# Measurements of high energy cosmic rays above 10 PeV with KASCADE-Grande

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

genehmigte

DISSERTATION

von

Dipl. Phys. Fabiana Cossavella

aus Torino

Tag der mündlichen Prüfung: 17.07.2009
Referent: Prof. Dr. J. Blümer, Institut für Experimentelle Kernphysik
Korreferent: Prof. Dr. G. Navarra, Dipartimento di Fisica Generale, Universitá degli Studi di Torino

## Abstract

In the present work, different aspects of the measurement and reconstruction of extensive air showers generated by cosmic rays in the energy range between $10^{16}$ eV and $10^{18}$ eV are investigated. KASCADE-Grande detects charged particles of extensive air showers at ground level. From the energy deposits and the arrival times of the particles in the detector stations, the main parameters of the extensive air showers are reconstructed: the impact point, the direction of the shower axis, the total number of electrons ($N_e$), and the total number of muons ($N_\mu$) of the shower at observation level. The numbers of electrons and muons are related to the mass and energy of the primary particle and are the basis for further analysis.

The shape of the electron number spectrum reflects the shape of the cosmic ray energy spectrum. Knowing the reconstruction accuracies, the $N_e$ spectrum of the KASCADE-Grande data is studied in order to identify possible spectral features. It is found, that observed structures are consistent with the detector resolution and the reconstruction uncertainties.

On the basis of the correlation between the total number of muons in the showers and the total number of charged particles, a composition estimation is carried out. A *k Nearest Neighbours* (kNN) classification procedure is developed: a measured air shower is classified as proton-like or iron-like by finding its nine nearest neighbours in the parameter space of the selected mass sensitive observables. The total number of charged particles ($N_{ch}$) and the total number of muons ($N_\mu$) are found to be suitable observables for this purpose. The parameter space is populated by a set of simulated proton and iron induced air showers and an air shower is assigned to the class most common among its neighbours. The misclassification errors, as obtained in the training phase, are taken into account.

An increase of the relative number of iron-like induced air showers is observed between $10^{16}$ eV and $3.2 \cdot 10^{16}$ eV, which could be associated with the decrease of the cosmic ray light component (He, CNO group) due to the break in its energy spectrum. Above $3.2 \cdot 10^{16}$ eV, about 70% of the events result to be iron-like and $\approx$ 30% proton-like. At the highest observed energies an increase of light primaries is indicated.

The results presented in this work constitute a first attempt to carry out composition analysis with KASCADE-Grande data.

## Zusammenfassung

## Messung hochenergetischer kosmischer Strahlung oberhalb 10 PeV mit KASCADE-Grande

In der vorliegenden Arbeit werden verschiedene Aspekte der Messung und Rekonstruktion ausgedehnter Luftschauern untersucht, die aus der primären kosmischen Strahlung im Energiebereich von $10^{16}$ eV bis $10^{18}$ eV entstehen. Das KASCADE-Grande Experiment auf dem Gelände des Forschungszentrum Karlsruhe misst die geladenen Teilchen des ausgedehnten Luftschauers. Aus der Energieverlust und den Ankunftszeiten der Teilchen in den Detektorstationen werden die wichtigsten Parameter des ausgedehnten Luftschauers rekonstruiert: die Position des Schauerzentrums und die Einfallsrichtungs des Luftschauers, die Anzahl der Elektronen ($N_e$), und die Anzahl der Myonen ($N_\mu$) des Schauers auf Detektorniveau. Die Zahl der Elektronen und Myonen stehen in Zusammenhang mit der Masse und Energie der primären Teilchen, und bilden die Grundlage für die weitere Analyse.

Die Form des Elektronenzahl-Spektrums spiegelt die Form des Energiespektrums der primären Strahlung wider. Mit Kenntnis der Rekonstruktionsgenauigkeiten wurde das $N_e$ Spektrum der KASCADE-Grande Daten auf mögliche spektrale Eigenschaften des Primärspektrums untersucht. Beobachtete Strukturen in den gemessenen Spektren sind dabei im Einklang mit der Auflösung und der Unsicherheit in der Rekonstruktion der Observablen.

Mit Hilfe der Korrelation zwischen der Myonenzahl im Schauer und der Anzahl aller geladenen Teilchen wird eine Abschätzung der Elementzusammensetzung der kosmischen Strahlung durchgeführt. Dazu wurde ein so gennantes *k-nächste Nachbarn* (kNN) Klassifikationsverfahren entwickelt: ein gemessener Luftschauer wird auf der Basis der Suche nach seinen neun nächsten Nachbarn im Parameter-Raum ausgewählter masseempfindlicher Observablen als proton- oder eisenähnlicher Luftschauer klassifiziert. Der Anzahl der geladenen Teilchen ($N_{ch}$) und die Myonenzahl ($N_\mu$) sind als für diesen Zweck geeignete Messgrössen identifiziert worden. Der Parameter-Raum ist durch ein Satz simulierter protonen- und eiseninduzierte Luftschauer besiedelt und der gemessene Schauer wird der Klasse seiner häufigsten Nachbarn zugeordnet. Die Wahrscheinlichkeiten der Fehlklassifikation, bestimmt in der Trainingsphase, werden bei der Analyse berücksichtigt.

Eine Zunahme des relativen Anteils der eiseninduzierte Luftschauer wird im Energeibereich von $10^{16}$ eV bis $3.2 \cdot 10^{16}$ eV beobachtet. Dies kann im Zusammenhang mit dem Rückgang der leichten Komponente der kosmischen Strahlung (He,CNO Gruppe) aufgrund ihres erwateten Knies im Energiespektrum stehen. Oberhalb einer Energie von $3.2 \cdot 10^{16}$ eV werden etwa 70% der Ereignisse als eisenähnlich und $\approx 30\%$ als protonähnlich klassifiziert. Bei den höchsten beobachteten Energien zeichnet sich dann wieder ein Anstieg des Anteils der leichten Komponente ab.

Die in dieser Arbeit vorgelegten Ergebnisse sind ein erster Versuch, die KASCADE-Grande Daten auf die Elementzusammensetzung der kosmischen Strahlung zu analysieren.

# Content

# 1. Introduction

High energy charged (nuclei, electrons, positrons) or neutral (photons, neutrinos) particles coming from space and continuously hitting our atmosphere are generally referred to as *cosmic rays*. About 98% of this cosmic radiation are hadrons, the remaining 2% are composed of electrons and photons. 87% of the hadronic component are protons, 12% helium nuclei and the rest corresponds to fully ionised nuclei of heavier elements.

The energy range of the cosmic particles extends over 11 orders of magnitudes (from 1 up to $10^{11}$ GeV). Their energy spectrum, summed up over all particle types, is described by a power law ${}^{dN}\!/\!{}_{dE} \propto E^{\gamma}$, whose spectral index $\gamma$ is nearly constant over the full energy range. However, at an energy $E \approx 3 \cdot 10^{15}$ eV, a steepening of the power law is observed, the spectral index changing from $-2.7$ to $-3.1$. This discontinuity, known as the *knee*, is caused by a related feature occurring in the spectrum of the lighter component of cosmic rays, first for proton primaries and at higher energies for helium primaries [1; 2]. The origin of the knee is not yet known. The most accredited astrophysical models, which could explain the existence of a knee, are related to either the acceleration [3] or the propagation [4; 5] mechanism of the cosmic rays in the Galaxy. Both models predict that the knee of each primary element occurs at a constant rigidity of the particles. Features in the spectra of individual elements reflect into structures of the all particle energy spectrum and in a variation of the observed composition. In order to confirm or exclude different models, energy spectra of the individual elements in the cosmic radiation have to be studied.

Above $10^{14}$ eV, direct measurements of the cosmic rays are not feasible due to the low flux. Hence, the study has to be carried out through the detection of the extensive air showers, i.e. cascades of particles generated by the interaction of a high energy cosmic particle with the atoms of the atmosphere. These showers are detected by extended detector arrays at ground level and interpreted by comparison with computer simulated air showers, which rely on hadronic interaction models. The collecting area of the arrays has to be larger the higher the energy range of interest to compensate for the decrease of flux. KASCADE-Grande is the extension of the experiment KASCADE [6], whose aim was the investigation of the cosmic ray flux in the knee region. The energy range between $10^{16}$ eV and $10^{18}$ eV is investigated with KASCADE-Grande. According to the models and the results of KASCADE, a break in the energy spectrum of the iron component should appear at energies around $10^{17}$ eV, which also reflects in a change of the mass composition of cosmic rays.

The purpose of this work is to study the electron number spectrum and to determine the fraction of proton and iron induced air showers. For that, first the reconstruction of air showers measured with KASCADE-Grande is established and cuts to the reconstructed data are evolved, applied and cross-checked. This procedure allows to determine the air shower geometry, i.e. its arrival direction, its impact point, and the total number of muons ($N_\mu$) and electrons ($N_e$) at observation level with high precision. The numbers of electrons and muons are related to the mass and energy of the primary particle and are the basis for further analysis. By analysing the reconstruction of simulated events, the accuracies of the reconstructed observables are estimated.

The shape of the total electron number spectrum reflects the shape of the cosmic ray energy spectrum, the shower size being related to the primary energy. Knowing the reconstruction accuracies, the $N_e$ spectrum of the KASCADE-Grande data is studied, to detect possible spectral features. Also, a statistical method to derive the mass of the cosmic rays detected by KASCADE-Grande is developed. The method is based on the correlation between the total number of electrons and muons in the shower and the mass of the primary particle. A classification of the primary particles into two mass groups, light (proton-like) and heavy (iron-like), is performed through the k nearest neighbours (kNN) method; on this basis, the relative amounts of air showers belonging to each group is presented as a function of the estimated energy.

# 2. Cosmic rays

Cosmic rays were first discovered in 1912 by V. Hess [7], who showed an increasing ionisation rate with altitude, and further studied by several physicists as B. Rossi and P. Auger [8], who first detected an Extensive Air Shower (EAS). When a high energy particle enters the atmosphere, it starts to interact with air nuclei initiating a cascade of secondary interactions whose result is a shower of photons, leptons, mesons and hadrons reaching the ground. The first interacting particle is then often called *primary particle*, while the other ones are generally referred to as *secondaries*.

## 2.1 Energy spectrum

The cosmic ray spectrum extends over 11 orders of magnitudes (from 1 up to $10^{11}$ GeV), where the flux $F$ of particles with energy $E$ follows a steep power law with spectral index $\gamma$:

$$F(E) \propto E^{\gamma}. \tag{2.1}$$

The flux changes from $\approx 1000\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}$, at $E \approx 1\,\mathrm{GeV}$, to $1\,\mathrm{km}^{-2}\,\mathrm{century}^{-1}$ at the highest energies. Therefore, the detection techniques are widely different, with direct detection with balloon and satellite based experiments for E<100 TeV and indirect detection by means of extended detector arrays for higher energies.

In fig. 2.1, an overview of the measured differential energy spectra from different experiments is shown. The cosmic ray spectrum is almost featureless, except for the existence of two clear structures. At an energy $E \approx 3 \cdot 10^{15}\,\mathrm{eV}$, a steepening of the power law is observed, with the spectral index changing from $-2.7$ to $-3.1$, while at an energy of about $4 \cdot 10^{18}\,\mathrm{eV}$ it shows a flattening of the power law with the spectral index becoming $\approx -2.8$. The former is referred to as *knee* and the latter as *ankle* of the cosmic ray spectrum. The ankle is presumably due to a take over from the galactic component to a harder extra-galactic component. At an energy of approximately $3$–$7 \cdot 10^{17}$ eV some experiments [9; 10; 11] report a further steepening in the spectrum, usually referred to as the "second knee". The tail of the cosmic ray spectrum above $10^{20}$ eV is scarcely populated. A cutoff is predicted by Greisen, Zatsepin and Kuzmin at $5 \cdot 10^{19}\,\mathrm{eV}$, due to the interaction of the primary particles with the photons of the cosmic microwave background radiation. The high energy region of the cosmic ray spectrum is subject of investigation by the Pierre

Auger Observatory (PAO) experiment [12], to understand whether the flux decrease is consistent with the Greisen-Zatsepin-Kuzmin (GZK) effect [13; 14].

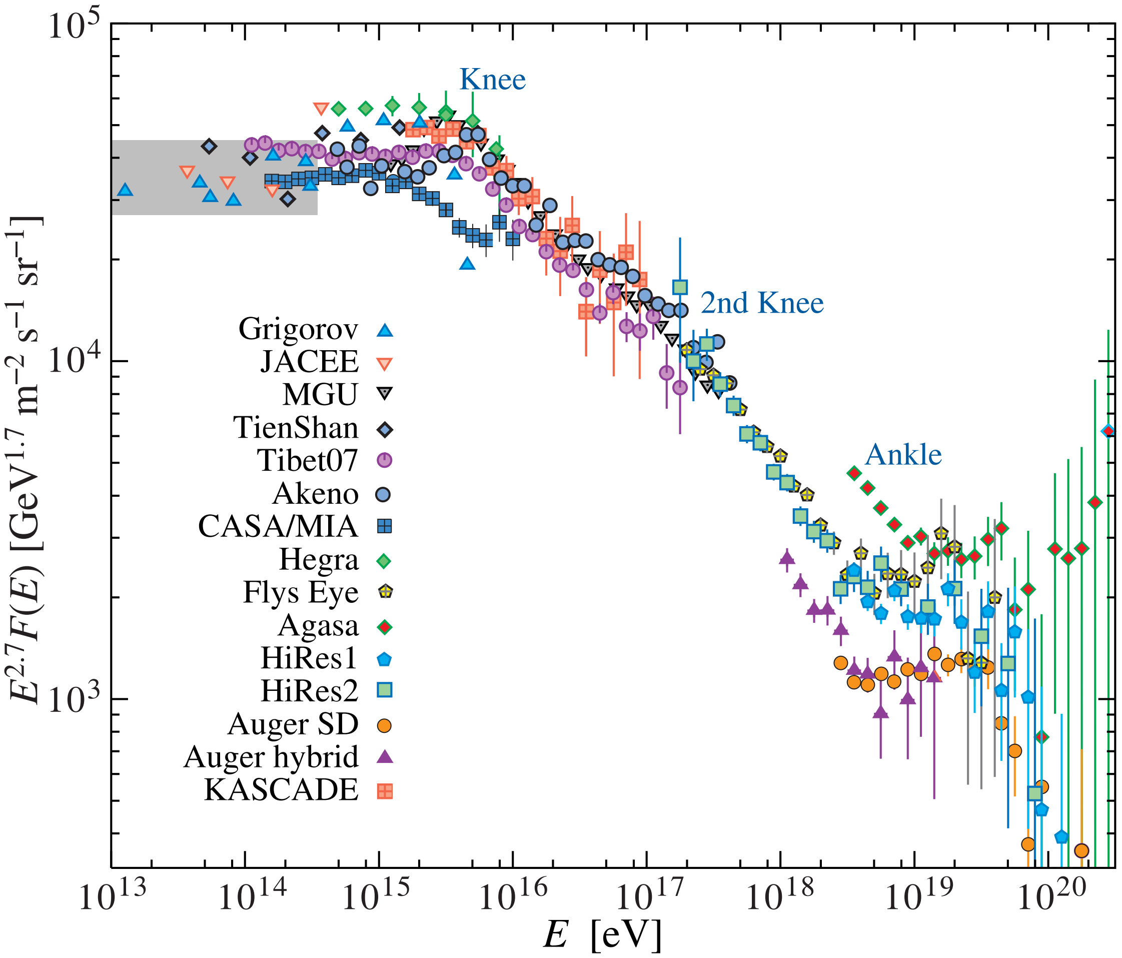


**Figure 2.1:** The all particle differential energy spectrum from air shower measurements. The shaded area shows the range of direct detection [15]

## 2.2 Composition

The chemical composition of cosmic rays is only known for energies below $10^{14}$ eV, as above this energy a direct measurement of the primary particles is practically impossible. The elemental distributions have been studied by satellite and balloon experiments [16] up to energies of 1-2 TeV/nucleon. About 98% of the cosmic radiation are hadrons, the remaining 2% is composed of electrons and photons. 87% of the hadronic component are protons, 12% helium nuclei and the rest corresponds to fully ionised nuclei of heavier elements.

Comparing the elementary abundances in the solar system with the one of cosmic rays, important information on the origin and propagation of the cosmic particles can be extracted (see next section). The analysis of the abundances of refractory nuclides has shown similarities between the derived abundances at the sources and the abundances in the solar system [17], suggesting that cosmic rays are accelerated out of a sample of well mixed interstellar matter. In cosmic rays, elements like Li, Be and B are overabundant, as well as all the groups with atomic mass lower than Fe. No significant difference is found, instead, in the abundances of the C, N, O and Fe groups. This suggests that, on their way through the galaxy, the latter nuclei further interact with the interstellar medium, generating lighter elements through *spallation* processes.

From the analysis of the relative abundances of these elements (and even isotopes) and on the basis of diffusion equations for the primary particles, the cosmic rays are found to cross on average a thickness of about $5\,\mathrm{g/cm^2}$, which corresponds to a path length of approximately 1 Mpc. Given that the thickness of the galactic disk is $\approx 1\,\mathrm{kpc}$, it means that the cosmic rays diffuse into the galaxy before escaping. An average escape time $\tau_{esc}$ can be estimated by the analysis of the ratio of radioactive secondaries and their decay products. The analysis of $^{10}\mathrm{Be}/^{9}\mathrm{Be}$ yields an escape time of approximatively $10^7$ years [18].

Above $10^{14}\,\mathrm{eV}$, information on the chemical composition of cosmic rays comes from studies of the extensive air showers. In the energy region around the knee, the composition changes towards heavier elements [1; 2]. At higher energies, Fly's Eye [19] and Yakutsk [20] report an evolution from iron dominated composition at $10^{17}\,\mathrm{eV}$ to a proton dominated composition above $10^{18}\,\mathrm{eV}$.

## 2.3 Acceleration and propagation

The observed cosmic ray energy density is

$$\epsilon_{cr} = \frac{4\pi}{c}\int_1^\infty E\cdot F(E)\,\mathrm{d}E \approx \frac{4\pi}{c}\int_1^\infty E^{-1.7}\mathrm{d}E\;\mathrm{GeV\,cm^{-3}} \approx 1\,\mathrm{eV/cm^3},$$

with $F(E)$ being the differential cosmic ray flux.

If this is assumed to be the typical density inside the galactic disk, then to maintain it, a power of $L_{cr} \approx {}^{\epsilon_{cr}\cdot V_{gal}}/_{\tau_r} \approx 5\cdot 10^{40}\,\mathrm{erg\ s^{-1}}$ is needed, where $V_{gal}$ is the galactic disk volume and $\tau_r \sim 10^7$ years is the confinement time of the cosmic rays in the galaxy.

The most probable acceleration sites for cosmic rays with energies above 1 TeV are the expansion shells of supernova remnants (SNR). With an explosion rate of three Supernovae per century in the Galaxy and assuming an average energy of $10^{51}$ erg per explosion, a total power of $L_{SN} \approx 3\cdot 10^{42}\,\mathrm{erg/s}$ would be released. Only 15% of the energy released by one Supernova is needed to explain the observed energy density of the cosmic radiation [21]. The underlying process of stochastic acceleration occurring on the shock front of the shells is called first order Fermi process [22]. During a supernova explosion, a shock wave is generated, which propagates at a supersonic speed of about $10^4\,\mathrm{km/s}$. As it propagates through the interstellar medium, particles from the surrounding medium can cross the shock front several times by diffusive collision-less scattering at magnetic irregularities. With each crossing they gain an amount of energy proportional to the speed of the shock itself. The maximum energy these particles can obtain depends on the shock front speed and on the number of reflections, which is limited by the lifetime of the shock front. This acceleration mechanism provides a power-law energy spectrum $E^{-2}$ at the source. Recent observations of the H.E.S.S experiment, which observes high energy $\gamma$-rays, come in support of the hypothesis of SNRs as cosmic particle accelerators: high energy photons are supposed to be produced by the collision of accelerated hadrons with the atoms and molecules of the interstellar cloud. H.E.S.S. has measured for the first time the energy spectrum of $\gamma$-rays in the TeV region from a supernova remnant [23]. The index of the power law is consistent with the expectation from Fermi acceleration at strong shocks.

Other environments for the occurrence of shock fronts are the termination shock of a stellar/galactic wind [24; 25]. Also, in binary systems with a pulsar, a shock front is generated due to the interaction between the pulsar wind and the atmosphere of the companion star [26].

Non-stochastic acceleration models foresee the presence of strong magnetic and electric fields. Such conditions can be found, for example, within relativistic jets from Active Galactic Nuclei (AGN) [27]. In this last case, an extra-galactic origin of the cosmic particles is assumed and primary energies above $10^{16}$ eV are predicted.

After acceleration, the particles diffuse through the Galaxy, being deflected by irregular magnetic fields [5]. They are not confined into the disk but they can propagate into the galactic halo as well. The evidence comes from the comparison of the chemical abundances of the cosmic rays with the cosmic abundances and from measured isotropy of the cosmic radiation. The particle's energy changes during propagation due to collisions, decays and energy loss processes. Moreover, the diffusion process is dominated by the galactic magnetic field. High energy nuclei can be diffused as long as their gyro-radius is smaller than the diffusing region, leading to an energy dependent probability of escape from the galaxy. All these processes cause the spectral index to change from $\gamma = -2$ to $\gamma = -2.7$.

## 2.4 The knee of the cosmic ray spectrum

This first distinct change of the spectral index occurring at a primary energy of $\approx 3 \cdot 10^{15}$ eV has been studied by several experiments in the past fifty years. The KASCADE [6] and EAS-TOP [28] experiments have investigated the energy range $10^{14}$–$10^{17}$ eV in order to determine the exact position of the knee and to answer the question whether it is caused by one primary mass group only or this feature is seen in each primary mass spectrum individually at different energies. The results of both experiments [1; 2] have shown a break in the energy spectrum of the lighter components, but not in those of the heaviest ones. The break happens at an energy proportional to the nuclear charge $Z$, as predicted under the assumption that the knee is a consequence of magnetic confinement occurring either at the acceleration region or during propagation. A break in the iron spectrum (referred to as iron-knee), according to this theory, should occur at $\approx 10^{17}$ eV.

However, the question of the origin of the knee is still unsolved. Several models have been proposed, which involve different aspects of the acceleration, propagation and interaction mechanisms of the cosmic particles.

### Acceleration

Assuming that acceleration happens at strong shocks, as first approximation the maximum energy attainable by a particle with charge $Z$ is determined by the magnetic field intensity $B$, by the dimension $L$ of the shock and by the speed of the plasma according to the following relation [3]:

$$E_{max} = 9.2 \cdot 10^{20}\ eV \cdot Z \cdot \frac{B}{\mathrm{G}} \cdot \frac{L}{\mathrm{pc}} \cdot \frac{V}{\mathrm{c}}.$$

Assuming typical values of $10\,\mu$G for the magnetic field, 1 pc for $L$ and $\beta = V/c \sim 1/300$, the upper limit of the energy is:

$$E_{max} \approx Z \times 3 \cdot 10^{15}\,\mathrm{eV} \tag{2.2}$$

Therefore, the position of the knee in the energy spectra of individual primary elements changes with the charge $Z$ of the primary particle.

### Propagation

Other astrophysical models interpret the knee as a propagation effect. The Larmor radius of a charged particle is proportional to its rigidity $R$ through the relation

$$r_L = R \sin\theta / Bc \qquad \text{with} \qquad R = \frac{pc}{Ze}. \tag{2.3}$$

The regular component of the galactic magnetic field will cause particles with charge $Z$ to describe helical trajectories with radius $r_L$. The knee may be the result of the interplay between the diffusion of particles along magnetic field lines and the Hall diffusion (i.e. a drift) perpendicular to the average azimuthal galactic magnetic field [4; 5]. With increasing momentum the escape probability from the Galaxy increases. Since $r_L \propto {}^1\!/_Z$, the leakage for light elements is expected to occur earlier than for heavy nuclei. Protons, thus, will escape first and subsequently all the other elements.

### Interaction

As already mentioned, for energies above 1 TeV direct detection of cosmic particles is not feasible. The only way is to sample the extensive air showers with arrays of detectors. To extrapolate the properties of the primary particle, a deep knowledge of the EAS development is fundamental. The interpretation of the measured events above 1 TeV relies on the comparison with simulations. These are based on interaction models, which extrapolate the result of accelerator experiments to much higher energies. At the highest energies, a new interaction channel could occur, with the production of a new heavy particle (see for example [29]) which might not be observable by an air shower experiment.

## 2.5 Extensive Air Shower

When a high energy particle interacts with the nuclei in the atmosphere, a sequence of processes starts, which leads to multiple production of secondary particles. An EAS is initiated by the first hadronic interaction. The interaction probability depends on the inelastic cross-section for the reaction of the primary nucleus with air, which in turn is a function of the primary energy. In addition, the height in the atmosphere of the first interaction at a fixed primary energy fluctuates. The interaction length for a proton in air is $\approx 80\,\mathrm{g/cm^2}$; about 12 consecutive interactions are possible when the first interaction happens at the top of the atmosphere, giving rise to what is called an *hadronic cascade*. If the primary particle is a heavy nucleus, it will be fragmented in the first interaction, which on average takes place higher in the atmosphere than for light nuclei, due to the smaller interaction length.

After the first interaction, the primary particle is fragmented and new hadrons are produced. If they have enough energy, they will continue their path through the atmosphere with further interactions. Mainly pions and kaons will be produced. The number of particles in the shower increases as long as the average energy per particle is above the energy threshold for further particle production. A maximum is then reached, after which the electrons, i.e. the most abundant particles in the shower, will start to loose energy in the atmosphere without any new particle creation.

An air shower is composed of many types of secondary particles, which can be classified in three principal components: the hadronic, the electromagnetic and the muonic component.

### The hadronic component

The hadronic component originates from the interaction of the primary particle and its fragments, as previously explained. The most abundant secondary hadrons are $\pi^+$, $\pi^-$ and $\pi^0$, but also kaons, and baryons like protons and neutrons are contributing. Charged pions above 100 GeV can interact, contributing to the development of the hadronic cascades, or decay into muons according to:

$$\pi^+ \longrightarrow \mu^+ + \nu_\mu \tag{2.4}$$

$$\pi^- \longrightarrow \mu^- + \bar{\nu}_\mu, \tag{2.5}$$

with a mean lifetime $\tau \sim 2.6 \cdot 10^{-8}\,\mathrm{s}$. The decays often occur in an early stage of the shower development, when the particles cross a region of low density so that charged mesons can decay before interacting.

### The muonic component

The muonic component of an EAS originates from the decay of pions and kaons. With a mean lifetime of $2.2 \cdot 10^{-6}\,\mathrm{s}$ and a Lorentz factor $\gamma \approx 90$ (for $E_\mu = 10\,\mathrm{GeV}$), muons will penetrate the whole atmosphere practically undisturbed. High energy muons (E>100 GeV) originate in the early stage of the shower development. They loose energy mainly through ionisation processes, as the bremsstrahlung emission is strongly suppressed compared to electrons. Low energy muons, descending from a later stage of the development, can decay into electrons and neutrinos according to:

$$\mu^+ \longrightarrow e^+ + \nu_e + \bar{\nu}_\mu \tag{2.6}$$

$$\mu^- \longrightarrow e^- + \bar{\nu}_e + \nu_\mu. \tag{2.7}$$

### The electromagnetic component

The electromagnetic component is sustained mainly by the $\pi^0$ decay into two photons, which happens about $10^{-16}\,\mathrm{s}$ after the production of the $\pi^0$, and partly by the decay of muons. The photons from the pions decays can create positron/electron pairs and these in turn can emit $\gamma$ particles via bremsstrahlung processes. Pair production and bremsstrahlung processes alternate, giving rise to an electromagnetic cascade.

### Longitudinal and lateral profiles

The three components feed each other while developing over an almost circular area around the shower axis. The air shower development is characterised by its *longitudinal* and *lateral development.* It is fundamental to the correct understanding of the evolution of the cascade, in order to trace back the primary particle properties.

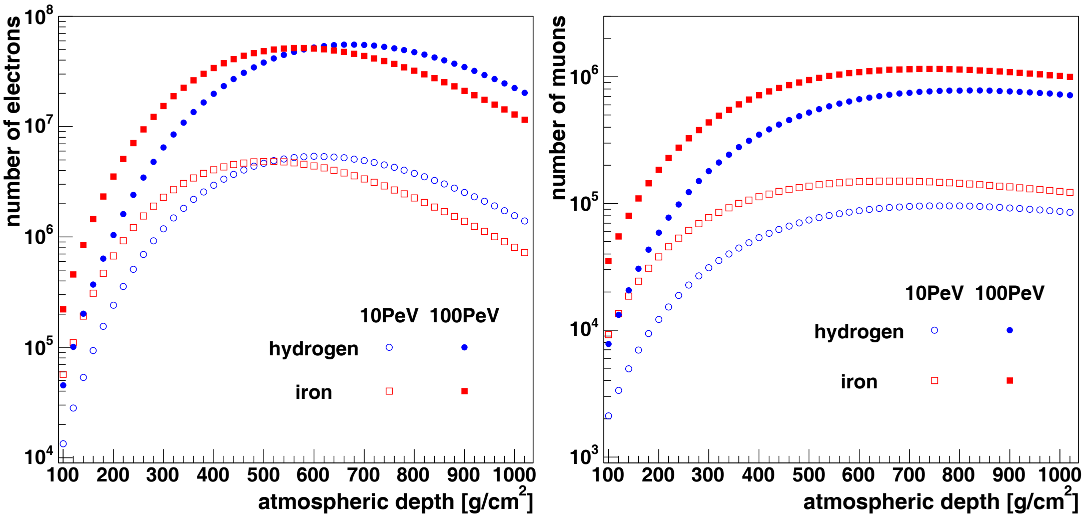
The *longitudinal development* of an air shower represents the way the number of secondaries changes as a function of the atmospheric depth. A critical energy $E_c$ can be defined as the energy at which decays or energy loss processes become more probable than the production of secondary particles, which reaches its maximum when the particles are generated with an energy of the order of the critical energy. After that the particles start to be "re-absorbed" and the cascade dies out.

The electromagnetic cascade develops as long as the energies of the photons exceed the pair production energy threshold and as long as the probability for bremsstrahlung processes for electrons is higher than the probability for ionisation losses, which happens at the critical energy of about 80 MeV. The basic properties of a cascade can be illustrated using a Heitler model [30]. It assumes that after each interaction length $X_0$ two particles with equal energy are generated from a parent particle. At an atmospheric depth $t = {}^{X}/{}_{X_0}$, the number of particles in the shower is $N(X) = 2^t$, with energy $E(X) = E_0/N(X)$ each. At the shower maximum, the number of particles is $N_{max}(X) = E_0/E_c$, with the position of the maximum itself being $X_{max} \approx X_0 \ln(E_0/E_c)$. If the primary particle is a nucleus with mass $A$, as first approximation it behaves in the atmosphere as a group of $A$ nucleons of energy $E_0/A$ interacting at the same time. Then, following a similarly simple model, the number of particles at the shower maximum will be the same for both a heavy nuclei or a proton of primary energy $E_0$. The position of the shower maximum, instead, depends on the primary mass according to $X_{max} \propto \ln(E_0/(E_c \cdot A))$. The longitudinal development of electrons foe proton and iron primaries is shown in fig. 2.2a. The total number of electrons in a shower at the observation level is called *electron shower size*. As neutral pions are continuously produced along the shower axis, the number of electromagnetic particles is determined by the superposition of different e.m. sub-cascades. For a primary photon of energy $E_0$, the evolution of the shower size as a function of the atmospheric depth is described by [31]:

$$N_e(E_0, t) = \frac{0.31}{\beta_0^{\frac{1}{2}}} \exp\left[t\left(1 - \frac{3}{2}\log s\right)\right], \tag{2.8}$$

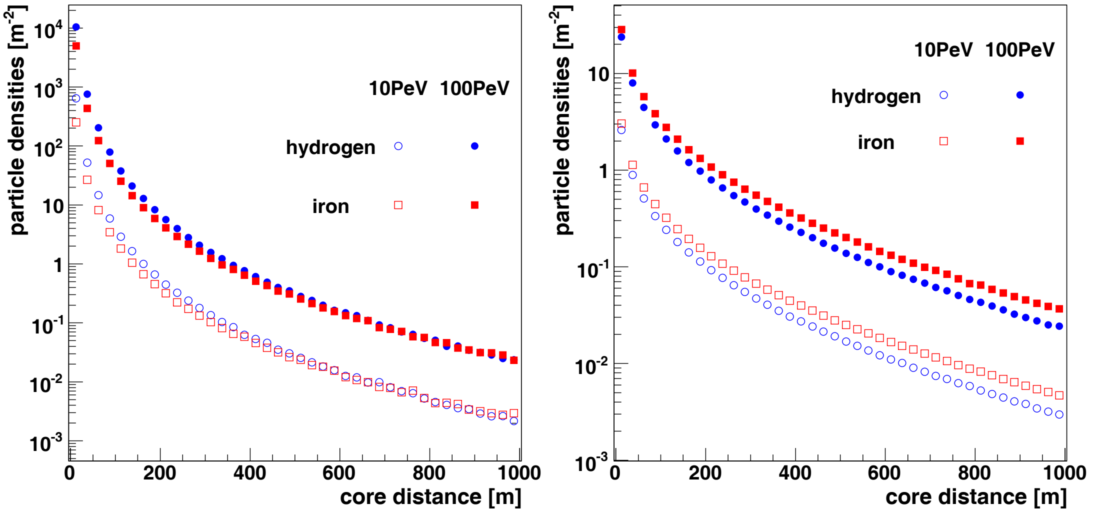
where $\beta_0 = \log(E_0/E_c)$ and $s = \frac{3t}{t+2\beta_0}$ is the shower *age*. At the maximum of the shower development $s = 1$, whereas before reaching the maximum $s < 1$ and after the maximum $s > 1$.

Due to the low decay and interaction probabilities, the longitudinal development of the muonic component shows a less pronounced maximum, which rather resembles a plateau where the number of muons decreases slowly. Their longitudinal profile is depicted in fig. 2.2b, for proton and iron primaries at two different primary energies.



(a) (b)

**Figure 2.2:** Average longitudinal distribution for electrons (a) and muons (b) of 50 simulated air showers for different energies and primaries [32].



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

**Figure 2.3:** Average lateral distribution for electrons (a) and muons (b) of 50 simulated air showers for different energies and primaries [32].

The *lateral distribution* describes the distribution of the particle densities as a function of the distance from the shower axis. The lateral spread shows different features according to the different shower components.

The **hadronic component** is concentrated close to the shower axis, due to the low transverse momentum ($p_T \sim 0.4\text{GeV/c}$) with which they are produced. If a hadron of momentum $p$ is generated at a height $h$ from the observation point, its lateral spread is $r = (p_T/p)h$. Thus, a shower core of high energy hadrons with a radius of about 20 m is formed and propagates in the direction of the incident primary. Only low energy hadrons such as neutrons can spread further.

The lateral distribution of the **electromagnetic component** is dominated by the multiple Coulomb scattering of the electrons. The Moliére radius, defined as $r_M = {}^{E_s}/_{E_c} \cdot X_0 = 9.50\,\text{g}\,\text{cm}^{-2}$ is the natural unit of the lateral development[1]. At see level $r_M \approx 100\,\text{m}$. The lateral distribution function for a pure electromagnetic cascade in a homogeneous atmosphere is well described by the Nishimura Kamata Greisen (NKG) formula [33], for $0 \leq s \leq 2$:

$$\rho(r) = C(s) \left(\frac{N_e}{r_M^2}\right) \left(\frac{r}{r_M}\right)^{s-2} \left(\frac{r}{r_M} + 1\right)^{s-4.5}, \tag{2.9}$$

with $C(s)$ being a normalisation constant. The parameters of this formula can be adjusted to better describe showers initiated by hadrons; a slightly modified version of the NKG function is discussed in chapter 4.

The lateral dispersion of the **muonic component** is the result of the transverse momenta of the parent pions and kaons, with a small contribution from multiple scattering which is suppressed by a factor $({}^{m_e}/_{m_\mu})^2 \approx 2 \cdot 10^{-5}$ compared to electrons. Because of this suppression and because muons are not affected by strong interaction, they travel almost

---

[1] $E_s$=21.2 MeV for electrons

straight through the atmosphere. In fig. 2.3, the lateral distributions of electrons and muons for proton and iron simulated showers is shown. At a given energy, the electron lateral distribution of iron initiated showers is flatter than for a proton induced shower and the total number of electrons is smaller (note the logarithmic scale in fig. 2.3a). For iron induced showers, the shower maximum is reached higher in the atmosphere as compared to the proton primaries. The electrons then are more scattered when reaching ground. An analogous behaviour is observed for the muonic component. If a primary particle interacts higher in the atmosphere, then the spread it causes in ground will be larger, considering that muons travel through the atmosphere undisturbed.

# 3. The KASCADE-Grande experiment

KASCADE-Grande is a multi-detector experiment for the measurement of air showers induced by primary cosmic rays in the energy range $10^{16} - 10^{18}$ eV. It is the extension of the original KASCADE experiment, whose aim was the investigation of the cosmic ray flux in the knee region. KASCADE-Grande measures the electromagnetic component of air showers with an array of 37 scintillator stations reassembled from the array detectors of the experiment EAS-TOP. The muonic component is measured with scintillation counters and tracking chambers at different particle energy thresholds, with the detectors of KASCADE.

In the following, the different detector components are described.

## 3.1 KASCADE

The components of the original KASCADE [6] experiment that are still active are the array, parts of the central detector and the muon tracking detector.

The **array** (fig.3.1) consists of 252 detector stations sensitive to electrons, photons and muons, laid out over a grid of 200×200 m$^2$, with 13 m spacing. It is organised in 16 clusters of 4×4 stations each, with four stations missing at the centre of the array. $e/\gamma$ particles are detected by an unshielded liquid scintillator[1] filled in a circular cone of 1 m diameter and 5 cm height. The light is diffused in vertical direction inside a cone and collected at the top by a 3" diameter photomultiplier[2] (PMT). There are two types of detector stations: in the inner four clusters four $e/\gamma$ detectors are installed, but no muon detectors. On the other hand the stations in the twelve outer clusters contain only two $e/\gamma$ detectors, but these are placed on a lead/iron absorber plate (10 cm Pb, 4 cm Fe), corresponding to 20 electromagnetic interaction lengths ($X_o$). Underneath the absorber plates, the muonic shower component is measured with 3 cm thick plastic scintillators[3], the light is coupled by wavelengths shifters and collected by four PMTs[4]. The screening results in an energy threshold for vertical muons of 230 MeV (fig.3.1).

The array provides a total detection area for $e/\gamma$ (${E_e}^{th} > 5\,\mathrm{MeV}$) equivalent to 490 m$^2$ and 622 m$^2$ for $\mu$.

The **central detector** is placed in the centre of the array, consisting of:

---

[1] 2 gl PMP (1-phenyl-3-mesityl-2- pyrazoline) in 80% paraffin and 20% pseudocumene

[2] Valvo XP3462

[3] Bicron BD-416

[4] type EMI 9902 or Valvo XP2081

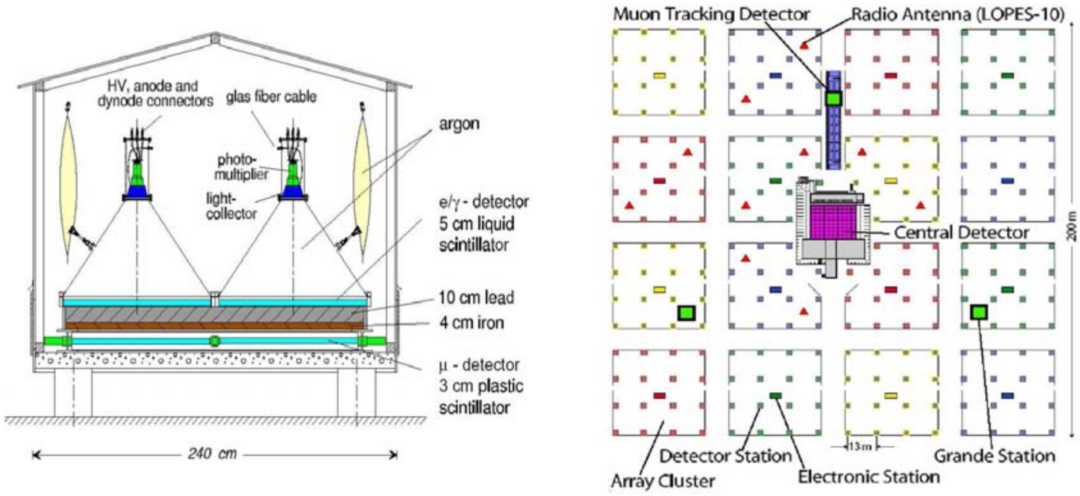


**Figure 3.1:** Left: sketch of an outer array station, with $e/\gamma$ detectors, the lead-iron absorber and the muon detector at the bottom [6]. Right: scheme of the KASCADE array. Three Grande stations are also shown (green boxes) [6].

- two layers of *multi-wire proportional chambers* (MWPC), positioned under the absorber of the (no longer operating) hadron calorimeter, for the detection of muons above an energy (for vertical particles) of 2.4 GeV. They track particles with an accuracy of 1.5° and cover an area of about $129{\times}2\,\mathrm{m}^2$;
- one layer of *streamer tubes*, built under the MWPC so to improve the reconstruction of muonic tracks for high particle densities ($> 2.0/\mathrm{m}^2$), covering an area of $258\,\mathrm{m}^2$.

On the north side of the central detector lies the **Muon Tracking Detector** (fig.3.2) consisting of 3 horizontal and 2 vertical layers of limited streamer tubes, installed in an underground tunnel of size 5.5 m×48 m, for a total of $128 \times 3\,\mathrm{m}^2$ detection area. Placed below an 18 radiation lengths shielding of soil, concrete and iron, its aim is to trace the muons ($E_{th}$ =800 MeV) penetrating the detector. With an accuracy of 0.5° on the reconstructed direction, it is possible to estimate the production height of the muons once the core position and the shower axis have been reconstructed from the data of other detector components [34].

## 3.2 Piccolo

Piccolo (fig.3.3) is an array of 8 scintillator stations, with 20 m spacing in between, forming an octagon over an area of $360\,\mathrm{m}^2$, placed between the centres of the Grande and KASCADE arrays. Each station contains 12 scintillators with an active area of $10\,\mathrm{m}^2$. The trigger is generated by a coincidence of ”two-out-of-eight” Piccolo stations. The trigger signals that Grande produces are too late for the muon tracking detector and the central detector in order to record the data, so a faster trigger is needed in this case, which is provided by Piccolo.

## 3.3 Grande

Grande is an array of 37 scintillator detector stations distributed over an area of $\approx 700 \times 700\,\mathrm{m}^2$, with an average spacing of 137 m between stations on average (fig.3.3).

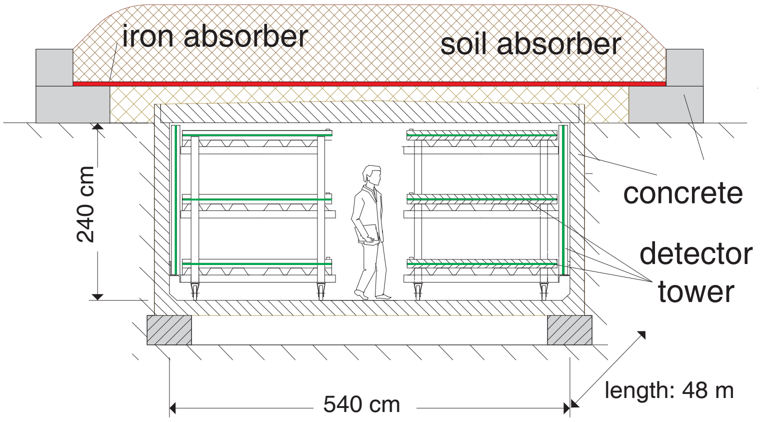


**Figure 3.2:** Muon Tracking Detector [6]

Each station is set up with 10 m$^2$ of plastic scintillator (NE102A), divided into 16 units (80×80 cm$^2$, 4 cm thick) ordered in a 4 × 4 matrix. All of the 16 modules are equipped with a high gain (1.6 pC/m.i.p.) photomultiplier[5] (in the following called Pb) for the measurement of low particle densities. In the four central units, the light is additionally collected by a second low gain (0.08 pC/m.i.p.) PMT (referred to as Pb1) for detection of high particle densities. A scheme of the station set up is shown in fig. 3.3.

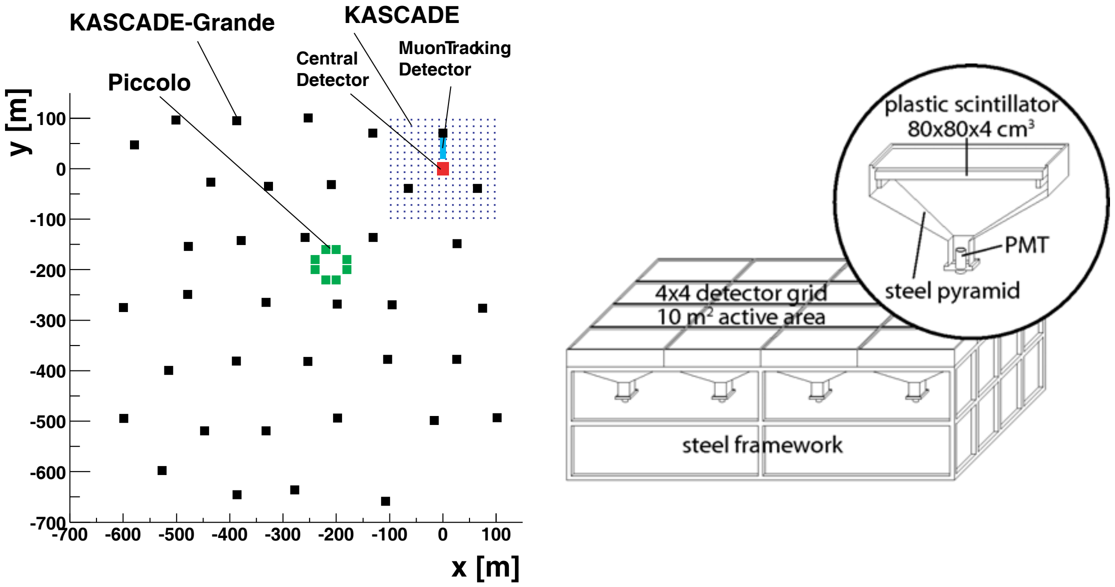


**Figure 3.3:** *Left*: layout of KASCADE-Grande [35]. The black squares represent Grande stations. At the upper-right is KASCADE. Piccolo is marked green. *Right*: sketch of the Grande detector stations set up [36].

---

[5] Philips XP3462B

Signals coming from each unit are pre-processed in the station before being transmitted to the central Grande Data AcQuisition station (DAQ).

In figure 3.4, a scheme of the electronic chain in a detector station is shown.

The 16 analog signals coming from high gain PMTs are added up by a mixer module (*CAEN 169 Analog Multiplexer*). A second mixer adds up the signals coming from the low gain PMTs. The Pb analog mixer has two outputs. One is sent to a double threshold discriminator in order to generate a trigger. The second output is passed to a shaping amplifier (*CAEN N442*), as is the output of the PB1 mixer. The shaping amplifier integrates the signals over 20 $\mu$s, producing pulses whose amplitude is proportional to the total charge collected.

The shaping amplifier has three outputs, corresponding to three different amplification factors: 2.5 mV/pC for Pb, a second output for the PB channel with 25 mV/pC (here called Pb×10) and 5 mV/pC for Pb1. There exists a third channel with an amplification of 50 mV/pC, used in the calibration procedure.

The first threshold of the discriminator is set to roughly -12.5 mV (≈0.1 m.i.p.) and is used for timing measurement, the second one is set to about -24 mV (≈0.3 m.i.p.) used for the trigger itself. When the signal passes both values, a NIM logical signal is generated (Logic Run, LR) and sent to the DAQ (fig. 3.5). The logic signal and the three analog signals obtained in this way (Pb×10, Pb and Pb1) are transmitted to the acquisition centre via coaxial cables 700 m long each.

In the DAQ, the three analog signals (covering a dynamic range from 1 to 200 m.i.p., 10 to 2000 m.i.p. and 200 to 30000 m.i.p, respectively) of all 37 stations are sent to four peak-ADC CAEN V785 modules, for digitisation corresponding to the energy deposit in each station.

Due to the attenuation in the cables between station and DAQ, the logic signal must be refreshed into a NIM signal and multiplied by means of a discriminator chain in order to be used for trigger, TDC and SCALER. The measurement of the relative time between the signals generated in the station, used for the reconstruction of the shower direction, is performed by a 128-channel TDC (CAEN V767). The TDC records all hits of all stations in a time window of $\pm 5\,\mu$s around the trigger signal, with a sensitivity of 0.78 ns/ch. The LR is also used to generate the trigger for single particle measurement in each station. It is sent to the SCALER modules (CAEN C257) for monitoring purposes and to the trigger generator, set up by 18 programmable logic units (CAEN C85) performing a Grande trigger by means of the coincidence of 7 Grande stations (see below). There exists also the Real Time Clock Module, which has as input 1 Hz and 5 MHz clocks and the event signal, in order to save the event time and create a veto system that forbids a new acquisition as long as the previous one is not terminated. The pattern units register for each event the trigger source.

The entire acquisition chain is initiated by a trigger signal, generated from Grande stations (internal trigger) or by a so called external trigger coming from Piccolo or KASCADE. The Grande array is divided for trigger purposes into hexagons of 7 stations (6 on the perimeter plus 1 in the centre), forming in total 18 clusters (fig.3.6). An *air shower* with $10^{16} < E_0 < 10^{18}$ eV will trigger a fourfold (4 adjacent stations in a trigger hexagon) or sevenfold (all the stations in a hexagon) coincidence of Grande detectors. The first fourfold coincidence is used as a common start for the time measurement. The counting rate for each detector station is $\approx 2$ kHz, the fourfold trigger rate for the whole Grande

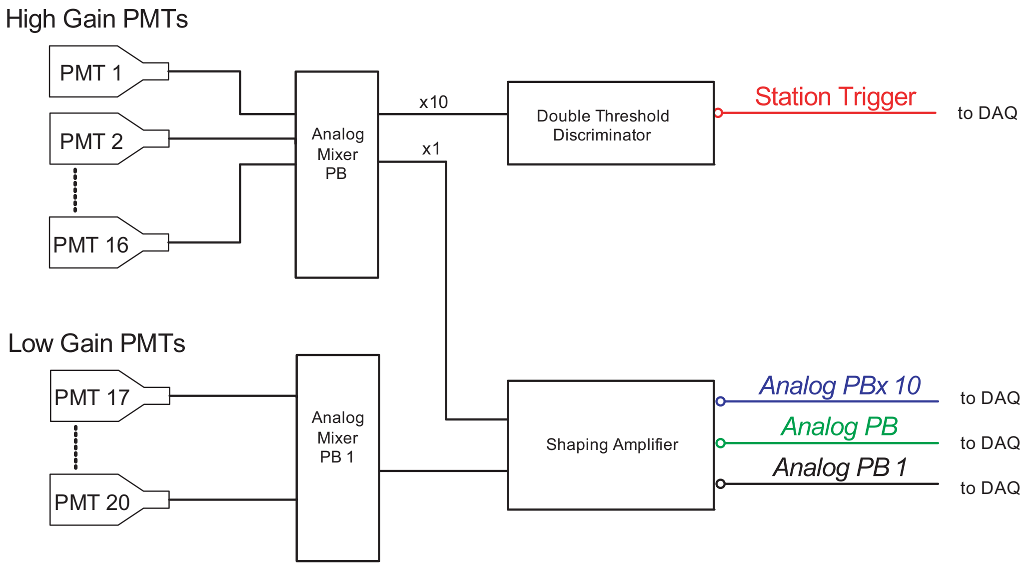


**Figure 3.4:** Scheme of the acquisition chain inside a Grande detector station [37]

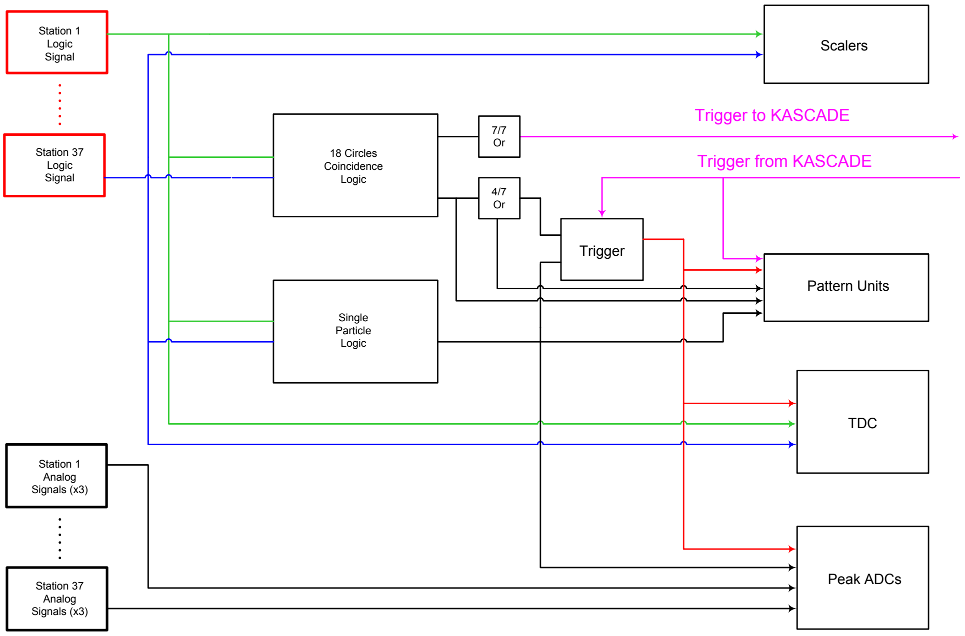


**Figure 3.5:** Scheme of the acquisition chain in the Grande DAQ [37]

array is ≈5 Hz, for the sevenfold ≈ 0.4-0.5 Hz. Cluster number 15 is formed by 6 stations only; a sixfold coincidence is thus required for this cluster instead of a sevenfold one.

Besides the air shower events, single particle and scaler events are also recorded in the Grande raw-data files. After every air shower, three single particle events are recorded, selecting basically one single muon hitting a detector. After approximately six hours of acquisition, a single particle spectrum is produced for each station. It is used mainly for calibration purposes. Moreover every second the scalers are read, providing the single station count rate, used for monitoring purposes only. Single particle and scaler events are recorded only in the Grande data-stream, while the air shower data are merged online with the KASCADE data. For each air shower event, the TDC and ADC read out for all channels of the 37 detector stations is provided, as well as the time-stamp (which is in common with the rest of the experiment), the event and run numbers and the trigger pattern. Four-fold triggers are mainly used for internal checks as, for example, the time calibration of Grande data.

Before the actual reconstruction of an air shower event, the signal received by the 37 detectors has to be calibrated. The ADC counts of the three analog chains are converted into an effective energy deposit through a detailed procedure, description of which can be found in [37], taking into account the behaviour of the station in the selected time period. Also the time recorded by the TDC has to be calibrated, due to possible multi-hits during the 10 mus acquisition window. If multi-hits are registered, the time is chosen as the value closest to the median time of the stations with a single hit. If the number of neighbouring detector stations with a valid time hit is less than 5, the median is calculated over all the detector with a single hit. Otherwise only those who are close to the analysed station are taken into account. In addition, there is a delay in the time recorded by the different detectors, due to the time the signal takes to be processed by the electronic and sent to the DAQ. These offsets are obtained on the base of a shower to shower analysis of the relative timing between the stations [39] and properly taken into account during the calibration procedure, so that the delay between stations can be after all attributed only to the shower propagation features. The value of the offsets spans over a wide range from -100 ns to +100 ns and is extremely sensitive to any change of the electronic of the stations.

In the unlikely case of a time stamp correlated to an energy deposit less than 2.6 MeV, i.e. the hardware threshold to generate a TDC trigger, the reconstruction relies on the measured time value and set the energy deposit in the station to the threshold value.

Further information on the acquisition chain and the station setup can be found in [38] and [37].

All the detector components are calibrated through specific procedures; a description can be found in [37] for the energy calibration of Grande detectors, and in [39; 36] for the time calibration of the Grande stations. For the KASCADE array, details can be found in [6; 40].

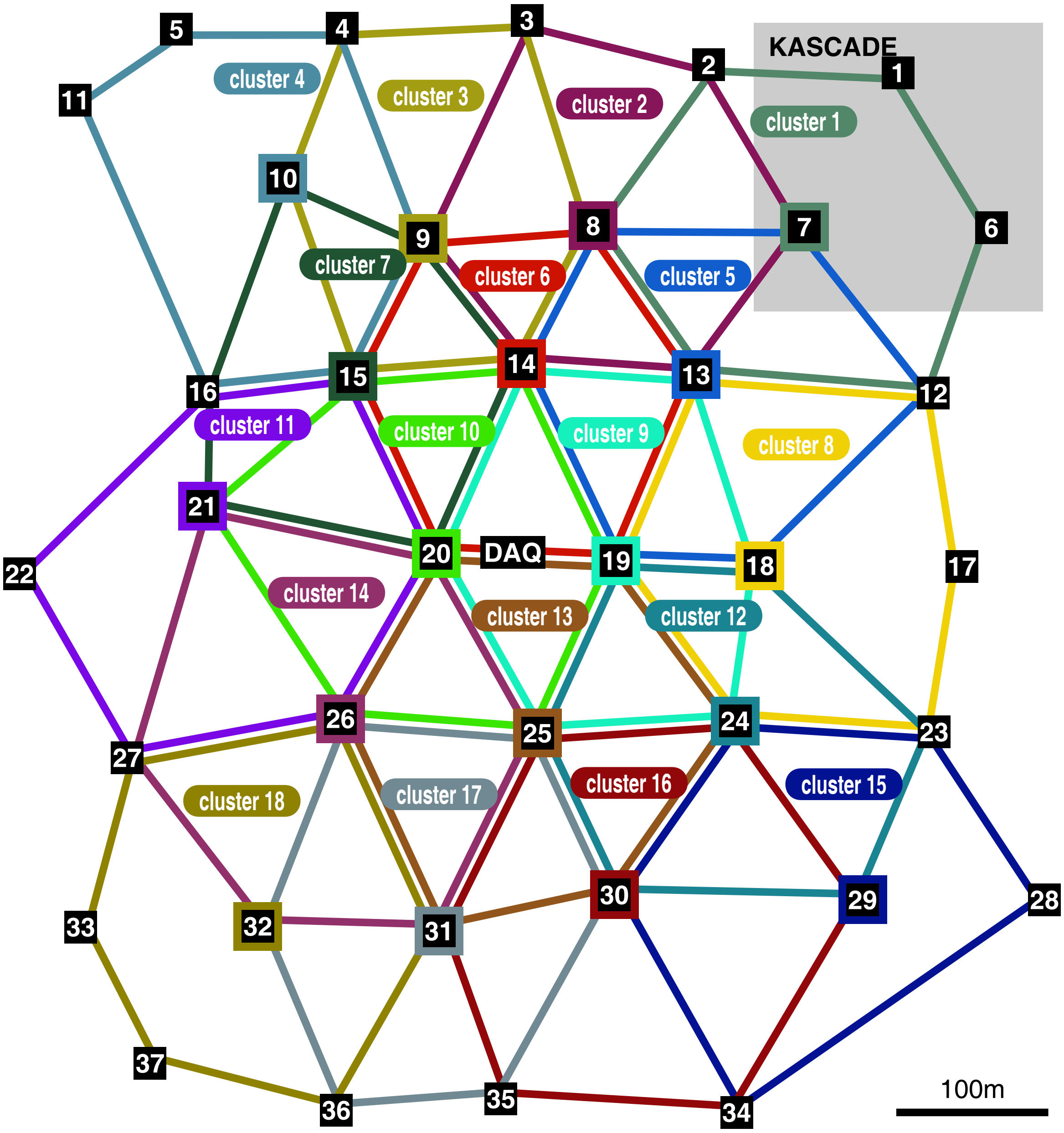


**Figure 3.6:** Trigger clusters in Grande [38]: the stations are represented by the numbers. The different colours represent the 18 trigger clusters.

# 4. Reconstruction procedure and accuracy

In KASCADE-Grande, the densities of electrons and muons on ground are measured, as well as their arrival times. From these measurements, the arrival direction of the shower, its impact point on ground (in the following referred to as shower *core*) and the total number of electrons and muons in the shower, i.e. the *electron* and *muon* shower *size* are reconstructed.

In this chapter a description of the shower reconstruction procedure, as adapted for the present work, as well as the efficiency of the detector and the reconstruction accuracies are given.

## 4.1 Reconstruction method

The reconstruction of measured air showers is realised in the Kascade Reconstruction for ExTensive Airshowers (KRETA) code, which reads the raw data, performs the calibration and reconstructs the basic shower observables as explained in the next paragraphs, storing all the results in the form of histograms and vectors of parameters.

As explained in the previous chapter, the Grande array is internally triggered by a fourfold (4/7) or a sevenfold (7/7) coincidence of the detector stations in one cluster. In this work, only the 7/7 coincidences are taken into account, that means, only events triggered by Grande (also sent to KASCADE) are analysed. In this way, a first selection of the raw data is done.

The reconstruction of the shower starts after the calibration and develops over three levels:

**Level 1:** the station classification is performed to identify the properly operating stations; the very first estimation of the shower core and arrival direction is carried out;

**Level 2:** the total number of charged particles in the shower is obtained through the fit of the lateral distribution and the shower axis is conclusively determined, as well as the muon size for the KASCADE array;

**Level 3:** the electron shower size is estimated.

| Iterative steps at level 2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | $\vartheta$ | $\varphi$ | $T_0$ | $N_{ch}$ | s | core x | core y |
| 1 | ● | ● | ● | | | | |
| 2 | | | | ● | ● | | |
| 3 | | | | | | ● | ● |
| 4 | ● | ● | ● | | | | |
| 5 | | | | ● | ● | | |

**Table 4.1:** Fitting procedure used at reconstruction level 2. Owing to the great numbers of parameters, the fitting is carried out in five consecutive steps, with different parameters being fitted while the others are being kept at fixed values. The bullets indicate which of the parameters are treated as free parameters in the different steps. Initial values are taken from previous steps.

### 4.1.1 Station selection

Stations are classified on the basis of the calibration output and of a status database, where malfunctioning due to electronics are reported [41]. **Good** are those which have a TDC count and an energy deposit above the threshold of 2.6 MeV (equivalent to a third of the energy deposit by a minimum ionising particle); **bad** are the ones so classified by the status database or for which the calibration procedure has failed; **silent** those which got no TDC hit but are properly functioning; **accidental** are the stations for which there is an energy deposit above threshold but whose time stamp is more than $3\,\mu$s away from the average time of the good ones, so it is possible that the measurement does not belong to the shower, and **saturated** are those where all the three ADC channels are saturated (12 bit ADC in saturation).

Only the arrival times of particles in station, which has correct TDC measurement (*good* and *saturated*), are used for the arrival direction fit (see 4.1.2). Regarding the lateral distribution function, the fit is done with the energy deposits measured in good, saturated and silent stations: the saturated stations are treated as lower limit and the silent ones as zero measurements.

### 4.1.2 Arrival direction

In the first level, the shape of the shower front is assumed to be planar. The arrival direction is estimated as the normal vector to the plane, as identified by the three detectors with the highest energy deposit. In a second step it is obtained by solving the system of linear equations of the minimization of a $\chi^2$, where the expected deviations (see equation 4.1) of the timing from a flat shower front are compared to the measured times. Outputs of the fit are the zenith and azimuth angles $(\vartheta, \varphi)$, which identify the shower axis, and the time $T_0$, which can be seen as the arrival time of the particles in the shower core.

The final determination is achieved in level 2, after an iterative fit procedure, with no more than three free parameters at a time [41]. The way this procedure develops is displayed in table 4.1. In each step, the fixed parameters assume the best value estimated so far.

The arrival direction is obtained with a ($\chi^2$)-fit of the times measured by each station, assuming a curved shape of the shower front. In order to do so, the residuals between a flat front and a curved one have been studied in simulations and found to follow [42]

$$\begin{aligned} \bar{t}(r) &= 2.43 \cdot \left(1 + \frac{r}{30\,m}\right)^{1.55}, & (4.1) \\ \sigma_t &= 1.43 \cdot \left(1 + \frac{r}{30\,m}\right)^{1.39}, & (4.2) \end{aligned}$$

with $r$ being the distance from the core in the shower disc frame of reference (in the following SDC). Equation 4.1 expresses the estimated mean arrival time, whereas equation 4.2 gives the spread of the underlying distribution. In order to extract the "first out of N particles" arrival time and its uncertainty, both values are then divided by $\sqrt{N}$, $N$ being the total number of particles hitting the detector station.

### 4.1.3 Core, N$_e$ and N$_\mu$ estimation

A rough estimation of the shower core is done in the first level by applying a centre of gravity method; the total number of charged particles ($N_{ch}$) and the total number of muons ($N_\mu$) in the shower are also estimated in a robust way. In order to convert the energy deposit in the detectors into number of charged leptons, the contribution from the photons and hadrons of the shower has to be subtracted. This is achieved through the application of a Lateral Energy Correction Function (LECF), which has been studied on simulated showers [43; 44]. The LECF gives the energy deposit of charged leptons as a function of the distance from the shower core $r$ in shower disc coordinate system. In fig. 4.1, the correction function is shown. After the correction for the photons and hadrons contributions, the total number of particles measured in the detectors ($n_p$) is equated to the one expected according to the lateral distribution function. If we assume that the measured densities with respect to the distance from the core ($r$) can be described by

$$F(r) = N_{ch} \cdot f(r), \text{where } f(r) = C \cdot \left(\frac{r}{r_0}\right)^{P_1} \cdot \left(\frac{r}{r_0}\right)^{2P_2} \cdot \left(\frac{r}{r_0}\right)^{3P_3}, \tag{4.3}$$

we can write

$$\sum_{i=1}^{n} N_{ch} \cdot f_i(r) \cdot A = \sum_{i=1}^{+n} n_{p_i}, \tag{4.4}$$

where $A$ is the effective area and $n$ the number of stations, excluding the bad and accidental ones. From Equation 4.4 we finally obtain

$$N_{ch} = \frac{\sum_{i=1}^{+n} n_{p_i}}{\sum_{i=1}^{+n} f_i(r) \cdot A}. \tag{4.5}$$

In equation 4.3, $r_0$=90 m is the scaling radius for charged leptons, $P_1 = -2.462$, $P_2 = -0.416$ and $P_3 = 0.098$ are the best parameters found from studies of simulated proton and iron initiated showers.

In an analogous way we can calculate $N_\mu$ from the KASCADE array measurements, using the following parametrisation for $f(r)$:

$$f(r) = \frac{0.28}{{r_0}^2} \cdot \left(\frac{r}{r_0}\right)^{-0.69} \cdot \left(1 + \frac{r}{r_0}\right)^{-2.39} \cdot \left(1 + \left(\frac{r}{10 \cdot r_0}\right)^2\right)^{-1}, \tag{4.6}$$

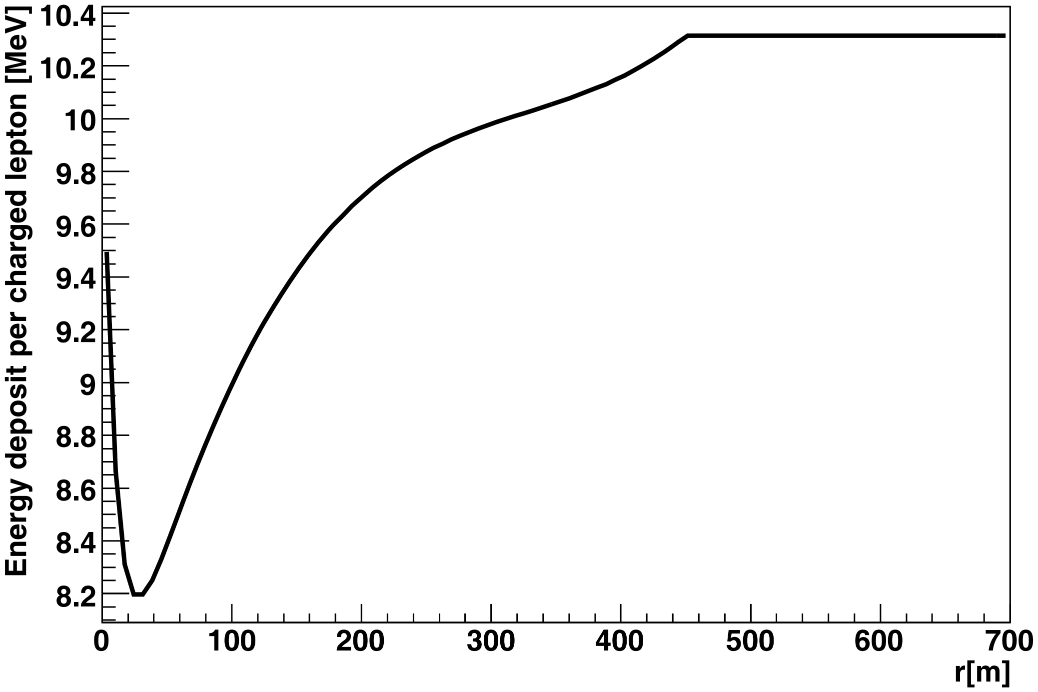


**Figure 4.1:** The LECF is shown as a function of the distance from the shower core, in the shower disk frame of reference.

where $r_0 = 320\,\text{m}$ and $N_\mu \cdot f(r)$ is a modified Lagutin function [45], whose parameters are found from studies of simulated air showers [32].

The values obtained in this first step are given as input parameters for the maximum likelihood fit realised in the second reconstruction level, where the shower geometry (i.e. core position and arrival direction) is determined and fixed.

The core is retrieved by fitting the lateral distribution of the charged particles with respect to 36 different potential core positions. These candidate core positions are arranged on a grid of 6×6 points with a 12 metre spacing. The core position that minimises the $\chi^2$ is kept. Once the shower impact point on ground is determined, the arrival direction is calculated, as well as $N_{ch}$ and the shower age. A last iteration is done to obtain the final values fitting first the core, then the direction and in the end the total number of particles and the shower age (see table 4.1).

Concerning the fit of the particle densities in Grande, a modified NKG [40] function has been found to best represent the lateral distribution of the particles.

$$\rho(r) = N_e \cdot \frac{\Gamma(\alpha+\beta-s)}{2\pi {r_0}^2 \Gamma(s-\alpha+2)\Gamma(\alpha+\beta-2s+2)} \cdot \left(\frac{r}{r_0}\right)^{s-\alpha} \cdot \left(1+\frac{r}{r_0}\right)^{s-\beta}, \qquad (4.7)$$

where $\Gamma$ represents the Gamma function.

The values of $\alpha$ and $\beta$ respectively are 1.6 and 3.4, while the best $r_0$ turns out to be 30 m for fitting individual showers. The parameter $s$ is often referred to as *shower age*, related in the unmodified NKG function (valid for pure electromagnetic showers) to the stage of development of the shower at the observation level: $s = 1$ means the shower is observed at the maximum of its development, for $s > 1$ and $s < 1$ we observe it after and before it has reached the maximum, respectively. With the present modification of the NKG function, the age has lost its meaning, and has become a pure mathematical variable whose value spans from -0.39 to 1.49, these limits being determined by the poles in the $\Gamma$-function plus a safety value of 0.01.

During the fitting procedure an uncertainty is assigned to each Grande measurement, following the relation given in [37]. The formulas are briefly summarised and depend on which ADC channel has contributed to the measurements:

$$\begin{aligned}
\textbf{PBx10}: \quad \sigma_{sta} &= \frac{1.049}{(n_{ex}\cdot\cos(\theta))^{0.507}} \quad \text{and} \quad \sigma_{sys} = 0.071,\\
\textbf{PB}: \quad \sigma_{sta} &= \frac{1.049}{(n_{ex}\cdot\cos(\theta))^{0.507}} \quad \text{and} \quad \sigma_{sys} = 0.093,\\
\textbf{PB1}: \quad \sigma_{sta} &= \frac{2}{(n_{ex}\cdot\cos(\theta))^{0.5}} \quad \text{and} \quad \sigma_{sys} = 0.102,
\end{aligned}$$

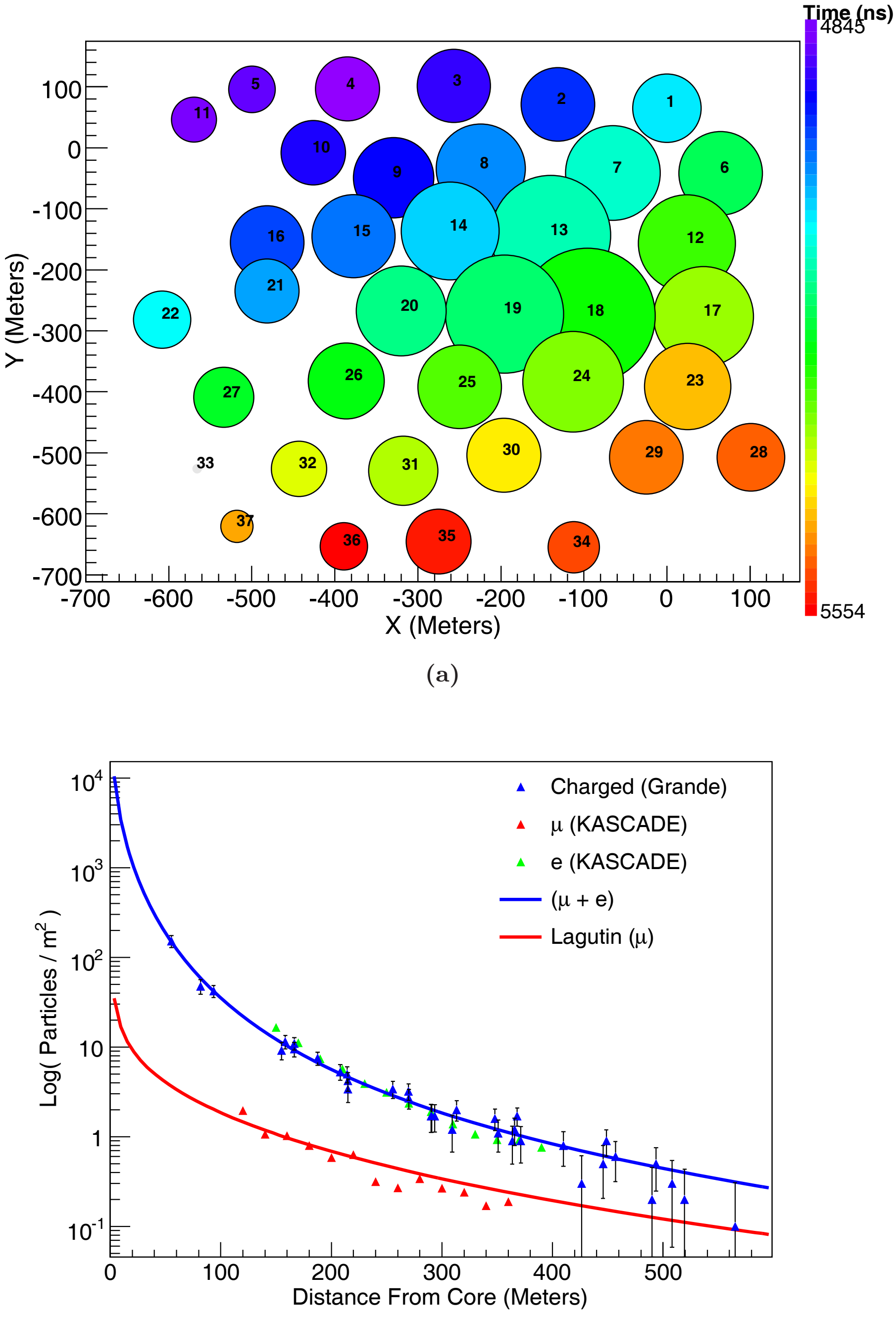
where $\sigma_{sta}$ and $\sigma_{sys}$ are the statistical and the systematic uncertainties associated with each ADC channel and $\mathrm{n_{ex}}$ the number of particles expected from the fit in the station. The error finally assigned to each measurement is

$$\sqrt{{\sigma_{sta}}^2 + {\sigma_{sys}}^2}\cdot n_{measured}. \tag{4.8}$$

An example of air shower event detected by Grande is shown in fig. 4.2. The event footprint on ground is depicted in (a): the size of the circles is proportional to the energy deposit in the station, while the measured arrival time is colour coded. The lateral distribution of this EAS is shown in fig. 4.2b: the red points represent the KASCADE array mean muon densities, in ranges of 20 m distance from the shower axis. Blue points stand for the charged particle densities measured in each Grande detector. Green points symbolise the electron densities measured by the KASCADE array for this event. Error bars are drawn for Grande measurements only, according to equation 4.8. The red line shows the best fitting Lagutin function to the muon measurements, while the best fitting NKG function to the charged particle densities measured by Grande is shown as a blue line. The different parametrisations used to describe the lateral distributions of the electromagnetic and the muonic component of a shower reflect directly the features of the air shower development. During the development of the air shower, the electromagnetic particles spread laterally due to Coulomb scattering and loose energy through ionisation losses in the medium. The lateral spread of muons is mainly due to the transverse momentum of the parent particles. Multiple scattering gives a negligible contribution in this case, as it is being suppressed by a factor of $({}^{m_e}/_{m_\mu})^2$. The overall effect is that muons travel almost undisturbed along a straight line through the atmosphere resulting in a flatter lateral distribution to be observed on ground as compared to the electromagnetic component (see chapter 2).

The muon number is finally calculated and fixed, by applying the formulas 4.3 to 4.6, now with the best parameters found for the shower geometry.

In order to obtain the total number of electrons ($N_e$) in the shower from $N_{ch}$ and $N_\mu$, a new fit of the lateral distribution is done in what is called the third reconstruction level. At each point the contribution of electrons (estimated on the basis of an NKG like function



(a)

(b)

**Figure 4.2:** Measured air shower reconstructed with $N_e = 10^7$, $N_\mu = 10^{5.7}$, $\theta = 11°$ and core position $(x, y) = (-115\,m, -223\,m)$. (a) The event footprint on ground: the size of the circles is proportional to the energy deposit in the station, the measured arrival time is colour coded. (b) The lateral distribution of the muonic and electromagnetic components. The red points show the mean muon density, in bins of 20 m distance from the shower axis. Blue points depict the charge densities measured in each detector. Green points give the electron densities measured by the KASCADE array. The red line represent the best fitting Lagutin function and blue is the sum of the best fitting NKG and Lagutin functions.

with $N_e$ and shower age $s$ as free parameters) and muons (from a Lagutin function with $N_\mu$ fixed by the previous step) to the measured densities is simultaneously taken into account, i.e.:

$$\rho_{ch}(r) = \rho_e(r) + \rho_\mu(r)$$

with

$\rho_{ch}$ = measured particle density at the shower core distance $r$,

$\rho_e = f(N_e, s)$, $f$ being the modified NKG function,

$\rho_\mu = f(N_\mu)$, $f$ being the Lagutin function.

## 4.2 Air shower simulation

Complete air shower simulations are necessary to interpret the recorded event in terms of primary energy and mass, since at these primary energies only indirect measurements of cosmic rays are possible. They are used to understand the connection between the measured quantities and the simulated properties of an air shower. Also more complicated analyses have to rely up to a certain level on the comparison between data and simulation. Unfortunately, the interactions of the shower particles happen at energies higher than what has been measured at accelerator sites, so air shower simulations have to rely on models. By comparing data and simulations, constraints can be put on the actual hadronic interaction models, which are the basis of the simulation scheme. For KASCADE-Grande, the simulation of an EAS is performed with COsmic Ray SImulation for KAscade (CORSIKA) [46]. It is a Monte-Carlo method that describes the development of extensive air showers by tracking nuclei, hadrons, muons, electrons and photons through the atmosphere until they undergo reactions with the nuclei of the air or decay. The most important decay channels are taken into account down to the 1% level. The primary particle's type, its energy and direction are upon the choice of the user, as well as the hadronic interaction models to be used. These can be divided into two groups, describing low and high energy hadron interactions with an adjustable transition energy, dependent on the reliability of the models. Electromagnetic processes are described by a special version of the shower program EGS4 (Electron Gamma Sower code version 4) [47]. In this work, the most used simulations for the KASCADE-Grande experiment have been produced with CORSIKA 6.307, the QGSJet-II-2 [48; 49] model, for the high energy hadronic interactions, and FLUKA 2002.4 [50] for the low energy ones. The transition energy between the former and the latter has been set to 200 GeV. The showers are simulated in an energy range from $10^{15}$ eV to $10^{18}$ eV, following a power law energy spectrum with spectral index $\gamma = -2$. The zenith angle of the incident particle spans from 0° to 42°, with an extension to 70° for some samples. Altogether, in order to study the composition of the cosmic ray spectrum, five primary particle types have been used: proton, helium, carbon, silicon and iron. Showers generated by CORSIKA are passed to Cosmic Ray Event Simulation (CRES), the detector simulation code based on the GEometry ANd Tracking 3 (GEANT3) [51; 52] tool. Since the full simulation of EAS with CORSIKA is time consuming, each simulated shower is used more than once (10, 15 or 20 times) to increase statistics. The simulated air shower is translocated to a different position spreading its core randomly over an area that is 1, 1.5 or 2 times the effective detector area. The response of CRES has the same

| sets | passes | area | zenith angle | number of events per set, for $10^{15}eV < E_0 < 10^{18}eV$ |
|---|---|---|---|---|
| 5 | 10 | $4.9 \cdot 10^5 m^2$ | 0-42° | 64180 |
| 1 | 15 | $7.35 \cdot 10^5 m^2$ | 0-42° | 96270 |
| 3 | 10 | $7.35 \cdot 10^5 m^2$ | 0-70° | 64180 |
| 1 | 20 | $9.8 \cdot 10^5 m^2$ | 0-70° | 128360 |

**Table 4.2:** Simulated air showers.

data format as measured data, so that after calibration we can reconstruct the events with the same KRETA code and be able to directly compare simulations and data. After the detector response simulations, we obtain per set roughly 64000 shower above $10^{15}$ eV for each primary particle. In table 4.2, a summary of the statistics used for this work is given.

## 4.3 Trigger efficiency and reconstruction accuracy

The study of the trigger efficiency of Grande is done on the basis of full CORSIKA and GEANT3 simulations. Simulated events have been selected with their reconstructed core inside a predefined fiducial area (fig. 4.4a) and only showers able to trigger a four-fold or a seven-fold coincidence have been considered. In fig. 4.3, it is clearly visible that 100% efficiency is reached in Grande for 7/7 triggers at a total true electron number of $10^{5.8}$, which corresponds to $\approx 10^{16}$ eV primary energy. It has to be pointed out that there is a small difference in the shower size threshold when looking at proton and iron primaries separately. The iron induced air showers start to trigger Grande at a smaller value of total electron number (but at higher primary energies), because they contain on average a higher number of muons than the proton initiated ones and Grande is triggered by charged particles, without any distinction between muons and electrons.

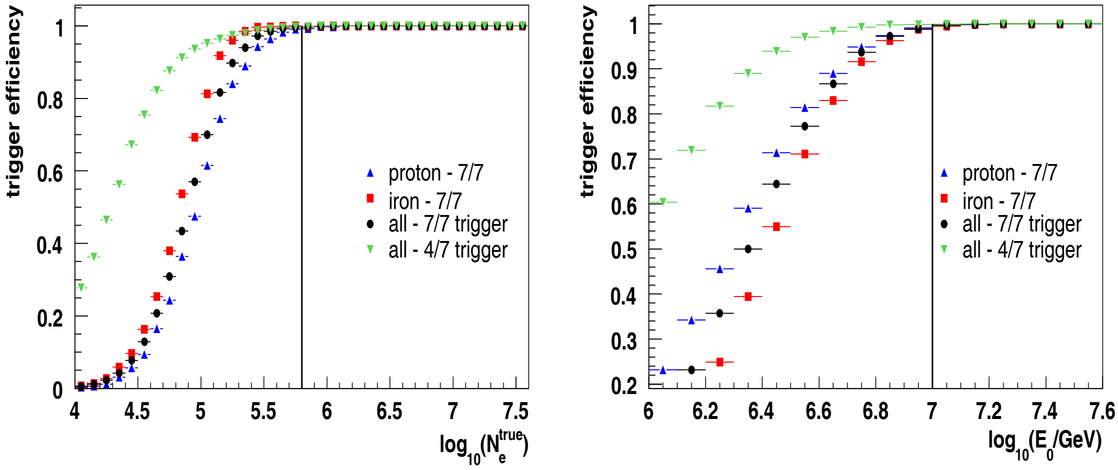


**Figure 4.3:** Left: trigger efficiency as a function of the total simulated electron number. Right: trigger efficiency as a function of the simulated primary energy.

| Hardware cut | | |
|---|---|---|
| Description | Variable | Function |
| 7/7 trigger | $T_7$ | $T_7 > 0$ |
| 18 active clusters | calculated from data | |
| KASCADE array active | $I_{act}$ | $I_{act}\& \; 1 = 1$ |
| no ANKA induced events | $F_{anka}$ | $F_{anka} < 4$ |
| maximum energy deposit in inner Grande station | $D_{max}$ | $D_{max} > 0$ |
| valid TDC hits in at least 12 stations | $N_{det}$ | $N_{det} > 11$ |

**Table 4.3:** List of applied hardware related cuts.

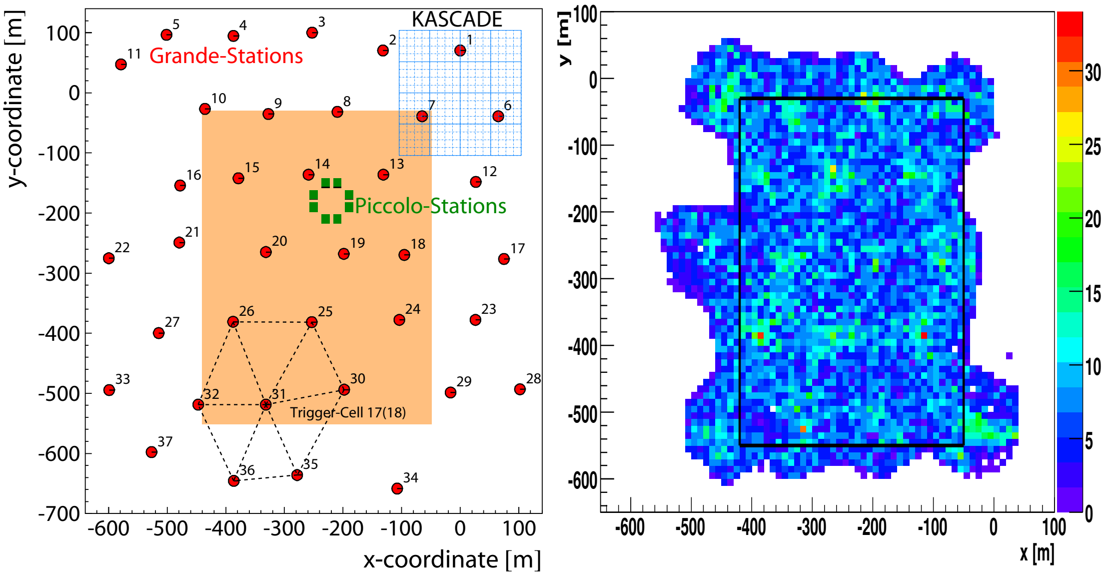
The reconstruction accuracy is also studied on the basis of simulations. Simulated shower observables are compared event by event to the corresponding reconstructed ones. A sequence of cuts is applied to the data in order to obtain high accuracy. They can be divided into two groups: **hardware related cuts** and **reconstruction cuts**. To the first category belong those which are independent from the reconstruction procedure, the second one is instead reliant on the reconstructed variables.

In tables 4.3 and 4.4, a short description of these two categories is given:

- only sevenfold (sixfold for cluster 15) coincidences are taken into account for trigger;
- only periods of full efficiency of the detectors are used: this means that, besides requiring always all the 18 Grande clusters to be active, also events for which there has been some technical error in the DAQ electronic are discarded;
- the data contain some events which are not air showers, but the result of the interference of the synchrotron radiation source ANKA, located in the south west region of Grande. During the injection of the electron beam and during beam dump, ANKA emits a particle cloud which hits enough Grande stations to fake a real shower trigger. All this information is stored in a database, accessible during the analysis [36];
- the requirement of having the KASCADE array always active in the event is necessary, since it provides the muon measurements;
- the requirement that the station with maximum energy deposit measured is not on the border of the Grande array ($D_{max}$ cut) has been chosen to exclude air showers that could have fallen outside our sampling area and of which we could see then only a tail;
- the $N_{det}$ cut, i.e the exclusion of events with less than 12 stations triggered, is necessary to cut some small events that are below our energy threshold and for which the reconstruction procedure is unreliable.

Going through the list of the reconstruction cuts:

- events are discarded where the minimisation procedure performed by MINUIT [53] has failed ($\approx 0.3\%$ of the sample);



(a)

(b)

**Figure 4.4:** (a) KASCADE-Grande scheme, in orange the fiducial area used. (b) Effect of the $D_{max}$ cut on the collecting area. The fiducial area has been chosen inside the boundary of the $D_{max}$ cut.

| Reconstruction cut | | |
|---|---|---|
| Description | Variable | Function |
| successful MINUIT fit | $N_{flag}$ | $N_{flag} > 0$ |
| Zenith angle < 40° | $\theta$ | $\theta < 40°$ |
| Fiducial area | $X_c$,$Y_c$ | 192400 $m^2$, see fig. 4.4a |
| Age limit discarded | $s_{age}$ | $-0.385 < s_{age} < 1.485$ |
| Total sampled energy deposit higher than a threshold value | $N_c^{tot}$ | $\log_{10}(N_c^{tot}) > a \cdot \log_{10}(N_e^{rec}) - b$ |

**Table 4.4:** List of applied reconstruction cuts.

- the analysis contains only events with a zenith angle less than 40° because the reconstruction procedure has been optimised to this angular range;

- the fiducial area (see fig. 4.4a) has a rectangular shape centred around the Grande DAQ; when applying the $N_{det}$ cut, the distribution of the shower core is not anymore homogeneous over the whole array. The fiducial area has thus been chosen to minimise this effect as shown in fig. 4.4b;

- reconstructions where the shower age value has reached one of the two limits imposed during the fit of the lateral distribution are rejected, since in this case the fit itself is not reliable;

- the cut on the total sample energy is fully discussed in the next section.

### Cut on the sum of energy deposit

The variable $\mathrm{N_c^{tot}}$ represents the sum of the energy deposits in the Grande scintillators by the shower's particles. Looking at the distribution of $\log_{10}(N_c^{tot}/8.5)^1$ as a function of the shower size (fig. 4.5), a group of events with small $\mathrm{N_c^{tot}}$ values whose reconstructed size is much larger than the true one is noticeable. Clearly this unphysical behaviour is due to misreconstruction of the events. The discrimination power of this observable has been found in simulated data. This group of showers can be discarded simply by finding

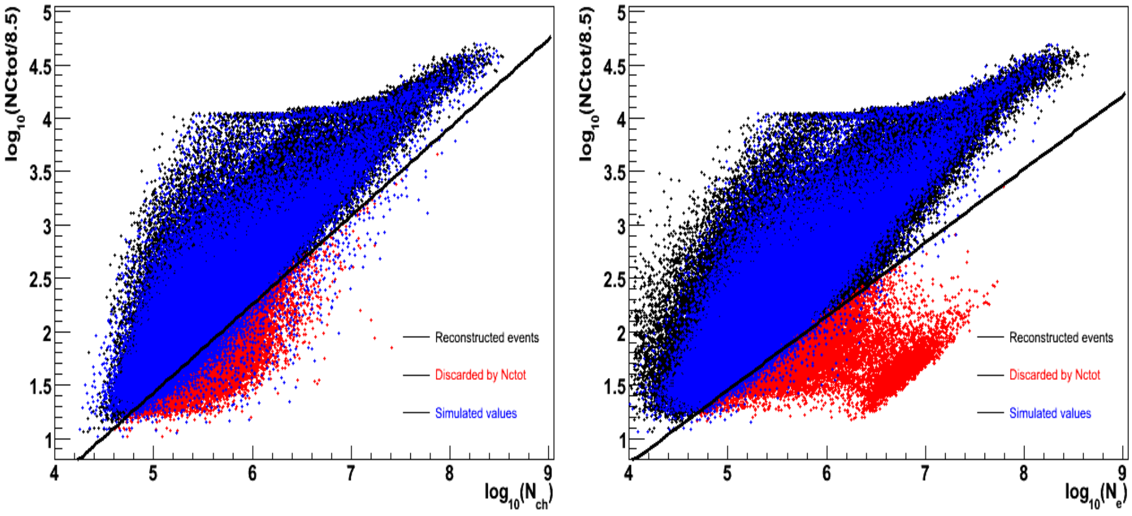


**Figure 4.5:** Distribution of total energy deposit as function of the shower size for simulations. Black are the reconstructed simulated values, blue the corresponding true simulated sizes and in red the events which are going to be discarded cutting along the black line. The distributions for reconstruction level 2 (left ) and for level 3 (right) are shown. The accumulation of events at $\log_{10}\mathrm{N_c^{tot}}/8.5 \approx 4$ is due to station saturation, that is simulated at 90 GeV.

---

[1] 8.5 MeV is the most probable energy deposit by a single particle in one station detector

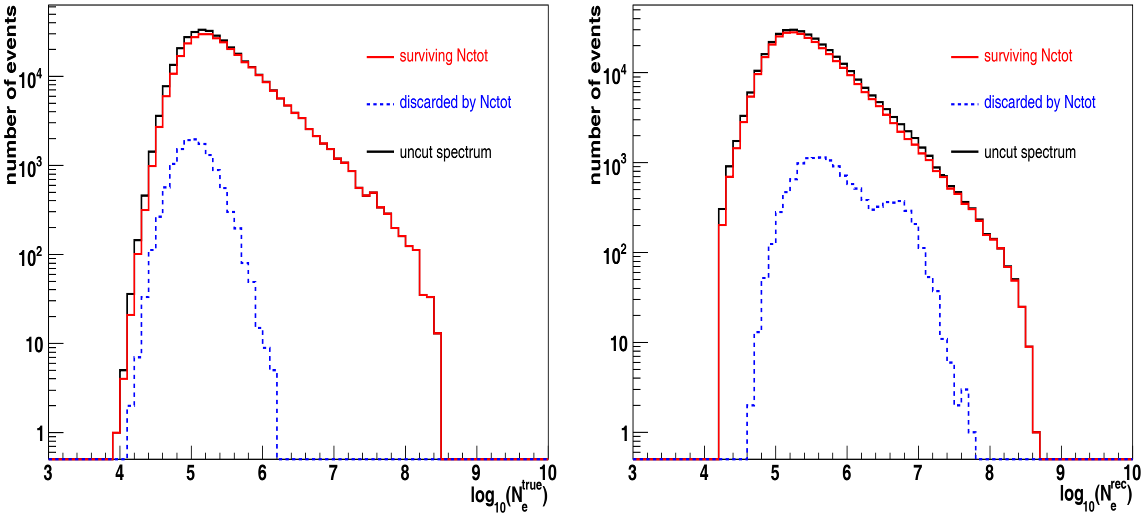


**Figure 4.6:** Effect of the $\mathrm{N_c^{tot}}$ cut on the distribution of the true electron number (left) and on the reconstructed electron number (right) at reconstruction level 3.

a straight line in this distribution, dividing good reconstructed events from the bad ones. This division is given by:

$$\log_{10}\left(\frac{\mathrm{N_c^{tot}}}{8.5}\right) = a \cdot \log_{10}(N_e) - b$$

with $a = 0.83$, $b = 2.71$ at reconstruction level 2, and $a = 0.69$, $b = 2$ at level 3, optimised by simulation and cross-checked on measured air showers. The parameters are different for level 2 and level 3, for two reasons: first, the meaning of shower size at level 2 is the total number of charged particles in the shower $N_{ch}$, whereas at level 3 it is the total number of electrons $N_e$; the second is that at level 3 the shower size reconstruction is affected by the total muon number reconstruction, thus being an additional source of reconstruction uncertainty on the electron size. In simulated air showers the electron shower size is the total number of electrons with energy higher than 3 MeV at observation level. In measured events it is the number of electrons in the shower, as estimated from the energy deposits in the detector after the correction for muons and photon contributions. The effect of this cut, after the application of all the others, is tested both on simulations and on real data. As seen in fig. 4.6, its application does not bias the true simulated spectrum above $\log_{10}(\mathrm{N_e}^{\mathrm{true}}) \geq 6$. In fig. 4.7, the analogous result for measured data of fig. 4.5 is displayed. Here it is also possible to see the misreconstructed events in the region with low values of total energy deposit, but with high reconstructed shower size; at both levels the slope of the dividing line has been optimised by looking at the reconstruction of the lateral distribution of the events with the highest reconstructed sizes, trying to avoid to discard those for which the lateral distribution fit is reliable.

From the reconstructed $N_e$ distribution (fig. 4.8) of discarded and accepted events, it appears clear that this cut is removing unphysical features from the distributions. This is as well confirmed by the distribution of the shower age for these events, where mainly events with reconstructed parameter $s$ in the negative range are discarded.

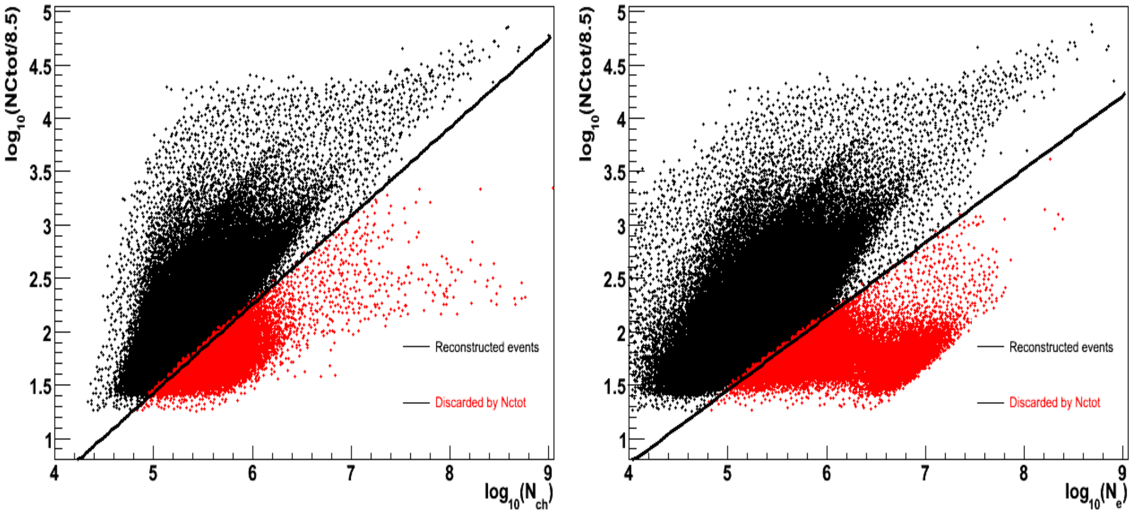


**Figure 4.7:** Distribution of $N_c^{tot}$ as function of the reconstructed shower size for data. Black are the events after all the cuts but the $N_c^{tot}$ one, red the events discarded by $N_c^{tot}$. On the left the distribution for level 2 is shown, on the right for level 3.

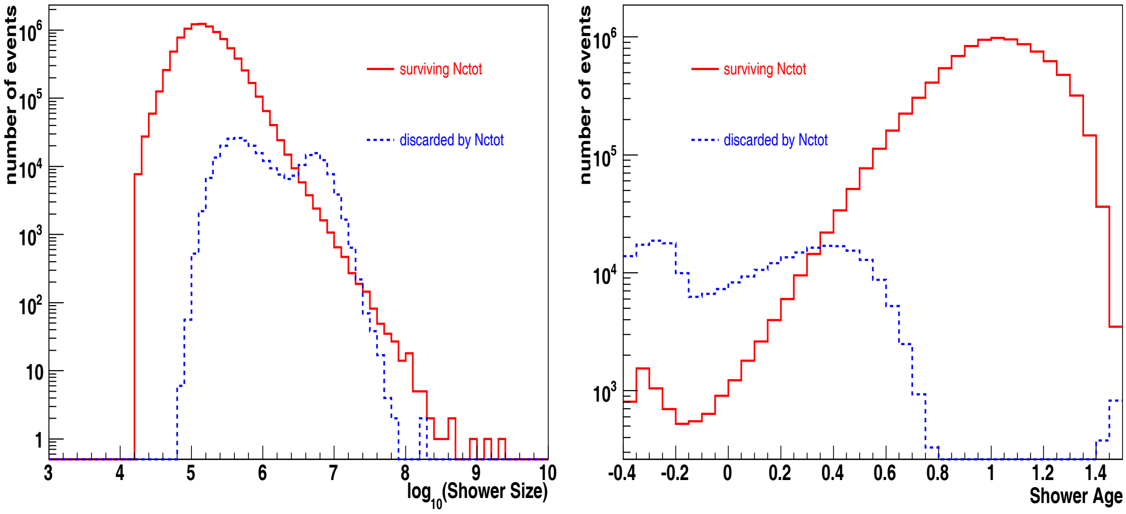


**Figure 4.8:** Left: shower size distribution for accepted (red) and discarded (blue) events at level 3. Right: the same for the shower age distribution

Finally, a check of the trigger and reconstruction efficiency is applied (fig. 4.9), including reconstruction and hardware cuts. The only cut having a strong influence on the efficiency is the requirement of at least 12 stations having triggered. Increasing this number would mainly shift the threshold, because we are indirectly increasing the primary energy threshold. As for the other cuts, no particular effect on the efficiency has been found; efficiency shortcomings are due to the fluctuation of shower cores into and out of the fiducial area at its edges.

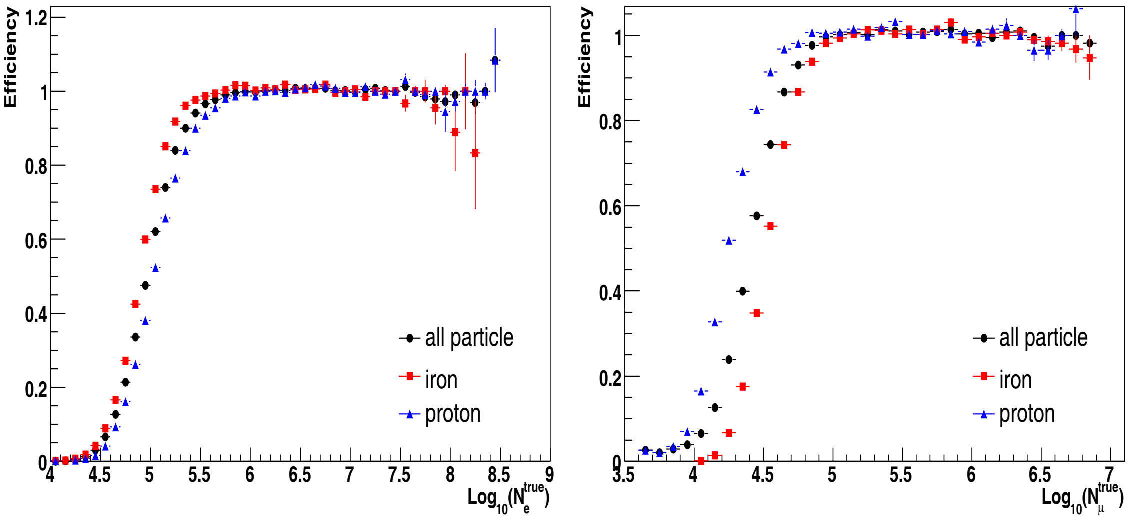


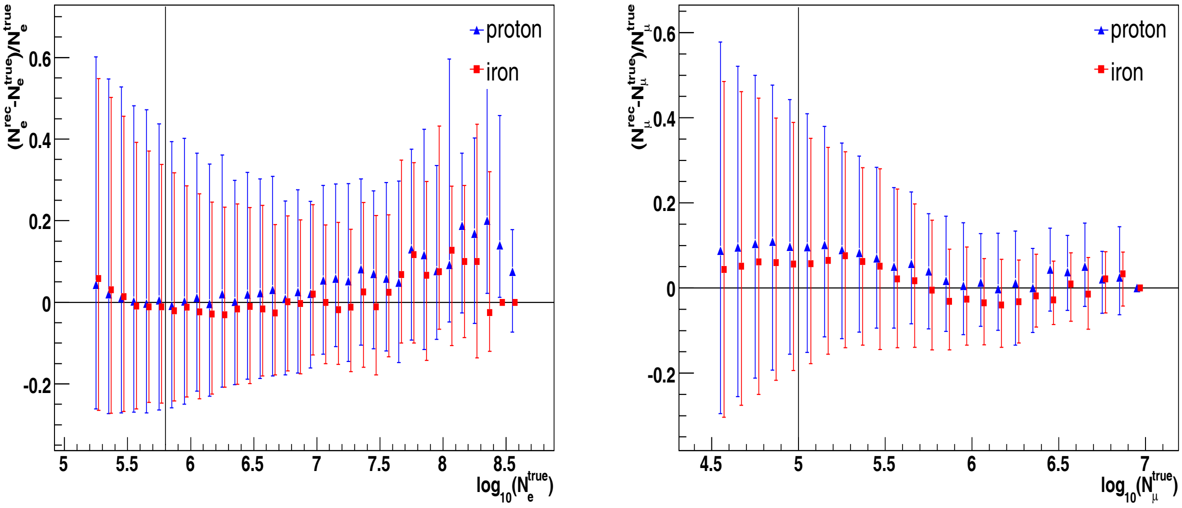
**Figure 4.9:** Reconstruction efficiency (including trigger efficiency) as a function of the total true electron number (left) and as a function of the total muon number (right)

The reconstruction accuracy after having applied all the cuts discussed above is shown in fig. 4.10 and fig. 4.12. The analysis is performed separately on the lightest and heaviest primary particles for which simulations are available, i.e. on protons (blue triangle) and iron (red squares), in order to see mass dependencies.

In fig. 4.10a, the reconstruction accuracy of $N_e$ at level 3 is shown. The relative difference between the true and the reconstructed value is calculated event by event. The points shown in the plot represent the median of the distribution, while the upper and the lower limit of the error bars are the quantiles at 84% and at 16%, respectively, so that 68% of the total number of events are included within these limits. The same procedure is applied in fig. 4.10b to calculate the accuracy on the total number of muons.

The plot indicates a reconstruction bias for $N_e$ lower than 5%, with an increase up to 15% starting at $N_e \approx 10^{7.5}$. No strong dependence on the mass of the primary particle is observed. The statistical fluctuation amounts to $+24\%/-20\%$ for proton primaries and $+20\%/-18\%$ for iron induced air showers. Concerning the total muon number reconstruction, the analysis of the simulated data shows a bias of 10% at the threshold ($\approx 10^5$), fluctuating around 0% at the highest energies, and a statistical uncertainty of 25%, decreasing to 10%.

In fig. 4.11 the correlations of the reconstruction biases on $N_e$ and $N_\mu$ are shown. The behaviour of the total muon number reconstruction biases is similar to what has been



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

**Figure 4.10:** Reconstruction accuracies at level 3 for shower size (a), total muon number (b). The error bars are the quantiles at 84% and at 16%, respectively, so that 68% of the total number of events are included within these limits. For better visualisation, a small shift of +0.02 along the x-axis is applied to the iron primaries (red squares).

observed in fig. 4.10b, with a stronger effect for iron showers. From fig. 4.11a a correlation between the overestimation of the shower size and the total muon number appears, although it follows the same features observed in fig. 4.10a. The correlation between these two observables is due to the estimation procedure applied in the third reconstruction level (see section 4.1.3).

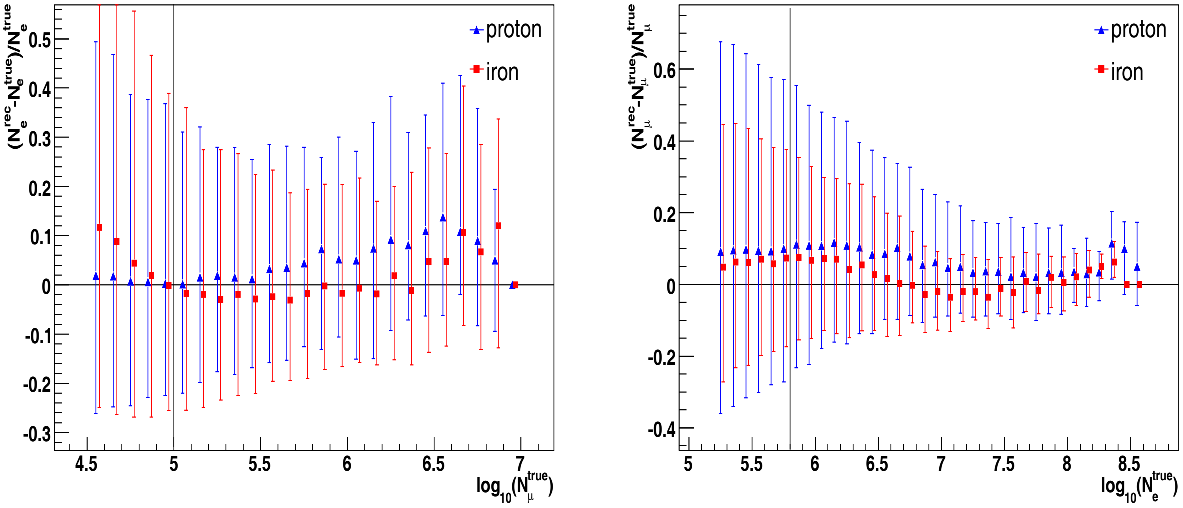
The angular and the core position accuracies as a function of the simulated shower size are shown in fig. 4.12a and 4.12b, respectively. Since these quantities follow a Rayleigh distribution, the mean value and the standard deviation of the best fitting Rayleigh function to our distribution are depicted in the figures. The position of the shower impact on ground is reconstructed with a precision of 6 m at full efficiency. The arrival direction has an accuracy better than 0.4° along the whole range of full efficiency of Grande.

All these accuracies, obtained from simulations, have been confirmed by investigating a subsample of events independently reconstructed by KASCADE and Grande [37].

Applying the discussed cuts, the two-dimensional spectrum $N_e$ as a function of $N_\mu$ for the measured data is obtained. In fig. 4.13 the spectrum is shown. Events above the experimental thresholds in the angular range 0° to 40° are considered. A total number of 344,902 events has been detected. The correlation between $N_e$ and $N_\mu$ gives information on the primary energy and mass, being thus subject to further analyses.

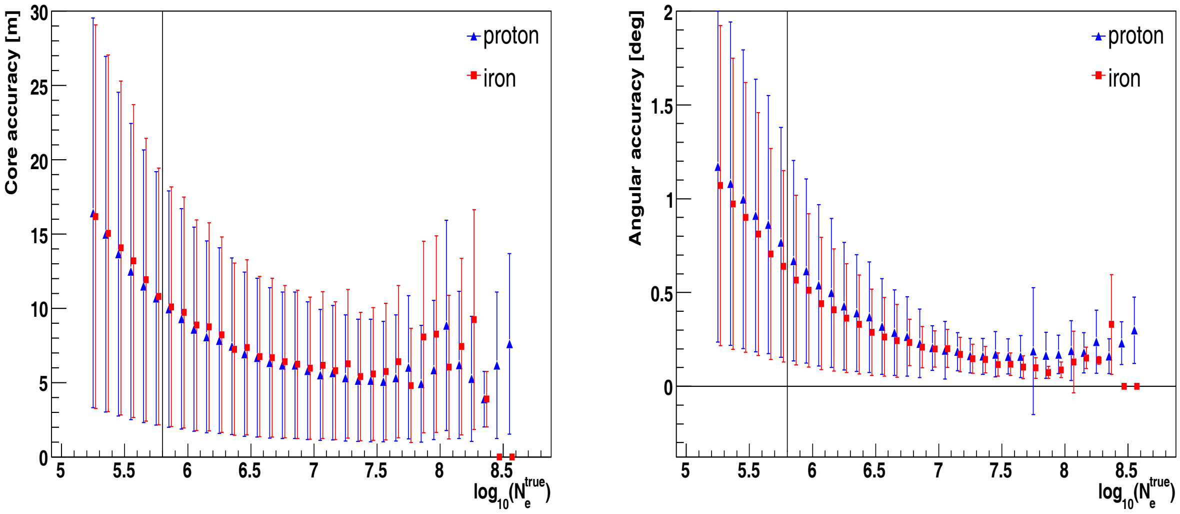
## 4.4 Not well reconstructed events

Although specific cuts have been developed in order to discard events which are not correctly reconstructed, a small fraction passes all of them.



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

**Figure 4.11:** Reconstruction biases at level 3 for shower size as a function of the muon number (a) and for the total muon number as a function of the shower size (b). The error bars represent the quantiles at 84% and at 16%. A shift of +0.02 along the x-axis is applied to the iron primaries (red squares).



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

**Figure 4.12:** Reconstruction accuracies at level 3 for core position (a) and arrival direction (b). The error bars represent the spread of the underlying distributions. A small shift of +0.02 along the x-axis is applied to the iron primaries (red squares).

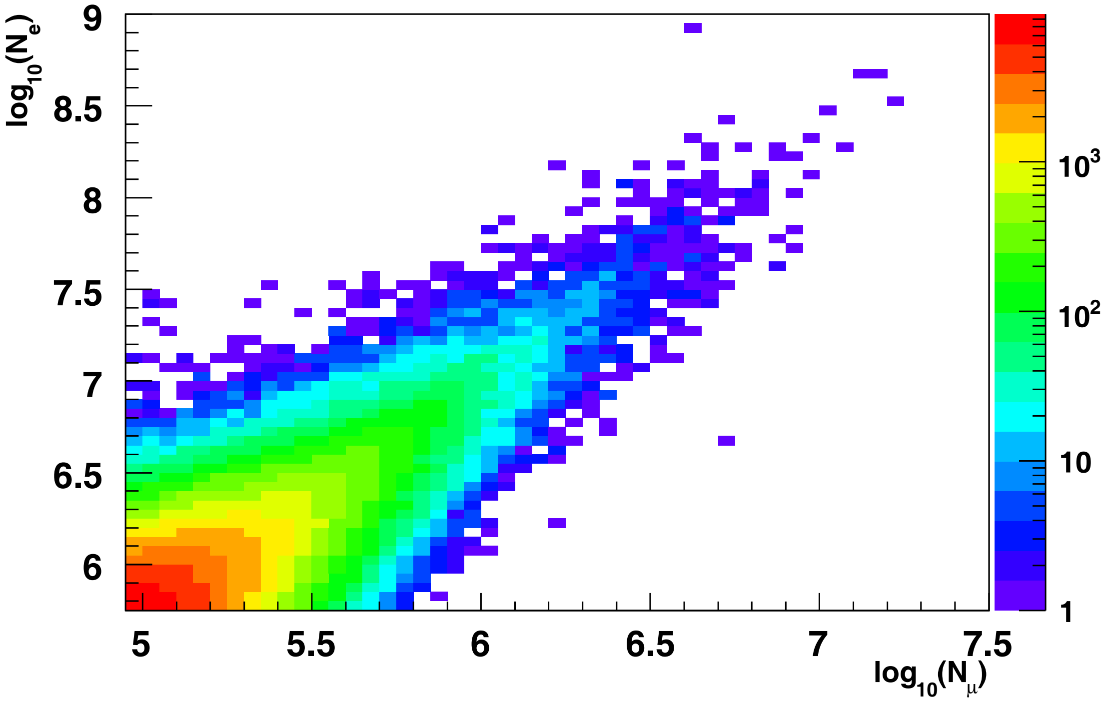


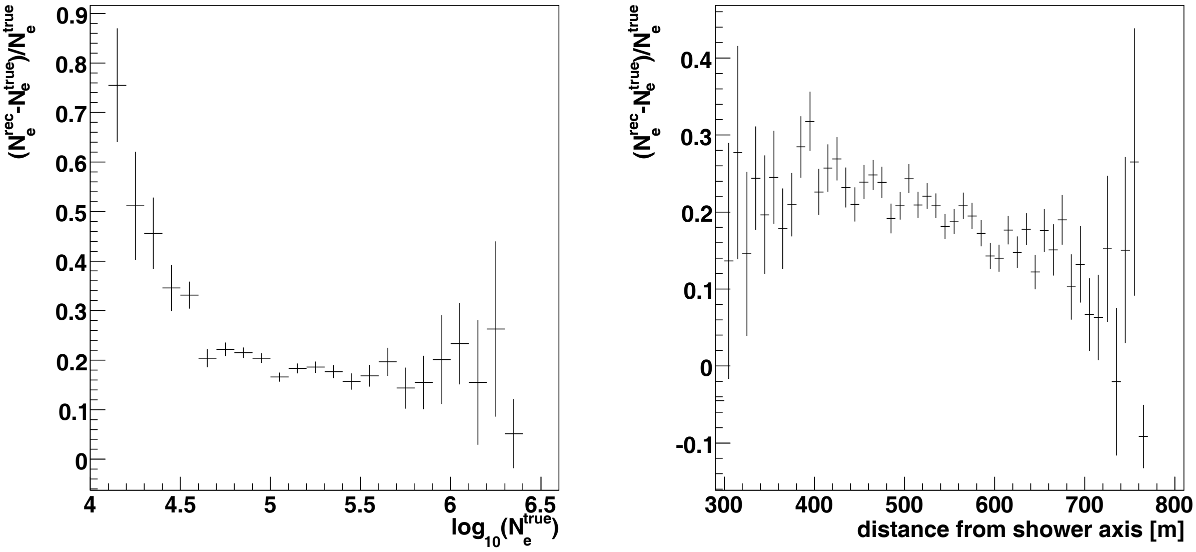
**Figure 4.13:** The two-dimensional shower size spectrum. It shows the correlation between $N_e$ and $N_\mu$. Air showers up to a zenith angle of 40° and above the thresholds in muon and electron numbers ($\log_{10}(N_e) > 5.8$ and $\log_{10}(N_\mu) > 5$) are considered.

It has been observed that such events contain one (or more) **outlier** station (in the following referred to as outliers), i.e. a single station whose energy deposit differs (usually exceeds) substantially from the lateral distribution. Depending on the distance of these detectors to the shower core, the NKG fit can be more or less biased. No correlation between the existence of an outlier and the reconstructed shower properties has been found (fig. 4.15). Moreover, such events can also be seen in simulations; from their study, it has been observed that the existence of an outlier station does not necessarily leads to a wrong reconstruction (see fig. 4.14).

Outlier stations may be caused by single hadrons, interacting very close to the detector itself and producing a small cascade of particles. Also, in the measured events flutuating electronic noise can lead to the same effect. In order to estimate the extent of these showers in the total data sample, a procedure to identify them has been developed in the reconstruction software. A detector station more than 100 m from the calculated core is identified as an outlier when the deviation $\delta$ of the measured density from the expected one exceeds 10 sigma:

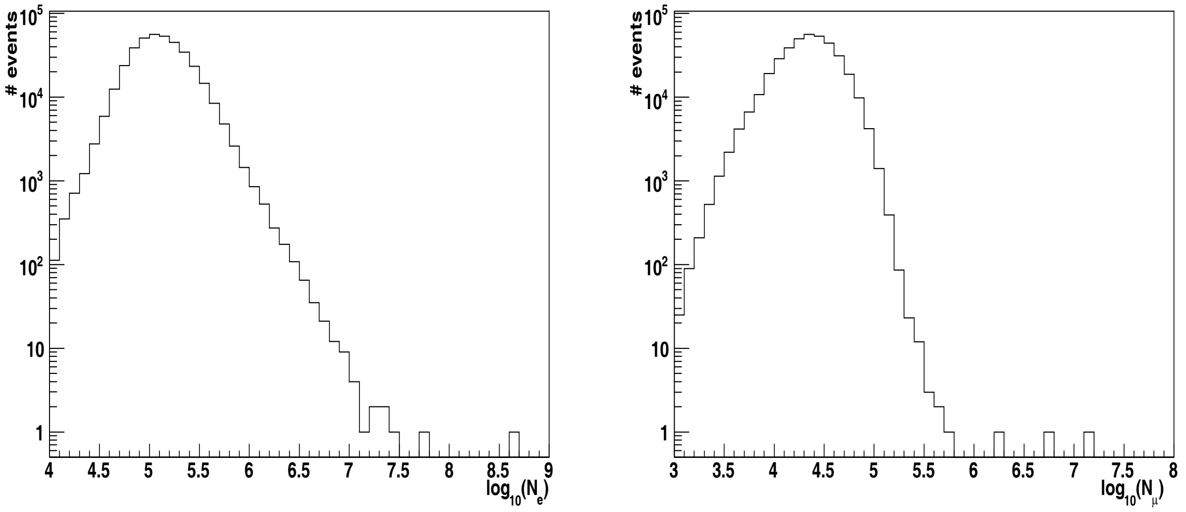
$$\delta = \frac{\rho_m - \rho(r)}{\sigma} > 10, \tag{4.9}$$

with $\rho_m$ being the measured density, $\rho(r)$ being the estimated particle density according to equation 4.7 at a distance $r$ from the shower axis, and $\sigma$ being the error as calculated in equation 4.8.



(a) (b)

**Figure 4.14:** Distribution of the reconstruction error on the shower size as a function of the simulated total electron number (a) and as a function of the distance of the outlier station from the shower axis (b) for events with one or more outlier stations (see text).



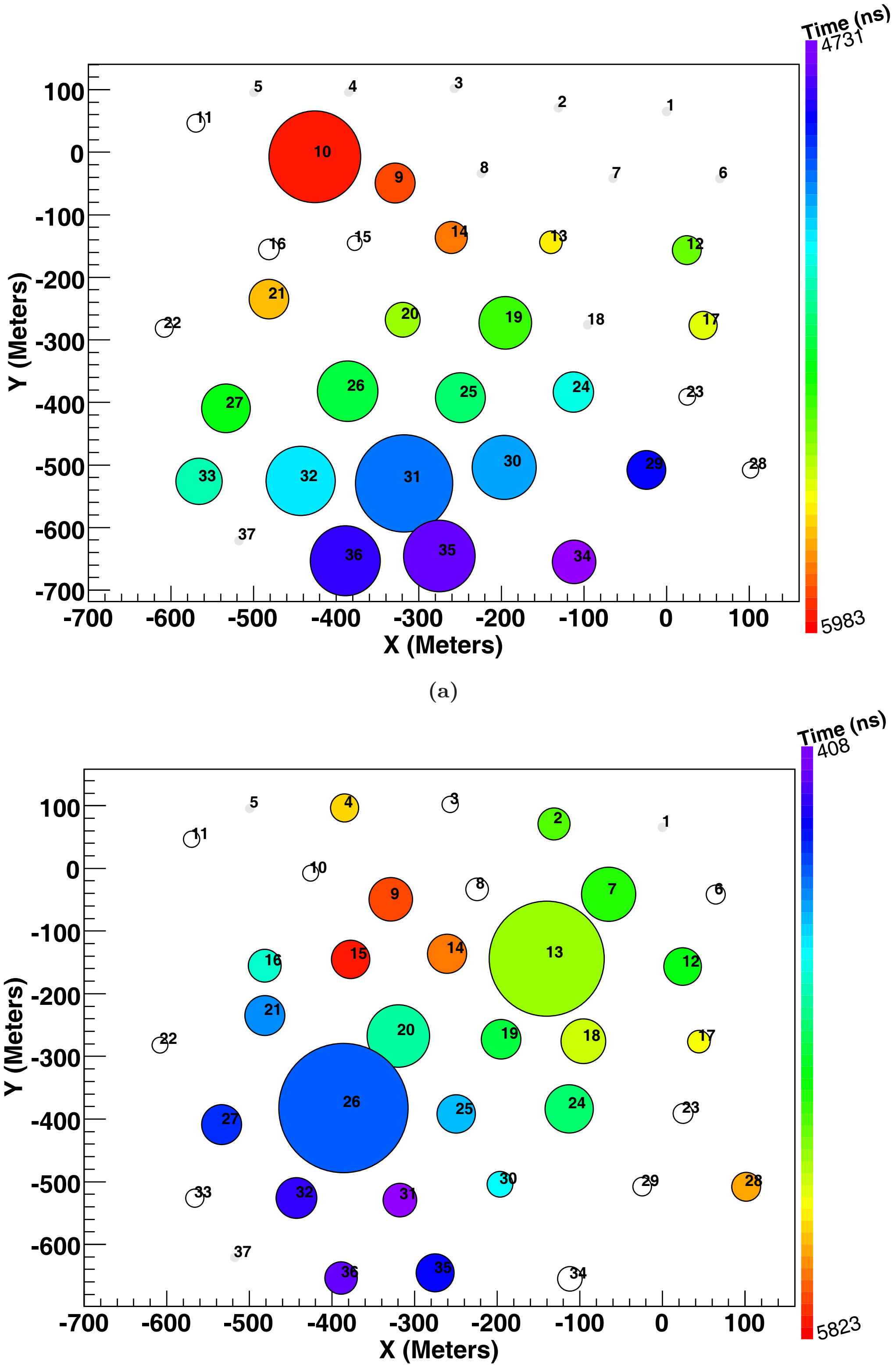
(a) (b)

**Figure 4.15:** Distribution of the electron shower size (a) and of the total muon number (b) for events with one or more outlier stations (measured data).

The distributions (fig. 4.15) of the electron and muon shower sizes for such events demonstrate that they concentrate mainly around the experimental threshold region. The number of events with one or more outliers is 1.5% of the total number of reconstructed showers above threshold, after applying all the cuts. It is lowered by a factor 10 when increasing the $N_{det}$ cut from 11 to 19 stations.

An iterative procedure to recover this type of events has been tested, consisting in the identification of the outlier station and its exclusion in a second step from the fit of the lateral distribution. This method is able to recover some events (see fig. 4.16a and 4.17), but it fails on others whose lateral distribution presents some not well understood features. An example of the latter case is shown in fig. 4.16b and 4.18, where by applying the foregoing procedure, the two stations with the highest energy deposits are subsequently identified as outliers and excluded from the fit, resulting in an almost flat lateral distribution and so in a shower age out of the accepted range.

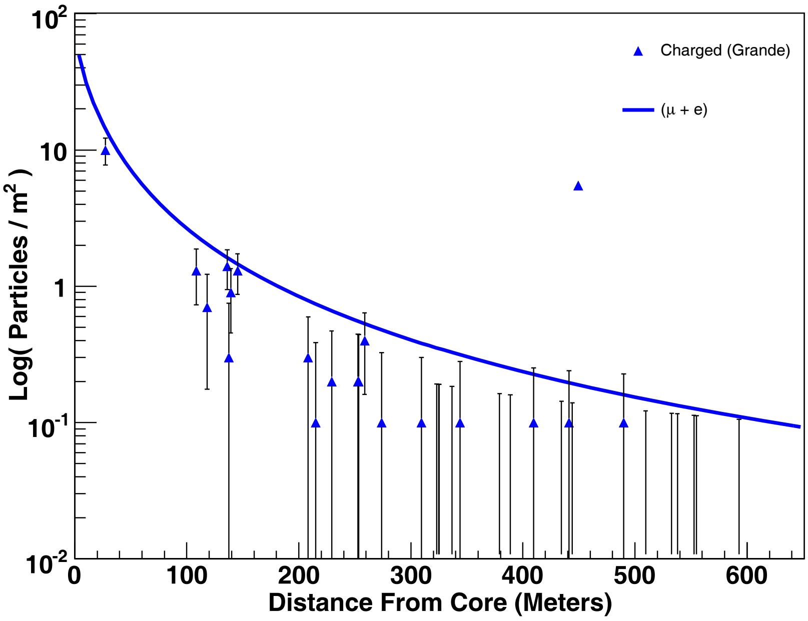
Due to the small fraction (1.5%) of events affected by this problem, which are mainly distributed around the threshold, and to the instability of the procedure itself, it has been decided not to recover them, since they do not really bias the final parameter distribution.



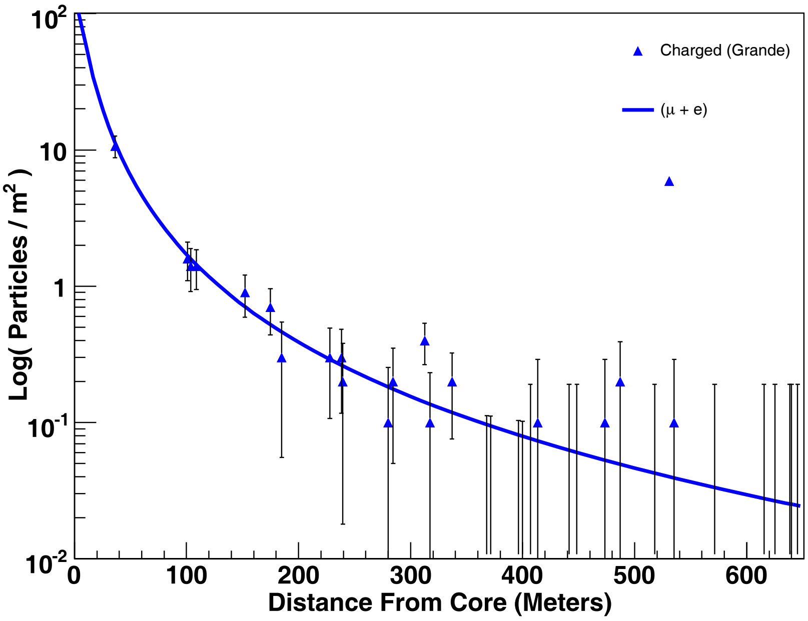
(a)

(b)

**Figure 4.16:** Footprints on ground of two events with an outlier station: (a) event with successful recovery procedure, see fig. 4.17 for the lateral distribution; (b) event with failing recovery procedure, see fig. 4.18 for the lateral distribution.

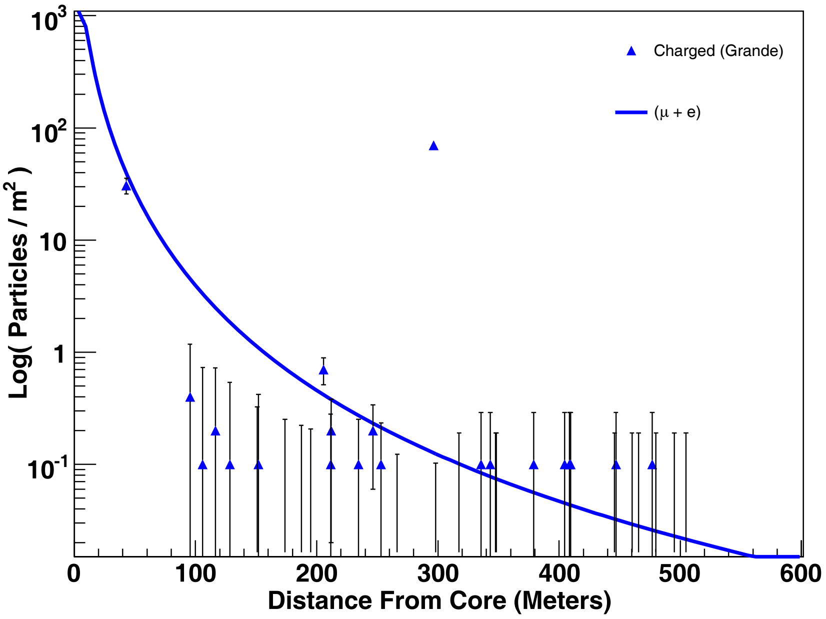


**(a)**

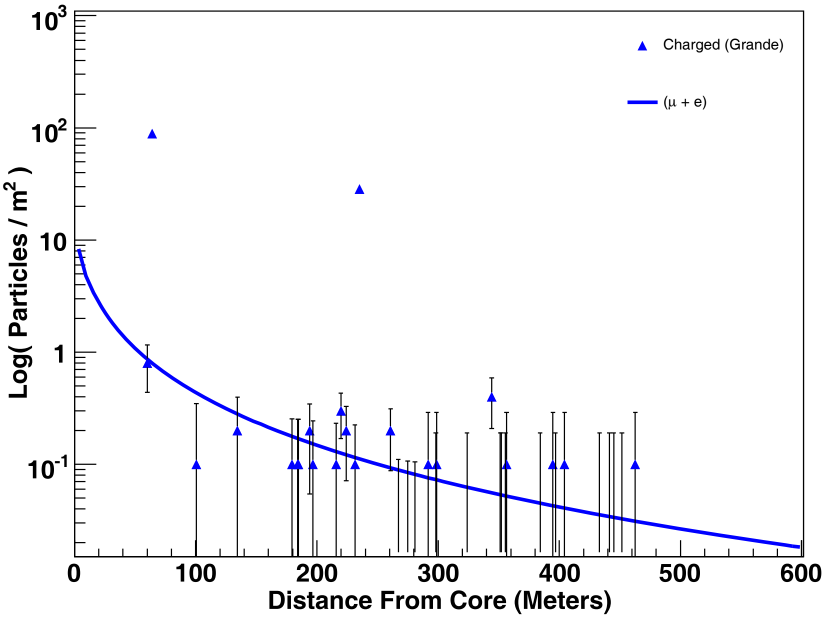


**(b)**

**Figure 4.17:** Example of an event with an outlier station and successful recovery procedure. (a) The lateral distribution of the electromagnetic component. Station 10 at 450 m distance is identified as outlier. Reconstructed parameters: $N_e = 10^{7.1}$, $N_\mu = 10^{4.88}$ and $s_{age} = 1.49$. (b) The final lateral distribution fit after the exclusion of the outlier station (now at 520 m distance). Reconstructed parameters: $N_e = 10^{5.59}$, $N_\mu = 10^{4.91}$ and $s_{age} = 1.14$.



(a)



(b)

**Figure 4.18:** Example of an event with outlier stations and failed recovery procedure. (a) The lateral distribution of the electromagnetic component. Station 13 (300 m distance) and 26 are successively identified as outliers. Reconstructed parameters: $N_e = 10^{6.27}$, $N_\mu = 10^{3.85}$ and $s_{age} = 0.7$. (b) The final lateral distribution fit after the exclusion of the two stations. Reconstructed parameters: $N_e = 10^{6.09}$, $N_\mu = 10^{4.79}$ and $s_{age} = 1.49$.

# 5. Shower size spectra

The main observables of KASCADE-Grande are the muon shower size and the electromagnetic shower size. The total muon and electron number spectra are related to the primary energy spectrum, therefore its spectral features will be present in the size spectra.

Information on the primary particles can be obtained by the analysis of the two-dimensional spectrum obtained from the correlation of the total electron and the total muon numbers.

In this chapter, the reconstructed shower size spectra will be discussed, together with an attempt to disentangle the true shower electron number spectrum from the reconstruction uncertainties.

## 5.1 Differential total electron number spectrum

Applying the data selection procedure described in the previous chapter, it is possible to build the differential electron size spectrum. The shape of the total electron number spectrum is reflecting the shape of the cosmic ray energy spectrum. Because the electromagnetic shower size $N_e$ develops proportionally to a power of the primary energy $E^{\alpha}$, at least over two decades with more or less a constant index, deviation from a simple power law in the primary energy spectrum is reflected into the $N_e$ spectrum. The shower size spectrum can be described by a power law:

$$I(N_e) = A \cdot \left(\frac{N_e}{N_e^0}\right)^{\gamma}, \tag{5.1}$$

where $I$ is the intensity, i.e. the number of measured events per second, unit area and steradian, $\gamma$ is the spectral index, and $A \cdot \left(N_e^0\right)^{-\gamma}$ is the normalisation factor.

Due to the propagation through the atmosphere, showers of the same primary mass and same primary energy are more attenuated the higher the inclination, so that shower sizes are reduced, i.e. less particles reach the observation level. To carry out the analysis of the shower size spectra, it is thus necessary to divide the data in zenith angular ranges.

The zenith angle ranges in this work have been chosen in order to cover the same solid angle each. As shown in table 4.4 the analysis is performed up to a zenith angle of 40°. Selecting a total number of bins $N$, the upper limit for the $i_{th}$ bin is given by

$$\theta_i = \arccos\left(1 - \frac{i}{N} \cdot (1 - \cos(\theta_{max}))\right),$$

with $\theta_{max} = 40°$. The five selected ranges $[\theta_{i-1}, \theta_i)$ are given in table 5.1.

| bin number | $\theta_{i-1}$ (deg) | $\theta_i$ (deg) |
|---|---|---|
| 1 | 0 | 17.60 |
| 2 | 17.60 | 24.99 |
| 3 | 24.99 | 30.73 |
| 4 | 30.73 | 35.63 |
| 5 | 35.63 | 40 |

**Table 5.1:** The five zenith angle ranges used in this work, each of which covers a solid angle of $\approx 0.294\,\mathrm{sr}$.

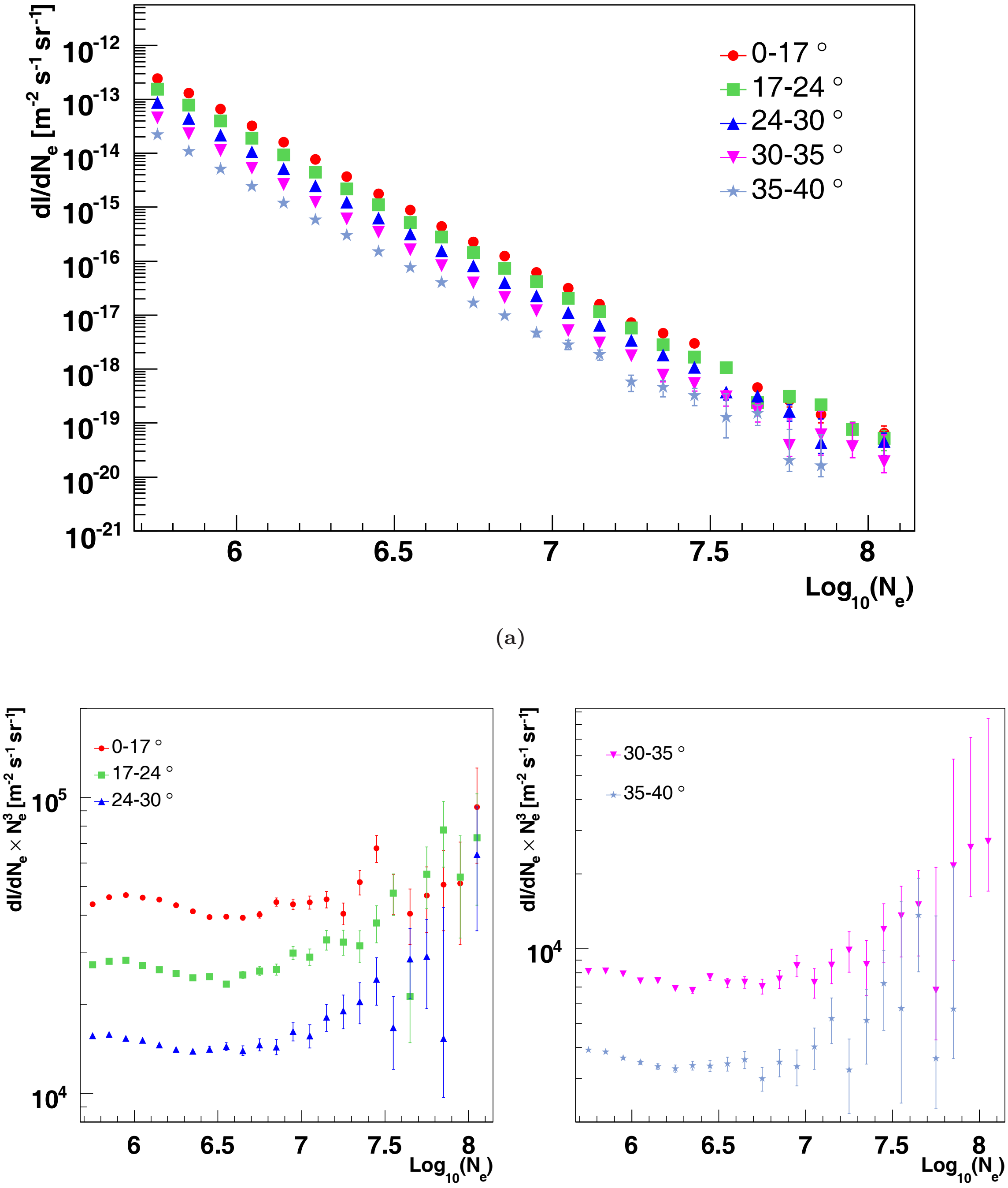
The differential flux as a function of the shower size is calculated as the number of measured events $N$ per unit area, unit time, steradian and $N_e$:

$$\frac{\Delta I}{\Delta N_e} = \frac{\Delta N}{\Delta \Omega d \Delta A \Delta t \Delta N_e}\,[\mathrm{m^{-2}s^{-1}sr^{-1}}]. \tag{5.2}$$

The shower size spectrum for the five zenith angle ranges is shown in fig. 5.1a, for a period of 987 days and a fiducial area of $\approx 0.2\,\mathrm{km}^2$. A total number of 877,233 events above the threshold ($\log_{10}(N_e) > 5.8$) have been collected. There are few events detected clearly above $10^{18}$ eV, i.e with $\log_{10}(N_e) > 8.1$. However a high number of saturated stations requires specific analysis which has not been implemented yet. Therefore, these events are rejected in this analysis. To enhance the visibility of possible spectral features the differential flux is multiplied by $N_e^{\gamma}$, with $\gamma = 3$. The corresponding spectra are depicted in fig. 5.1b and fig. 5.1c. The spectra show structures in all zenith angular bins. However, these structures are more pronounced for less inclined showers. An increase above $\log_{10}(N_e) \approx 7$ is evident in all the five zenith angle ranges, but starting at slightly different electron numbers. The corresponding muon number spectra, not corrected for the reconstruction uncertainties and multiplied by $N_\mu^3$, is shown in fig. 5.2. A change of slope in the muon size spectra is observed at $5 \cdot 10^5 < N_\mu < 8 \cdot 10^5$. This muon number range correspond roughly to the electron number range $3 \cdot 10^6 - 10^7$. The observed structures in the muon spectrum may be related to the uncertainties. Since the total electron number is calculated as $N_{ch} - N_\mu$ (see chapter 4), the uncertainties on $N_\mu$ are propagated on $N_e$.

As shown in fig. 5.3, the total electron number reconstruction uncertainty depends on the zenith angle. The bias, defined as $(N_e^{rec} - N_e^{true})/N_e^{true}$, evolves from 6% to −2% in the case of iron induced showers and an almost constant value is visible for proton primaries up to 30°. The statistical fluctuations are also zenith angle dependent and oscillate between 30% and 20% for both primaries.

In fig. 5.4 the dependence of the uncertainties as a function of the true electron number is shown again as in fig. 4.10a. For better understanding, the median values and the upper/lower errors are plotted separately (see chapter 4 for the error definition). There is a clear overestimation of the reconstructed shower size for $N_e^{true} > 10^7$ for both primaries. The statistical fluctuations, identified by the upper/lower errors, reflect the asymmetries in the accuracy distributions. Fluctuations towards higher $N_e$ of up to 40% are observed at $N_e^{true} \approx 10^6$ for proton induced showers, whereas being only 20% for the iron induced showers. Structures in the resolution of the reconstructed electron size may cause the



(a)

(b)

(c)

**Figure 5.1:** (a) Shower size spectrum for five zenith angle ranges, corresponding to equal solid angle, (b) and (c) shower size spectrum multiplied by $N_e^3$ in order to enhance visibility of possible spectral features.

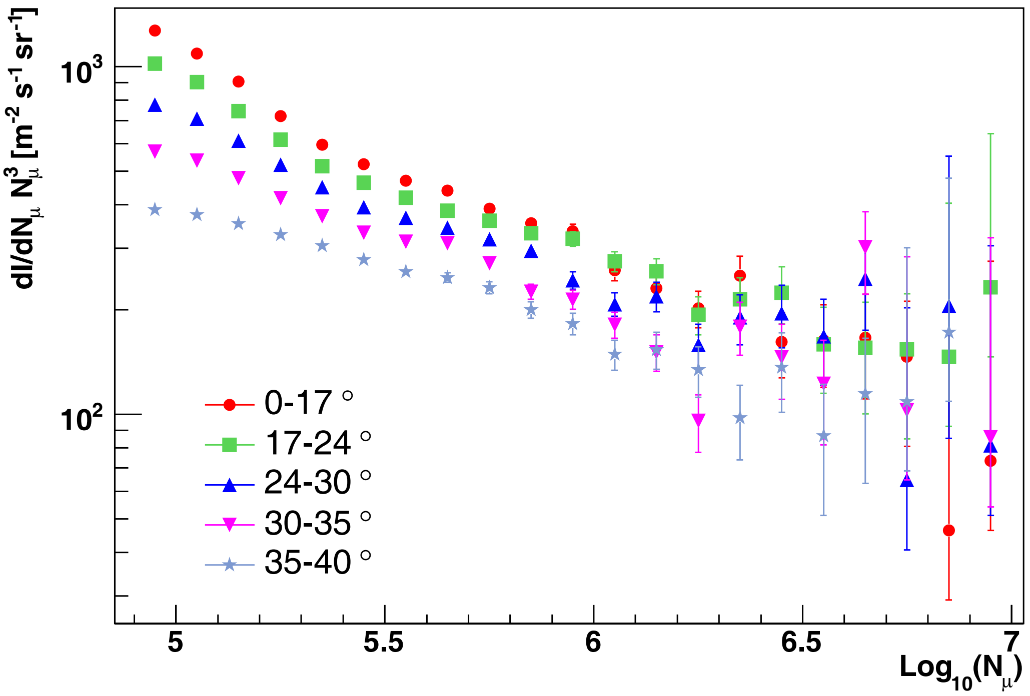


**Figure 5.2:** Muon size spectrum for five zenith angle ranges, corresponding to equal solid angle, multiplied by $N_\mu^3$ in order to enhance the visibility of possible spectral features.

feature observed in the measured $N_e$ spectrum. This has to be investigated before any physical conclusion can be drawn from the observed size spectra.

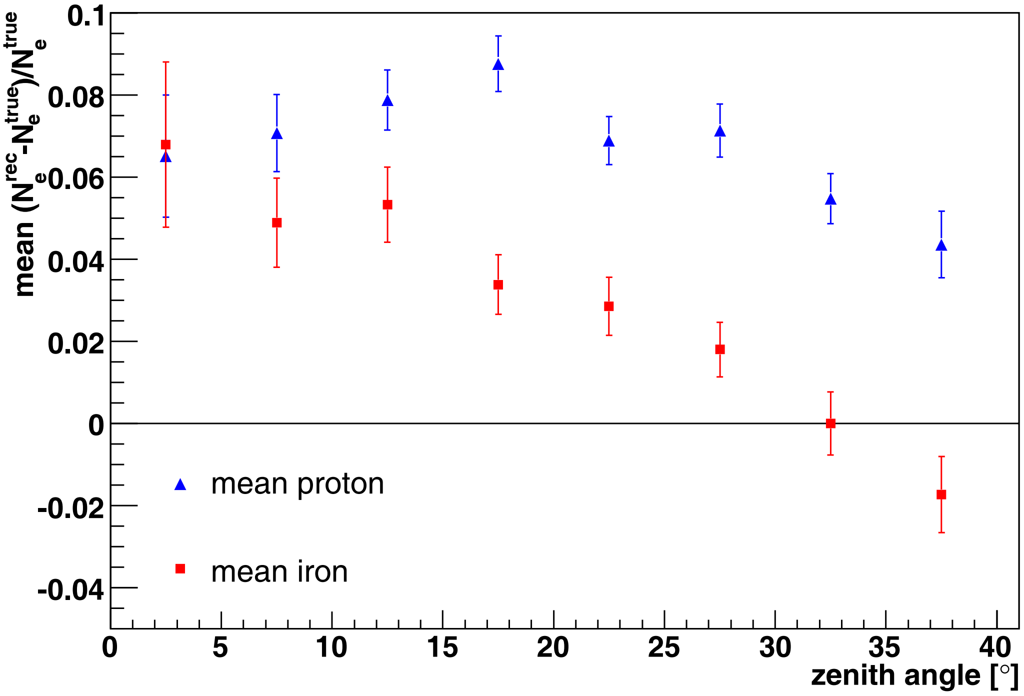
## 5.2 Forward folding

In order to understand whether the observed spectral features are physical properties or the results of the reconstruction uncertainties known from the simulation studies, a forward folding procedure is applied to disentangle the underlying true shower size spectrum from the reconstruction uncertainties.

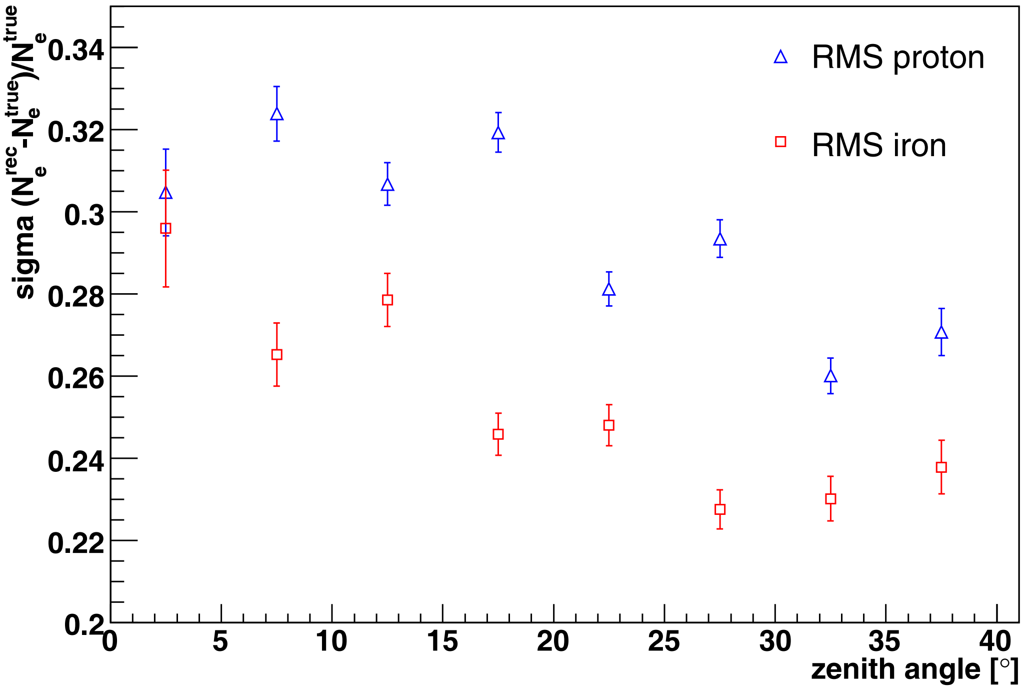
For performing this analysis, some assumptions on the true spectrum $I_{true}$ and the mass of the primary particles have to be made. Features in the total electron number spectrum can be produced by a change of the spectral index of the injected spectrum or by a change of composition in the energy range considered. In this work, a simple power law function is assumed, as described by equation 5.1 with $N_e^0 = 10^7$. The function is tested for a pure proton, a pure iron and a mixed composition in which proton, helium, carbon, silicon and iron are combined with equal proportions.

The measured flux can be described by

$$I_{meas} = \mathbf{T} \cdot I_{true}, \tag{5.3}$$



(a)



(b)

**Figure 5.3:** Bias (a) and statistical fluctuations (b) of the reconstructed total electron number as a function of the zenith angle. Showers simulated with QGSJetII-2 above full efficiency threshold are considered.

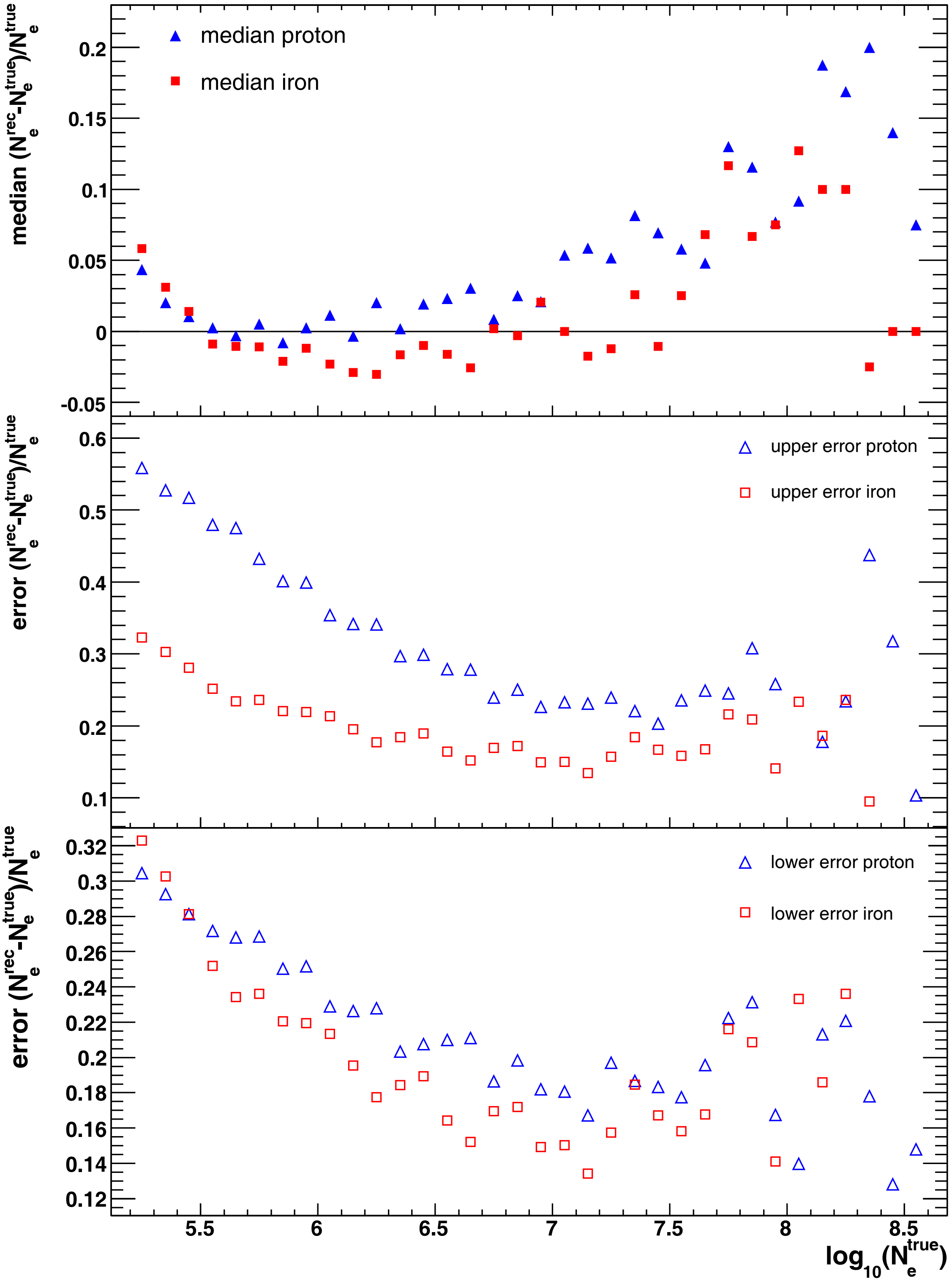


**Figure 5.4:** $N_e$ uncertainties as a function of $N_e^{true}$. *Top*: median of the $({N_e^{rec}}/{N_e^{true}} - 1)$ distribution as a function of $N_e^{true}$. *Centre*: upper limit of the spread of the underlying distribution, defined as its 0.84 quantile, as a function of $N_e^{true}$. *Bottom*: lower limit of the spread of the underlying distribution, defined as its 0.16 quantile, as a function of $N_e^{true}$. Showers simulated with QGSJetII-2 in the angular range 0° to 40°.

where **T** is the transfer matrix, whose elements are built from simulation studies as the probabilities to measure a shower with reconstructed size $N_e^{rec}$, given a true shower size $N_e^{true}$:

$$T_{ij} = P({N_e^{rec}}_i | {N_e^{true}}_j) = \frac{n_i(rec)}{n_j(true)}, \tag{5.4}$$

with $n_i(rec)$ being the number of events reconstructed at size $N_e^{rec}$ and $n_j(true)$ the number of events simulated at total electron number $N_e^{true}$. The transfer matrix is built by applying the same data selection as for the measured air showers, so trigger efficiencies are already accounted for. In fig. 5.5 the transfer matrix for each zenith angle range is shown as an example, for the pure proton and pure iron cases.

A $\chi^2$ fit to the measured data is performed to estimate the spectral index and the normalisation of the considered power law . The whole analysis is carried out separately for the five angle ranges. The fit is performed not on the differential $N_e$ spectrum but on the measured $N_e$ distributions and, as consequence, the absolute value of the spectral index $\gamma$ is one unit smaller than the value expected for the differential spectrum.

### CONEX simulations

The attenuation of air showers measured at different atmospheric depths might contribute, in addition to the reconstruction uncertainties, to a change of the slope of the spectra.

In order to check whether the spectral index of the electron spectrum stays constant for vertical and inclined showers, a set of simulations of proton and iron induced showers has been produced with CONEX[1] [54] code. A total of 110,000 air showers spread on an energy spectrum following a $E^{-3}$ power law have been produced, with $10^{16}$ eV $\leq$ E $\leq$ $10^{18}$ eV distributed over a zenith angle range from 0° to 40°. The value of the total electron number has been selected for each shower at five different slant depths, corresponding to the five zenith angle ranges used in the main analysis. The total electron number distributions in the case of primary protons and primary iron nuclei are shown in fig. 5.6a and fig. 5.6b, respectively. The response of a fit with a simple power law is reported in table 5.2 for both cases. The starting point of the individual fits is chosen in order to take into account threshold effects, whereas the end point is set to an intensity of minimum 10 events to avoid large poissonian fluctuations. For both, proton and iron induced showers, the fits give consistent results inside the uncertainties at all slant depths. Average spectral indices $\gamma = -1.77$ and $\gamma = -1.67$ are obtained for proton and iron, respectively. Therefore, structures in the shower size spectrum are not caused by attenuation of the electromagnetic particles in the atmosphere, but can hint to a change of composition.

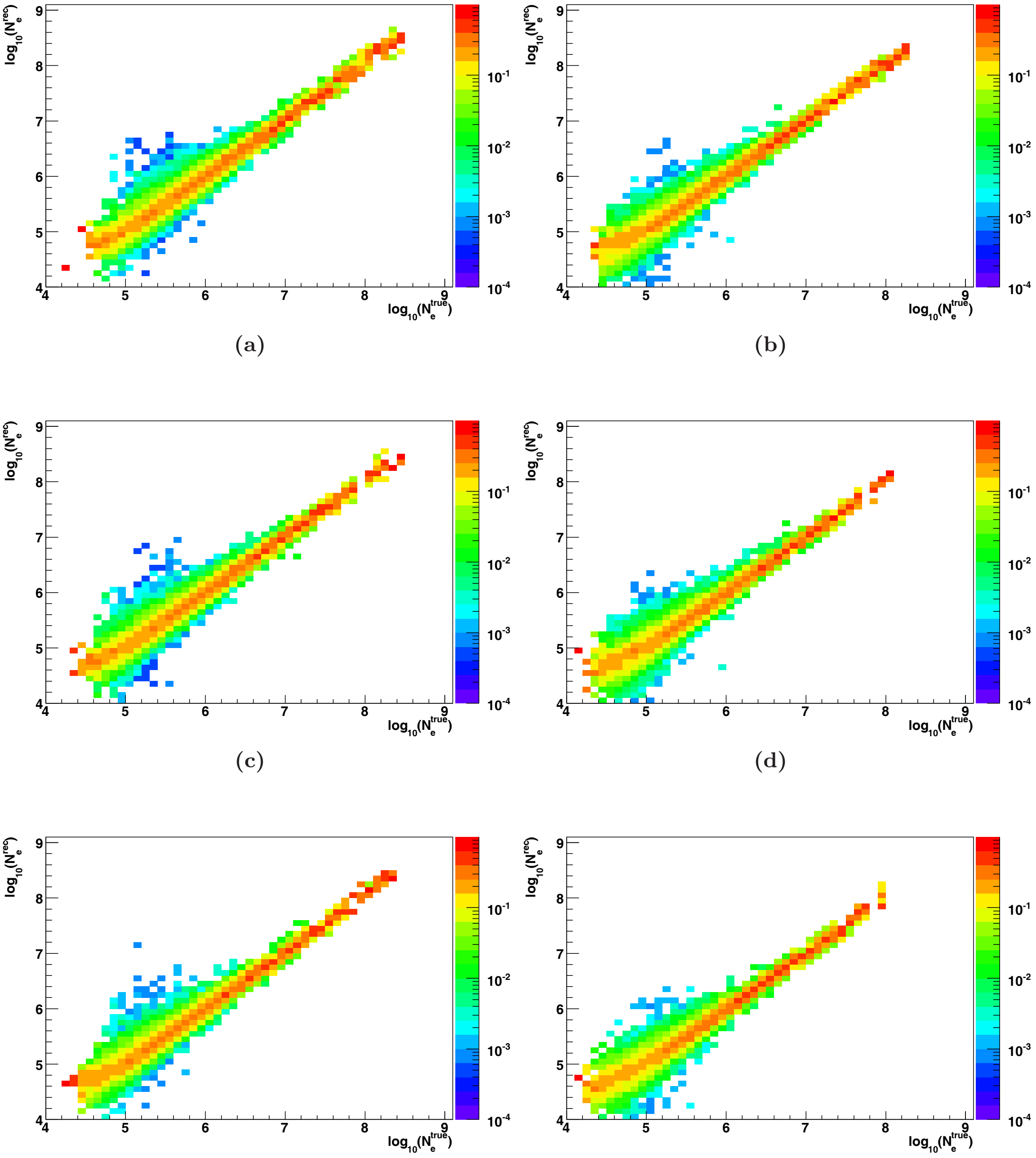
### Pure proton composition

The best fit parameters in the case of pure proton composition are reported in table 5.3. The true flux $I_{true}$ is folded with the uncertainties shown in fig. 5.5 and compared with the measured spectrum. For each angular range, the comparison of the measured size spectrum (black) with the expectation from the folding analysis (blue) is shown in fig.5.7.

From the reduced $\chi^2$ values, it appears clear that a pure proton composition cannot describe the measured data under the assumption of a simple power law. Moreover it is

---

[1] CONEX is a fast and efficient one-dimensional hybrid simulation scheme for ultra-high energy air showers, which combines an explicit MC simulation of particles cascade at energies above some chosen $E_{thr}$ and a solution of nuclear-electro-magnetic cascade equations for sub-cascades of smaller energies.

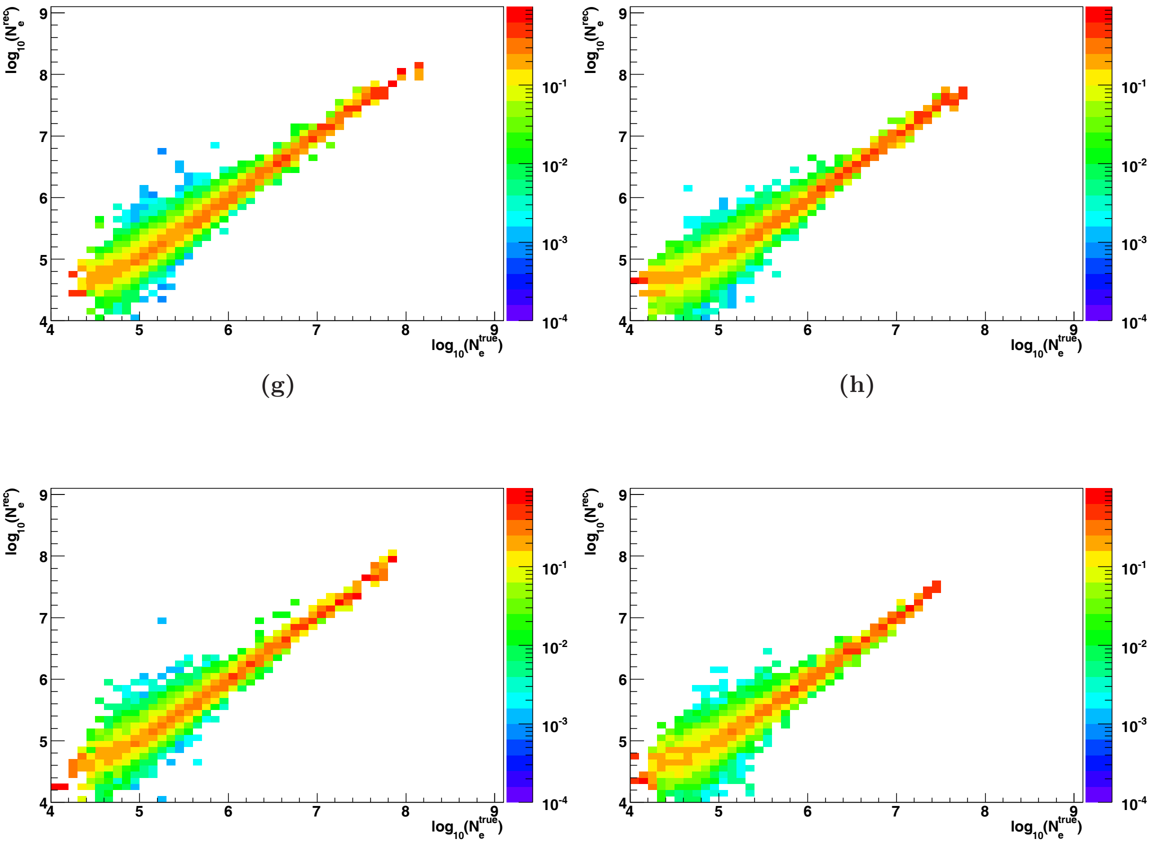


(a) (b)

(c) (d)

(e) (f)

**Figure 5.5:** Transfer matrices for proton (left column) and iron (right column) induced showers, for three zenith angle bins: (a) and (b) 0° – 17.6°, (c) and (d) 17.6° – 24.99°,(e) and (f) 24.99° – 30.73°.



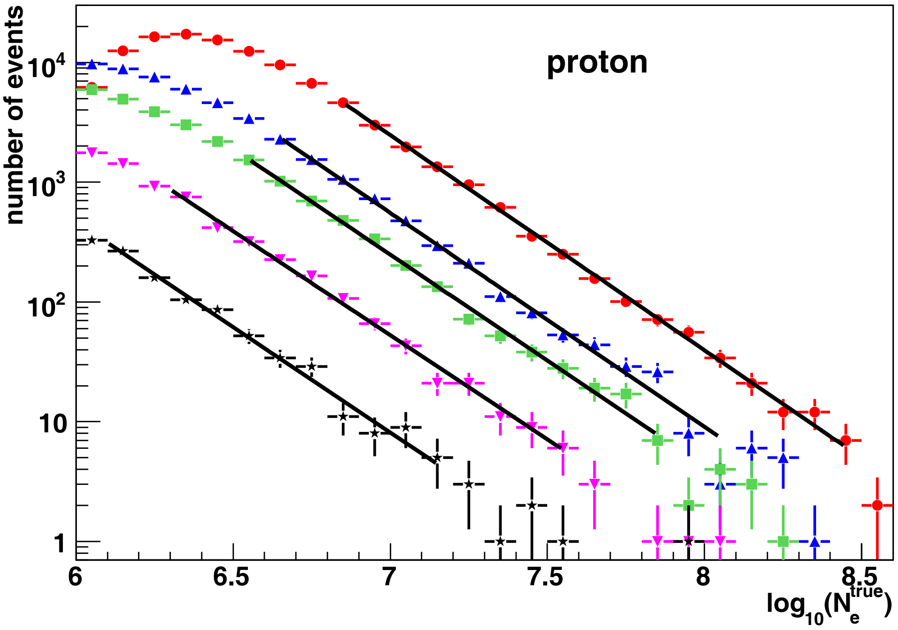
**(g)** **(h)**

**(i)** **(j)**

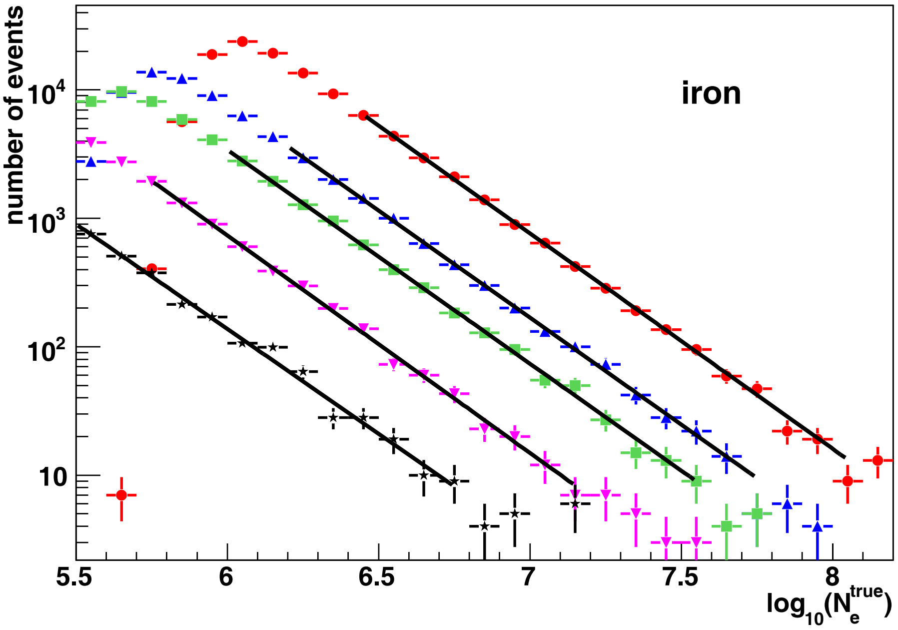
**Figure 5.5:** (continued) Transfer matrices for proton (left column) and iron (right column) induced showers, for two zenith angle bins:(g) and (h) 30.73° – 35.63°, (i) and (j) 35.63° – 40°.

| | Proton | | Iron | |
|---|---|---|---|---|
| Depth (g·cm$^{-2}$) | $\gamma$ | $\chi2/d.o.f.$ | $\gamma$ | $\chi2/d.o.f.$ |
| 1020 | -1.80 ± 0.02 | 1.0 | -1.68 ± 0.01 | 0.9 |
| 1130 | -1.79 ± 0.02 | 1.8 | -1.67 ± 0.02 | 0.8 |
| 1180 | -1.77 ± 0.03 | 1.0 | -1.66 ± 0.02 | 0.8 |
| 1260 | -1.74 ± 0.04 | 1.6 | -1.69 ± 0.02 | 0.6 |
| 1320 | -1.76 ± 0.07 | 0.8 | -1.63 ± 0.04 | 1.4 |

**Table 5.2:** Best fit spectral indices of simulated $N_e$ spectra at five different slant depths for proton and iron induced showers with the respective reduced $\chi^2$.



(a)



(b)

**Figure 5.6:** Shower size distributions for proton (a) and iron (b) induced showers at five different observation levels: 1020 g·cm$^{-2}$ (red bullets), 1130 g·cm$^{-2}$ (blue triangles), 1180 g·cm$^{-2}$ (green squares), 1260 g·cm$^{-2}$ (pink triangles) and 1320 g·cm$^{-2}$ (black stars). The continuous lines represent the best fits of the distributions. The showers are simulated with CONEX, with QGSJetII as high energy hadronic interaction model.

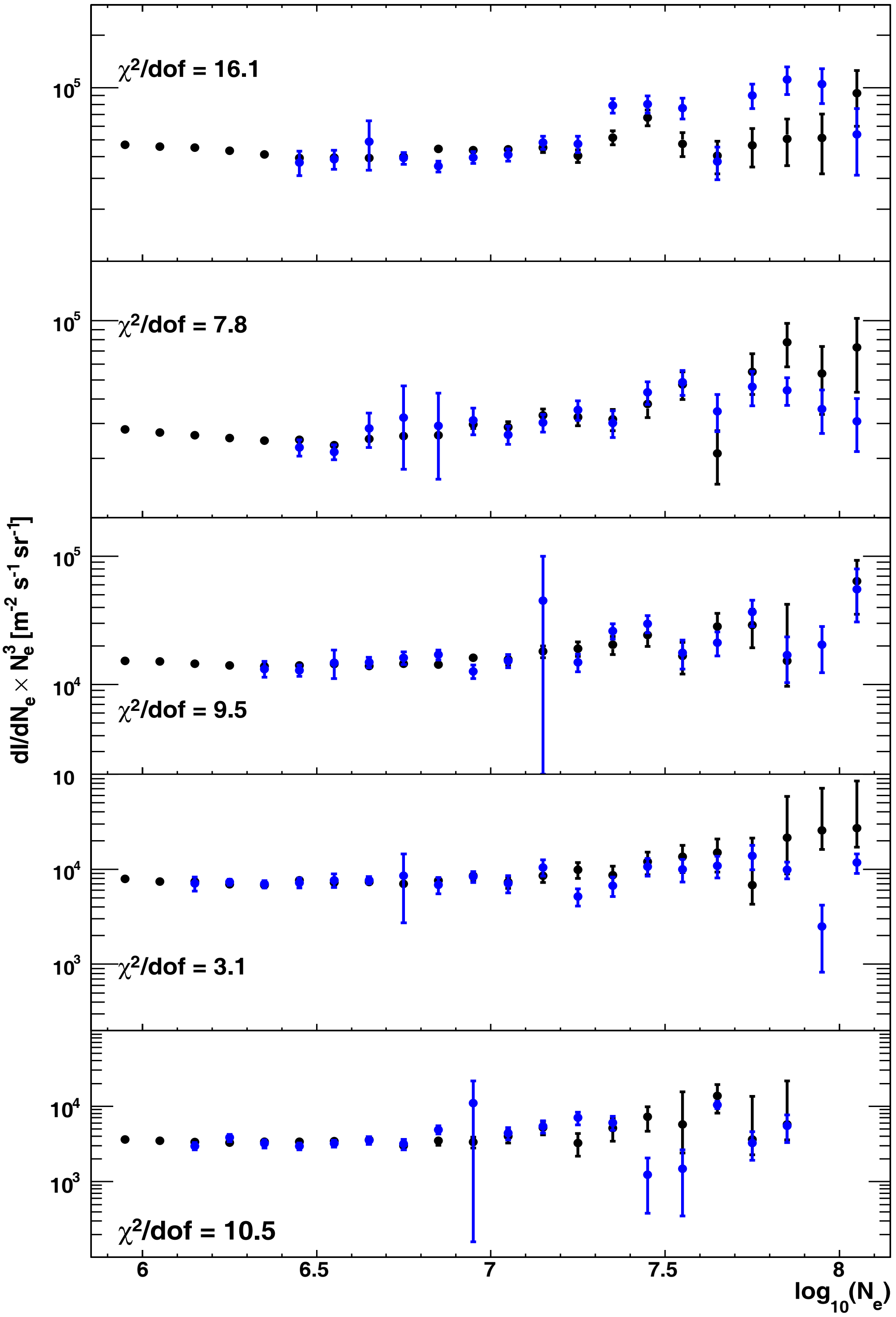


**Figure 5.7:** Comparison of the measured spectrum (black) and the forward folded one (blue), assuming a pure proton composition. The errors of the blue points are obtained by the propagation of the error of the best fit parameters of the true spectrum and the statistical error of the transfer matrix. The comparisons for the five zenith angle ranges used in this work (see table 5.1) are shown from top to bottom.

| zenith bin | A | $\gamma$ | $\chi^2/d.o.f.$ |
|---|---|---|---|
| 1 | 333 ± 6 | -1.63 ± 0.01 | 16.1 |
| 2 | 203 ± 5 | -1.69 ± 0.02 | 7.8 |
| 3 | 115 ± 3 | -1.70 ± 0.02 | 9.5 |
| 4 | 47 ± 2 | -1.88 ± 0.02 | 3.1 |
| 5 | 26 ± 1 | -1.80 ± 0.02 | 10.5 |

**Table 5.3:** Forward folding assuming pure proton composition: the best value for the normalisation factor A and the spectral index $\gamma$ are given for each zenith angle range.

not possible to reproduce the behaviour of the measured spectrum just above threshold and the tail of the distribution at the same time, especially for the vertical showers with $0° \leq \theta < 25°$. As can be seen in fig. 5.5, there are showers simulated at $N_e^{true} \approx 10^5$ which fluctuate above $N_e^{true} \approx 10^6$. Such fluctuations, for which the probability is around 0.1%, can heavily bias the procedure when the fit range is lowered from $\log_{10}(N_e^{rec}) = 6.4$ to $\log_{10}(N_e^{rec}) = 6$. These fluctuations become less important for inclined events, allowing an extension of the fit range to $\log_{10}(N_e^{rec}) = 6.1$. An analysis of such "outliers" shows that these fluctuating events are perfectly reconstructed, they have passed all the cuts. An example is given in fig. 5.8: the shower has been simulated with an electron number of $1.5 \cdot 10^5$ and reconstructed at $N_e^{rec} = 1.4 \cdot 10^7$, with an inclination of 29.8°. The lateral distribution fit for this event does not show any evident bias. If Monte Carlo fluctuations do not correspond to the real shower fluctuations, the forward folding will not perform correctly.

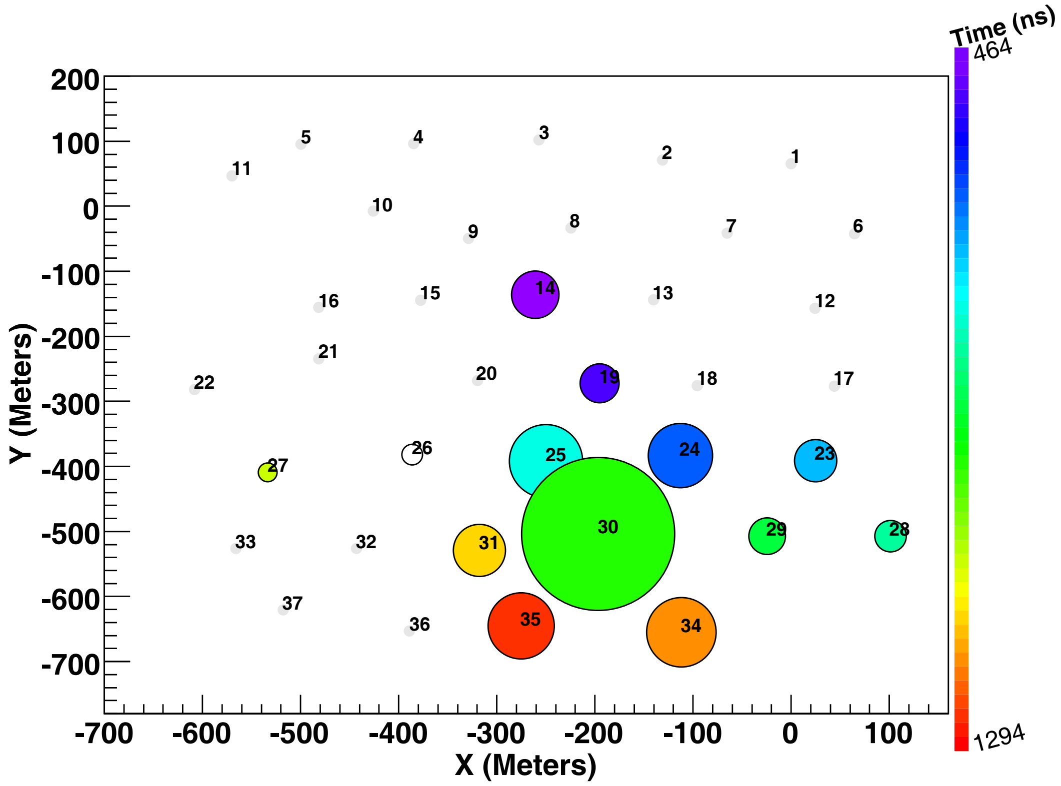
### Pure iron composition

In the case of a pure iron composition, the best fit parameters are given in table 5.4. In fig.5.9 the comparison of the measured size spectrum (black) with the folded spectrum (red) is shown for the pure iron composition.

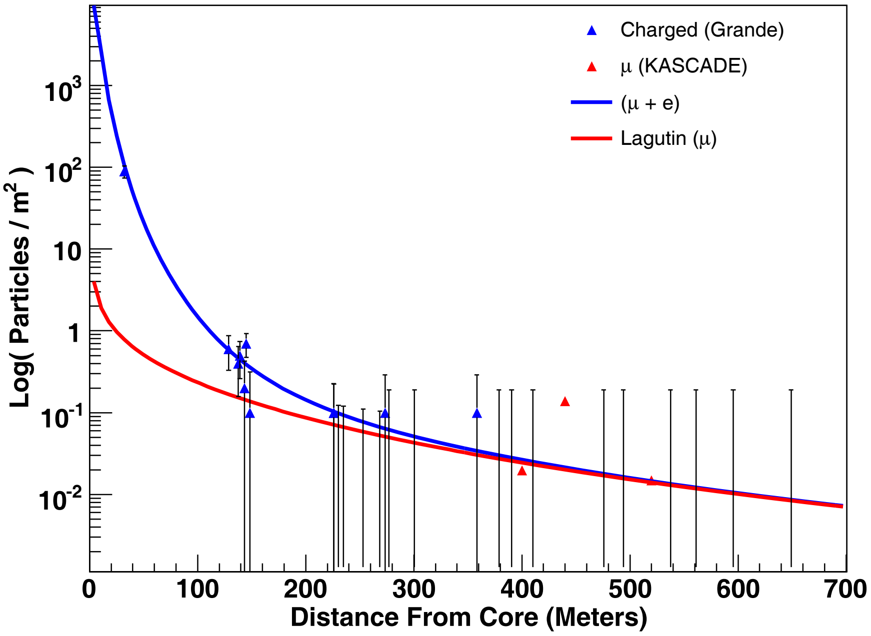
| zenith bin | A | $\gamma$ | $\chi^2/d.o.f.$ |
|---|---|---|---|
| 1 | 336 ± 6 | -1.74 ± 0.01 | 9.3 |
| 2 | 235 ± 5 | -1.73 ± 0.02 | 5.6 |
| 3 | 123 ± 3 | -1.82 ± 0.02 | 4.7 |
| 4 | 89 ± 2 | -1.56 ± 0.01 | 23.6 |
| 5 | 36 ± 1 | -1.66 ± 0.02 | 11.2 |

**Table 5.4:** Forward folding assuming pure iron composition: best fit parameters for each zenith angle range.

The $\chi^2/d.o.f.$ values of the fits indicate, also in this case, a bad description of the measured data of iron induced showers. It must be noticed that the number of showers used to build the transfer matrices (after the cuts) in the iron case is 20% less than in the proton case for vertical showers, and 35% less for the most inclined ones. For events reconstructed at an angle $\theta > 35°$, no information for an electron shower size higher than



(a)



(b)

**Figure 5.8:** (a) Footprint of a simulated shower on ground. (b) Lateral distribution of the shower: muon densities (red triangles) and charged particle densities (blue triangles) are depicted. The best fit (red line) of the muon lateral distribution is shown. The sum (blue line) of the NKG and Lagutin functions as fit of the charged particle lateral distribution is also shown. The shower has been reconstructed with $N_e = 1.4 \cdot 10^7$, whereas the true electron number is $1.5 \cdot 10^5$.

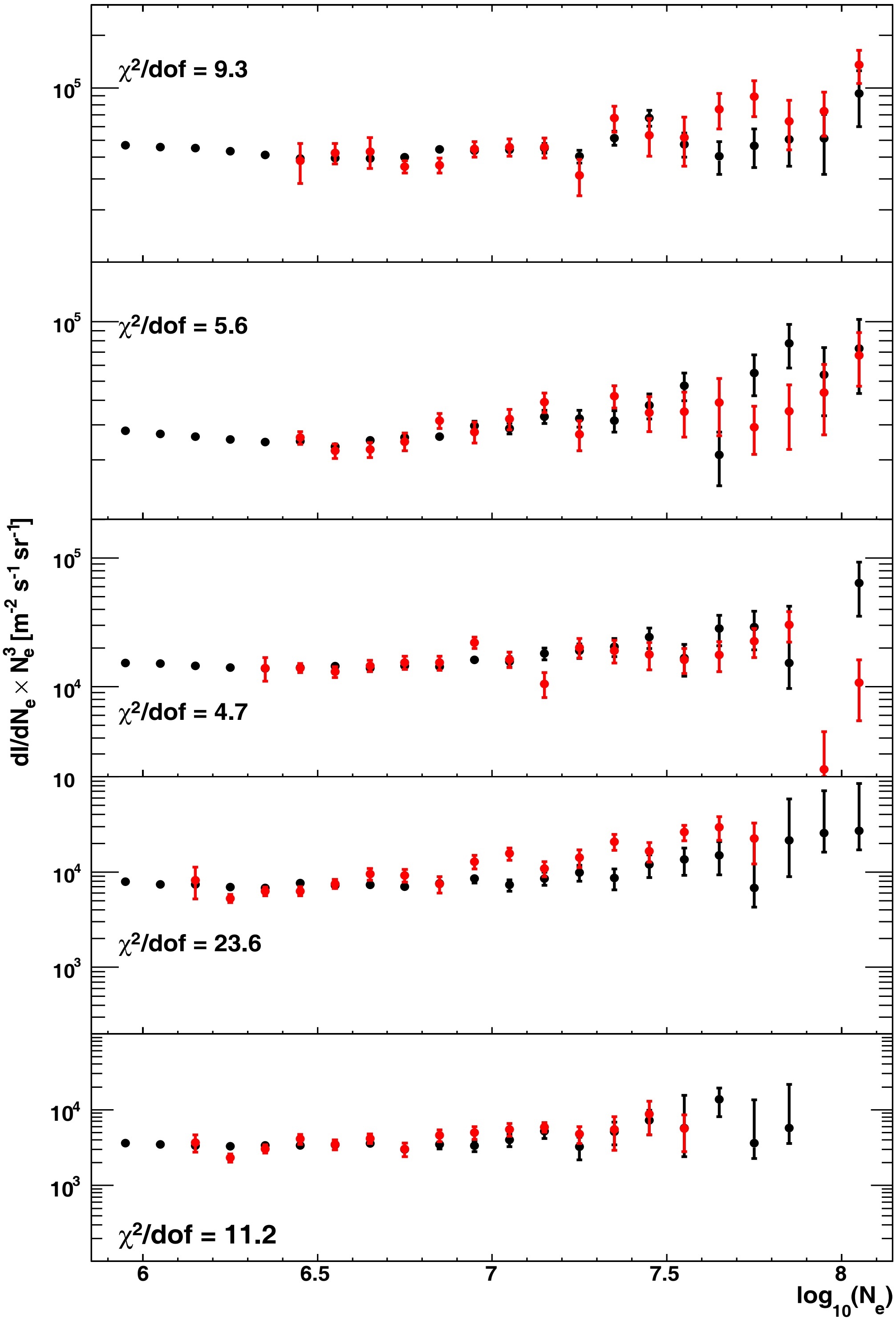


**Figure 5.9:** Comparison of the measured spectrum (black) and the forward folded one (red) assuming a pure iron composition. The errors of the red points are obtained by the propagation of the error of the best fit parameters of the true spectrum and the statistical error of the transfer matrix. From top to bottom, the comparisons for the five zenith angle ranges used in this work are shown (see table 5.1)

$3.1 \cdot 10^7$ is available. Therefore, the forward folding analysis has to be limited in the region covered by the simulations.

### Mixed composition

As a further possibility, a mixed composition is used, where five primaries are taken into consideration with equal proportions. The best fit parameters are listed in table 5.5. In fig.5.10, the comparison of the measured size spectrum (black) with the folded spectrum (green) is shown.

| zenith bin | A | $\gamma$ | $\chi^2/dof$ |
|---|---|---|---|
| 1 | 327 ± 6 | -1.75 ± 0.01 | 3.3 |
| 2 | 201 ± 5 | -1.76 ± 0.02 | 4.2 |
| 3 | 108 ± 4 | -1.81 ± 0.02 | 2.2 |
| 4 | 57 ± 2 | -1.78 ± 0.01 | 1.2 |
| 5 | 25 ± 1 | -1.81 ± 0.02 | 1.8 |

**Table 5.5:** Forward folding assuming a mixed composition: best fit parameters for each zenith angle range.

This seems to be the best description of the measured spectrum. The reduced $\chi^2$ obtained from the fit is smaller than the corresponding values obtained for proton and iron in each zenith angle range. Also, it has to be pointed out, that for the mixed composition case the number of available simulated showers is much larger than for each single component. For vertical showers up to a zenith angle of 20°, the number of showers available for the five primaries is 99,640, a factor of 5 higher than in the single component analysis. The procedure is sensitive to the statistics used. With a larger number of simulated showers, the effect of showers fluctuating to higher electron numbers is reduced. A test performed with a mixed composition, but only one fifth of the total statistics, leads to different spectral indices and higher $\chi^2$ values. The effect of the primary mass on the performance of the forward folding cannot clearly be evaluated, since the effect of statistics seems to be more important.

For this last test case, the true shower size spectra obtained from the forward folding analysis are shown in fig. 5.11. These spectra can be compared with fig. 5.1; it is obvious that a large amount of the structures seen in fig. 5.1 come from the reconstruction bias and fluctuations. The shape of the corrected spectra is now more similar in the different angular ranges.

The relative difference between the true intensity and the measured one is calculated. The results are shown in fig. 5.12. The error band shows the propagated uncertainties on the best fitting parameters, the error bars represent the statistical fluctuations of the measurements. For each angular range, the relative deviation is of $-40\%$ at $N_e \approx 10^6$, whereas it is less than $-20\%$ at $N_e \approx 10^7$. Above an electron number of a few times $10^7$, large fluctuations are evident. They reflect the limitation of the method to reproduce the tail of the spectra on the base of the estimated reconstruction uncertainties and of the available statistics.

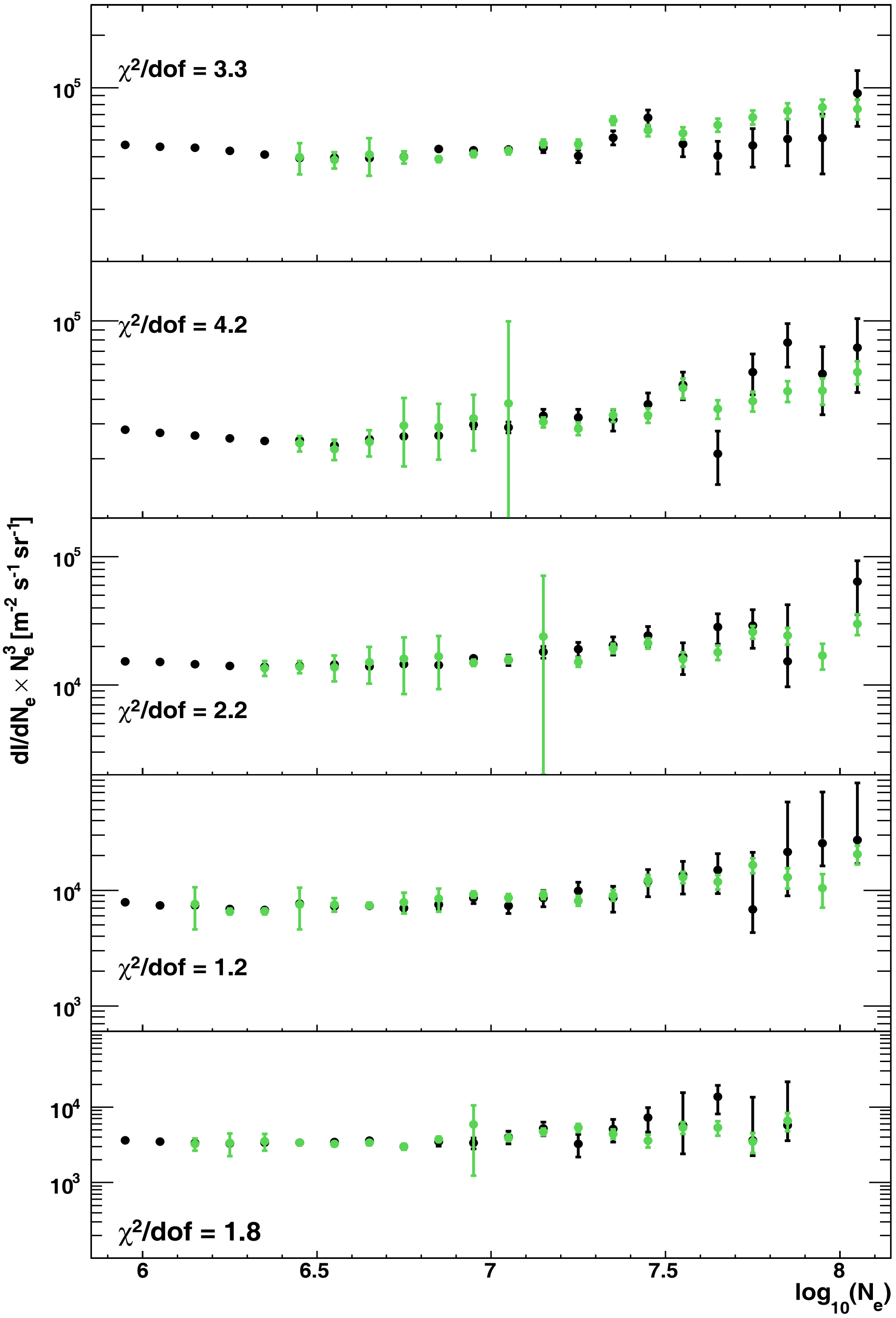


**Figure 5.10:** Comparison of the measured spectrum (black) and the forward folded one (green) assuming a mixed composition. The errors of the green points are obtained through propagation of the error of the best fit parameters of the true spectrum and the statistical error of the transfer matrix. From top to bottom, the comparisons for the five zenith angle ranges used in this work are shown (see table 5.1)

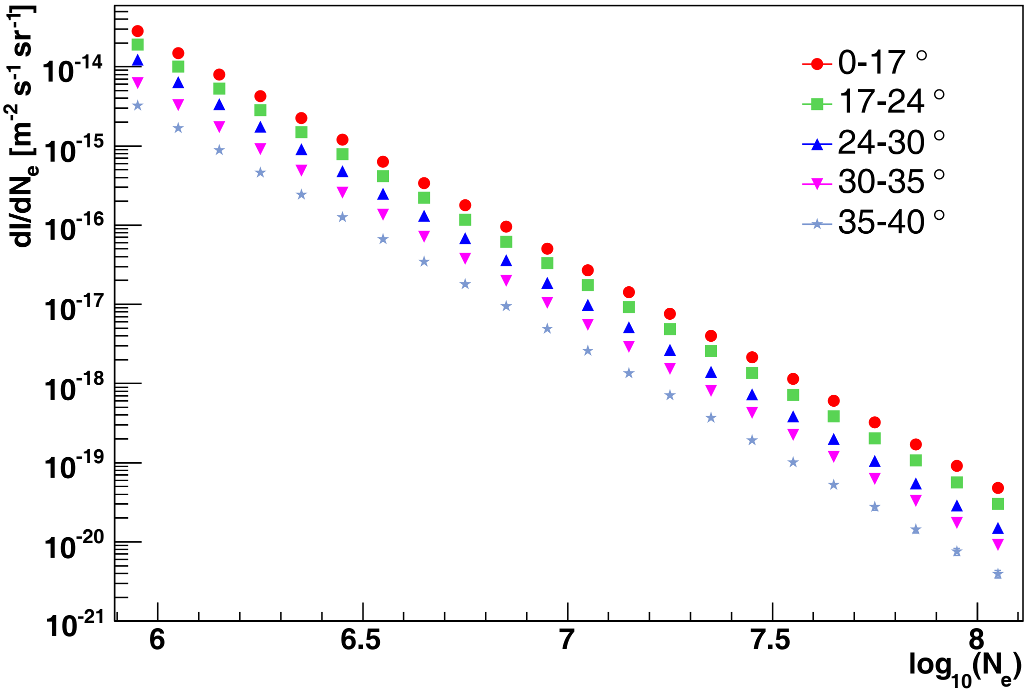


**Figure 5.11:** The true shower size spectra for five zenith angle ranges, as obtained from the forward folding procedure, assuming a mixed composition.

## 5.3 Conclusion

The applied technique of the forward folding, for all the cases, can only work if the simulation really describes the nature. The mismatch for vertical showers can be explained either by the wrong assumption of $I_{true}$, or by the incapability of the simulation to correctly represent the shower to shower fluctuations and the attenuation of the shower particles in the atmosphere. In general, it has been observed that large fluctuations at low energies, i.e. at small electron numbers, can heavily bias the method.

The applied procedure relies completely on the accuracy of the simulations used as reference, on the available statistics and on the reliability of the assumed true spectrum. An attempt to use a broken power law to describe $I_{true}$ has been carried out, without any improvement on the overall result.

The observed structures in the measured $N_e$ spectrum are compatible with the propagation of the reconstruction uncertainties along a $N_e^\gamma$ true spectrum. A spectral index $\gamma = -2.78$, averaged over the five angular ranges, has been obtained for the unfolded spectra. Given:

$$\begin{array}{rcl} N_e & \propto & E^\alpha, \\ \dfrac{dN}{dN_e} & \propto & N_e^\gamma, \\ \dfrac{dN}{dE} & = & \dfrac{dN}{dN_e} \cdot \dfrac{dN_e}{dE}, \end{array}$$

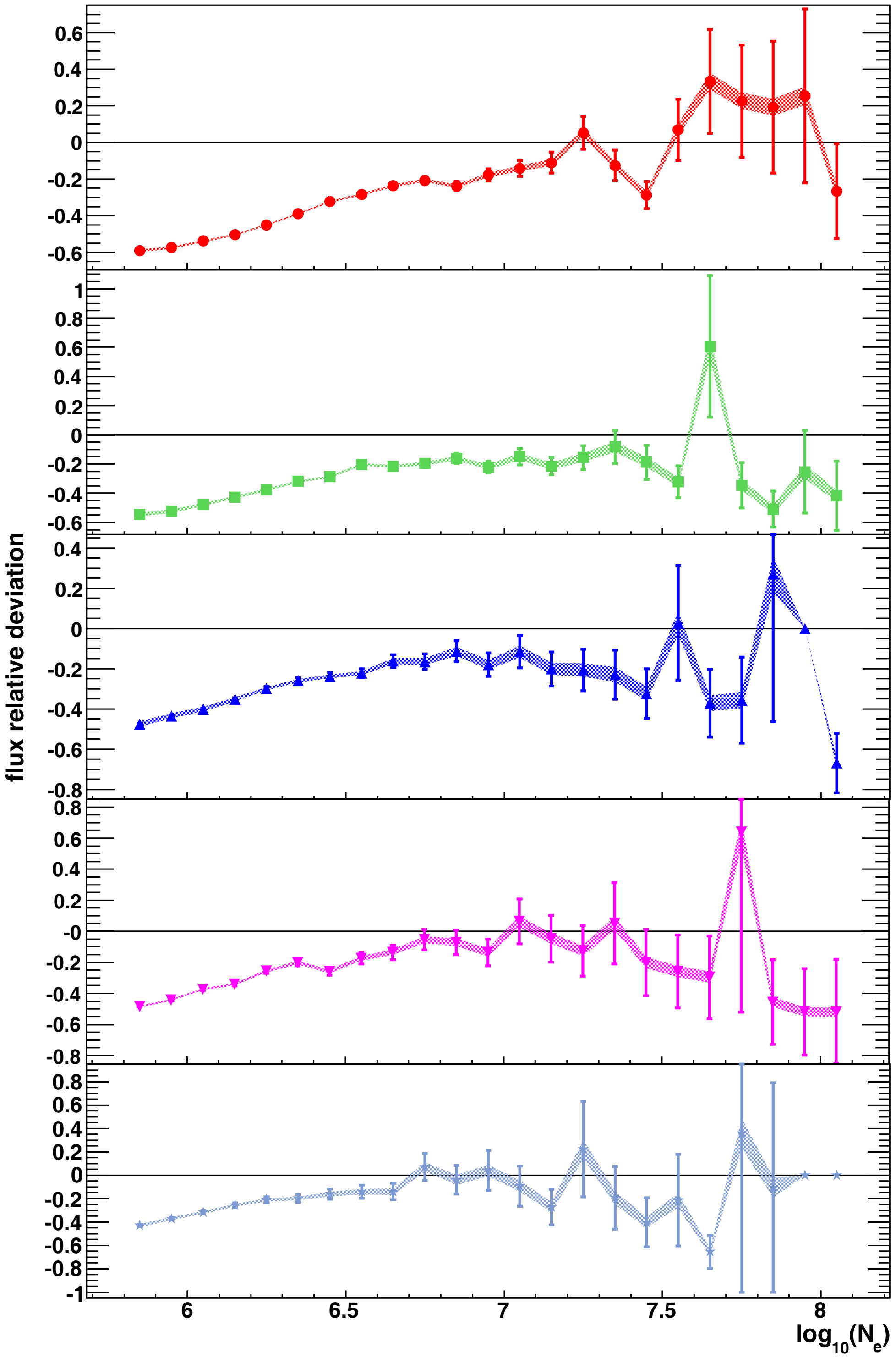


**Figure 5.12:** Relative difference between the measured shower size spectrum and the true $N_e$ spectrum, as calculated from the unfolding analysis under the assumption of a simple power law. The error band shows the propagated uncertainties on the best fitting parameters, the error bars represent the statistical fluctuations of the measurements. The flux deviations for the five zenith angle ranges are shown from top to bottom.

being $N$ the number of observed events and being $\alpha$ found in simulation to be slightly above one, the spectral index of the energy spectrum results to be around $-3.1$.

On the present analysis basis, the spectral features cannot be attributed to a change in the mass composition of the cosmic rays. This effect is of a higher order compared to the effect of the statistics used in the folding procedure.

The mean relative difference between the true spectrum and the measured spectrum, as a function of the reconstructed electron number, has been obtained in each angular range, under these assumptions.

# 6. Composition study

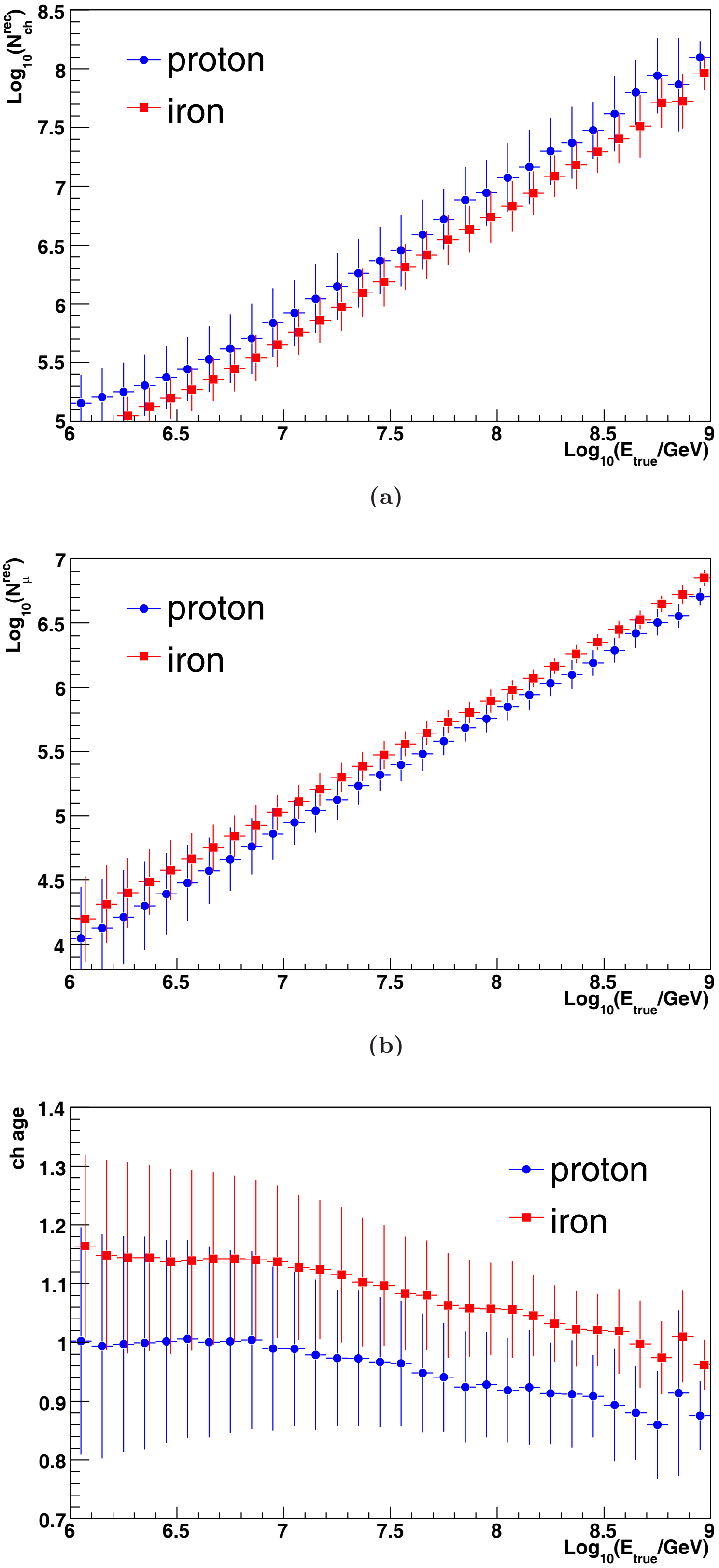
The KASCADE experiment has shown that the knee in the all particle energy spectrum is caused by a change of the spectral index in the spectrum of the light elements. According to the most accredited theories of acceleration and propagation of cosmic rays, a break in the energy spectrum of the heaviest elements is expected at a primary energy of $\approx 10^{17}$ eV. However, depending on the astrophysical model (e.g. [3]), the iron knee is not necessarily visible in the all-particle energy spectrum. The puzzle of the existence of this "iron knee" in the cosmic ray energy spectrum can be solved only by separating the spectra of the single elemental components.

In this chapter, a technique to estimate the mass of the primary particles generating the observed air showers will be presented. Due to the limited statistics available at the highest energies, a procedure which performs well under this condition is developed. Measured EAS, with zenith angle below 40°, are classified as belonging to a certain primary mass group according to the reconstructed numbers of muons and charged particles. The relative number of proton-like and iron-like showers is obtained, showing the general tendency of the composition evolution with increasing energy.

## 6.1 Composition sensitive observables

Given the impossibility to measure directly the nature of the primary particle, it is compulsory to find measured or reconstructed observables sensitive to the mass of the primary ion. Taking into account the correlation of the observables, they can be combined in a multivariate analysis to estimate the mass with accuracy as good as possible.

From simulation studies, the total charged particle and muon numbers, the shower age and the curvature radius of the shower front are found to be composition sensitive parameters. In fig. 6.1, the mean values for simulated $N_{ch}$, $N_\mu$ and the shower age $s$ are shown, as a function of the true energy. The error bar associated with each point represent the spread of the underlying distribution. Although each observable has a certain discrimination power, due to shower to shower fluctuations it is not possible to estimate the mass of the primary particle from any of them alone. It is, thus, necessary to combine them in one unique procedure. For all observables to be considered, the agreement between simulated and measured events has to be checked. It is not advisable to carry on with the analysis if the simulations are not reliable.



(a)

(b)

(c)

**Figure 6.1:** Average value of $N_{ch}$ (a), $N_{\mu}$ (b) and the shower age $s$ (c) as a function of the simulated energy, for proton (blue dots) and iron (red squares) induced showers. The error bars represent the spread of the underlying distribution. A shift of +0.02 in $\log_{10}(E_{est}/GeV)$ is applied to the iron induced showers.

### Total muon number

In order to investigate the compatibility of simulated and measured reconstructed muon numbers, their distributions as functions of the shower size $N_e$ (parameter available for both simulated and observed air showers) are compared.

Since the distribution of $N_\mu$ per bin of $N_e$ is not symmetric and to prevent a bias from the distribution's tail, the median value is considered instead of the mean. The asymmetric spread of the distribution is described by the 0.16 and 0.84 quantiles, so that 68% of the events are included inside these limits.

In fig. 6.2, the median value of $N_\mu$ as a function of the total electron number is shown. A spectral index $\gamma = -3$ is assumed for simulated showers, as it is very close to the measured spectrum. As expected, the data points lie between the two extreme cases of pure proton (blue) and pure iron (red) compositions. Therefore, the correlation between the reconstructed muon number and the shower size is proven to be sensitive to the primary mass. Following this correlation alone and believing in the QGSJETII hadronic model, in the considered energy range the composition seems to start mixed and becoming heavier.

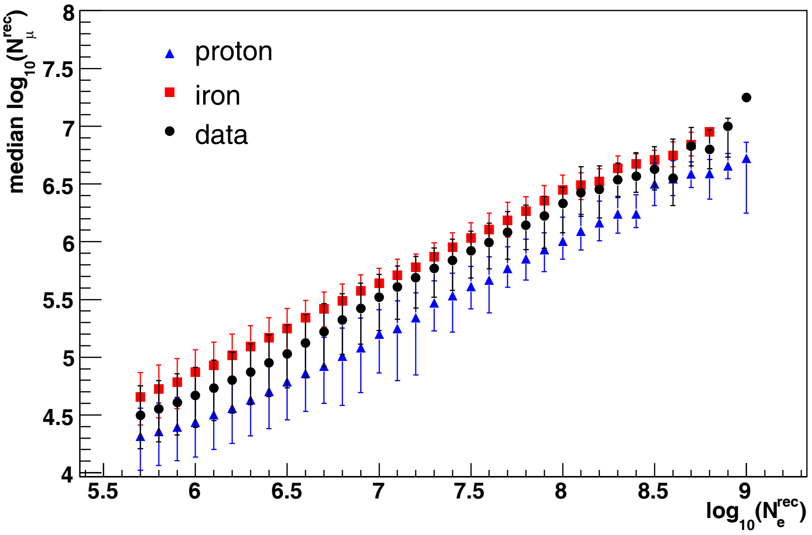


**Figure 6.2:** Median values of $N_\mu$ distributions in bins of the reconstructed total electron number, for proton (blue triangle) and iron (red square) induced showers. Black dots symbolise the observed events. The error bars represent the spread of the distribution, a spectral index $\gamma = -3$ is assumed for simulated showers.

### Total charged particle number

The total number of reconstructed charged particles of iron and proton induced showers is also compared with measured data. In fig. 6.3, the median value of $N_{ch}$ is shown as a function of $N_e$, showing a linear correlation of the two variables. If the composition drastically changes in the measured range, the correlation would also change, as iron and proton have different muon content.

### Shower age

In fig. 6.4, the shower age $s$ (reconstruction level 3) of simulated and measured EAS are compared. Above the experimental threshold of $N_e \approx 10^6$, the median age is constantly

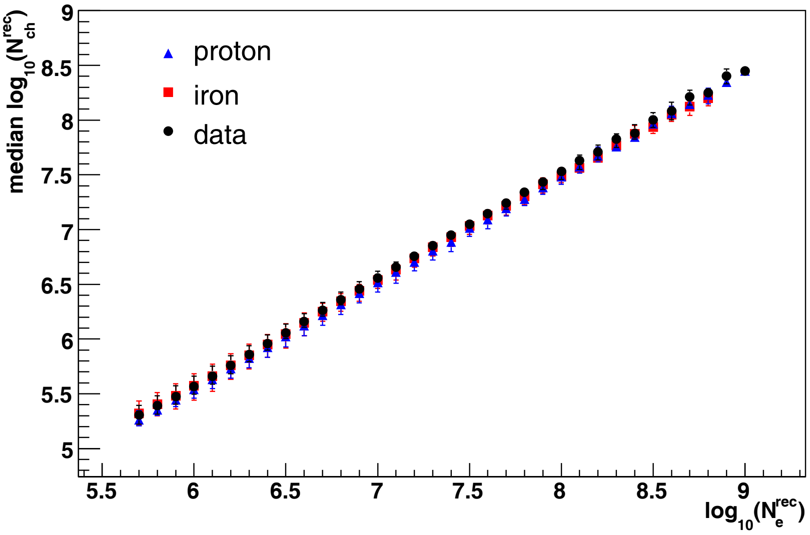


**Figure 6.3:** Median values of $N_{ch}$ distributions in bins of the reconstructed total electron number, for proton (blue triangle) and iron (red square) induced showers. Black dots symbolise the observed events. The error bars represent the spread of the distribution, a spectral index $\gamma = -3$ is assumed for simulated showers.

larger for the detected showers than for both, proton and iron induced showers. In this case, the simulations seem not to be reliable and consistent with the data. If $s$ was to be used in the analysis as mass parameter, the results would be constantly biased towards primary particles heavier than iron.

As mentioned in chapter 2, the meaning of the parameter $s$ is that of "age" of the shower. In chapter 4, it has been shown that the lateral distribution function used in this work is a slightly modified version of the original NKG formula. In this case, the parameter $s$ has lost its original meaning, but can be interpreted as a descriptor of the shape of the distribution. Smaller values of $s$ correspond to steeper distributions than larger values. On average, iron induced showers reach their maximum higher in the atmosphere than proton initiated showers for a given energy. Therefore, due to the Coulomb scattering of the particles (see chapter 2), an iron induced shower shows a flatter lateral distribution than a proton induced shower.

The observed discrepancy between simulated and measured EAS can be explained through a closer view of the high energy hadronic interaction model, employed in the Monte Carlo simulation. Grande measures charged particles. The distinction between muons and electrons is introduced at reconstruction level 3, by subtracting in each station the muon number expected from the Lagutin function, obtained from muon measurements at the KASCADE muon detectors. The parameters of the Lagutin function have been optimised in simulations. If the number of muons produced in the shower is underestimated by QGSJetII-2, then the expected fraction of electrons measured in each Grande station is locally overestimated. This overestimation is more important in the tail of the measured distribution, where the muon contribution is higher than the electron contribution. The effect will be a flatter shape of the electron lateral distribution, i.e. a larger value of the $s$

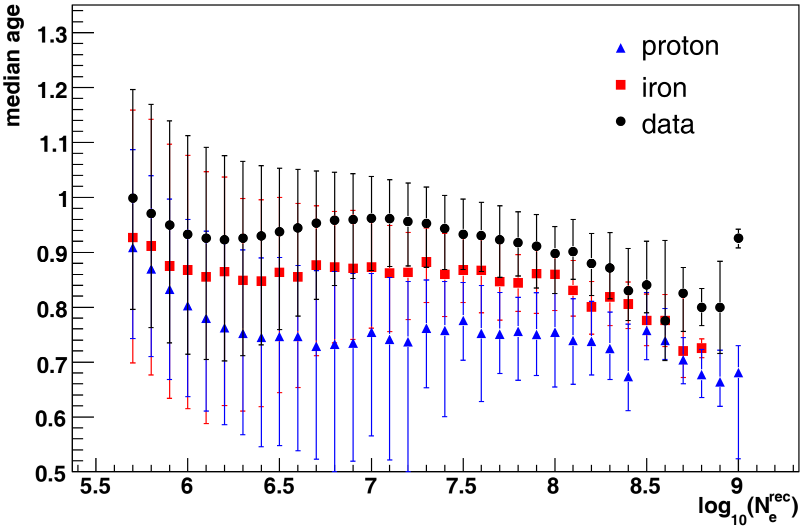


**Figure 6.4:** Median values of the shower age $s$ (reconstruction level 3) distribution, in bins of the reconstructed total electron number, for proton (blue triangle) and iron (red square) induced showers. Black dots symbolise the observed events . The error bars represent the spread of the distribution and not the error on the median. A spectral index $\gamma = -3$ is assumed for simulated showers

parameter. A cross-check of this interpretation comes from the analysis of the shower age at reconstruction level 2, where the charged particle distribution is directly fitted with the modified NKG function, with no muon subtraction. In fig. 6.5, the correlation between the shower age $s$ and the total number of charged particles $N_{ch}$ is shown. Compared to fig. 6.4, the simulations match the measured data better, since no muon information is required to fit the charged particle lateral profile. However, between $\log_{10}(N_{ch}) = 6.5$ and $\log_{10}(N_{ch}) = 7.5$ the data distributions is still not consistent with the simulated values. Therefore, it seems that the QGSJETII model predicts a too flat lateral distribution of the particles, which might be related to the extrapolation of the hadronic cross-section at high energies and/or to the simulated pion multiplicity generated at the hadronic interactions, as explained in the next section. To understand the origin of this discrepancy, a detailed study is needed, which has not been performed yet.

This comparison has demonstrated that the shower age is not a good parameter for composition analysis, hence it will not be considered anymore in the following composition analysis.

### Radius of curvature

Adopting a spherical description of the shower front, it is possible to calculate not only the shower direction, but also the radius ($R_c$) of the sphere used to fit the front. $R_c$ is referred to as *curvature radius.*

The shape of the shower front, as detected on ground, reflects the scatter of the secondary particles. An EAS started higher in the atmosphere will result in a flatter shower front, i.e. in a larger radius of curvature, as compared to a shower initiated deeper in the atmosphere. For the same primary energy, iron induced air showers will originate higher

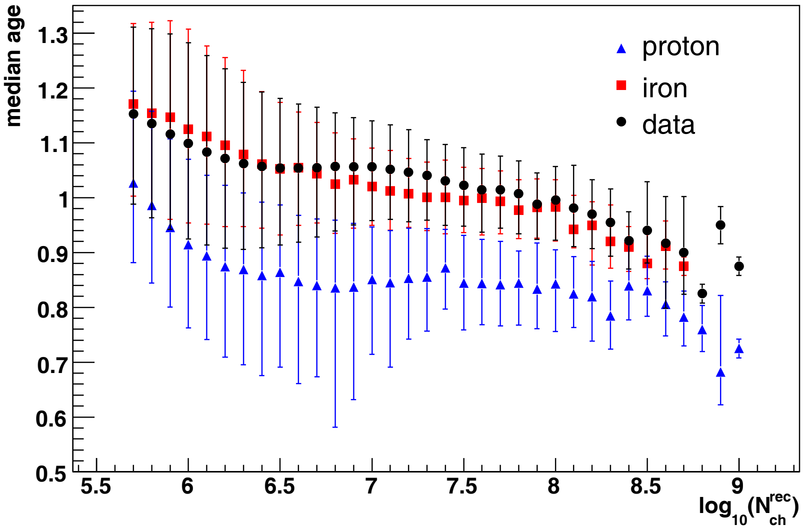


**Figure 6.5:** Median values of the shower age $s$ (reconstruction level 2) distribution in bins of the reconstructed total electron number for proton (blue triangle) and iron (red square) induced showers and for observed events (black bullet). The error bars represent the spread of the distribution.

in the atmosphere than those of proton primaries. Therefore, the curvature radii of iron showers are usually larger than for proton initiated showers of the same energy.

This procedure (see appendix B for details) has been applied to the set of well reconstructed events. The observed and simulated median values have been compared as a function of the reconstructed total electron number. The result is shown in fig. 6.6. As in the case of the shower age parameter, there is some disagreement between simulations and observed showers: the median values for real events are, along the whole range of $N_e$, smaller than for simulated showers. The indication given by the analysis of the curvature radii will lead to a composition lighter than protons alone, which contradicts with what has been found in the analysis of the shower age distribution. The origin of the mismatch has to be sought in the features of the hadronic interaction model QGSJetII-2.

In fig. 6.8, the extrapolated pion cross-section from the QGSJetII model is compared with the available direct measurements. At high energies, the model predicts a cross-section $\sigma_\pi$ that is higher than the last available measurements (SELEX collaboration [55]). Therefore, the whole shower develops fast, with muons and electrons which arrive at ground with less energy and more scattered. Also, high values of multiplicity contribute to a fast shower development. In fig. 6.7, the multiplicity as a function of the energy is shown for different hadronic interaction models. Up to an energy of about 1 TeV, a good agreement between the models is observed; at higher energies an increase can be seen, stronger for QGSJetII than for the other available models. The combined effect of an overestimation of $\sigma_\pi$ and a high multiplicity number can lead to the observed discrepancies between simulation and measured events: showers simulated with QGSJetII as hadronic interaction model develop faster and, consequently, they show a flatter shower front shape.

The analysis of all the candidates as mass composition sensitive observables, with the present knowledge and on the basis of QGSJetII, shows that only the total number of

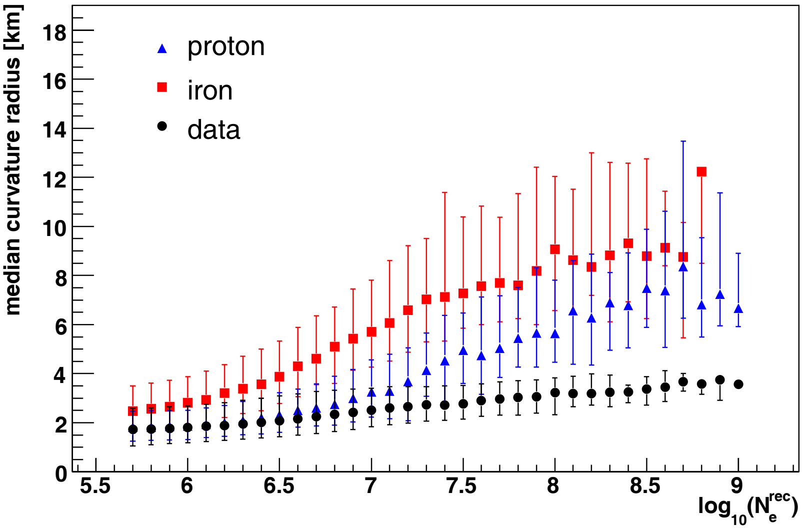


**Figure 6.6:** Median values of the radius of curvature $R_c$ distribution, in bins of reconstructed total electron number. Blue triangles and red squares represent proton and iron primaries induced showers respectively. Black dots symbolise observed events. The error bars represent the spread of the distribution.

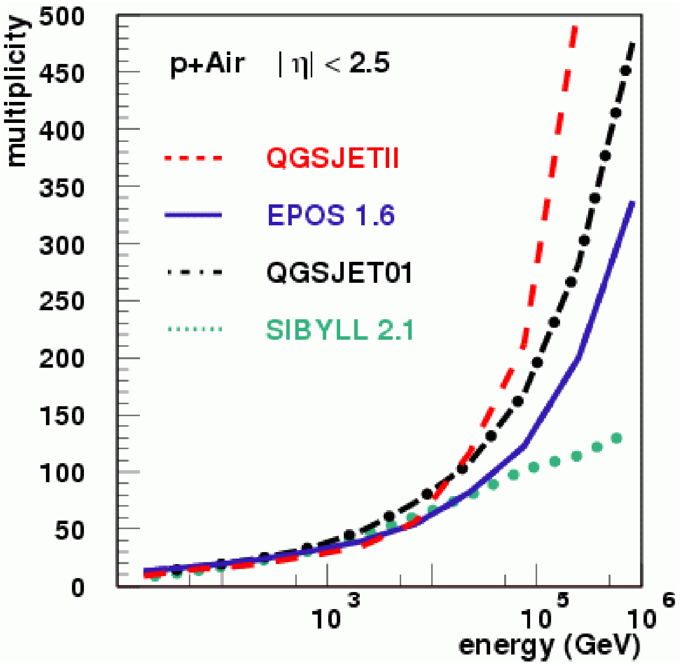


**Figure 6.7:** Multiplicity for p-Air interaction as a function of energy. The red dashed line was obtained with the QGSJetII model, while the blue solid line represents the EPOS 1.6 model. A tendency to overestimate the multiplicity is evident in the case of QGSJetII compared to EPOS, e.g. [56].

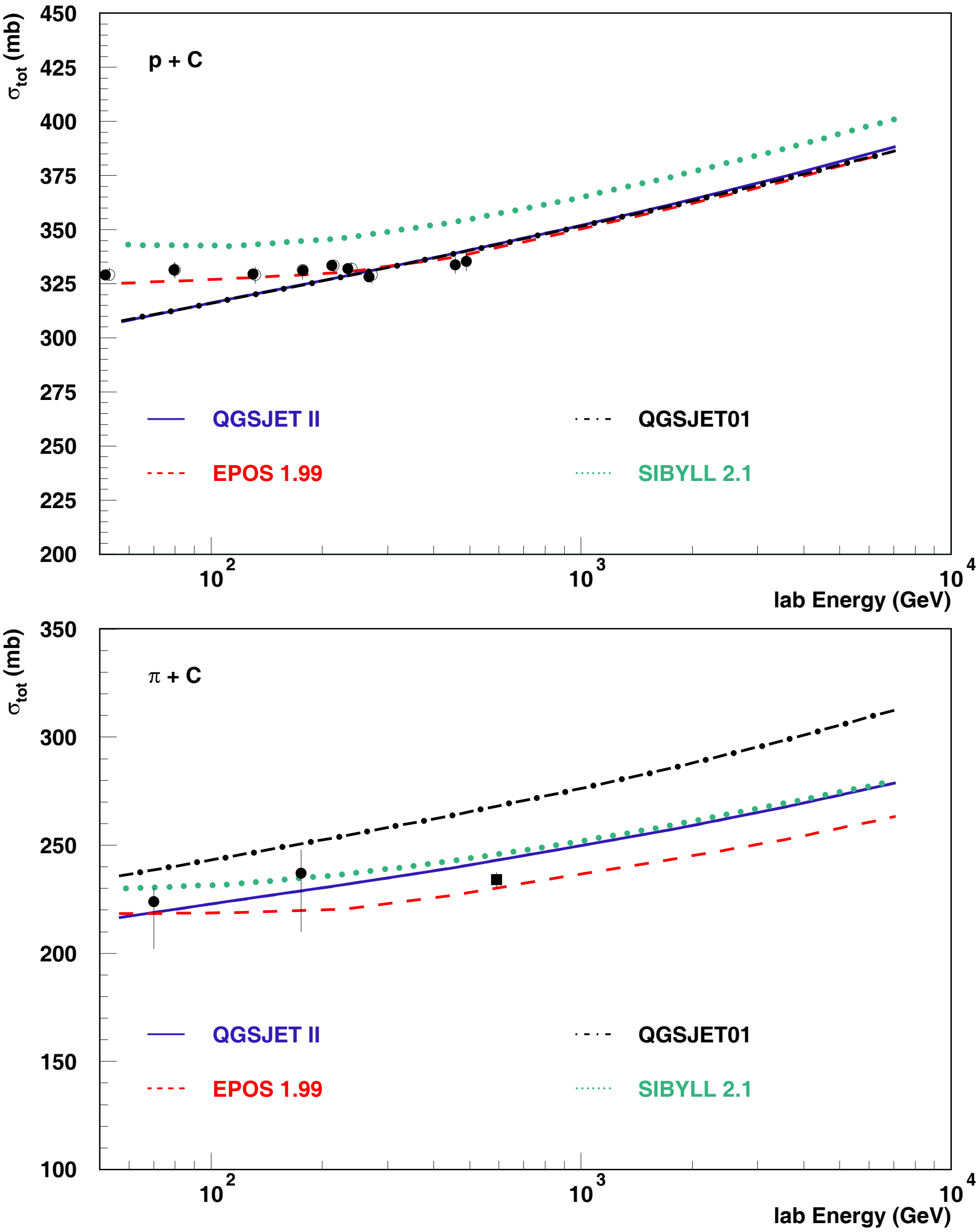
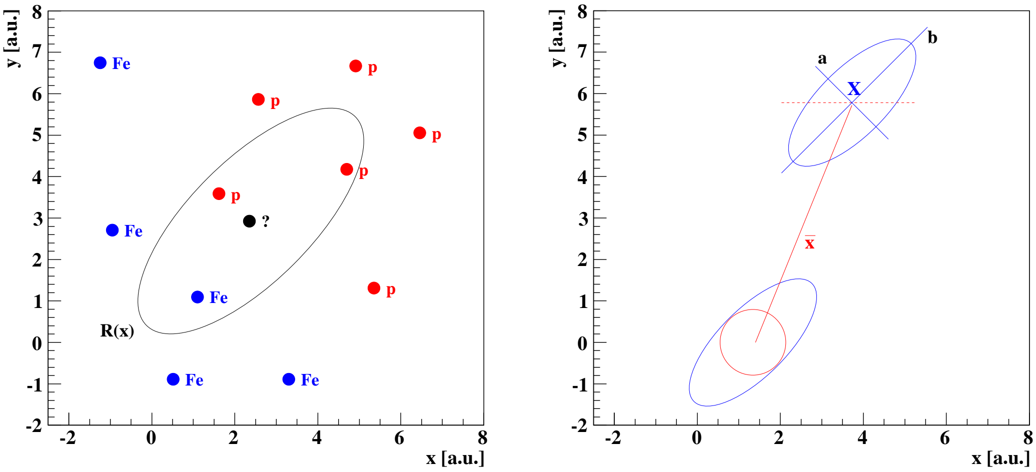


**Figure 6.8:** Comparison of proton-C (top) and pion-C (bottom) cross-section direct measurements with the expectations of the different interaction models, as a function of energy. The QGSJetII model (blue solid line) shows a tendency to overestimate the pion cross-section at high energies [57].

charged particle $N_{ch}$ and the total number of muons $N_\mu$ in a shower are reliable parameters to be used as starting point for the study of the primary mass of the observed EAS. Due to the power law dependences of the electron and muon shower sizes to the primary energy, failures in the hadronic interaction model would, for these observables, only change the assigned energy, but not the relative mass dependencies.

## 6.2 K Nearest Neighbours classification

To infer the mass of the primary particle, a classification procedure named k Nearest Neighbours (kNN) [58] is applied. This method has already been applied to cosmic ray analysis by the CASA-MIA experiment [59; 60]. According to this algorithm, an object is assigned to the class most common among its $k$ nearest neighbours (see fig. 6.9a), with $k$ being a positive integer number. In a two class classification, as it can be the proton/iron one, it is helpful to choose $k$ to be an odd number to avoid tied votes. The neighbours are taken from a set of objects (referred to as *training set*), for which the correct classification is known. In order to identify the neighbours, each object is represented by a vector in the parameter space. A distance between points in this space has to be calculated.



(a)

(b)

**Figure 6.9:** (a) Example [61] of 3NN classification. The black dot is to be classified. It has two proton events and one iron event in its neighbourhood (R(x)). According to the 3NN algorithm, it will be classified as a proton. (b) Transformation of a distribution through the application of the Mahalanobis metric.

An Euclidean distance could be used, but it would not take into account the correlation between the parameters. In addition, if the parameters describing the multidimensional space are not measured with the same metric scale, they have to be converted into units of standard deviations from the mean.

Therefore, given a point $A$ defined by the column vector $\mathbf{x}_A$ and a point $B$ defined by the column vector $\mathbf{x}_B$, the square distance between $A$ and $B$ is calculated as:

$$d^2(\mathbf{x}_A, \mathbf{x}_B) = [\mathbf{x}_A - \mathbf{x}_B]^T \mathbf{\Sigma}^{-1} [\mathbf{x}_A - \mathbf{x}_B], \tag{6.1}$$

where $\mathbf{\Sigma}^{-1}$ is the inverse of the covariance matrix of the population which B belongs to, and $^T$ indicates the transposed vector. The expression in 6.1 is known as *Mahalanobis* *distance*[62]. The vectors $\mathbf{x}_A$ and $\mathbf{x}_B$ will be also referred to as *feature vectors.* An example of the application of the Mahalanobis metric is shown in fig. 6.9b. A given distribution with centroid $\mathbf{X}$ is transformed into a normal distribution by the application of the Mahalanobis metric. The point $\mathbf{X}$ will be shifted by a quantity $\overline{x}$ and scaled and rotated according to the inverse of the covariance matrix. In the given example, the distribution will be scaled according to $b = 1/\sigma_x^2$ and $a = 1/\sigma_y^2$.

In fig. C.2, distributions of the Mahalanobis distances for measured air showers are shown. For each event, the average Mahalanobis distance of the test point to its nine nearest neighbours is calculated. A Mahalanobis distance of $\approx 1$ indicates that the neighbours are on average at a distance of one standard deviation. For increasing energy, a slightly higher Mahalanobis distance is found: the neighbours are found further away from the test point. The large tail that appears at low energies is made up of observed events, which lie outside both the iron and the proton distributions.

Nine sets of simulated proton, carbon and iron induced showers have been chosen as *training set* for the method. Approximatively 21,000 showers per primary particle have been selected, with energies above $6.3 \cdot 10^{15}$ eV, inside the zenith angular range $0° - 40°$. An independent set of roughly 2,800 air showers per primary particle is used as *testing set.* With both, the selection of well reconstructed events is done, as explained in chapter 4. Elements of the training (or testing) set are in the following referred to as *items* or *units.*

The classification has been tested in two ways: first, the measured events have been classified into proton or iron-like air showers. Then, a third class, corresponding to carbon primaries, has been added.

The training and testing sets are divided in ranges of zenith angle, as done in the shower size spectra analysis (see table 5.1). Each sample is characterised by the reconstructed number of charged particles in the shower $N_{ch}$ and the reconstructed muon number $N_\mu$. $N_{ch}$ has been chosen instead of $N_e$, because of the independent reconstruction of $N_{ch}$ from the $N_\mu$ estimation (see chapter 4). For each zenith angle bin and for each of the three training classes under investigation, the corresponding covariance matrices are calculated. To classify a test item, its Mahalanobis distance to each of the training items is calculated and the ten closest ones are stored. A majority vote is then applied to assign the test item to one of the training classes.

A set of weights is applied to the simulated energy spectrum, to reproduce the spectral index $\gamma = -3$, which is expected for the measured data. Classification problems may occur at the highest energies, where only few samples are available due to the limited statistics in the training set and to the steepness of the energy spectrum after the weighting. To limit the deficit of statistics, the training set is *locally* weighted [59]: the training unit closest to the test item is given a weight of 1, whereas the following units obtain a weight

$$w = \left(\frac{E_{item}}{E_{closest}}\right)^{\gamma_1 - \gamma_2},$$

where $\gamma_1$ and $\gamma_2$ are the spectral indices of the original and desired spectrum respectively. This implies that, after the weighting procedure, the closest events will be counted as $1 \cdot w$. The total number $k$ of nearest neighbours will then not be an integer anymore.

### 6.2.1 Training of the method

For the kNN algorithm, the training phase consists mainly in storing the feature vectors and the class labels of the training items. One unique parameter has to be fixed, the number $k$ of neighbours that we consider for the classification of the test item. $k$ is chosen in order to minimise the classification error of the samples.

**Hit rate estimation**

The Leave One Out for a time method (LOO) [62] is applied, to estimate the hit and misclassification rates. This is a two steps process: firstly one unit is removed from the training set and the covariance matrices are calculated with the remaining $N-1$ units; in the second step, these matrices are used to calculate the Mahalanobis distance of the deleted unit from the $N-1$ training items and classify it into one of the training classes. The process is carried out $N$ times, once for each unit of the sets. The probability that a unit of the set $\omega_i$ is classified as member of the class $\omega_k$ is expressed as:

$$P_{\omega_i \to \omega_k} = \frac{n_{\omega_i \to \omega_k}}{n_i}, \tag{6.2}$$

with $n_i$ being the total number of items of the class $i$, and $n_{\omega_i \to \omega_k}$ the number of units belonging to $\omega_i$ but assigned to $\omega_k$. The $P_{\omega_i \to \omega_i}$ provides the classification hit-rate, while $1 - P_{\omega_i \to \omega_i}$ the classification error (i.e. the misclassification rate).

A possible drawback of the LOO method is that it may yield hit-rate estimations with high variability over repeated sampling, due to the fact that the $N$ covariance matrices are derived from nearly identical samples and that the statistics available in the simulations is limited.

To estimate the hit-rate error, a *bootstrap analysis* is performed. In the simplest form of bootstrapping, subsets of the given training set are repeatedly analysed. Each subset (*bootstrap sample*) is a random sample with replacement from the full sample: the artificial sets will have the same size as the original set, with typically $\approx 37\%$ of the original events being substituted with duplicates. In the present work, 100 bootstrap samples are drawn for each of the proton, carbon and iron original sets. Each bootstrap sample is used to design the kNN classifier, the units of the original set which have not been re-sampled are used for testing. The mean value and the variance of the hit and misclassification rates are obtained from the analysis of these 100 sets, providing a measure of the statistical uncertainties of the method.

**Influence of the weighting procedure**

Tests are performed with toy Monte Carlo simulations, to see to what extent the hit-rate improves, by using a $\mathrm{E}^{-2}$ spectrum locally weighted to $\mathrm{E}^{-3}$ instead of the original $\mathrm{E}^{-2}$ spectrum as training set. The results are shown in fig. 6.10, using a test set with spectral index $\gamma = -3$. The red and blue triangles represent the classification errors obtained by using a spectrum with index $\gamma = -2$ as training set, with and without the application of the local weights, respectively. Green circles are obtained by using a spectral index $\gamma = -3$ in the training set as well. The weighting algorithm shows the improvement of the classification for proton induced showers and an odd number of $k$ nearest neighbours: for any odd $k > 5$, the classification error is totally consistent with the error obtained by using the correct spectral index in the training set. For iron induced showers, the misclassification rate is lower in the weighted training set than in the case of a $\mathrm{E}^{-3}$ spectrum, for any $k$. The shower to shower fluctuations, which are larger for proton primaries, are less incisive when using $\gamma = -2$ than $\gamma = -3$. Therefore, the probability to misclassify the iron induced showers as proton induced ones is lower.

The effect of the weighting algorithm is tested on full simulated training set as well. The study of the classification error and the best number of neighbours $k$ is carried out

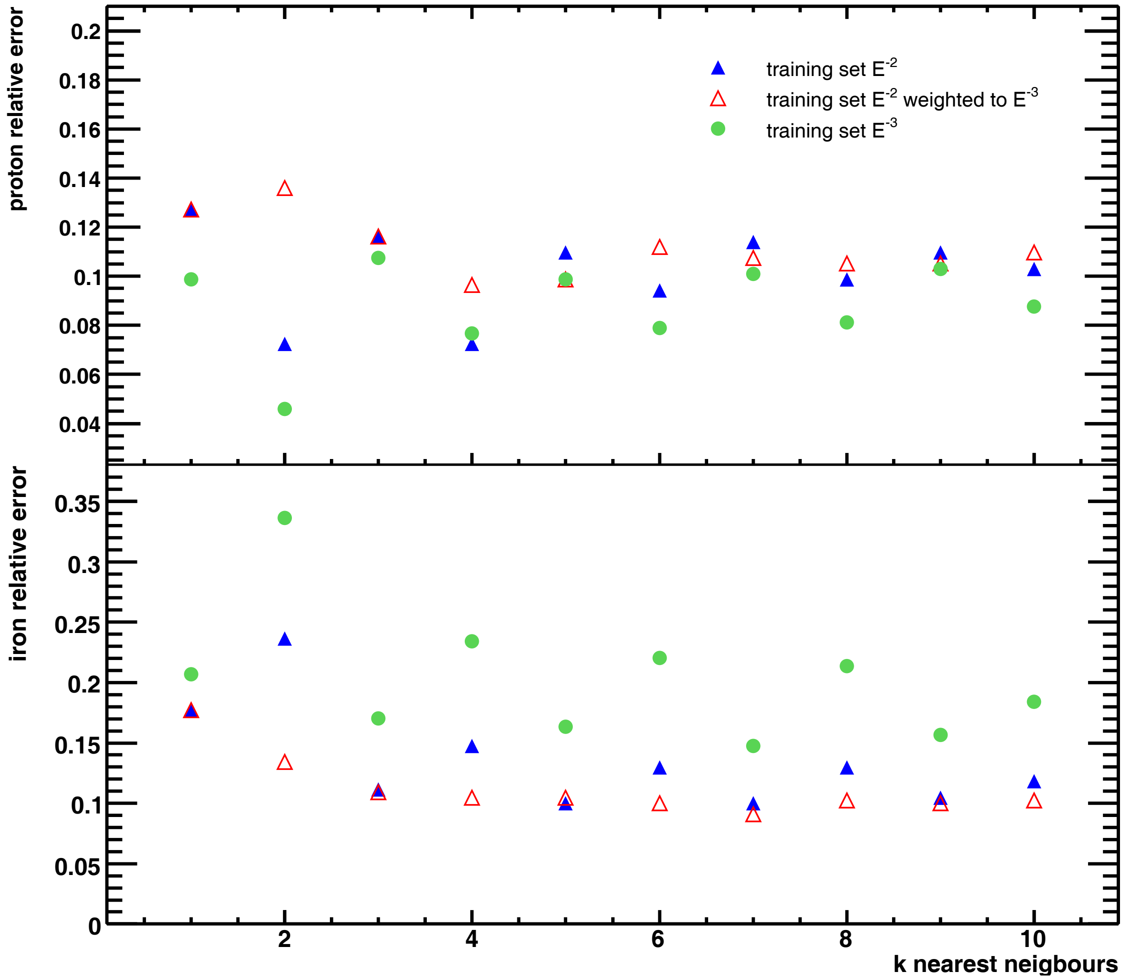


**Figure 6.10:** Classification error obtained for proton (top) and iron (bottom) induced showers, sampled on a $E^{-3}$ spectrum, as a function of the number of neighbours $k$ used to classify the event. Blue triangles represent the errors obtained without weighting the simulated spectra, red open triangle those obtained in the case of a locally weighted spectrum and green circles the errors obtained using in the training set a $E^{-3}$ power law.

twice, with and without applying weights. For proton and iron primaries, the classification error, as obtained with the LOO method, is shown as a function of $k$ in fig. 6.11 . The

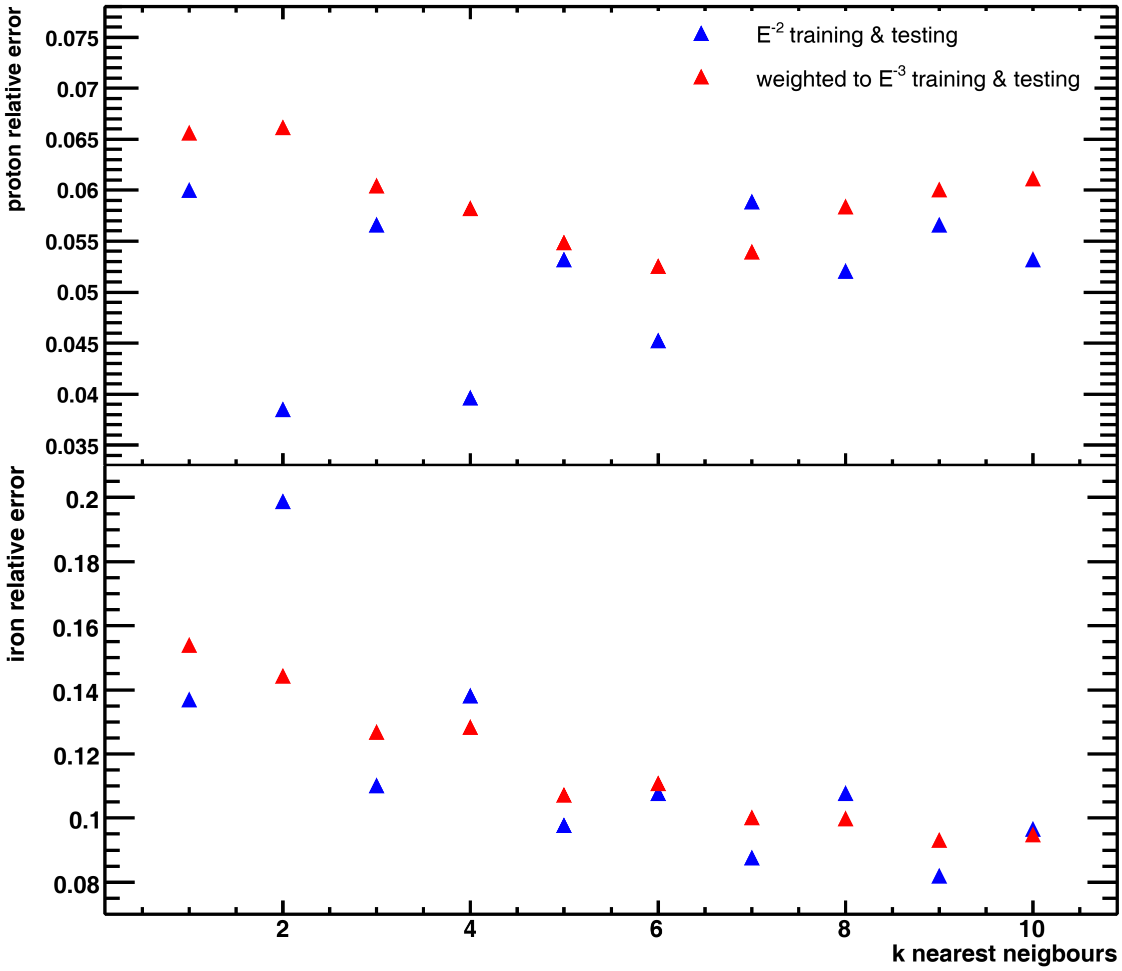
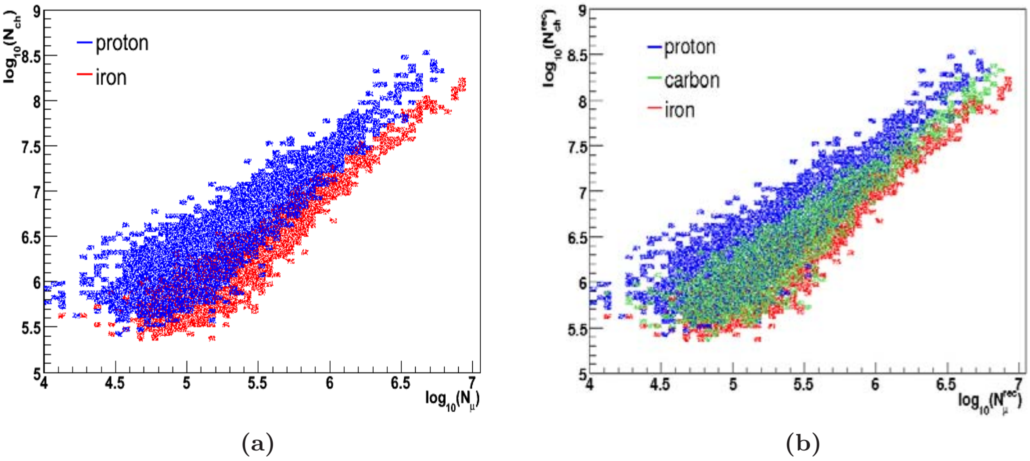


**Figure 6.11:** Classification error obtained from the LOO method, for proton (top) and iron (bottom) induced showers, as a function of the number of neighbours $k$. Errors obtained with no weights on the training set are represented in blue, whereas those obtained in the case of a locally weighted spectrum are shown in red. Showers with zenith angles in the range between 17.6° and 24.99° are considered.

blue triangles represent the values without the application of any weight, while the errors obtained adopting the weighting technique are displayed in red. In both cases, the errors decrease with higher values of the $k$ parameter, but no significant difference is seen for any $k$ above 6. Errors are always bigger for iron primaries, when $k$ is a even number, and smaller for proton induced showers. Being a two class classification, the method seems to have a preference to classify events as proton like, due to the larger fluctuation of $N_{ch}$ for proton induced air showers (see fig. 6.12a). For iron primaries, the introduction of weights worsens the classification error for odd $k > 6$ by about 1%. For proton induced showers, a similar effect, but of only 0.5%, is observed.



(a) (b)

**Figure 6.12:** Distributions in $N_{ch}-N_{\mu}$ plane of the training samples, for vertical showers. (a) distribution of proton (blue) and iron (red) samples, (b) the analogous distribution for the carbon set (green) is added.

### Estimation of the best $k$

The whole error analysis is repeated, dividing the test sets into 4 energy bins, from $\log_{10}(E_{est}/\mathrm{GeV}) = 7.0$ to $\log_{10}(E_{est}/\mathrm{GeV}) = 9.0$ in steps of $\log_{10}(E_{est}/\mathrm{GeV}) = 0.5$. The energy of the primary particle is estimated on the basis of the reconstructed electron number $N_e$ and of the reconstructed muon number $N_{\mu}$, according to the following expression:

$$r = \frac{\log_{10}\left(\frac{N_e}{N_\mu}\right) - \log_{10}\left(\frac{N_e}{N_\mu}\right)_p}{\log_{10}\left(\frac{N_e}{N_\mu}\right)_{Fe} - \log_{10}\left(\frac{N_e}{N_\mu}\right)_p}, \tag{6.3}$$

$$\log_{10}(E_{est}) = (a_p + \Delta a \cdot r) \cdot \log_{10}(N_e^0) + (b_p + \Delta b \cdot r) \tag{6.4}$$

For a detailed explanation of the derivation of the formula and its parameters see appendix A.

With application of the weighting procedure, the evolution of the error as a function of $k$ for each energy bin and for each zenith angle range is checked. In fig. 6.13, an example for vertical showers ($0° \leq \theta < 17.6°$) is shown. For proton primaries, a decrease of the error is observed for high energy events. For $k = 9$ the proton misclassification rate improves from 13%, in the lowest energy, to an error of 2% for the highest energy. For iron primaries the error fluctuates between 8% and 14%. A $k$ equal to 9 seems to minimise the misclassification error of the iron induced showers, especially in the energy range $8.5 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 9.0$, where the available simulation statistics decreases. The number $k$ of neighbours will, thus, be fixed to nine for the rest of the analysis. In fig. 6.14, the misclassification rates as a function of the estimated energy are shown: the rates for proton and iron induced showers for the five zenith angle ranges are displayed. For sake of clarity the error, obtained from the bootstrap analysis, is not drawn . With increasing zenith angle, a deterioration of the misclassification rate is observed, as expected from the decreasing statistics available for inclined showers. The numerical values of $P_{\omega_i \to \omega_i}$ and $P_{\omega_i \to \omega_k}$ (equation 6.2), whose errors are estimated with the bootstrap analysis, are reported

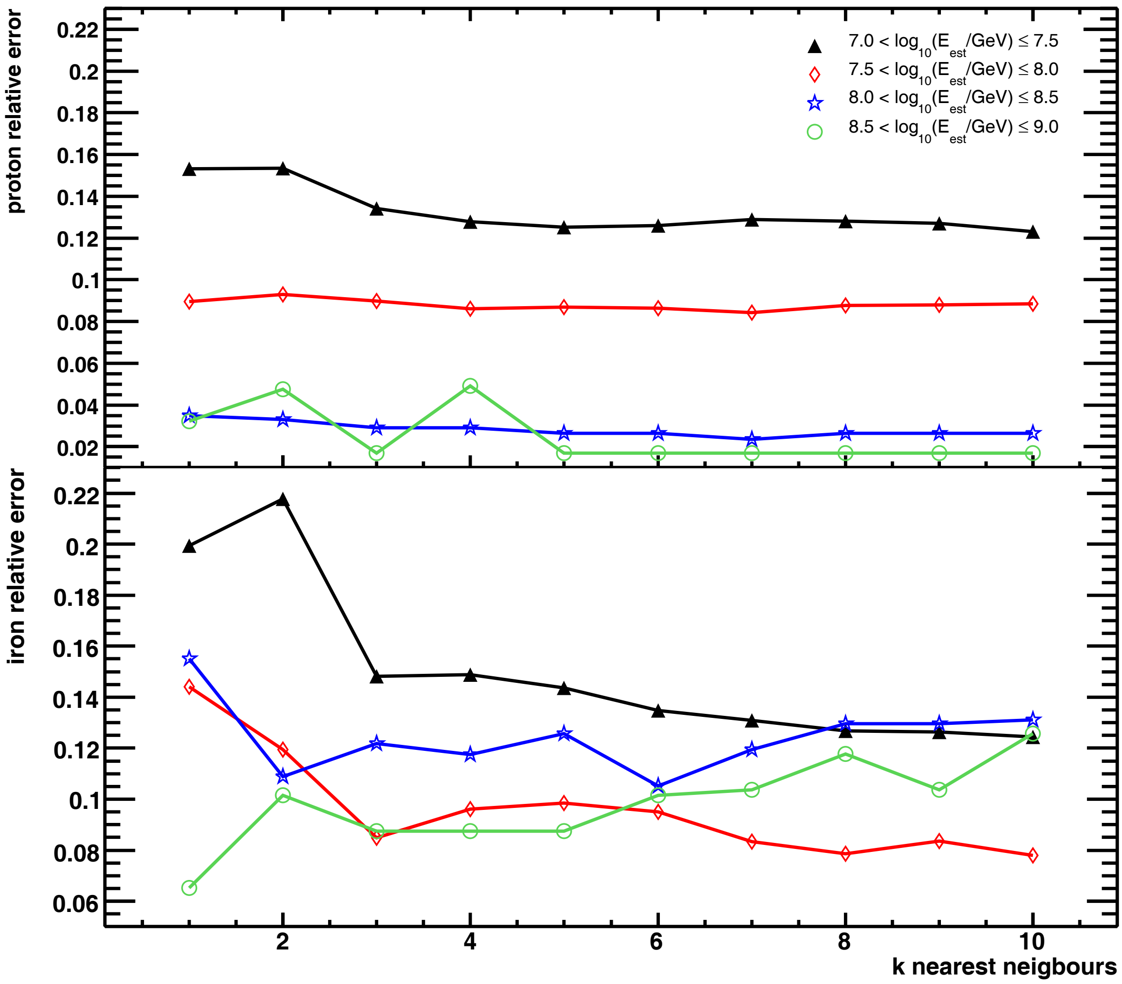


**Figure 6.13:** Classification error obtained from the LOO method, for proton (top) and iron (bottom) induced showers, as a function of the number of neighbours $k$. Different markers and colours represent the four energy ranges considered in this work: $7.0 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 7.5$ (black triangles), $7.5 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 8.0$ (red diamonds), $8.0 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 8.5$ (blue stars), $8.5 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 9.0$ (green circles). Only showers in the zenith angle range from 0° to 17.6° are shown. To guide the eyes, straight lines are drawn.

in table 6.1 for proton and iron primaries. As measurement of the global separation capability of the method, a new quantity is introduced, the *separability index* $G$ [61]. It is defined as the geometric mean of the probabilities $P_{\omega_i \to \omega_i}$. A separability index lower than 50% hints to the incapacity of the kNN to correctly perform the classification. The results, for each zenith angle and energy range, are reported in table 6.2. For the two groups classification, a separability better than 80% is observed, with a higher discrimination power for $E_{est} > 3 \cdot 10^{16}$ eV, well above the detection energy threshold range.

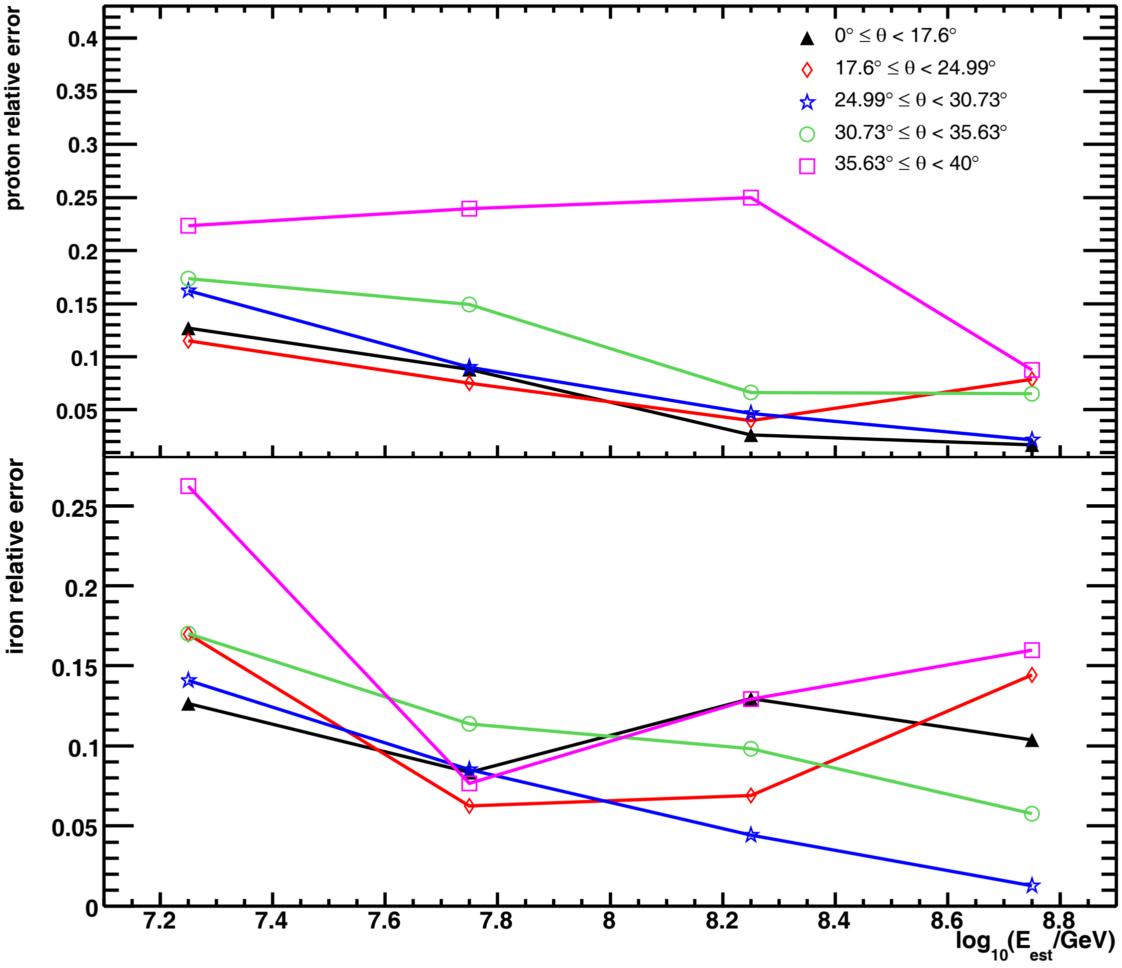


**Figure 6.14:** Classification error obtained from the LOO method for proton (top) and iron (bottom) induced shower, as a function of the estimated energy for $k = 9$. Different markers and colours represent the five zenith angle ranges considered in this work: $0° \leq \theta < 16.6°$ (black triangles), $16.6° \leq \theta < 24.99°$ (red diamonds), $24.99° \leq \theta < 30.73°$ (blue stars), $30.73° \leq \theta < 35.63°$ (green circles), $35.63° \leq \theta < 40°$ (pink squares) . Straight lines are drawn to guide the eyes.

### Classification into three primary mass groups

The classification of an item to iron-like or proton-like has been considered, so far. In the detected events, however, there are also contributions from other primary mass groups, such as helium, carbon and silicon. An attempt to classify KASCADE-Grande events into three groups is performed, by adding the carbon class to those of proton and iron. An event is classified as proton, carbon or iron induced shower on the basis of a majority vote. As done for the two class classification, the hit-rate and classification errors are calculated for this case. In fig. 6.15, the classification error as a function of $k$ is shown for vertical showers and for each energy bin. For all the three primaries, a reduction of the error for high energy events is observed. However, compared to the two classes case, the

| $\theta$ (deg) | $\log_{10}(E_{est}/\mathrm{GeV})$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 7.0-7.5 | | 7.5-8.0 | | 8.0-8.5 | | 8.5-9.0 | |
| | p→p | p→Fe | p→p | p→Fe | p→p | p→Fe | p→p | p→Fe |
| 0-17.6 | 87±1 | 13±1 | 91±2 | 9±2 | 97±2 | 3±2 | 98±4 | 2±4 |
| 17.6-24.99 | 88±1 | 12±1 | 92±2 | 8±2 | 96±3 | 4±3 | 92±4 | 8± 4 |
| 24.99-30.73 | 84±2 | 16±2 | 91±2 | 9±2 | 95±2 | 5±2 | 98±4 | 2±4 |
| 30.73-35.63 | 83±2 | 17±2 | 85±3 | 15±3 | 93±3 | 7±3 | 93±5 | 6±5 |
| 35.63-40 | 78±2 | 22±2 | 76±5 | 24±5 | 75±12 | 25±12 | 91±7 | 9±7 |
| | Fe→p | Fe→Fe | Fe→p | Fe→Fe | Fe→p | Fe→Fe | Fe→p | Fe→Fe |
| 0-17.6 | 13±2 | 87±2 | 8±2 | 92±2 | 13±5 | 87±5 | 10±7 | 90±7 |
| 17.6-24.99 | 17±2 | 83±2 | 6±2 | 94±2 | 7±3 | 93±3 | 13±11 | 86±11 |
| 24.99-30.73 | 14±2 | 86±2 | 9±2 | 91±2 | 4±2 | 96±2 | 1±2 | 99±2 |
| 30.73-35.63 | 17±2 | 83±2 | 11±3 | 89±3 | 10±4 | 90±4 | 6±5 | 94±5 |
| 35.63-40 | 26±2 | 73±2 | 8±3 | 92±3 | 13±9 | 87±9 | 16±11 | 84±11 |

**Table 6.1:** Hit-rate (p→p, fe→fe) and misclassification rate (p→fe, fe→p) in percentage, for proton and iron induced showers. The probabilities for five zenith angle ranges and four energy bins in the range $7.0 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 9.0$ are given.

| $\theta$ (deg) | $\log_{10}(E_{est}/\mathrm{GeV})$ | | | |
|---|---|---|---|---|
| | 7.0-7.5 | 7.5-8.0 | 8.0-8.5 | 8.5-9.0 |
| 0-17.6 | 87 | 91 | 92 | 94 |
| 17.6-24.99 | 85 | 93 | 94 | 89 |
| 24.99-30.73 | 85 | 91 | 95 | 98 |
| 30.73-35.63 | 83 | 87 | 91 | 93 |
| 35.63-40 | 75 | 84 | 81 | 87 |

**Table 6.2:** Separability indices, in percentage, for the classification into proton and iron induced showers.

classification is worse. For $k = 9$, the frequency of proton misclassifications is between 10% and 35%, for iron primaries it is between 20% and 40% and for carbon between 30% and 65%. In fig. 6.16 the classification errors for proton, carbon and iron induced showers as a function of the estimated energy are shown. For sake of clarity, the errors obtained from the bootstrap analysis are not shown. The capability to classify the data set into the three classes is clearly poor. For an estimated primary energy lower than $10^{17}$ eV, the probability to classify a carbon initiated shower correctly is never above 50%, improving to about 70% for higher energies. Moreover also $P_{p \to p}$ and $P_{fe \to fe}$ are smaller than in the two group classification. In table 6.3, the separability indices for this case are given. In appendix C, all the probabilities $P_{\omega_i \to \omega_k}$, with $i$ and $k$ representing the proton, carbon or iron class, are listed (table C.2). For estimated energies lower that $10^{17}$ eV and for all the zenith angles considered, a separability index of less then 65% is obtained. It improves to about 70% at higher energies, with high variability between the 5 zenith angle ranges investigated.

Carbon induced showers can fluctuate towards both heavier and lighter primary masses, since their distribution in the $N_{ch} - N_\mu$ space overlaps with both the proton and iron primary distributions (fig. 6.12b). On the contrary, proton (iron) induced showers have no lighter (heavier) neighbour; their misclassification rate is then reduced. The error is smaller at high energies, i.e. larger muon and charged particle numbers, where the overlap between the three primary distributions is smaller. With the current characterisation of events through their $N_{ch}$ and $N_\mu$ numbers, the method shows to be not sensitive enough for a classification into three primary groups, which therefore will not be applied to the measured data set.

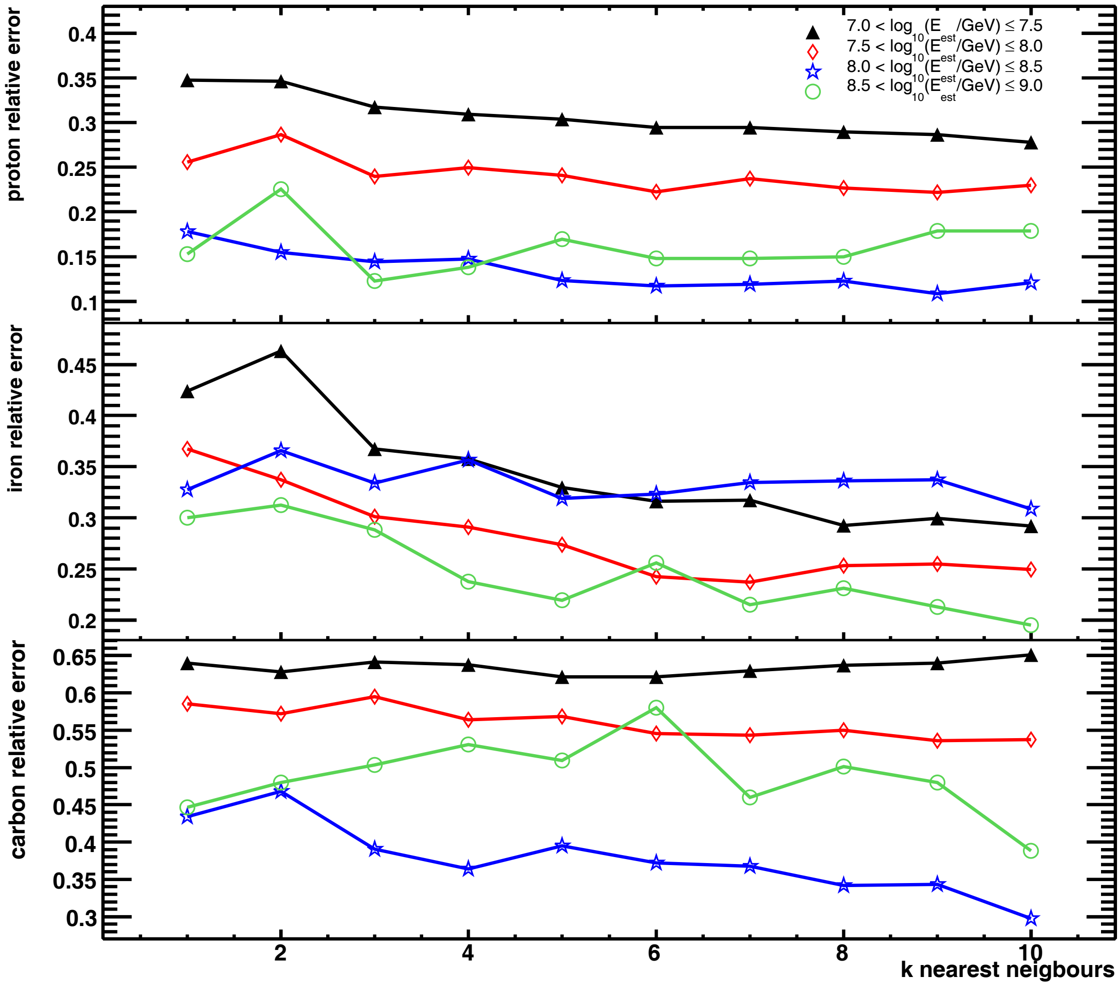


**Figure 6.15:** Misclassification rates as a function of the number of neighbours $k$, as obtained from the LOO method, for proton (top) and iron (middle) and carbon (bottom) induced showers. Different markers and colours represent the four energy ranges considered in this work: $7.0 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 7.5$ (black triangles), $7.5 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 8.0$ (red diamonds), $8.0 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 8.5$ (blue stars), $8.5 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 9.0$ (green circles) . Showers with zenith angle in the range 0° to 17.6° are considered. The connecting lines are drawn to guide the eyes.

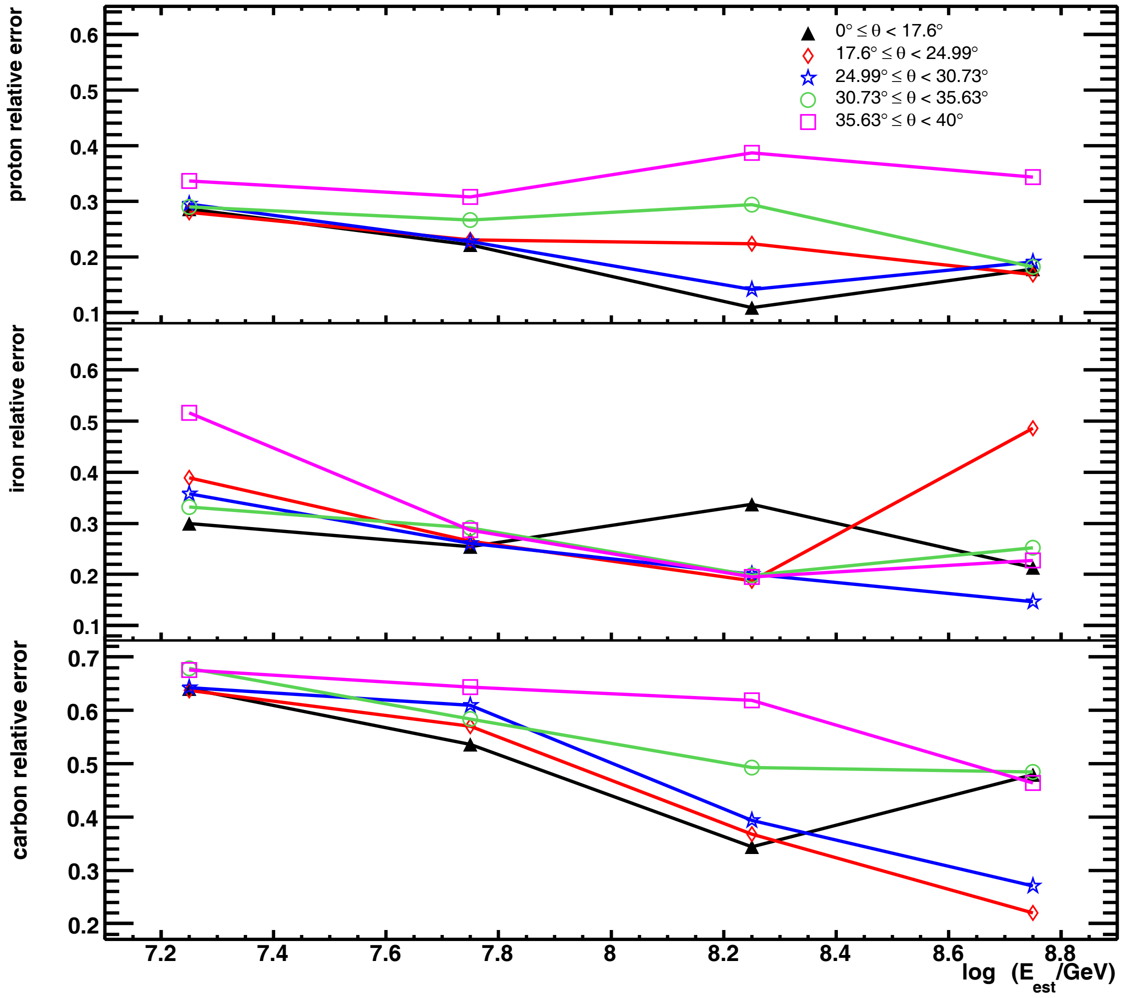


**Figure 6.16:** Classification error as a function of the estimated energy for $k = 9$, as obtained from the LOO method, for proton (top) and iron (middle) and carbon (bottom) induced showers. Different markers and colours represent the five zenith angle ranges: $0° \leq \theta < 16.6°$ (black triangles), $16.6° \leq \theta < 24.99°$ (red diamonds), $24.99° \leq \theta < 30.73°$ (blue stars), $30.73° \leq \theta < 35.63°$ (green circles), $35.63° \leq \theta < 40°$ (pink squares) . The lines are drawn to guide the eyes.

### 6.2.2 Testing kNN

After the establishment of the number of $k$ neighbours, and after the estimation of the correct classification and misclassification probabilities, the whole procedure has to be tested. For this purpose, an independent set of simulations, with the same features of those used in the training set, has been used. Two sets have been created with different composition: the first contains $\approx 25\%$ proton and $\approx 75\%$ iron primaries (set $A$), while the second is made up by $\approx 75\%$ proton and $\approx 25\%$ iron induced showers (set $B$). The events have been injected into the kNN classifier and the identification capability evaluated.

Due to the classification errors, the number of events $n_i$, which has been found by the kNN to belong to the class $\omega_i$, does not correspond to the true number $n_i^*$ belonging to it.

| $\theta$ (deg) | $\log_{10}(E_{est}/\mathrm{GeV})$ | | | |
|---|---|---|---|---|
| | 7.0-7.5 | 7.5-8.0 | 8.0-8.5 | 8.5-9.0 |
| 0-17.6 | 56 | 64 | 72 | 69 |
| 17.6-24.99 | 54 | 62 | 73 | 69 |
| 24.99-30.73 | 54 | 60 | 75 | 79 |
| 30.73-35.63 | 53 | 60 | 65 | 68 |
| 35.63-40 | 47 | 56 | 67 | 65 |

**Table 6.3:** Separability indices, in percentage, for the classification into proton, carbon and iron induced showers.

In fact, the misclassified events $P_{\omega_i \to \omega_k}$ will decrease $n_i^*$ and the misclassified events from class $\omega_k$ to class $\omega_i$ will increase $n_i^*$. In the case of a classification into two classes, it can be written:

$$\begin{pmatrix} P_{\omega_i \to \omega_i} & P_{\omega_k \to \omega_i} \\ P_{\omega_i \to \omega_k} & P_{\omega_k \to \omega_k} \end{pmatrix} \begin{pmatrix} n_i^* \\ n_k^* \end{pmatrix} = \begin{pmatrix} n_i \\ n_k \end{pmatrix}. \tag{6.5}$$

By solving this linear system and propagating the error of the misclassification matrix components, an estimation of the true number of items in the classes $\omega_i$ and $\omega_k$ is finally obtained.

The results of the test for the sets $A$ and $B$ are compiled in tables 6.4 and 6.5, respectively. The top block of the tables shows the true relative frequency of proton and iron induced showers in the set, the second part shows $n_p$ and $n_{fe}$ , as obtained from the classification, and the third section of the table gives the estimated $n_p^*$ and $n_{fe}^*$. By comparing the true and the estimated frequencies, a good agreement (within two standard deviations) is found, for both classes. For the most inclined showers ($\theta > 35°$) and energies above $10^{17}$ eV, big fluctuations are observed, due to the lower statistics available in the training set, which result in larger fluctuations in the estimation of $P_{\omega_i \to \omega_k}$.

### 6.2.3 kNN classification of the measured data set

Having tested the classification method and knowing the uncertainties connected to it, it is possible to classify the KASCADE-Grande data.

The events are divided into the proton-like and iron-like groups. In tab. 6.6, the relative frequencies of the two classes, already corrected for the misclassification errors, are reported. The values for each zenith angle and energy bin are given. The fraction of proton-like and iron-like air showers is shown in fig. 6.18. In fig. 6.17, the distribution for measured data in the energy ranges $3 \cdot 10^{16}$ eV $- 5.6 \cdot 10^{16}$ eV and $3 \cdot 10^{17}$ eV $- 5.6 \cdot 10^{17}$ eV are shown as contour lines. The distributions for the training set are shown as well. It can be already seen that a large fraction of events lies on the iron "side".

In order to avoid possible efficiency effects at detection threshold, the analysis has been repeated, by shifting of $\log_{10}(E_{est}/\mathrm{GeV}) = 0.25$ towards higher values the energy ranges selected. In this case, however, the statistics in the last energy bin ($8.75 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 9.25$) is drastically reduced, with no more than 29 events per zenith bin. The results are compiled in table 6.7. In appendix C the corresponding plot is shown (fig. C.1), while in table C.1 the $P_{\omega_i \to \omega_k}$ probabilities for these energy ranges are given.

| $\log_{10}(E_{est}/\mathrm{GeV})$ | 7.0-7.5 | | | 7.5-8.0 | | | 8.0-8.5 | | | 8.5-9.0 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta$ (deg) | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot |
| 0-17.6 | 73 | 27 | 478 | 68 | 32 | 160 | 97 | 3 | 47 | 97 | 3 | 28 |
| 17.6-24.99 | 81 | 19 | 461 | 76 | 24 | 178 | 60 | 40 | 34 | 100 | 0 | 11 |
| 24.99-30.73 | 70 | 30 | 485 | 76 | 29 | 192 | 83 | 17 | 58 | 100 | 0 | 20 |
| 30.73-35.63 | 71 | 29 | 467 | 66 | 34 | 177 | 84 | 16 | 54 | 100 | 0 | 7 |
| 35.63-40 | 75 | 25 | 359 | 68 | 32 | 116 | 80 | 20 | 37 | 82 | 18 | 15 |

| $\log_{10}(E_{est}/\mathrm{GeV})$ | 7.0-7.5 | | | 7.5-8.0 | | | 8.0-8.5 | | | 8.5-9.0 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta$ (deg) | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot |
| 0-17.6 | 65 | 36 | | 66 | 34 | | 93 | 7 | | 95 | 5 | |
| 17.6-24.99 | 80 | 20 | | 73 | 26 | | 62 | 38 | | 100 | 0 | |
| 24.99-30.73 | 59 | 41 | | 72 | 28 | | 74 | 26 | | 100 | 0 | |
| 30.73-35.63 | 64 | 36 | | 65 | 35 | | 87 | 12 | | 100 | 0 | |
| 35.63-40 | 63 | 37 | | 46 | 54 | | 75 | 25 | | 86 | 14 | |

| $\log_{10}(E_{est}/\mathrm{GeV})$ | 7.0-7.5 | | | 7.5-8.0 | | | 8.0-8.5 | | | 8.5-9.0 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta$ (deg) | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot |
| 0-17.6 | 70±3 | 30±3 | | 69±5 | 31±5 | | 95±5 | 5±5 | | 97±6 | 3±6 | |
| 17.6-24.99 | 88±3 | 12±3 | | 78±4 | 22±4 | | 62±10 | 38±10 | | 110±6 | -10±6 | |
| 24.99-30.73 | 65±4 | 35±4 | | 77±4 | 23±4 | | 77±7 | 23±7 | | 102±4 | -2±4 | |
| 30.73-35.63 | 72±4 | 28±4 | | 72±6 | 28±6 | | 93±6 | 7±6 | | 107±6 | -7±6 | |
| 35.63-40 | 71±6 | 29±6 | | 56±8 | 44±8 | | 100±22 | 0±22 | | 92±15 | 8±15 | |

**Table 6.4:** The result of the kNN test with a set with $\approx 75\%$ proton and $\approx 25\%$ iron is given. In the first block the true relative frequency $n_p^*$ and $n_{fe}^*$ of proton and iron primaries, for five zenith angle ranges and four energy bins, is shown. In the second block $n_p$ and $n_{fe}$, as they result from the kNN classification, are given. In the third block the expectations for the $n_p^*$ and $n_{fe}^*$, with the propagated error, after solving the linear system 6.5 is given.

| $\log_{10}(E_{est}/\text{GeV})$ | 7.0-7.5 | | | 7.5-8.0 | | | 8.0-8.5 | | | 8.5-9.0 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta$ (deg) | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot |
| 0-17.6 | 22 | 78 | 557 | 5 | 95 | 166 | 0 | 100 | 32 | 32 | 68 | 37 |
| 17.6-24.99 | 23 | 77 | 493 | 17 | 83 | 197 | 0 | 100 | 61 | 25 | 75 | 21 |
| 24.99-30.73 | 16 | 84 | 519 | 25 | 75 | 149 | 8 | 92 | 46 | 0 | 100 | 11 |
| 30.73-35.63 | 23 | 77 | 412 | 24 | 76 | 150 | 38 | 62 | 52 | 0 | 100 | 14 |
| 35.63-40 | 23 | 77 | 347 | 25 | 75 | 125 | 26 | 74 | 35 | 13 | 87 | 7 |

| $\log_{10}(E_{est}/\text{GeV})$ | 7.0-7.5 | | | 7.5-8.0 | | | 8.0-8.5 | | | 8.5-9.0 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta$ (deg) | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot |
| 0-17.6 | 29 | 71 | | 14 | 86 | | 12 | 88 | | 38 | 62 | |
| 17.6-24.99 | 32 | 68 | | 23 | 77 | | 11 | 89 | | 30 | 70 | |
| 24.99-30.73 | 23 | 77 | | 28 | 72 | | 7 | 93 | | 0 | 100 | |
| 30.73-35.63 | 33 | 67 | | 33 | 67 | | 37 | 63 | | 0 | 100 | |
| 35.63-40 | 33 | 67 | | 27 | 73 | | 14 | 86 | | 40 | 60 | |

| $\log_{10}(E_{est}/\text{GeV})$ | 7.0-7.5 | | | 7.5-8.0 | | | 8.0-8.5 | | | 8.5-9.0 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta$ (deg) | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot | $n_p$/tot | $n_{Fe}$/tot | tot |
| 0-17.6 | 22±3 | 78±3 | | 7±4 | 93±4 | | -1±9 | 101±9 | | 32±11 | 68±11 | |
| 17.6-24.99 | 20±4 | 80±4 | | 20±4 | 80±4 | | 4±5 | 96±5 | | 20±17 | 80±17 | |
| 24.99-30.73 | 12±4 | 88±4 | | 23±5 | 77±5 | | 2±5 | 98±5 | | -1±2 | 101±2 | |
| 30.73-35.63 | 24±4 | 76±4 | | 30±6 | 70±6 | | 33±9 | 67±9 | | -6±6 | 106±6 | |
| 35.63-40 | 13±6 | 87±6 | | 28±7 | 72±7 | | 1±17 | 99±17 | | 32±27 | 68±27 | |

**Table 6.5:** The result of the kNN test by a set with ≈ 25% proton and ≈ 75% iron is given. In the first block the true relative frequency $n_p^*$ and $n_{fe}^*$ of proton and iron primaries, for five zenith angle ranges and four energy bins, is shown. In the second block $n_p$ and $n_{fe}$, as they result from the kNN classification, are given. In the third block the expectations for the $n_p^*$ and $n_{fe}^*$, with the propagated error, after solving the linear system 6.5, is given.

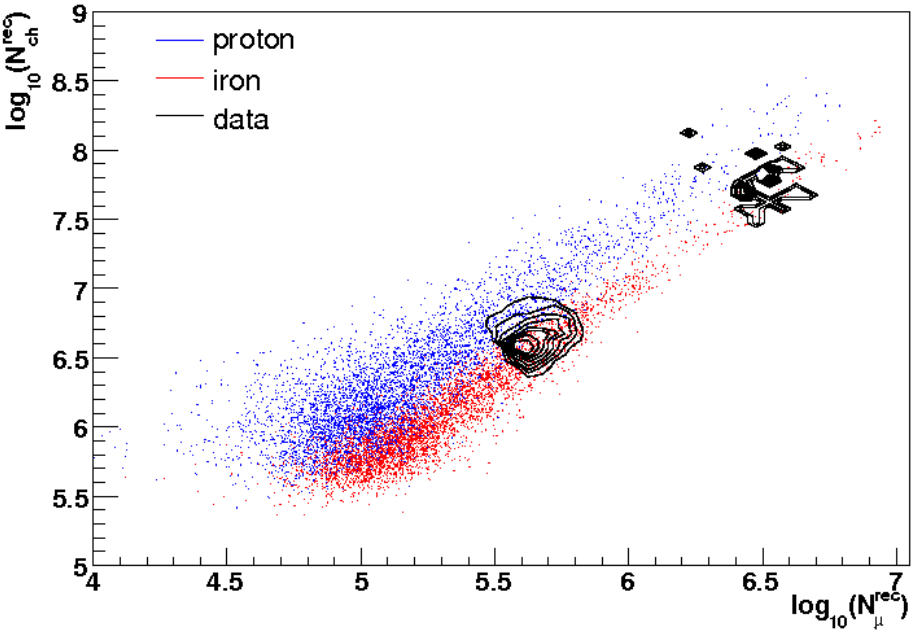


**Figure 6.17:** Comparison of p, Fe and data distributions in the $N_{ch} - N_{\mu}$ space. Proton and iron induced showers are in blue and red indicated, respectively. The black lines represent the contour distribution of the measured data: for $3{\cdot}10^{16}\,\mathrm{eV} < E_{est} \leq 5.6{\cdot}10^{16}\,\mathrm{eV}$ (centre) and $3{\cdot}10^{17}\,\mathrm{eV} < E_{est} \leq 5.6{\cdot}10^{17}\,\mathrm{eV}$ (top right). Only vertical showers are depicted.

| | $\log_{10}(E_{est}/\mathrm{GeV})$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\theta$ (deg) | 7.0-7.5 | | 7.5-8.0 | | 8.0-8.5 | | 8.5-9.0 | |
| | $n_p^*$ | $n_{Fe}^*$ | $n_p^*$ | $n_{Fe}^*$ | $n_p^*$ | $n_{Fe}^*$ | $n_p^*$ | $n_{Fe}^*$ |
| 0-17.6 | 40±2 | 60±2 | 29±2 | 71±2 | 30±5 | 70±5 | 37±10 | 63±10 |
| 17.6-24.99 | 42±2 | 58±2 | 33±2 | 67±2 | 33±3 | 67±3 | 32±12 | 68±12 |
| 24.99-30.73 | 41±2 | 59±2 | 28±2 | 72±2 | 31±3 | 69±3 | 31±6 | 69±6 |
| 30.73-35.63 | 43±2 | 57±2 | 33±3 | 67±3 | 31±4 | 69±4 | 26±8 | 74±8 |
| 35.63-40 | 48±3 | 52±3 | 41±4 | 59±4 | 33±12 | 66±12 | 38±14 | 62±14 |
| mean | 41±1 | 59±1 | 30±1 | 70±1 | 32±2 | 68±2 | 31±4 | 69±4 |

**Table 6.6:** Percentage of p and Fe showers measured by KASCADE-Grande, as classified by the kNN method. The analysis is carried out on five zenith angle ranges and four energy bins, in the range $7.0 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 9.0$.

Excluding the most inclined showers, for which the testing phase has shown the largest fluctuations, the weighted mean of the relative frequency of proton and iron-like air showers is computed over the remaining four zenith angle ranges (see tables 6.6, 6.7). The results for both cases, i.e starting the classification at an energy of $10^{16}\,\mathrm{eV}$ or of $1.8{\cdot}10^{16}\,\mathrm{eV}$, are shown in figures 6.19a and 6.19b. A change towards heavier elements is observed, starting from the energy threshold at $E = 10^{16}\,\mathrm{eV}$, up to $E = 3 \cdot 10^{16}\,\mathrm{eV}$. The same

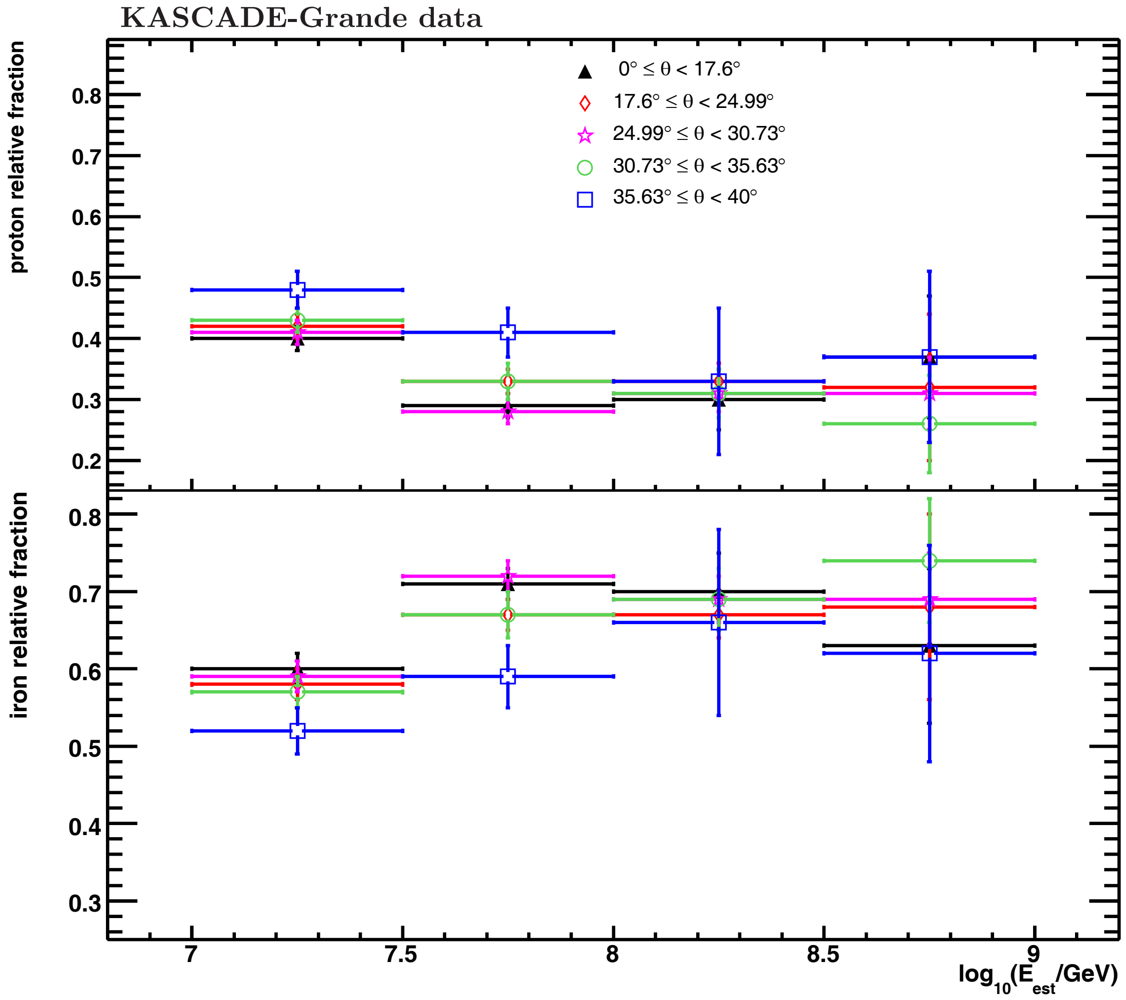
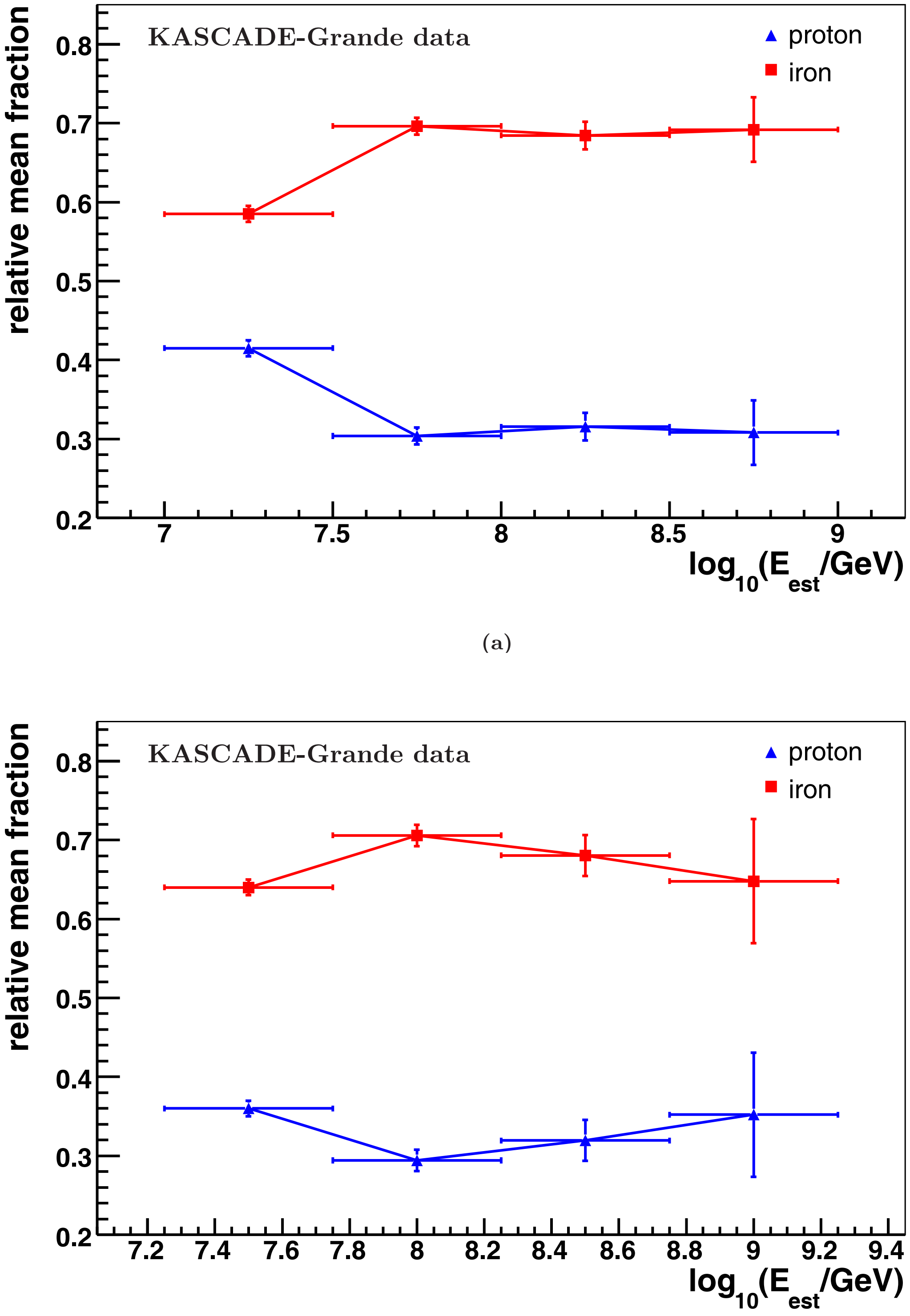


**Figure 6.18:** Relative frequency of proton and iron induced showers by kNN method. Markers and colours represent five zenith angle ranges: $0° \leq \theta < 16.6°$ (black triangles), $16.6° \leq \theta < 24.99°$ (red diamonds), $24.99° \leq \theta < 30.73°$ (blue stars), $30.73° \leq \theta < 35.63°$ (green circles), $35.63° \leq \theta < 40°$ (pink squares).

trend is observed when starting the classification at $1.8{\cdot}10^{16}$ eV, which excludes efficiency problems. At higher energies, no evolution towards a lighter or heavier composition is seen. About 30% of the measured air showers are classified as proton-like, whereas $\approx 70\%$ as iron-like. If the classification is started at an energy of $1.8{\cdot}10^{16}$ eV, a change in the relative amount of iron-like showers is observed above $\approx 6 \cdot 10^{17}$ eV. In order to better understand the nature of this change, the classification is repeated by dividing the data set into energy bins of width $\log_{10}(E_{est}/\mathrm{GeV}) = 0.25$, assuming as valid the misclassification rates calculated for a bin width of $\log_{10}(E_{est}/\mathrm{GeV}) = 0.5$. The weighted average is calculated also in this case (see table 6.8), the results being shown in fig. 6.20. Above an energy of $10^{18}$ eV, the statistics is low, with no more than six events being observed in each zenith angular range; the error on the estimated frequencies are totally dominated by the statistics. The general trend observed in the previous analysis is confirmed. In the range



(a)

(b)

**Figure 6.19:** Relative frequency of p and Fe induced showers, averaged over four zenith angle ranges. In (a) the mean fraction of iron (red square) and proton (blue triangle) primaries in the energy range $7.0 < \log_{10}(E_{est}/\text{GeV}) \leq 9.00$ is shown. In (b) the same quantities are reported for an energy range $7.25 < \log_{10}(E_{est}/\text{GeV}) \leq 9.25$.

| | $\log_{10}(E_{est}/\mathrm{GeV})$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\theta$ (deg) | 7.25-7.75 | | 7.75-8.25 | | 8.25-8.75 | | 8.75-9.25 | |
| | $n^*_p$ | $n^*_{Fe}$ | $n^*_p$ | $n^*_{Fe}$ | $n^*_p$ | $n^*_{Fe}$ | $n^*_p$ | $n^*_{Fe}$ |
| 0-17.6 | 32±2 | 68±2 | 31±3 | 69±3 | 28±7 | 72±7 | 44±14 | 56±14 |
| 17.6-24.99 | 35±2 | 65±2 | 30±2 | 70±2 | 37±5 | 63±5 | -13±59 | 113±59 |
| 24.99-30.73 | 37±2 | 63±2 | 24±3 | 76±3 | 30±4 | 70±4 | 32±12 | 68±12 |
| 30.73-35.63 | 40±2 | 60±2 | 34±4 | 66±4 | 32±6 | 68±6 | 33±16 | 67±16 |
| 35.63-40 | 50±3 | 50±3 | 31±5 | 69±5 | 29±9 | 71±9 | -2±67 | 102±67 |
| mean | 36±1 | 64±1 | 29±1 | 71±1 | 32±2 | 68±2 | 35±8 | 65±8 |

**Table 6.7:** Relative number (in percent) of p and Fe showers measured by KASCADE-Grande, as classified by the kNN method. The analysis is carried out on five zenith angle ranges and four energy bins in the range $7.25 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 9.25$.

$3 \cdot 10^{16}\,\mathrm{eV} < E_{est} \leq 10^{18}\,\mathrm{eV}$, the relative frequencies of the two classes are compatible, within their uncertainties, with a constant composition, with about 69% of the shower being iron-like. An increase of the relative number of iron-like showers is observed between $E_{est} \approx 10^{16}\,\mathrm{eV}$ and $E_{est} \approx 3 \cdot 10^{16}\,\mathrm{eV}$, with a significance of 12 standard deviations in the first case and 5 standard deviations in the second case. Above $E_{est} \approx 10^{18}\,\mathrm{eV}$, an average increase of the frequencies in the proton class is observed, the percentage of proton-like of events being $50\% \pm 13\%$. In any case, less than six air showers per zenith angle range have been detected above an energy of $\approx 10^{18}\,\mathrm{eV}$. The measurement is consistent within 1.5 standard deviations with a constant composition of 69% iron-like showers. Therefore, due to the statistical uncertainty associated with the measurements, this change in the relative composition is not significant. The evidence for such an increase requires a higher statistics, which is presently not available.

| $\log_{10}(E_{est}/\mathrm{GeV})$ | p-like [%] | fe-like [%] |
|---|---|---|
| 7.00-7.25 | 43±1 | 57±1 |
| 7.25-7.50 | 36±1 | 64±1 |
| 7.50-7.75 | 31±1 | 69±1 |
| 7.75-8.00 | 31±1 | 69±2 |
| 8.00-8.25 | 31±2 | 68±5 |
| 8.25-8.50 | 32±3 | 72±6 |
| 8.50-8.75 | 32±5 | 69±5 |
| 8.75-9.00 | 28±7 | 62±7 |
| 9.00-9.25 | 50±13 | 50±13 |

**Table 6.8:** Average proton/iron frequencies from $E_{est} = 10^{16}\,\mathrm{eV}$ to $E_{est} = 10^{18}\,\mathrm{eV}$.

### 6.2.4 EPOS test

In order to estimate systematic uncertainty, due to the hadronic interaction model used for the training sample, the same analysis should be performed assuming showers simu-

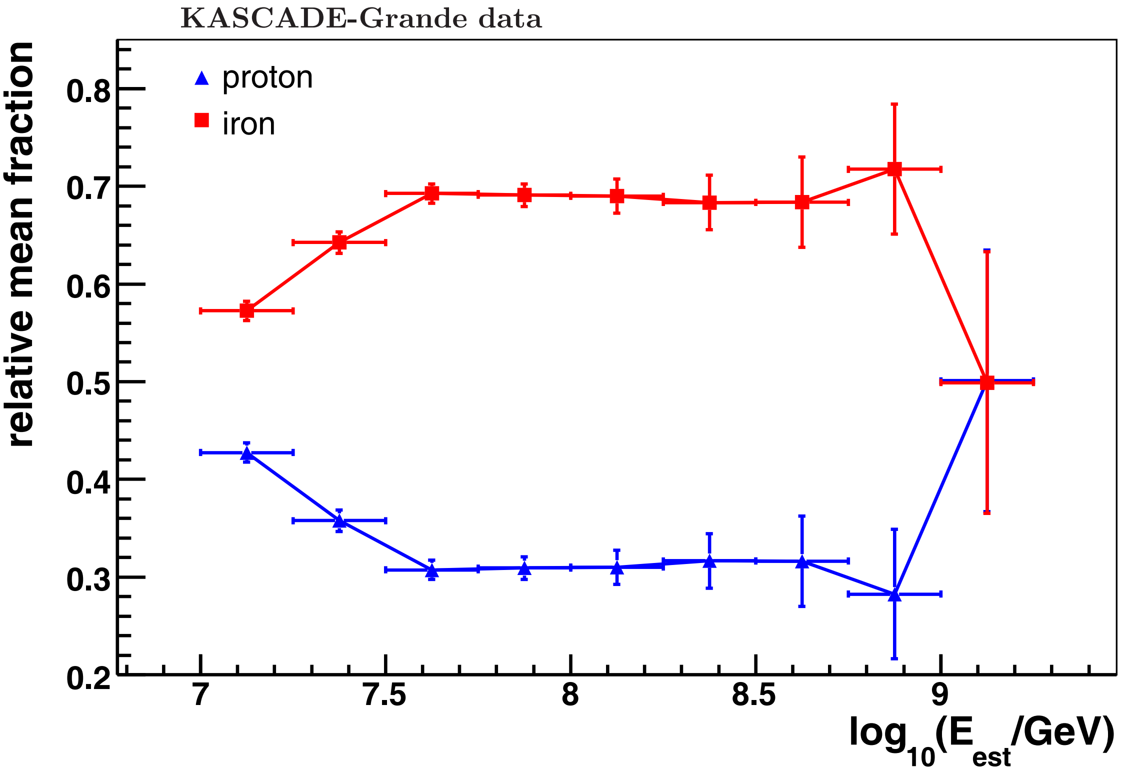


**Figure 6.20:** Relative frequency of p and Fe induced showers, averaged over four zenith angle ranges. The mean fraction of iron (red square) and proton (blue triangle) primaries in the energy range $7.0 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 9.00$, in bins of $\log_{10}(E_{est}/\mathrm{GeV}) = 0.25$, is shown.

lated with other hadronic interaction models as reference set, for example EPOS [63; 64]. The statistics available for full simulated showers with EPOS1.6 as high energy hadronic interaction model is limited to $\approx 8,000$ showers, for each of the proton and iron primaries. It is, thus, not possible to carry out the kNN tests with EPOS showers as training set. An alternative is to use them as a test sample, to see how they would be classified with the already developed classifier. Three test cases are built: a set with 100% proton induced showers (TEST A), a set with 100% iron induced showers (TEST B) and one with 50% iron and 50% proton induced showers (TEST C). For each of them, $n_p^*$ and $n_{Fe}^*$ are estimated; the results are reported in table 6.9. There is a marked bias toward iron-like showers, with an average of 50% of the proton primaries being assigned to the iron class and with a constant overestimation of the iron fraction for TEST B. If EPOS1.6 describes the reality, the classification of the measured events would be systematically shifted towards the proton class, resulting in a lighter average composition along the whole energy range.

The comparison of EPOS1.6 simulated showers with the KASCADE data set [65] has shown that EPOS does not deliver enough hadronic energy down to the observation level and the energy per hadron seems to be too small. In the $N_e - N_\mu$ plane the EPOS showers are shifted to lower electron and/or higher muon numbers relative to QGSJETII (see fig.6.21). When the mass composition of cosmic rays is derived from measured values this effect leads to a relatively light mass composition. The incompatibility of the EPOS1.6 predictions with the KASCADE measurements is caused by too high inelasticity and too high inelastic cross sections for hadronic interactions implemented in the EPOS1.6 code [64].

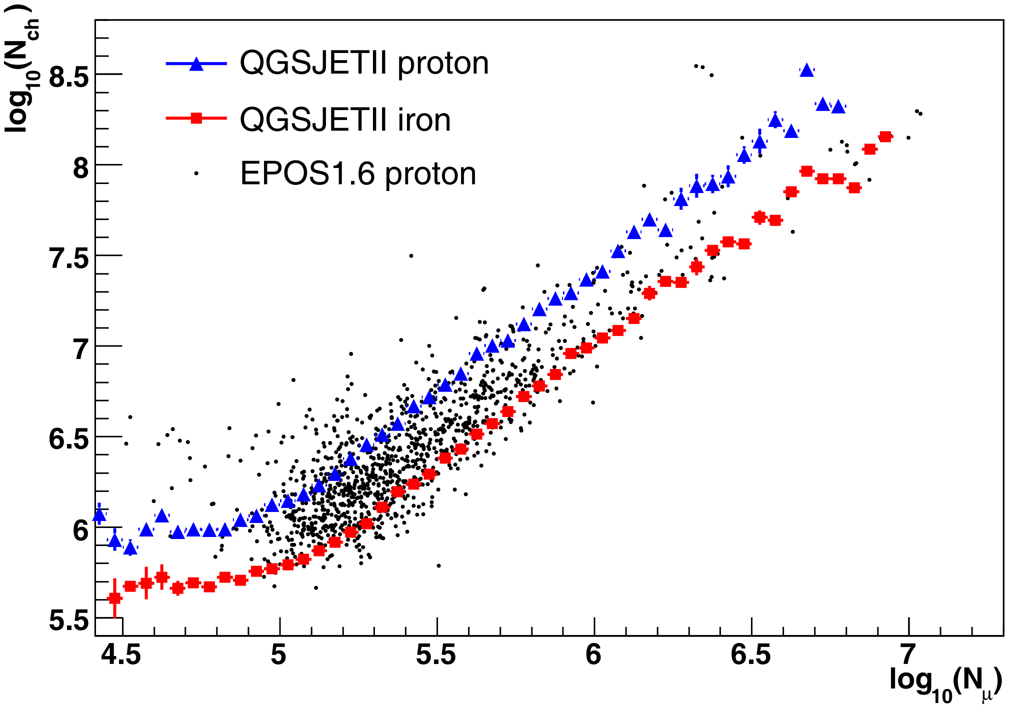


**Figure 6.21:** $N_{ch} - N_{\mu}$ distribution for EPOS1.6 proton induced showers (black points). Red squares represent the profile distribution of iron induced showers, simulated with QGSJetII-2. Blue triangles symbolize the same profile distribution for proton induced showers with QGSJetII-2.

## 6.3 Discussion of the result

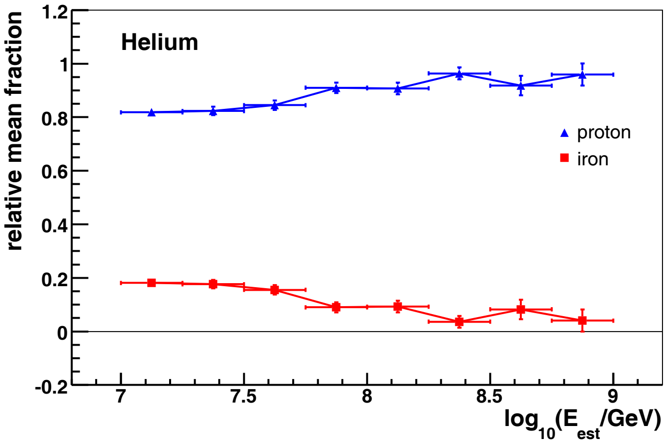
The classification of the KASCADE-Grande events does not consider other primary mass groups, which are anyway contributing to the measured events. In order to understand the extent of this bias, simulated helium, carbon and silicon induced air showers have been classified by the kNN method, according to proton-like and iron-like events. The results, averaged over the zenith angular ranges and corrected for the misclassification errors, are shown in fig. 6.22. Up to $E_{est} = 10^{18}$ eV, helium primaries contribute with more than 80% of the showers to the proton class. Carbon primaries contribute almost in equal proportions to both classes, with a strong fluctuations above energies of $6 \cdot 10^{16}$ eV. Silicon induced air showers are mainly ($\approx 80\%$) classified as iron-like, with a trend to increase the iron fraction with increasing energy. As expected, light primaries, as helium, are classified as proton-like air showers and heavy primaries, as iron-like air showers. It is not possible on the basis of the present analysis to disentangle the contribution of each primary mass group to the measured data, so the classification so far performed gives an indication of the ratio “light to heavy” primaries.

| | $\log_{10}(E_{est}/\mathrm{GeV})$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| bin $\theta$ | 7.0-7.5 | | 7.5-8.0 | | 8.0-8.5 | | 8.5-9.0 | |
| | $n^*_p$ | $n^*_{Fe}$ | $n^*_p$ | $n^*_{Fe}$ | $n^*_p$ | $n^*_{Fe}$ | $n^*_p$ | $n^*_{Fe}$ |
| 1 | 51±3 | 49±3 | 44±4 | 56±4 | 66±8 | 34±8 | 46±15 | 54±15 |
| 2 | 54±3 | 46±3 | 56±4 | 44±4 | 59±8 | 41±8 | 47±13 | 54±13 |
| 3 | 53±3 | 47±3 | 50±4 | 50±4 | 37±6 | 63±6 | 38±8 | 62±8 |
| 4 | 41±3 | 59±3 | 38±5 | 62±5 | 68±8 | 32±8 | 47±11 | 53±11 |
| 5 | 36±6 | 64±5 | 44±7 | 56±7 | 78±18 | 22±18 | 86±25 | 14±25 |
| | $n^*_p$ | $n^*_{Fe}$ | $n^*_p$ | $n^*_{Fe}$ | $n^*_p$ | $n^*_{Fe}$ | $n^*_p$ | $n^*_{Fe}$ |
| 1 | -14±3 | 114±3 | -10±3 | 110±3 | -15±7 | 115±7 | -12±9 | 112±9 |
| 2 | -18±3 | 118±3 | -7±3 | 107±3 | -8±4 | 108±4 | -18±17 | 118±17 |
| 3 | -14±3 | 114±3 | -10±3 | 110±3 | -5±3 | 105±3 | -13±2 | 113±2 |
| 4 | -13±3 | 112±3 | -10±4 | 110±4 | -12±5 | 112±5 | -6±6 | 106±6 |
| 5 | -20±7 | 120±7 | -7±5 | 107±5 | -21±18 | 121±18 | -21±18 | 121±18 |
| | $n^*_p$ | $n^*_{Fe}$ | $n^*_p$ | $n^*_{Fe}$ | $n^*_p$ | $n^*_{Fe}$ | $n^*_p$ | $n^*_{Fe}$ |
| 1 | 14±3 | 86±3 | 9±4 | 91±4 | 37±10 | 63±10 | 30±13 | 70±13 |
| 2 | 17±3 | 83±3 | 24±4 | 76±4 | -1±5 | 101±5 | 27±16 | 72±16 |
| 3 | 20±3 | 80±3 | 24±4 | 76±4 | 17±5 | 83±5 | 33±7 | 67±7 |
| 4 | 13±3 | 87±3 | 9±5 | 91±5 | 38±8 | 62±8 | 15±10 | 85±10 |
| 5 | 15±6 | 85±6 | 28±6 | 72±6 | 27±16 | 73±16 | 15±32 | 84±32 |

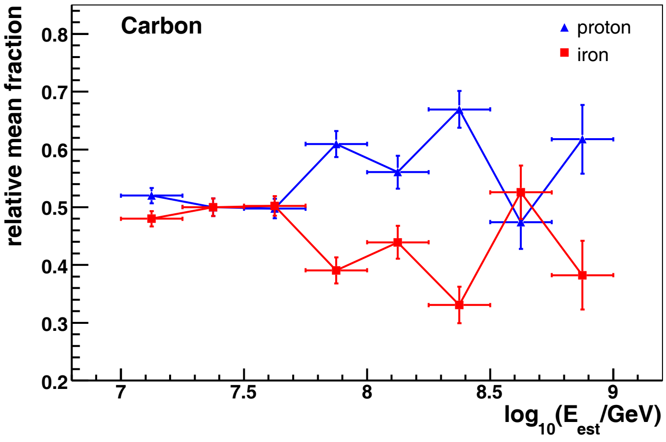
**Table 6.9:** Relative number of p and Fe showers (in percentage) as classified by the kNN method. The test sets are 100%proton (top), 100% iron (centre) and 50% iron+50% proton (bottom) induced showers, simulated with EPOS hadronic interaction model instead of QGSJetII. The analysis is carried out on five zenith angle ranges and four energy bins in the range $7.0 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 9.0$. For the zenith angular ranges, see chapter 4.

The present classification is completely dependent on the hadronic interaction models used in the shower simulations. As seen in the case of the EPOS1.6 test, different hadronic interaction models lead to a different correlation of the total number of charged particles and muons in the shower. The systematic uncertainties associated with the hadronic models have not been estimated yet. The test case performed with EPOS1.6 as hadronic interaction model, shows biases up to 50% in the classification of a known composition.

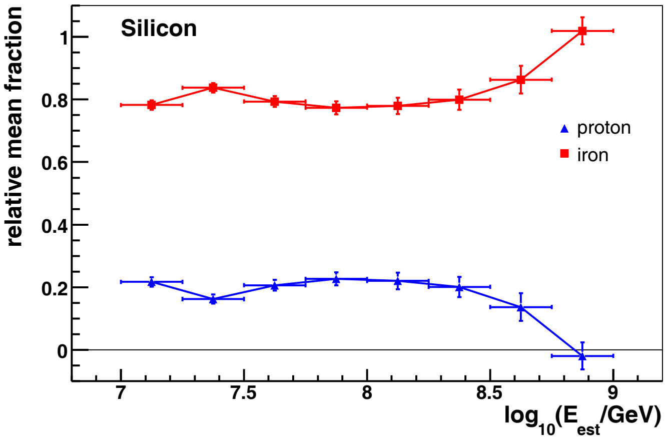
Also, the energy estimation applied in this work is not completely independent from the classification itself, both being based on nearly the same observables. If the uncertainties in the assignment of the energy are dependent on the primary mass, a systematic uncertainty in the fraction of proton/iron like air showers is expected. However, only a small difference is seen in the mean energy assignment (see fig. A.1), with the iron primary mean energy having a bias of $\approx -2\%$, while for the proton primary it is of $\approx -6\%$. This difference translates, on a logaritmic scale, into a maximum shift of $\log_{10}(\Delta E) \approx -0.02$. Being the minimum energy bin width used in the classification analysis of $\log_{10}(\Delta E) \approx 0.25$, a systematic shift of $\log_{10}(\Delta E) \approx 0.02$ will not affect the classification.



(a)



(b)



(c)

**Figure 6.22:** Classification into proton and iron, for helium (a), carbon (b) and silicon (c) induced air showers. The average frequencies for $\theta \leq 35°$ are shown. Applying the misclassification procedure, undershoots and overshoots can be introduced in the result, as observed for the silicon case at $E = 10^{18}$ eV, where a negative percentage is given.

# 7. Summary

In this work different aspects of the measurement and reconstruction of extensive air showers, generated by cosmic rays in the energy range between $10^{16}$ eV and $10^{18}$ eV, are investigated.

KASCADE-Grande detects at ground level the charged particles of extensive air showers. From the energy deposits and the arrival times of the particles in the detector stations, the main parameters of the extensive air showers are reconstructed: the impact point, the direction of the shower axis, the total number of electrons ($N_e$), and the total number of muons ($N_\mu$) of the shower at observation level.

The reconstruction procedure is carefully checked. By studying full simulated air showers, the threshold for full detection efficiency is determined at $N_e = 6 \cdot 10^5$. With the current reconstruction, a set of cuts is defined in order to select properly reconstructed events. Estimators for $N_e$ and $N_\mu$, the impact point position and the arrival direction of the shower, valid for shower sizes above the full efficiency threshold, are determined, with the following accuracies:

- the impact point of the shower at ground level is determined with a precision of 6 m;
- the arrival direction is reconstructed with an accuracy better than 0.4°;
- the reconstruction of the total electron number shows a bias of less than +5% up to $N_e = 3.2 \cdot 10^6$, for higher energies an increase to up to +15% is observed. The bias shows a dependency on the zenith angle, evolving from $\approx +6\%$ to $\approx -2\%$ in the case of iron primaries, whereas it is almost constant for proton induced air showers. The statistical uncertainty is in the order of $+24\%/-20\%$ for proton primaries and $+20\%/-18\%$ for iron induced air showers;
- the total muon number reconstruction shows a bias of +10% at the threshold ($N_\mu \approx 10^5$), fluctuating around 0% at the highest energies. A statistical uncertainty of 25% is observed, decreasing to 10% at higher energies.

With the KASCADE-Grande data taken from March 2004 until February 2009, the $N_e$ spectra for five different zenith angular ranges up to 40°, are constructed. Due to the propagation in the atmosphere, the observed flux is more attenuated for more inclined showers. In all zenith angular ranges structures are observed in the spectra, which are found to be consistent with the detector resolution and the reconstruction uncertainties.

In order to estimate the true flux, a forward folding procedure is separately applied in each zenith range. A simple power law $\mathrm{d}I/\mathrm{d}N_e \propto N_e^{\gamma}$ is folded with the reconstruction responses, as obtained with air shower and detector simulations, based on a mixed composition (20% for each p, He, C, Si and Fe). The spectral indices $\gamma$ are obtained through $\chi^2$-fits of the folding results to the reconstructed spectra. A spectral index $\gamma = -2.76 \pm 0.01$ is obtained for showers up to 25°, whereas for more inclined showers a $\gamma = -2.80 \pm 0.02$ is calculated. The true size spectrum, obtained after the folding procedure, is along the whole energy range structureless. However, the absence of features in the size spectrum does not necessarily imply that the corresponding all-particle energy spectrum is featureless as well, since at the same energy a higher $N_e$ is observed for proton than for iron induced showers, the electron size spectrum being thus dominated by the light component.

In order to estimate the relative number of proton and iron induced air showers in the measured data set, a *k Nearest Neighbours* (kNN) classification procedure is developed. A measured air shower is classified as proton-like or iron-like by finding its nine nearest neighbours in the parameter space of selected mass sensitive observables. The parameter space is populated by a set of simulated proton and iron induced air showers and an air shower is assigned to the class most common among its neighbours. The total number of charged particles ($N_{ch}$) and the total number of muons ($N_{\mu}$) are found to be suitable observables for this purpose. Other shower parameters have been analysed, such as the radius of curvature of the shower front and the shower age. Although they show a sensitivity to the primary mass, inconsistencies between the simulated and measured distributions of these parameters have been observed. It is found that simulations, based on QGSJetII-2, cannot correctly describe, in all its aspects, the development of an extensive air shower. Underestimation of the total muon number and a too high pion cross-section, together with a high value of the multiplicity, can lead to the observed discrepancies.

On the basis of the correlation between the total number of muons in the showers and the total number of charged particles, a composition estimation is carried out. The misclassification errors, as obtained in the training phase for each zenith angular range and energy bin, are taken into account.

An increase of the relative number of iron-like induced air showers is observed between $10^{16}$ eV and $3.2 \cdot 10^{16}$ eV, evolving from (57±1)% to (69±1)%. Detection efficiency problems are excluded, since the measurements are performed already above the full efficiency threshold. Above $3.2 \cdot 10^{16}$ eV, about 70% of the events result to be iron-like and $\approx 30\%$ proton-like. Above $E_{est} \approx 10^{18}$ eV, an increase of the frequencies in the proton class to 50% is observed. However, with less than six air showers per zenith angle range above an energy of $\approx 10^{18}$ eV, the measurement is still consistent within 1.5 standard deviations with a constant composition of 69% iron-like showers.

A classification into three primary mass groups has been tested. Due to shower-to-shower fluctuations and the measured reconstruction uncertainties, on the basis of the relation $N_{ch} - N_{\mu}$, the presented method is not sensitive enough for a classification into three groups.

The results depend on the hadronic interaction model used to create the simulated training sets. Different models, with different predictions for the correlation of the number of charged particles and the number of muons, will lead to a different classification. However, as the available models do not differ drastically in the ratio of electron to muon numbers, no strong change is expected in the relative composition.

The results presented in this work constitute a first attempt to carry out composition analyses with KASCADE-Grande data. The stable composition observed up to an energy of $10^{18}$ eV implies that the transition between galactic to extragalactic cosmic rays is not happening below $10^{18}$ eV [3; 66; 67]. A mixed composition is observed over the whole energy range, with an increase of the heavy component fraction from (57±1)% up to (69±1)% below $3.2 \cdot 10^{16}$ eV, which could be associated with the decrease of the light component (He, CNO group) due to the break in its energy spectrum.

# A. Energy estimation

The estimation of the energy on a event by event basis is performed in the following way: the main idea [68] is to use the reconstructed electron number N$_e$ as energy estimator, whereas the ratio N$_e$/N$_\mu$ is used to take into account shower-to-shower fluctuations and to average over different primary masses.

The reconstructed N$_e$ for a shower reconstructed at zenith angle $\theta$ is converted into the equivalent shower size N$_e^0$ for a vertical shower according to the formula:

$$N_e(0) = N_e(\theta) \cdot 10^{\frac{(sec\theta - 1)\cdot x_0}{\Lambda_e \cdot \log_{10} e}}, \tag{A.1}$$

with x$_0$ $= 800\, g/cm^2$ and $\Lambda_e = 219\, g/cm^2$ [69]. Taking the logarithm of equation A.1 and substituting the numerical values for $x_0$ and $\Lambda_e$, the following conversion is obtained:

$$\log_{10} N_e^0 = \log_{10}(N_e(\theta)) + 1.57 \cdot (sec\theta - 1). \tag{A.2}$$

In a second step, from simulation studies the relations $N_e/N_\mu$ and $E(N_e)$ for the two extreme cases of proton and iron primary are parametrised according to:

$$\log_{10}\left(\frac{N_e^0}{N_\mu}\right)_{p,Fe} = c_{p,Fe} \cdot \log_{10}(N_e^0) + d_{p,Fe} \tag{A.3}$$

$$\log_{10}\left(E_{est}/\text{GeV}\right)_{p,Fe} = a_{p,Fe} \cdot \log_{10}(N_e^0) + b_{p,Fe} \tag{A.4}$$

In order to estimate the energy of each single event on an event-by-event basis, an expression $E(N_e, N_\mu)$ which takes into account the different energy dependence of the electron and muon sizes of each element has to be taken into account. For this reason a new parameter $r$ is defined through the following expression:

$$r = \frac{\log_{10}(N_e/N_\mu) - \log_{10}\left(N_e/N_\mu\right)_p}{\log_{10}\left(N_e/N_\mu\right)_{Fe} - \log_{10}\left(N_e/N_\mu\right)_p}, \tag{A.5}$$

where $\log_{10}\left(N_e/N_\mu\right)_{p,Fe}$ are obtained from equation A.3.

The expression of the energy dependence on the electron size is finally parametrised as it follows:

$$\log_{10}(E_{est}/\text{GeV}) = (a_p + \Delta a \cdot r) \cdot \log_{10}(N_e) + (b_p + \Delta b \cdot r), \tag{A.6}$$

| $\theta$(deg) | c$_p$ | d$_p$ | c$_{Fe}$ | d$_{Fe}$ | a$_p$ | b$_p$ | $\Delta a$ | $\Delta b$ |
|---|---|---|---|---|---|---|---|---|
| 0-17.75 | 0.204 | -0.086 | 0.206 | -0.344 | 0.943 | 0.981 | -0.035 | 0.561 |
| 17.75-24.62 | 0.182 | 0.240 | 0.166 | -0.065 | 0.976 | 0.795 | -0.023 | 0.468 |
| 24.62-29.52 | 0.164 | 0.372 | 0.178 | -0.164 | 0.951 | 0.951 | -0.004 | 0.404 |
| 29.52-39.72 | 0.106 | 0.766 | 0.207 | -0.347 | 1.051 | 0.393 | -0.137 | 1.251 |

**Table A.1:** Parameter of the energy estimation formula

with $\Delta a = a_{Fe} - a_p$ and $\Delta b = b_{Fe} - b_p$.

All the parameters are derived considering showers with $\log_{10}(N_e^0) > 5.5$, corresponding to a threshold energy for iron of $5 \cdot 10^5$ eV, and for four different zenith angle ranges.

In table A.1 the numerical values for $c_{p,Fe}, d_{p,Fe}, a_p, b_p, \Delta a$ and $\Delta b$ are reported.

The distribution of the relative uncertainties on the estimated energy is shown in fig. A.1. Simulated data have been divided in bins of true energy and the distributions of the relative differences between reconstructed and true energy have been calculated. The case for proton and iron primaries are shown as continuous and dashed lines, respectively. For proton primaries, the energy resolution (represented by the RMS of such distributions) is about 30% at threshold and decreases to less than 20% at the highest energy. For iron induced showers the resolution is about 10% better as compared to proton primaries. A small offset in the mean values of the distributions at low energies is necessary to take into account the effect of shower fluctuations on a steep spectrum.

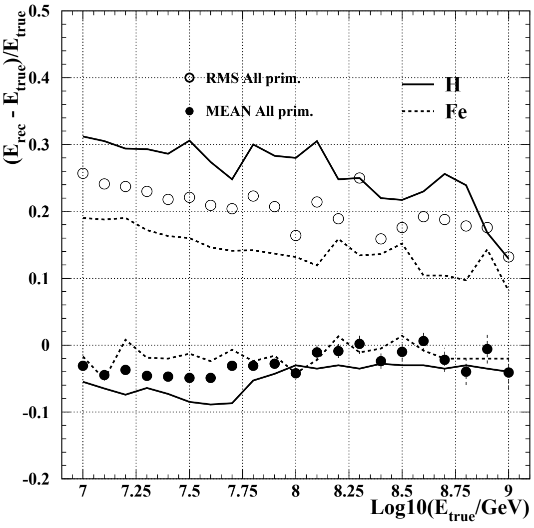


**Figure A.1:** Resolution in the energy assignment for a mixture of primaries of the 5 simulated mass groups (relative abundance of each group 20%), for H and Fe. The full dots show the offset of the reconstructed energy $E_{rec}$ in bins of true energy $E_{true}$. The open dots show the RMS of such distributions.

# B. Geometry of the shower front reconstruction with a sphere

In the following the equations for the reconstruction of the shower front, assuming a spherical description, are discussed. Let us consider figure B.1.

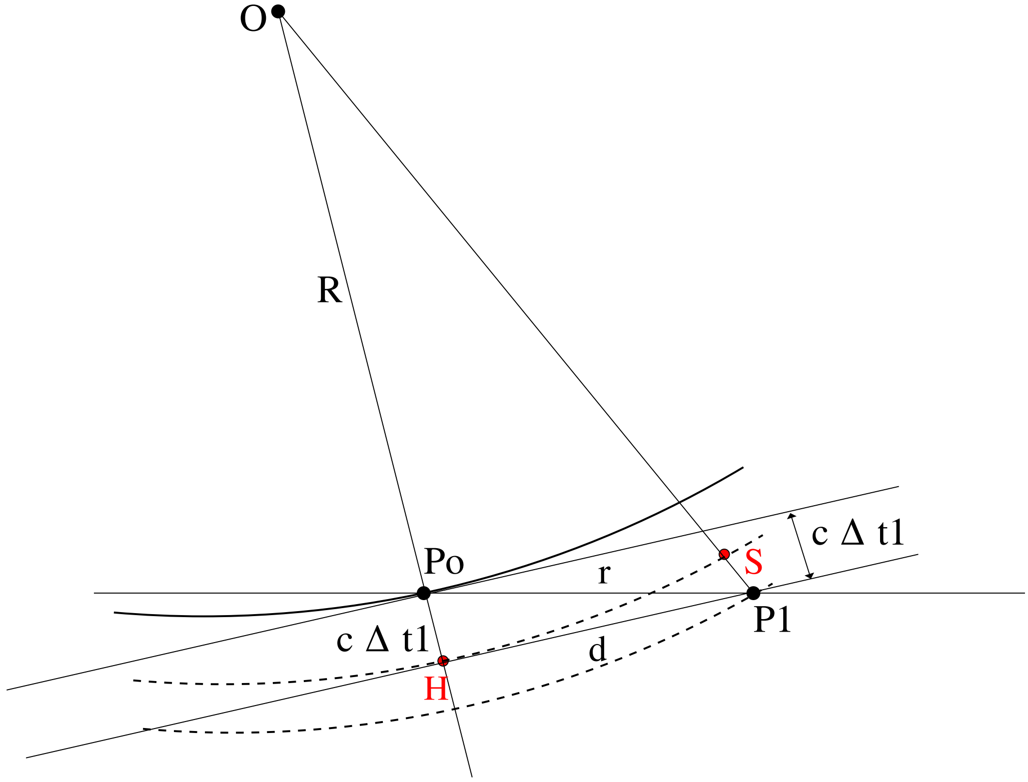


**Figure B.1:** Geometry of the shower front reconstruction by means of a sphere.

Let $P_0$ be the shower core and let us consider the station $P_1$, with $\overline{P_1P_0} = r$. The time difference $\Delta t_1$ taken by the particles to reach $P_0$ and $P_1$ as first estimation can be equal to the distance between $P_1$ and the plane tangential to the spherical front in $P_0$, with cosine directors $l$, $m$, $n$:

$$c \cdot \Delta t_1 = l\,(x_1 - x_0) + m\,(y_1 - y_0) + n\,(z_1 - z_0)$$

Assuming a spherical shape of the shower front, with curvature radius R, the time delay between $P_1$ with respect to $P_0$ is given by the sum of $\Delta t_1$ and the time taken to cover the distance $\overline{P_1 S}$. With few geometrical consideration we obtain:

$$\begin{aligned} d^2 &= |\overrightarrow{P_1 H}|^2 = r^2 - c^2 \cdot \Delta {t_1}^2 \\ |\overrightarrow{OP_1}|^2 &= d^2 + (R + c \cdot \Delta t_1)^2 \\ |\overrightarrow{P_1 S}| &= |\overrightarrow{OP_1}| - |\overrightarrow{OS}| = \sqrt{d^2 + (R + c \cdot \Delta t_1)^2} - (R + c \cdot \Delta t_1) \\ &\approx \frac{d^2}{2 \cdot (R + c \cdot \Delta t_1)} \end{aligned}$$

The expected delay between $P_1$ and $P_0$ is

$$t_1 - t_0 = \Delta t_1 + \frac{d^2}{2 \cdot c \cdot (R + c \cdot \Delta t_1)}.$$

A $\chi^2$ can be built as the difference between the observed arrival time of the particles in the detectors and the expected one and it can be minimised with respect to $l$, $m$, $n$, $R$.

Considering all the detectors involved in a EAS, the $\chi^2$ will be the sum of these differences over all $N$ detector stations considered:

$$\chi^2 = \sum_{i=1}^{N} \left( \frac{t_i - \left[ t_0 + \frac{l\,(x_i - x_0) + m\,(y_i - y_0) + n\,(z_i - z_0)}{c} + \frac{{d_i}^2}{2\,c\,(R + c \cdot \Delta t_i)} \right]}{\sigma_{ti}} \right)^2 .$$

From the $\chi^2$ fit the values of $R$, $l$, $m$, $n$ are thus obtained as well as the time $t_0$ at which the particles have reached the centre of the shower at ground.

# C. kNN

## kNN classification

Misclassification rates in the energy range $1.8 \cdot 10^{16}$ eV $< E < 1.2 \cdot 10^{18}$ eV for the kNN analysis are compiled in table C.1. In table 6.7, the relative frequencies of proton-like and iron-like air showers for KASCADE-Grande data are compiled. The fractions for each zenith angular and energy bin are shown in fig. C.1.

| | $\log_{10}(E_{est}/\mathrm{GeV})$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 7.25-7.75 | | 7.75-8.25 | | 8.25-8.75 | | 8.75-9.25 | |
| $\theta$ (deg) | p→p | p→Fe | p→p | p→Fe | p→p | p→Fe | p→p | p→Fe |
| 0-17.6 | 90±1 | 10±1 | 95±2 | 5±2 | 99±2 | 1±2 | 100±1 | 0±1 |
| 17.6-24.99 | 90±2 | 10±2 | 96±2 | 4±2 | 96±2 | 4±2 | 94±6 | 6±6 |
| 24.99-30.73 | 85±2 | 15±2 | 91±3 | 9±3 | 98±2 | 2±2 | 100±5 | 0±5 |
| 30.73-35.63 | 83±2 | 17±2 | 88±3 | 12±3 | 94±5 | 6±5 | 84±12 | 16±12 |
| 35.63-40 | 78±3 | 22±3 | 82±6 | 18±6 | 84±7 | 16±7 | 97±3 | 3±3 |
| | Fe→p | Fe→Fe | Fe→p | Fe→Fe | Fe→p | Fe→Fe | Fe→p | Fe→Fe |
| 0-17.6 | 10±2 | 90±2 | 10±3 | 90±3 | 13±6 | 87±6 | 5±7 | 95±7 |
| 17.6-24.99 | 14±2 | 86±2 | 7±2 | 93±2 | 6±4 | 94±4 | 43±25 | 57±25 |
| 24.99-30.73 | 12±2 | 88±2 | 8±3 | 92±3 | 3±3 | 97±3 | 8±6 | 92±6 |
| 30.73-35.63 | 14±2 | 86±2 | 10±4 | 90±4 | 8±4 | 92±4 | 10±9 | 90±9 |
| 35.63-40 | 18±3 | 82±3 | 9±5 | 91±5 | 7±7 | 93±7 | 35±39 | 65±39 |

**Table C.1:** Hit-rate (p→p, fe→fe) and misclassification rate (p→fe, fe→p) in percentage, for proton and iron induced showers. The probabilities for five zenith angle ranges and four energy bins in the range $7.25 < \log_{10}(E_{est}/\mathrm{GeV}) \leq 9.25$.

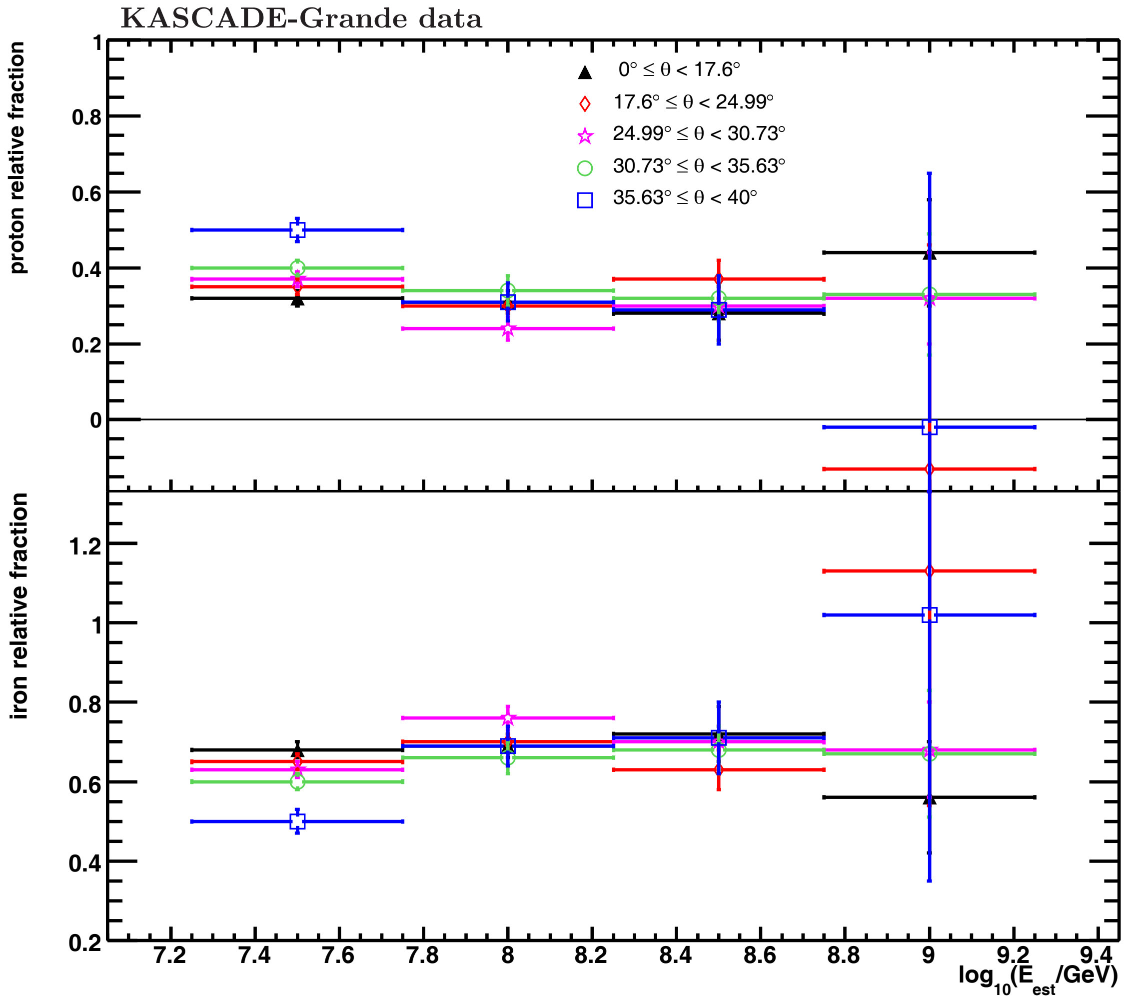


**Figure C.1:** Relative frequency (in percentage) of proton and iron induced showers by kNN method. Markers and colours represent the five zenith angle ranges considered in this work: $0° \geq \theta < 16.6°$ (black triangles), $16.6° \leq \theta < 24.99°$ (red diamonds), $24.99° \leq \theta < 30.73°$ (blue stars), $30.73° \leq \theta < 35.63°$ (green circles), $35.63° \leq \theta < 40°$ (pink squares) .

| $\theta$ (deg) | $\log_{10}(E_{est}/\mathrm{GeV})$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 7.0-7.5 | | | 7.5-8.0 | | | 8.0-8.5 | | | 8.5-9.0 | | |
| | p→p | p→C | p→Fe | p→p | p→C | p→Fe | p→p | p→C | p→Fe | p→p | p→C | p→Fe |
| 0-17.6 | 71±2 | 21±2 | 8±1 | 78±3 | 19±3 | 3±1 | 89±4 | 11±4 | 0±0.6 | 82±9 | 18±9 | 0±1 |
| 17.6-24.99 | 72±2 | 22±2 | 6±1 | 77±3 | 19±3 | 4±1 | 78±6 | 22±6 | 0±1 | 83±8 | 17±8 | 0±2 |
| 24.99-30.73 | 70±2 | 20±2 | 10±1 | 77±3 | 18±3 | 5±1 | 86±4 | 13±4 | 1±1 | 81±10 | 19±11 | 0±1 |
| 30.73-35.63 | 71±2 | 17±2 | 12±1 | 73±4 | 19±3 | 8±3 | 70±7 | 27±7 | 3±3 | 82±9 | 15±8 | 3±4 |
| 35.63-40 | 66±2 | 21±2 | 13±2 | 69±5 | 13±4 | 18±5 | 61±11 | 15±8 | 24±12 | 66±11 | 28±11 | 6±7 |
| | Fe→p | Fe→C | Fe→Fe | Fe→p | Fe→C | Fe→Fe | Fe→p | Fe→C | Fe→Fe | Fe→p | Fe→C | Fe→Fe |
| 0-17.6 | 6 ±1 | 24±2 | 70±2 | 3±1 | 23±4 | 74±3 | 3±2 | 31±8 | 66±8 | 3±4 | 19±10 | 79±10 |
| 17.6-24.99 | 8 ±1 | 31±3 | 61±2 | 2±1 | 24±4 | 74±4 | 1±2 | 18±5 | 81±5 | 2±5 | 46±12 | 52±11 |
| 24.99-30.73 | 7 ±1 | 29±2 | 64±3 | 3±2 | 23±4 | 74±3 | 1±1 | 19±5 | 80±5 | 0±1 | 15±11 | 85±11 |
| 30.73-35.63 | 10 ±2 | 23±2 | 67±2 | 5±2 | 24±4 | 71±4 | 3±2 | 17±6 | 80±6 | 1±3 | 24±9 | 75±9 |
| 35.63-40 | 16±2 | 36±3 | 48±3 | 6±3 | 23±5 | 71±5 | 9±9 | 10±7 | 81±10 | 14±11 | 9±12 | 77±12 |
| | C→p | C→C | C→Fe | C→p | C→C | C→Fe | C→p | C→C | C→Fe | C→p | C→C | C→Fe |
| 0-17.6 | 33±2 | 36±2 | 31±2 | 28±4 | 46±4 | 26±3 | 21±6 | 65±7 | 14±5 | 29±14 | 52±15 | 19±11 |
| 17.6-24.99 | 33±2 | 36±2 | 31±2 | 29±4 | 43±4 | 28±4 | 22±7 | 63±8 | 15±5 | 18±7 | 78±10 | 4±6 |
| 24.99-30.73 | 29±2 | 36±2 | 35±2 | 27±3 | 39±4 | 34±4 | 12±5 | 61±8 | 27±7 | 12±7 | 73±13 | 15±8 |
| 30.73-35.63 | 29±2 | 32±2 | 39±2 | 29±3 | 42±3 | 29±3 | 22±8 | 51±8 | 27±10 | 26±12 | 52±13 | 22±12 |
| 35.63-40 | 31±2 | 32±3 | 37±2 | 24±4 | 36±5 | 40±6 | 38±8 | 38±8 | 24±6 | 23±15 | 54±15 | 23±12 |

**Table C.2:** Hit-rate (p→p, Fe→Fe, C→C) and misclassification rates for proton (top), iron (middle) and carbon (bottom) induced showers, for five zenith angle ranges and four energy ranges. The rates are expressed in percentage.

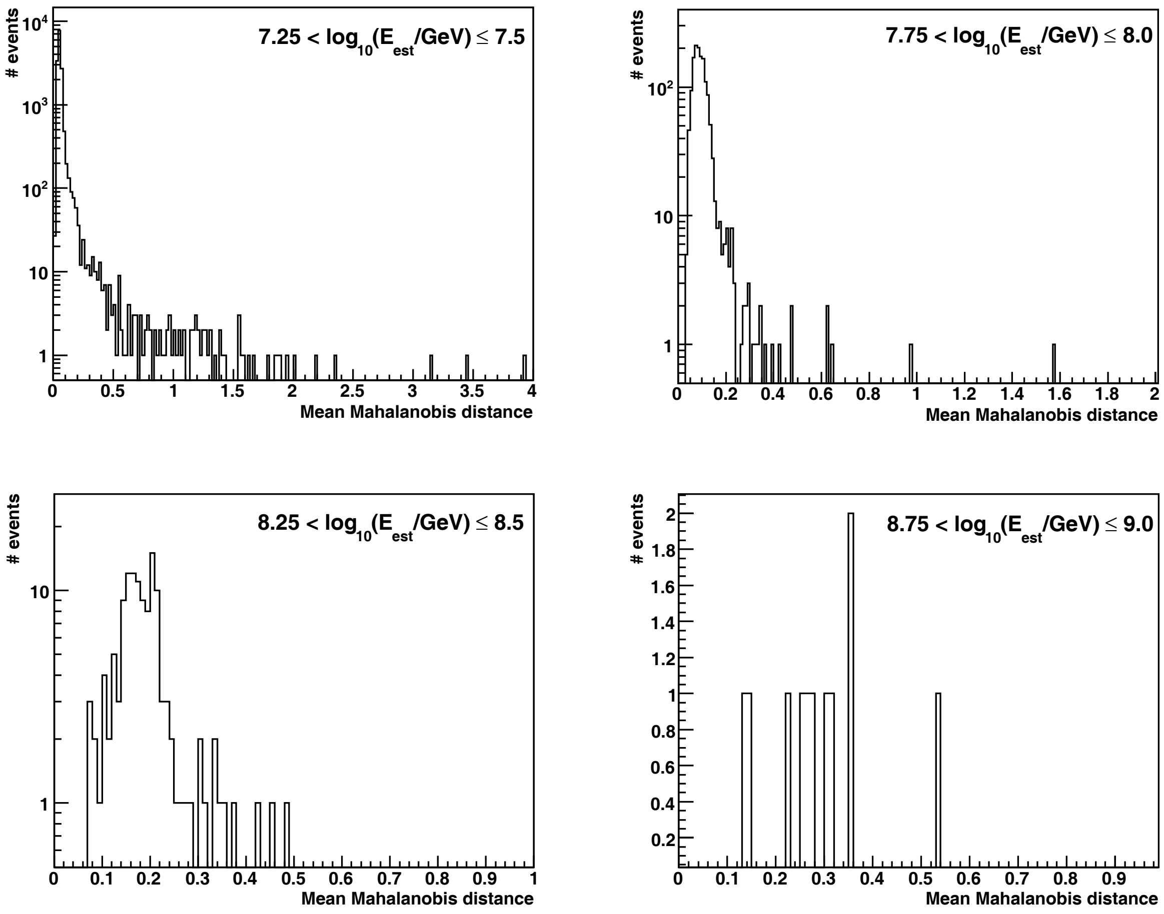


**Figure C.2:** Mahalanobis distances for measured air showers, in four estimated energy ranges. On the abscissa is the average Mahalanobis distance of the query points to the nine nearest neighbours. Due to the lowest density of showers in the training set at higher energies, a small increase of the mean distance is observed with increasing energy.

# List of Figures

# List of Tables

# Acronyms

| | |
|---|---|
| **AGN** | Active Galactic Nuclei |
| **PAO** | Pierre Auger Observatory |
| **GZK** | Greisen-Zatsepin-Kuzmin |
| **SNR** | supernova remnants |
| **KRETA** | Kascade Reconstruction for ExTensive Airshowers |
| **CORSIKA** | COsmic Ray SImulation for KAscade |
| **CRES** | Cosmic Ray Event Simulation |
| **GEANT3** | GEometry ANd Tracking 3 |
| **DAQ** | Data AcQuisition station |
| **NKG** | Nishimura Kamata Greisen |
| **EAS** | Extensive Air Shower |
| **kNN** | k Nearest Neighbours |
| **LOO** | Leave One Out for a time method |
| **LECF** | Lateral Energy Correction Function |

# Bibliography

[1] T. Antoni et al. KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems. *Astropart. Phys.*, 24:1–25, 2005.

[2] G. Navarra et al. Study of Cosmic Ray Primaries between $10^{12}$ and $10^{16}$ eV from EAS-TOP. *Proc. 28th ICRC, Tsukuba*, 1, 2003.

[3] A. M. Hillas. Can diffusive shock acceleration in supernova remnants account for high-energy galactic cosmic rays? *J. Phys.*, G31:R95–R131, 2005.

[4] Julian Candia, Silvia Mollerach, and Esteban Roulet. Cosmic ray drift, the second knee and galactic anisotropies. *JHEP*, 12:032, 2002.

[5] V. S. Ptuskin, S. I. Rogovaya, L. G. Zirakashvili, V. N. an d Chuvilgin, G. B. Khristiansen, E. G. Klepach, and G. V. Kulikov. Diffusion and drift of very high energy cosmic rays in galactic ma gnetic fields. *Astron. Astrophys.*, 268:726–735, February 1993.

[6] T. Antoni et al. The Cosmic ray experiment KASCADE. *Nucl. Instrum. Meth.*, A513:490–510, 2003.

[7] V. Hess. Über Beobachtungen der durchdringenden Strahlung bei siebe n Freiballonfahrten. *Physikalische Zeitschrift*, 13:1084–1091, 1912.

[8] P. Auger, P. Ehrenfest, R. Maze, and R. A. Daudin, J. an d Fréon. Extensive Cosmic-Ray Showers. *Reviews of Modern Physics*, 11:288–291, July 1939.

[9] V. P. Egorova et al. The spectrum features of UHECRs below and surrounding GZK. *Nucl. Phys. Proc. Suppl.*, 136:3–11, 2004.

[10] T. Abu-Zayyad et al. Measurement of the cosmic ray energy spectrum and composition from $10^{17}$ eV to $10^{18.3}$ eV using a hybrid fluorescence technique. *Astrophys. J.*, 557:686–699, 2001.

[11] D. J. Bird et al. The Cosmic ray energy spectrum observed by the Fly's Eye. *Astrophys. J.*, 424:491–502, 1994.

[12] J. Abraham et al. Properties and performance of the prototype instrument for the Pierre Auger Observatory. *Nucl. Instrum. Meth.*, A523:50–95, 2004.

[13] G.T. Zatsepin and V.A. Kuzmin. Upper limit of the spectrum of cosmic rays. *JETP Lett.*, 4:78–80, 1966.

[14] K. Greisen. End to the cosmic ray spectrum? *Phys. Rev. Lett.*, 16:748–750, 1966.

[15] C. Amsler and others(Particle Data Group). Review of Particle Physics. *Physics Letters*, B667(1), 2008.

[16] J. A. Simpson. Elemental and isotopic composition of the galactic cosmic rays. *Ann. Rev. Nucl. Part. Sci.*, 33:323–382, 1983.

[17] M. E. Wiedenbeck et al. Refractory Nuclides in the Cosmic-Ray Source. *Proc. 28th ICRC*, 2003.

[18] N. E. Yanasak et al. Measurement of the Secondary Radionuclides $^{10}$Be, $^{16}$Al, $^{36}$Cl, $^{54}$Mn, and $^{14}$C and Implications for the Galactic Cosmic-Ray Age. *Astrophys. J.*, 563:768–792, 2001.

[19] D. J. Bird et al. Evidence for correlated changes in the spectrum and composition of cosmic rays at extremely high-energies. *Phys. Rev. Lett.*, 71:3401–3404, 1993.

[20] S. P. Knurenko, A. A. Ivanov, M. I. Pravdin, Artem V. Sabourov, and I. Ye. Sleptsov. Recent results from Yakutsk experiment: Development of EAS, energy spectrum and primary particle mass composition in the energy region of $10^{15}\,\mathrm{eV} - 10^{19}\,\mathrm{eV}$. *Nucl. Phys. Proc. Suppl.*, 175-176:201–206, 2008.

[21] Baade, W. and Zwicky, F. . Remarks on Super-Novae and Cosmic Rays. *Phys. Rev.*, 46(1):76–77, Jul 1934.

[22] Fermi, E. . On the Origin of the Cosmic Radiation. *Phys. Rev.*, 75(8):1169–1174, Apr 1949.

[23] F. A. Aharonian et al. High-energy particle acceleration in the shell of a supernova remnant. *Nature.*, 432:75–77, 2004.

[24] C. J. Cesarsky and T. Montmerle. Gamma rays from active regions in the galaxy: The possible contribution of stellar winds. *Space Sci. Revs.*, 36:173–193, 1983.

[25] J. R. Jokipii and G. Morfill. Ultra-high-energy cosmic rays in a galactic wind and its terminati on shock. *Astrophys. J.*, 312:170–177, 1987.

[26] A.K. Harding and T.K. Gaisser. Acceleration by pulsar winds in binary systems. *Astrophys. J.*, 358:561–574, 1990.

[27] Protheroe, R. J. and Szabo, A. P. High energy cosmic rays from active galactic nuclei. *Phys. Rev. Lett.*, 69(20):2885–2888, Nov 1992.

[28] M. Aglietta et al. UHE cosmic ray event reconstruction by the electromagnetic detector of EAS-TOP. *Nucl. Instrum. Meth.*, A336:310–321, 1993.

[29] D. Kazanas and A. Nicolaidis. Cosmic rays and large extra dimensions. *Gen. Rel. Grav.*, 35:1117–1123, 2003.

[30] W. Heitler. *The quantum theory of radiation.* Oxford, 1954.

[31] K. Greisen. The Extensive Air Showers. *Progess in Elementary Particle and Cosmic Ray Physics*, 3, 1956.

[32] J. van Buren. *Investigations of the Muon Component of Extensive Air Showers measured by KASCADE-Grande.* PhD thesis, Universität Karlsruhe, 2006.

[33] K. Kamata and J. Nishimura. The Lateral and the Angular Structure Functions of Electron Sh owers. *Prog. Theoret. Phys. Suppl*, 6, 1958.

[34] P. Doll et al. Muon tracking detector for the air shower experiment KASCADE. *Nucl. Instrum. Meth.*, A488:517–535, 2002.

[35] A. Haungs et al. The KASCADE-Grande Experiment. *Proc. 28th ICRC, Tsukuba*, 2:985–988, 2003.

[36] Stümpert, M. *Suche nach Anisotropie in der kosmischen Strahlung mit KASCADE-Grande.* PhD thesis, Universität Karlsruhe, 2007.

[37] F.Di Pierro. *Measurement of particle energy above $10^{16}$ eV: technique and uncertainties of the cosmic ray experiment KASCADE-Grande.* PhD thesis, Universitá di Torino, 2007.

[38] S. Over. Development and Commissioning of the Data Acquisition Systems for the KASCADE-Grande Experiment. Master's thesis, Universität Siegen, 2004.

[39] F. Cossavella. Misura delle direzioni di arrivo dei raggi cosmici di alta energia in KASCADE-Grande. Master's thesis, Universitá di Torino, 2005.

[40] D. Apel et al. Comparison of measured and simulated lateral distributions for electrons and muons with KASCADE . *Astropart. Phys.*, 24:467–483, 2006.

[41] A. Chiavassa. Private communication.

[42] R. Glasstetter et al. Shower Size Reconstruction at KASCADE-Grande. *Proc. 29th ICRC*, 2005.

[43] H. J. Mayer. Private communication.

[44] R. Glasstetter et al. Shower Reconstruction Performance of KASCADE-Grande. *Proc. 28th ICRC*, 2003.

[45] A. A. Lagutin and R. I. Raikin. Lateral distribution of electrons in EAS at superhigh energies: predictions and experimental data. *Nucl. Phys. Proc. Suppl.*, 97:274–277, 2001.

[46] D. Heck, G. Schatz, T. Thouw, J. Knapp, and J. N. Capdevielle. CORSIKA: A Monte Carlo code to simulate extensive air showers. 1998. FZKA-6019.

[47] W. Ralph Nelson, H. Hirayama, and David W. O. Rogers. THE EGS4 CODE SYSTEM. 1985. SLAC-0265.

[48] N. N. Kalmykov, S. S. Ostapchenko, and A. I. Pavlov. Quark-gluon string model and EAS simulation problems at ultra-high energies. *Nucl. Phys. Proc. Suppl.*, 52B:17–28, 1997.

[49] S. S. Ostapchenko. QGSJET-II: towards reliable description of very high energy hadronic interactions. *Nucl. Phys. Proc. Suppl.*, 151:143–146, 2006.

[50] A. Fasso, A. Ferrari, P. R. Sala, and J. Ranft. FLUKA: Status and prospects for hadronic applications. Prepared for International Conference on Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications (MC 2000), Lisbon, Portugal, 23-26 Oct 2000.

[51] GEANT Detector Desc. and Sim. Tool. CERN Program Library Long Writeup W5013, 1993.

[52] S. Agostinelli et al. GEANT4: A simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.

[53] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975.

[54] T. Bergmann et al. One-dimensional hybrid approach to extensive air shower simulation. *Astropart. Phys.*, 26:420–432, 2007.

[55] U. Dersch et al. Total cross section measurements with pi-, Sigma- and protons on nuclei and nucleons around 600 GeV/c. *Nucl. Phys.*, B579:277–312, 2000.

[56] T. Pierog. Review of hadronic interaction models. Corsika school 2008.

[57] T. Pierog. Private communication.

[58] A. A. et al Chilingarian. On the possibility of a multidimensional kinematic information analysis by means of nearest-neighbour estimations of dimensionality. *Nucl. Instrum. Meth.*, A 281:388, 1989.

[59] M. M. Kennedy Glasmacher. *Cosmic ray composition studies with CASA-MIA*. PhD thesis, University of Michigan, 1998.

[60] M. A. K. Glasmacher et al. The cosmic ray composition between $10^{14}$ eV and $10^{16}$ eV. *Astropart. Phys.*, 12:1–17, 1999.

[61] M. Roth. Analysis of air shower measurements of the KASCADE experiment by applying statistical classification methods for the determination of the energy and type of the primary cosmic radiation. (In German). FZKA-6262.

[62] Carl J. Huberty. *Applied Manova and Discriminant Analysis*. Wiley-Interscience, second edition, 2006.

[63] T. Pierog and K. Werner. Muon Production in Extended Air Shower Simulations. *Phys. Rev. Lett.*, 101:171101, 2008.

[64] T. Pierog and K. Werner. EPOS Model and Ultra High Energy Cosmic Rays. 2009.

[65] D. et al. Apel. A test of the hadronic interaction model EPOS with air shower data. *J. Phys. G.: Nucl. Part. Phys.*, 36:035201, 2009.

[66] Berezinsky, V.S. and Grigoreva, S.I. and Hnatyk, B.I. Extragalactic UHE proton spectrum and prediction of flux of iron-nuclei at $10^8$ GeV - $10^9$ GeV. *Nucl. Phys. (Proc. Suppl.)*, 151:497–500, 2006.

[67] Allard, D. and Olinto, A.V. and Parizot, E. Signatures of the extragalactic cosmic-ray source composition from spectrum and shower depth measurements. *submitted to Astron. Astrophys.*, 2007. astro-ph/0703633.

[68] M. Bertaina. Energy spectrum from $N_e - N_\mu$. Internal collaboration communication.

[69] M. Aglietta et al. The EAS size spectrum and the cosmic ray energy spectrum in the region $10^{15}$ eV - $10^{16}$ eV. *Astropart. Phys.*, 10:1–9, 1999.