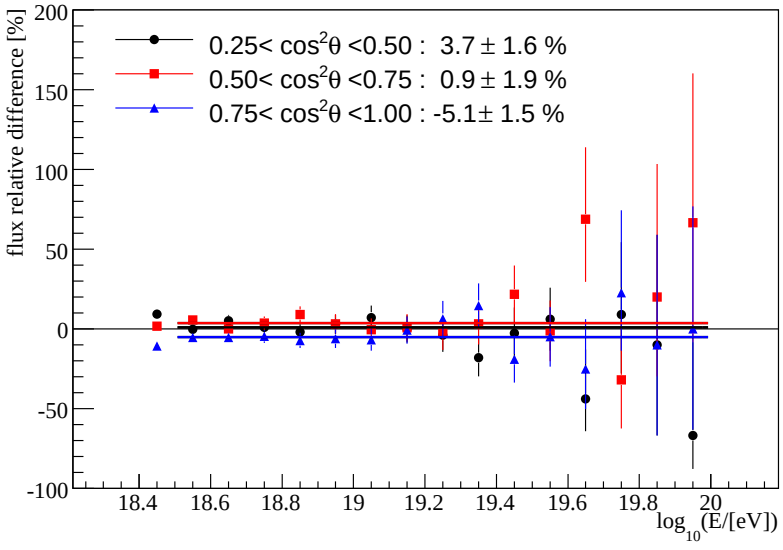


(c) *S(1000 m) difference*

Figure A.6: Impact on acceptance and S(1000 m). The rejection has no influence on the acceptance. The number of stations rejected is ≈ 300 out of ≈ 900.000 events. The cleaning of data at PMT level is important only for individual events.

Appendix B

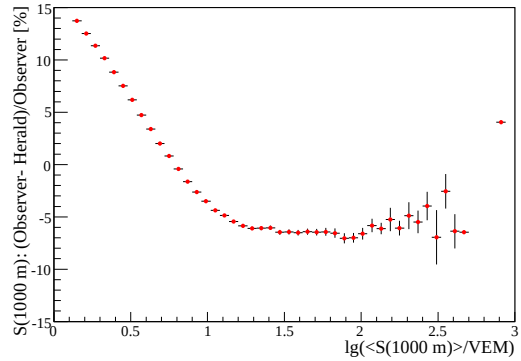
Energy spectrum: cross-checks and tables



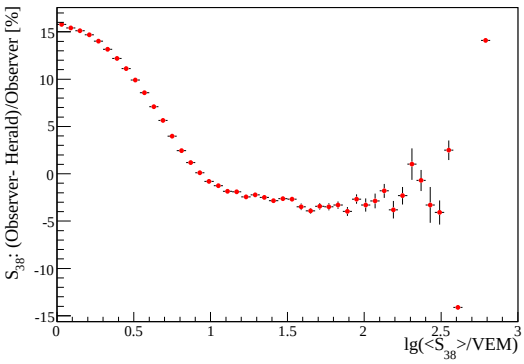
(a) *zenith angle*

Figure B.1: Cosmic ray flux: cross check for a zenith angle dependency. The 5% difference is within the systematic uncertainties deduced from the attenuation curve.

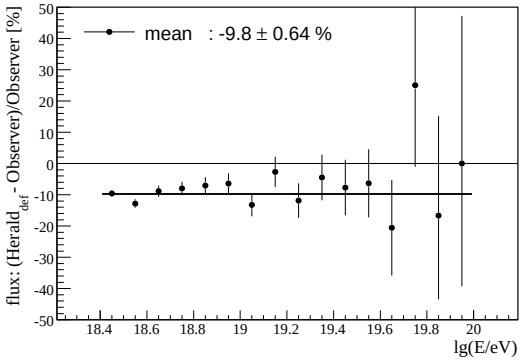
Different reconstruction algorithms Two reconstructions are compared: Observer, based on the Offline reconstruction algorithm and Herald, based on CDAS algorithm. The Herald reconstruction has been used in the anysotropies analysis, while the Observer was tested and used in this work. The main difference between the two algorithms is the different parameterization of the lateral distribution function and the usage of a different calibration procedure. The effect is seen mostly for the low energetic events, with only three triggered stations where S(1000 m) is completly determined by the shape of the LDF. The difference in S(1000 m) of 5% is compensated partially by applying different attenuation curves, becoming a 4% difference in S_{38}.



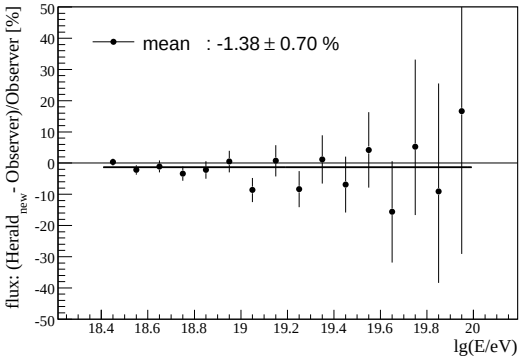
(a) *S(1000 m) difference*



(b) S_{38} *difference*



(c) *before recalibration*



(d) *after recalibration*

Figure B.2: Different reconstruction algorithms: Observer, based on the Offline reconstruction algorithm and Herald, based on CDAS algorithm. The Herald reconstruction has been used in the anysotropies analysis, while the Observer was partialy developped and used in this work.

Using the same energy conversion the flux difference is 10% the measured flux with Herald being higher. Applying a newly deduced energy calibration to the Herald data, the relative difference in flux becomes negligible, of 1%.

Table B.1. The T5 trigger efficiency. In brackets the systematic uncertainty, as deduced from simulations is given.

lg(E/eV)	N_{FD}	N_{SD}	θ	$\cos^2\theta$	$P\pm$stat (sys)
18.15	75	62	57.58	0.2875	$0.83 \pm 0.044(-0.14)$
18.15	101	80	52.98	0.3625	$0.79 \pm 0.04(-0.074)$
18.15	133	109	48.59	0.4375	$0.82 \pm 0.033(-0.072)$
18.15	152	133	44.28	0.5125	$0.88 \pm 0.027(-0.078)$
18.15	174	158	39.96	0.5875	$0.91 \pm 0.022(-0.069)$
18.15	131	115	35.52	0.6625	$0.88 \pm 0.029(-0.052)$
18.15	145	124	30.82	0.7375	$0.86 \pm 0.029(-0.058)$
18.15	144	117	25.66	0.8125	$0.81 \pm 0.033(-0.1)$
18.15	144	124	19.6	0.8875	$0.86 \pm 0.029(-0.069)$
18.15	142	111	11.17	0.9625	$0.78 \pm 0.035(-0.12)$
18.25	61	55	57.58	0.2875	$0.90 \pm 0.038(-0.032)$
18.25	83	75	52.98	0.3625	$0.90 \pm 0.032(-0.046)$
18.25	97	91	48.59	0.4375	$0.94 \pm 0.024(-0.027)$
18.25	95	88	44.28	0.5125	$0.93 \pm 0.027(-0.026)$
18.25	96	91	39.96	0.5875	$0.95 \pm 0.023(-0.019)$
18.25	126	119	35.52	0.6625	$0.94 \pm 0.02(-0.034)$
18.25	95	87	30.82	0.7375	$0.92 \pm 0.028(-0.027)$
18.25	91	82	25.66	0.8125	$0.90 \pm 0.031(-0.054)$
18.25	85	75	19.6	0.8875	$0.88 \pm 0.035(-0.042)$
18.25	83	76	11.17	0.9625	$0.92 \pm 0.031(-0.091)$
18.35	46	44	57.58	0.2875	$0.96 \pm 0.03(0.0048)$
18.35	76	70	52.98	0.3625	$0.92 \pm 0.031(0.0036)$
18.35	76	71	48.59	0.4375	$0.93 \pm 0.028(-0.023)$
18.35	75	74	44.28	0.5125	$0.99 \pm 0.013(-0.011)$
18.35	65	63	39.96	0.5875	$0.97 \pm 0.021(-0.015)$
18.35	77	76	35.52	0.6625	$0.99 \pm 0.013(0.0027)$
18.35	82	78	30.82	0.7375	$0.95 \pm 0.024(-0.012)$
18.35	66	62	25.66	0.8125	$0.94 \pm 0.029(-0.029)$
18.35	68	65	19.6	0.8875	$0.96 \pm 0.025(-0.015)$
18.35	51	49	11.17	0.9625	$0.96 \pm 0.027(-0.046)$
18.45	44	44	57.58	0.2875	$1.00 \pm 0(-0.0098)$
18.45	40	39	52.98	0.3625	$0.97 \pm 0.025(-0.0017)$
18.45	55	55	48.59	0.4375	$1.00 \pm 0(-0.0083)$
18.45	56	56	44.28	0.5125	$1.00 \pm 0(-0.0046)$
18.45	57	57	39.96	0.5875	$1.00 \pm 0(0)$
18.45	50	50	35.52	0.6625	$1.00 \pm 0(0)$
18.45	54	52	30.82	0.7375	$0.96 \pm 0.026(0)$
18.45	40	38	25.66	0.8125	$0.95 \pm 0.034(-0.0071)$
18.45	43	42	19.6	0.8875	$0.98 \pm 0.023(-0.014)$
18.45	33	31	11.17	0.9625	$0.94 \pm 0.042(-0.025)$