Henryk Niewodniczański Institute of Nuclear Physics,
Polish Academy of Sciences,
Kraków, Poland

The top-down method of extensive air shower reconstruction: its feasibility and prospects for testing nuclear interaction models

Czesław Porowski

Thesis submitted for the Doctor degree prepared under the supervision of
prof. dr hab. Henryk Wilczyński
and auxiliary supervision of
dr Tanguy Pierog (Karlsruhe Institute of Technology)

Kraków 2018

Abstract

The subject of this thesis is to present extension of so called top-down reconstruction method, used for reconstruction of extensive air showers induced by the high energy cosmic rays. To reconstruct an extensive air shower the measurements of the longitudinal and lateral distributions of particles are used, as well as the simulations of shower development. Such simulations require models of hadronic interactions. Since the energies at which the models are used exceed the energy of accelerator data, the extrapolations of interaction models are used. Such extrapolations are not calibrated. A use of the extrapolations and fact that the results of the air shower analyses are not in agreement with other types of shower analyses, causes that the calibration tests of extrapolated interaction models are required. The top-down method may turned to be helpful in such calibrations. An improved version of the top-down method is presented in this thesis. Using generated Monte Carlo data the tests of intrinsic consistency of improved version of the top-down method are shown. It was shown that the improved version provides consistent results. Subsequently the method was applied to the real data. It was shown that the results are inconsistent and these inconsistencies are in agreement with the results obtained with the previous versions of the top-down method.

Streszczenie

Przedmiotem niniejszej pracy jest przedstawienie metody rekonstrukcji wielkich pęków atmosferycznych indukowanych przez promieniowanie kosmiczne najwyższych energii, tak zwanej metody top-down. Wielkie pęki atmosferyczne są rekonstruowane na podstawie pomiarów podłużnego i poprzecznego rozkładu cząstek w pęku, z wykorzystaniem symulacji rozwoju tych pęków. Do symulacji rozwoju pęków są stosowane modele oddziaływań jądrowych. Ponieważ są one stosowane w zakresach energii wykraczających poza energie dostępne w akceleratorach, wykorzystuje się ekstrapolacje własności modeli. Własności modeli po zastosowaniu ekstrapolacji nie są kalibrowane. Użycie w/w ekstrapolacji, jak i fakt, że wyniki analizy pęków zrekonstruowanych za pomocą tych modeli są niespójne z wynikami innych analiz, skutkuje potrzebą testowania modeli oddziaływań jądrowych. Uważa się, że metoda top-down wychodzi naprzeciw tym oczekiwaniom. W pracy tej zostały zaprezentowane własności ulepszonej metody top-down, tj. wyniki testów spójności używanej wersji metody top-down. Dane użyte do testów zostały wygenerowane z użyciem metody Monte Carlo. Zostało wykazane, że ulepszona wersja metody daje poprawne wyniki dla danych Monte Carlo. Następnie metoda ta została użyta do rekonstrukcji danych rzeczywistych. Pokazano, że wyniki rekonstrukcji top-down danych rzeczywistych są niekonsystentne, a uzyskane rozbieżności są analogiczne do rozbieżności uzyskanych w innych analizach z wykorzystaniem poprzednich wersji metody top-down.

Contents

1 Introduction

Cosmic rays, a stream of energetic, mostly charged particles incoming to the Earth from the space, were discovered more than 100 years ago by Victor Hess. He found that the level of air ionization increases with altitude, which was interpreted as a result of air bombarding by the cosmic rays. For this discovery, which led to formulation of cosmic rays physics, V. Hess was awarded the Nobel prize in 1936.

The cosmic ray particles (primary particles) interact with the air molecules, hence the primary particles do not reach the Earth's surface. The thickness of the atmosphere corresponds to ~ 11 hadronic interaction lengths of proton. So, for the primary cosmic rays, the probability of reaching the ground is near zero. In order to detect the primary cosmic rays directly, the detector equipment would have to be located outside the atmosphere (satellites) or in stratosphere (balloons).

Due to the interactions with the atmosphere, the cosmic rays create cascades of secondary particles, so-called extensive air showers. The showers, discovered by Pierre Auger in 1938, occur when the cosmic rays with energies above 10^{14} eV propagate through the Earth's atmosphere. Such energy of primary is sufficient to sustain the production of many generations of the secondary particles which propagate approximately along the direction of the primary. Some of the secondary particles may reach the ground and be detected. The charged particles may also excite the molecules of atmospheric nitrogen leading to fluorescence light emission. This light can also be used for detection of air showers using optical detectors.

The energy spectrum of primary cosmic rays extends over many orders of magnitude: from $\sim 10^7$ eV until at least $\sim 10^{20}$ eV. The flux of cosmic rays drops fast with energy. For example, the number of PeV cosmic rays passing through 1 m^2 surface is about one per year. At the most energetic part of the spectrum the number of incident particles drops below $1/km^2/century$. So a direct detection of primaries with energies above $\sim 10^{15}$ eV is not feasible due to a low flux of the cosmic rays. Only detection of extensive air showers allows us to register indirectly such energetic cosmic rays.

A detection of cosmic rays through observation of the air shower provides information about superposition of interaction products which arose during many generations of particle production. The energy and mass of the primary are linked with the general properties of a given shower, but to retrieve the information about the primary from air shower interpretation, the simulations of air shower development are necessary. Due to its nature, the air shower development is a stochastic process with fluctuations which introduce uncertainties in interpretation of the shower. Besides, for ultra-high energies the properties of nuclear interactions are not known, since the accelerator data on particle production are available only at much lower energies. So to perform simulations of air shower development, the extrapolations of hadronic interaction properties are used. Such extrapolations as well as the stochastic character of the shower development are the sources of uncertainties in air shower simulations and interpretation of the results.

For detection of air showers the detectors placed on the ground are used, among which two widely used types of detectors can be distinguished: fluorescence detector (used for registration of longitudinal development of a shower), and surface detector (used for registration of lateral distribution of particles of the shower in many locations). A simultaneous use of the two types of detectors allows a so-called hybrid reconstruction of showers. This type of reconstruction is characterised by more reliable estimation of primary energy when compared with similar estimation using the two types of detectors separately. However, any reconstruction relies on simulations and some additional assumptions which, taking into account the fluctuations of shower development and simulation uncertainties mentioned above, decreases accuracy of the air shower reconstruction and consequently, the accuracy of the final scientific results.

A method of shower reconstruction called top-down uses multiple simulations of air shower development, which is performed to choose a simulation which fits best the data. It is assumed that such best simulation roughly reproduces a set of fluctuations which occurred in the real air shower, so such approach of reconstruction may help us to reduce uncertainties in air shower interpretation. Many shower simulations are necessary to reconstruct one data shower, which is time consuming. To accomplish the reconstruction in a reasonable amount of time, a fast and simplified 1-dimensional initial simulations of shower development are performed. Then, for the best matching 1-D profile, a precise resimulation is performed.

In this thesis I present the top-down reconstruction method, its test using the generated data set and the top-down reconstruction of the real data events from the Pierre Auger Observatory. The version of the method used here is improved by applying a more accurate simulation tool used for simulations of air shower development. Also some extensions of the method were introduced, like using of atmosphere model from fit to data instead of seasonal mean atmosphere. Another extension is required by applied method of comparison of the reconstruction results. While the usual comparison of the top-down reconstruction results rely on the fit to detector signals, in the thesis I perform also a comparison of the signals in individual detector stations. To realise such comparison, another effect is taken into account during top-down reconstruction, i.e. removing an instrumental effect of shower axis shifts on detector level. Such shifts can occur after standard reconstruction and may impact the comparison between the detector stations. The main goal of the thesis is to show that the signals in the detectors, which are a basis for estimation of hadronic part of shower, are influenced mainly by the characteristics of the hadronic interaction.

The structure of the text is as follows. In chapter 2 the general properties of cosmic rays are shortly presented, with special interest of composition. Next chapter 3 includes a description of a detector from which data are used (Pierre Auger Observatory). Chapter 4 presents the methods used in air showers simulations. A precise description of the top-down reconstruction is presented in chapter 5. A test of the method accuracy is shown in chapter 6. The results of application of the top-down method to the real data events in

presented in chapter 7. Chapter 8 includes conclusions and outlook.

2 General properties of cosmic rays

2.1 Energy spectrum

The energy of cosmic rays ranges from 10^7 eV (solar cosmic rays) to 10^{21} eV (extragalactic cosmic rays). The differential "all particle" spectrum of primary cosmic rays is presented in figure 1. The flux of cosmic rays drops with energy increase. For less energetic cosmic rays (energies below ∼10 GeV, not included in figure 1) the flux depends on time and location on the Earth. This is due to the geomagnetic field, which causes also a flux cutoff below energy value which depends on the arrival direction of the particle relative to the geomagnetic field, and on the location of observer on Earth. Also the 11-year solar magnetic activity cycle influences the cosmic rays flux for energies up to few hundreds of MeV [1]. The solar cycle results in a characteristic anticorrelation between the solar activity and the flux of cosmic rays. The anticorrelation is caused by deceleration of the MeV cosmic rays by stronger solar wind. Such deceleration occurs when the cosmic rays pass through the Solar system during the time of Solar maximum [2].

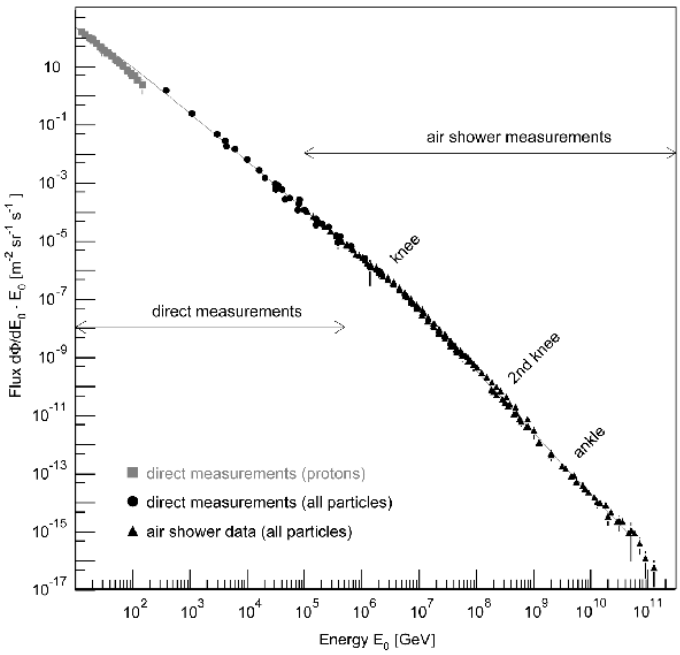

Figure 1: Energy spectrum of cosmic rays [3]. The data points are taken from different experiments. Energy range where the direct and indirect measurements are performed is denoted.

The flux of the GeV cosmic rays is disturbed by the mentioned above factors, causing the flux is time-dependent and place-dependent. At the top of atmosphere the mean number of cosmic rays is of the order of $1000/\mathrm{m}^2/\mathrm{sec}$ in GeV energy region. More energetic cosmic rays penetrate the heliosphere and geomagnetic field deeper. Their number is smaller, e.g. in PeV energy region is $1/\mathrm{m}^2/\mathrm{year}$. At the most energetic part of the spectrum the number of incident particles is below $1/\mathrm{km}^2/\mathrm{century}$. Due to low flux of the more energetic cosmic rays their direct detection is inefficient. Therefore the investigation of

cosmic rays relies on detection and analysis of the extensive air showers (i.e. particle cascades which develop in the atmosphere).

Above the GeV energy region the cosmic rays flux can be described by a power-law:

$$\mathrm{d}N/\mathrm{d}E \sim E^{-\alpha},$$

where α is the spectral index. The slope of the spectrum is not constant, but slightly changes with energy. To enhance visibility of the slope changes, the spectrum is plotted with flux rescaled by factor $E^{2.5}$, where E-energy (see fig. 2). The spectrum is getting steeper at energies: $\sim 4 * 10^{15}$ eV (α changes from ~2.7 to $\sim$ 3.1), and $\sim 4 * 10^{17}$ eV (α changes to $\sim$ 3.3) [4]. These changes of the slope are called a knee and a second knee respectively (marked in fig. 1). At energy about 4*10^{18} eV another change of the spectral index appears, called the ankle, where the spectrum becomes more flat (α changes to value 2.7). The cosmic rays in the ankle region and above are usually called ultra high-energy cosmic rays (UHECR). For energies above 5*10^{19} eV a suppression of UHECR flux in the spectrum is visible, known as GZK cutoff [5].

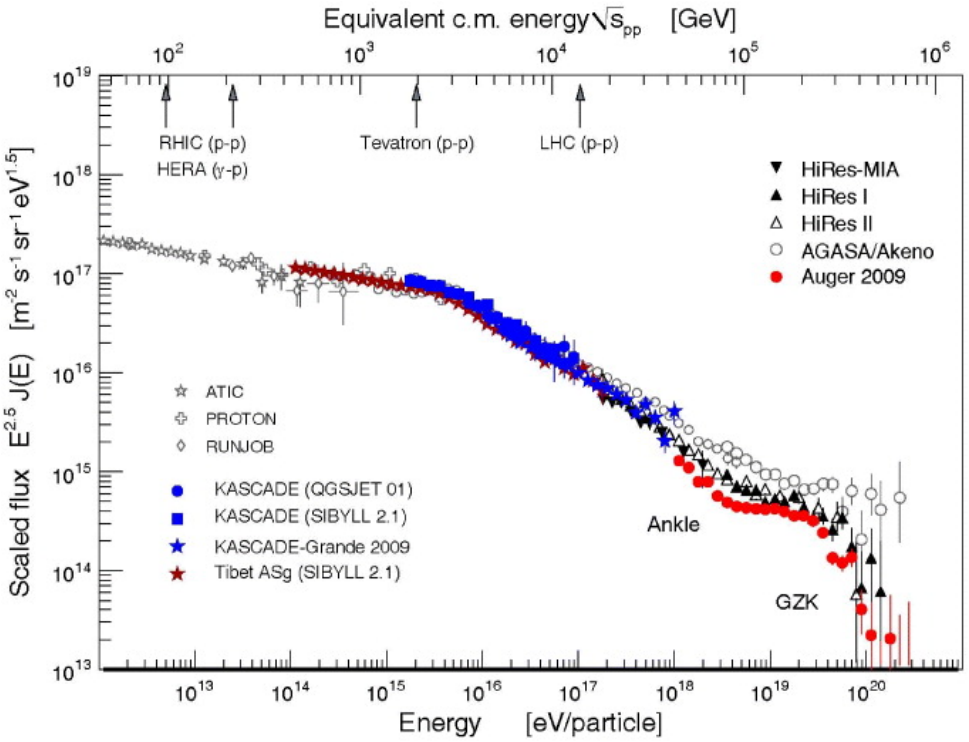

Figure 2: Observed cosmic rays energy spectrum [6]. The data points are taken from different experiments. The flux is scaled by factor $E^{2.5}$ to emphasise the change of slopes in the spectrum. Energy corresponding to the latest LHC measurements is marked. Also characteristic features like ankle and GZK suppression are marked.

Origins of these spectral features are still not determined. It is expected that the two-knee structure is linked with a change of properties of the Galactic cosmic rays (i.e.

cosmic rays which sources are located in Galaxy). One of the most regarding interpretation, known as the "many knees model" [7], explains the knees as a superposition of separate cutoffs, which correspond to components with different mass Z. Other interpretation relates the knees with the properties of particle drift in the global magnetic field of the Galaxy, which is expected to be composed of regular and turbulent parts. In the cited model the turbulent Galactic magnetic field is assumed to exist simultaneously with the regular, mainly toroidal Galactic magnetic field. For ultrarelativistic energies the propagation of charged particles in regular part of magnetic field depend on particle energy to charge ratio E/Z, which lead to cutoff dependence, and so to the knee features [8].

The ankle is traditionally regarded as a transition from the Galactic to extragalactic component of the cosmic rays, but it is not clear where exactly the transition occurs. Assuming that the ankle is an intersection of the steep Galactic and more flat extragalactic components, the Galactic cosmic rays should reach energies above EeV [9]. As the sources of the Galactic cosmic rays the supernova remnants are regarded. In such sources the particles gain energy in diffusive shock acceleration mechanism (see section 2.3), which is characterised by the maximum obtainable energy for a given type of particle. For supported by observations lighter composition of cosmic rays at knees region, the maximum obtainable energy, up to which Galactic cosmic rays are accelerated in supernova remnants, is below the EeV. So currently it is not excluded, that other explanations like e.g. a "dip" in the spectrum are correct. Such "dip" could be caused by the electron-positron pair production in spectrum dominated by the extragalactic protons [10]. However, proton dominance in the extragalactic cosmic rays is also not confirmed in observations.

The suppression of cosmic rays at the most energetic end of the spectrum was predicted by K. Greisen, G. Zatsepin and V. Kuzmin (hence GZK cutoff). According to the theory, the GZK cutoff may appear due to energy losses of ultra high-energy protons via production of the Δ resonance. This process occurs when the UHE protons interact with the cosmic microwave background (CMB) photons (GZK process) [11]. The GZK process results in pion production at cost of proton energy decrease. Also energy of more massive nuclei is decreased, additionally their photodisintegration occur. Such energy losses cause that the number of particles with energies above the threshold for Δ production is reduced significantly with the distance traveled. In effect it is expected that we should not register cosmic rays from the sources located beyond distance $\sim$ 200Mpc (so-called GZK horizon, see fig. 3).

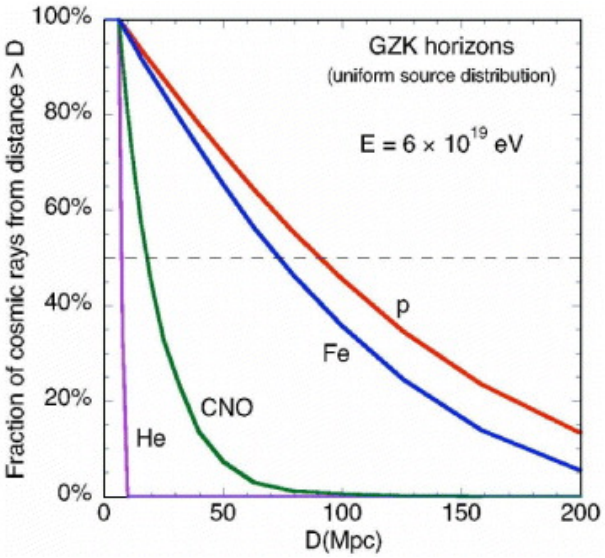

Figure 3: The GZK horizon for different primaries with energy 10^{19} eV [6].

However, results of observation analysis suggests heavier composition of cosmic rays above energy 3 EeV (see chapter 2.2). So proton domination in UHECR is not confirmed, and so the flux cutoff as a result of the GZK process. Therefore other models of flux suppression are proposed. For example, in work [12] authors argue that the flux suppression is linked with reduction of acceleration efficiency at the sources. In this model it is assumed that cosmic rays with energies 1 EeV to 3 EeV are dominated by UHE protons, which is supported by observations. If the rigidity-dependent acceleration mechanism in sources is present, the model indicates that the maximum energy obtainable by proton will be around 10 EeV, so too small for interaction with CMB photons (threshold for interactions is ∼50 EeV). In that case the GZK process is absent, and the cutoff in spectrum is linked rather with reduction of acceleration efficiency at the sources. Also the pion photoproduction on CMB photons in extragalactic space should be absent, resulting in small UHE photon flux. Identification of UHE photon sources, as well as mass determination of cosmic rays, is expected to constrain models of the flux suppression (see chapters 2.2, 2.4) [13].

2.2 Composition

The composition of cosmic rays is determined more precisely for the lower energies. The mass spectrum of the cosmic rays for energy 1 GeV per nucleon is presented in figure 4.

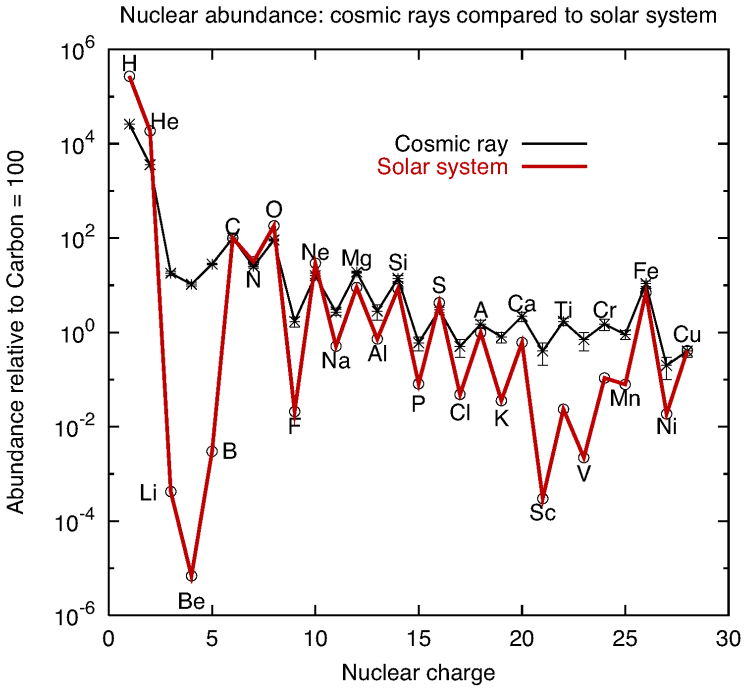

Figure 4: The composition of low energy cosmic rays compared with the composition of Solar System [14].

In the energy range up to ∼100 TeV about 98% of the cosmic rays are composed of protons and stable nuclei with lifetime at least 10^6 years (in which protons abundance is 87%, helium 12%, 1% are heavier nuclei, including even iron). The remaining ∼2% includes electrons and positrons [15]. The cosmic rays composition is similar to the composition of the Solar system, which indicate stellar origin of the cosmic rays. Comparison of the two compositions indicate that lithium, beryllium and boron are more abundant in cosmic rays. Similarly for the "iron group" as seen in figure 4. These differences arise due to spallation process which occur during propagation of more abundant CNO nuclei and iron through the Galaxy [4]. Hence the spallation products are usually referred to as secondary nuclei. The saw-like shape of the spectrum is due to weaker bounds of nuclei with odd atomic mass which cause they are less numerous products of nucleosynthesis reactions.

Due to a low flux of high energy cosmic rays their detection is possible only by observation of extensive air showers. So determination of cosmic rays composition rely on interpretation of the air showers. It was shown that different compositions and astrophysical scenarios may reproduce the observed ankle and flux suppression in the energy spectrum [16]. Therefore to understand the acceleration properties of cosmic rays and origins of the spectral features, a determination of UHECR mass composition is necessary. Many experiments concerning determination of cosmic rays composition above PeV energy region were performed, or are still in progress (e.g. Pierre Auger Observatory [17], KASCADE-Grande [18], Telescope Array [19] or HiRes [20]).

The energy spectra of individual elements for 1-100 PeV range were determined e.g. using KASCADE-Grande detector [21] (fig, 5). It is seen in the spectra that proton and

helium dominate below the energy corresponding to the knee, and their abundance drops fast above the knee energy. On the other side the more massive elements, including iron, start to dominate above the knee. This seems to confirm the expected knee interpretation as a result of the limit of acceleration efficiency for light elements in cosmic rays of galactic origins.

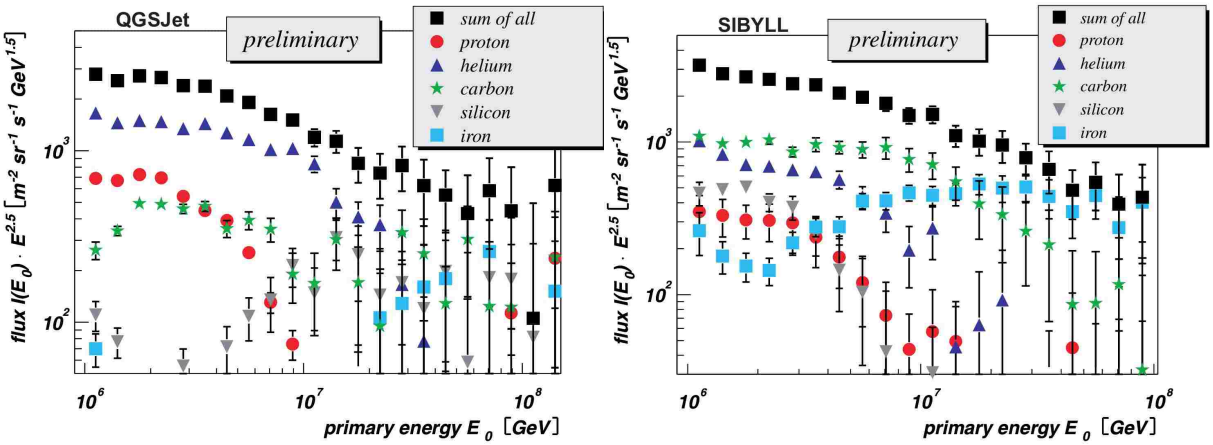

Figure 5: Composition of cosmic rays in PeV region [21]. Fit of chosen elements (proton, helium, carbon, iron) using QGSJet and SIBYLL interaction model is shown. Increasing fraction of more massive elements with energy increase is seen for both models.

At higher energies the composition of UHECR, derived from Auger data, changes with energy (fig. 6) [22]. While the light elements appear to dominate in cosmic rays up to energy $10^{18.3}$ eV (which is ~ 2 EeV), above this energy the results suggest that heavier elements are more abundant. A fit of individual elements to data of a large statistics indicate domination of intermediate-mass elements (helium, nitrogen) for energies above $10^{18.3}$ eV (fig. 7).

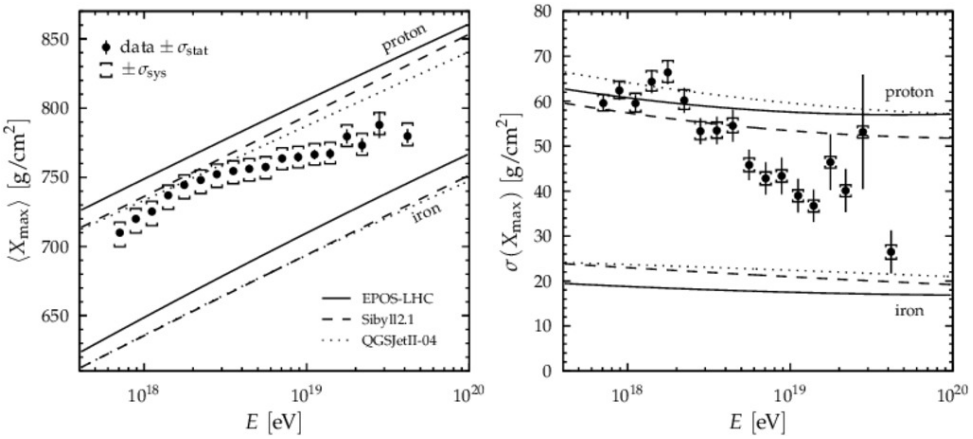

Figure 6: Mean mass evolution of cosmic rays [22]. The lines corresponding to proton and iron pure compositions are shown. The data points show evolution of mean mass between the pure proton and iron compositions.

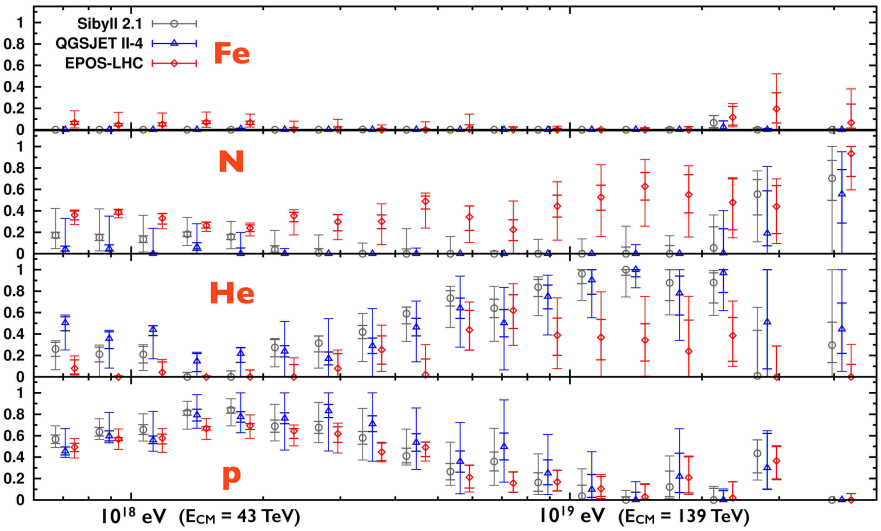

Figure 7: The four component fractions of cosmic rays fitted to $X_{\max}$ distribution in Auger data [23].

Despite that the above results, obtained with different methods, the experimental results seem to confirm a general conclusion that UHECR are not composed of protons only, but include heavier elements. Besides, the results clearly show that the interaction models have to be improved [23].

Among less abundant components of the UHECR, the UHE photons (i.e. photons with energies above 1 EeV), neutrinos and magnetic monopoles are considered. Their abundance in the UHECR is a valuable indicator of UHECR production scenarios. Since the so-called top-down acceleration mechanisms (see section 2.3) are characterised by production of UHE photons and neutrinos, a determination of abundance of these particles in cosmic rays may confirm or exclude this type of acceleration mechanisms. The GZK process also leads to production of neutral pions, which subsequently decay into UHE photons. So, the GZK mechanism may be confirmed by determination of UHE photons flux.

Photon-induced air showers develop slower than showers induced by hadrons. In effect photon-induced showers have the shower maximum about 200 g/cm^2 deeper in the atmosphere than hadronic showers (in the EeV energy range). Also the lateral distribution of particles in photon-induced showers have steeper distribution than in hadronic showers. These observables enable separation of possible photon-induced showers in the data, and to estimate the diffusive flux limit of UHE photons. Such limit was estimated in many experiments, like e.g. Telescope Array, Yakutsk and Auger observatory. In figure 8 (left) a summary of different measurements, including the recent upper limit in Auger data [24], are presented. Auger UHE photon limit, determined from data sample collected with a large exposure, has the lowest measured value up to now, when compared with the other models of UHE photon emission in exotic processes. This implies that production of UHECR in exotic processes, in which extensive UHE photon production is present, are currently disfavored. Only the GZK photons are possible to be reached in measurements

within next years, when a larger exposure will be available (shaded areas in figure 8). So confirmation of GZK process by means of the UHE photons detection is still an opened question.

Registration of cosmic rays with energies above 10^{20} eV motivated scientists to search for high-energy neutrinos in cosmic rays. The neutrinos are expected to be produced during interactions between the propagating UHECR and the cosmic microwave background (CMB) photons [25] (so-called cosmogenic neutrinos), and in interactions of cosmic rays with matter [26]. The intensity of neutrino production depends on the density of matter in source and UHECR source spectrum. Therefore measurements of the neutrino flux could provide informations about nature of the primaries. Additionally a determination of the neutrino flux may provide some hints about evolution of the unknown neutrino sources in function of redshift, and so can help with their identification [26].

Gamma-Ray Bursts (GRB) and Active galactic Nuclei (AGN) are regarded among other possible neutrino sources. High-energy neutrino production in these sources is expected to be associated with production of high-energy protons [27]. The protons interact with the radiation field of in the source producing the charged pions, which subsequently decay into neutrinos and other decay products, depending on decay channel. The neutrino production in AGN may occur in optically dense cores of these objects, but AGN jets are also regarded as the source of neutrinos [28]. It was shown that the observation-based estimation of neutrino upper flux bound is model-independent (Waxman-Bahcall bound [27]), and is lower than expected flux of neutrinos produced in AGN jets. So determination of neutrino flux can be useful in discrimination between the possible neutrino production scenarios and may help to confirm models of high-energy processes in astrophysical sources.

Many experiments provided limits of neutrino flux, e.g. ANITAII [29] and IceCube [30]. Also data analysis in Pierre Auger Observatory was oriented for neutrino search. Neutrino flux determination from Auger data is possible since the neutrino can induce extensive air shower, despite that the neutrinos have small cross section for interaction. Detection of neutrino-induced shower is focused on very inclined showers only (i.e. showers with zenith angle $\Theta > 60^{\circ}$). Neutrino cross section for interaction is very low, so the air shower induced by neutrino can start to develop deep in the atmosphere, unlike the hadron induced air shower, which starts to develop in the atmosphere at high altitude. When neutrino shower starts to develop deep in the atmosphere, the electromagnetic component traverse smaller attenuation lengths than in a shower induced at high altitude. So in neutrino induced air shower, which started to develop at low altitudes, the relative electromagnetic component is less attenuated on the ground, when compared with the hadron induced shower. In effect the detectors will register larger than usually electromagnetic component, indicating that the registered inclined shower may be induced by neutrino.

Latest Auger data show no neutrino events registered [31]. The neutrino flux is close to the upper flux bound to be below Waxman-Bahcall bound (fig. 8, right). Larger statistics of collected data may allow us for discrimination between the neutrino production scenar-

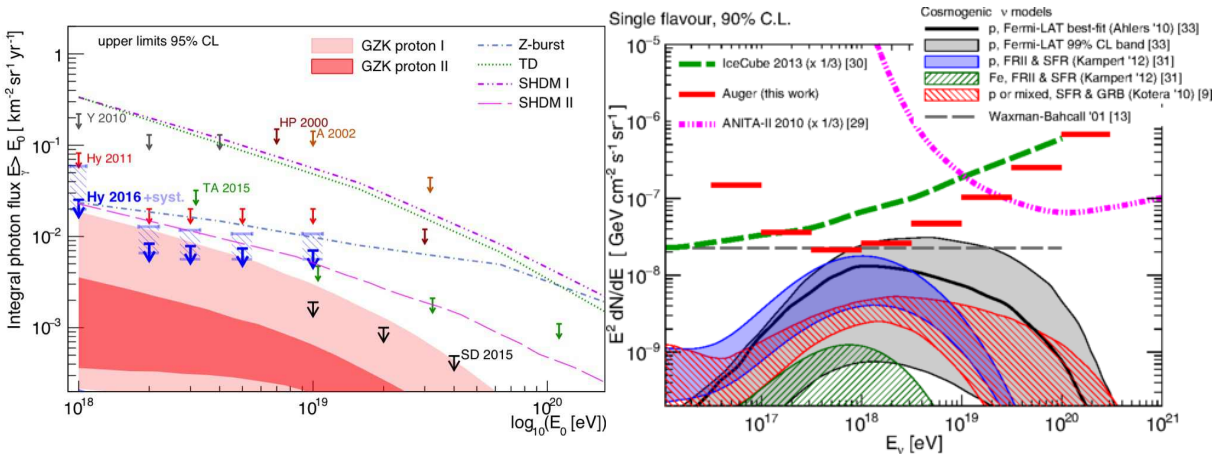

Figure 8: Left panel: A summary of photons flux at 95% confidence level [24], measured in many experiments: Auger hybrid (Hy), Auger SD (SD), Telescope Array (TA), Yakutsk (Y), Haverah Park (HP) (taken from [33]). The shaded area indicates the flux prediction for photons originated in GZK process. The lines show flux prediction from top-down models. Right panel: Upper limit to the normalisation of the diffuse flux of UHE neutrinos at 90% confidence level [31]. Also the corresponding limits from other experiments are shown: ANITAII [29], IceCube [30]. Also the Waxmann-Bahcall bound is marked.

ios. Also the cosmogenic neutrino production models, which assume a pure primary proton composition of injected particles at the sources and strong evolution of neutrino sources, are disfavoured. The latest Auger result [31] exclude also a wide range of exotic models of neutrino production, for which flux limits were estimated in [32], with confidence level larger that 99%.

The magnetic monopole, a particle required by Grand Unified Theory, is also of interest in UHECR detection. Energy of intermediate-mass magnetic monopole (i.e. with mass 10^{11} to 10^{20} eV/c^2) is expected to be comparable with the energy of UHECR hadronic primary. More massive magnetic monopoles should be bound gravitationally to the Galaxy or Sun, so they are not able to reach relativistic velocities. Lighter magnetic monopoles can gain relativistic velocities through acceleration in Galactic or intergalactic magnetic fields, or astrophysical objects like e.g. neutron stars [34][35]. During propagation in the Earth's atmosphere the relativistic monopole will dissipate its energy by interacting with air molecules. In effect production of secondary subshowers, initiated along the path of propagating monopole, will occur. The energy losses of monopole would arise due to collisional mechanisms or pair production, and due to photo-nuclear interaction mechanisms for higher Lorenz factors ($\gamma > 10^4$). Because of high mass of the magnetic monopole, when compared to electron, almost no bremsstrahlung losses would be present. The example simulated longitudinal profile of shower induced by the magnetic monopole is presented in figure 9 (left). Such profile can be easily distinguished in the data. It was shown that among the Auger data collected in years 2004-2012 no candidate event was found [36]. Taking into account the time period of data collection, the upper limit of monopole flux was estimated to have value $10^{-20}/\mathrm{cm}^{-2}/\mathrm{s}/\mathrm{sr}$ for Lorenz factor $\log(\gamma) \leqslant 10$. This result provides the strongest limitation for intermediate magnetic monopole flux when compared

with other measurements (figure 9 right).

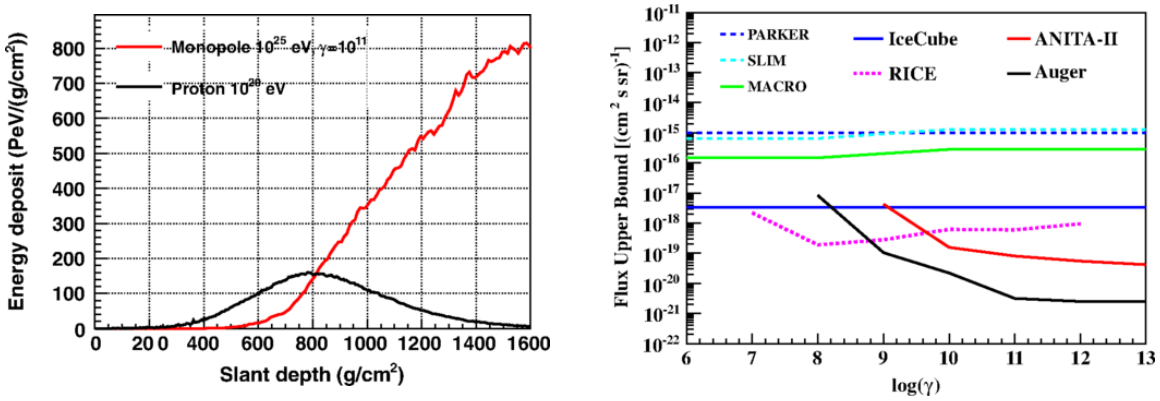

Figure 9: Left: an example longitudinal profile of a shower induced by the magnetic monopole [36]. Right: Auger upper flux limit of the intermediate-mass magnetic monopole with 90% confidence level (black line) [36]. Parker limit [37] (i.e. expected limit of magnetic monopoles which would be accelerated by galactic magnetic fields) is shown as dark-blue dashed line. Fluxes of other experiments are also presented (MACRO [38], IceCube [39], ANITAII [40], RICE [41]).

2.3 Acceleration processes and production mechanisms of cosmic rays

While the power-law behaviour of the UHECR energy spectrum indicates a non-thermal origin of cosmic rays, the wide energy range of primary cosmic rays suggests their different sources of acceleration. Two main mechanisms of acceleration are considered: so-called bottom-up and top-down [42].

In the bottom-up scenarios the charged particles gain energy during interactions with electromagnetic fields. The maximum attainable energy, which particle with mass Z is able to gain during acceleration, depends on the strength of magnetic field in place of acceleration, and the size of the acceleration region. For the region with dimension L kpc, in which a magnetic field B (in microGauss) is present, the maximum obtainable energy is:

$$E_{\max} = (10^{18}\text{eV})ZBL.$$

This dependence can be plotted in the so-called Hillas plot (fig. 10) [43], where $E_{\max}$ is limited by the condition in the equation above. The possible sources which correspond to given strength of magnetic field and size of acceleration region are presented in the Hillas plot. Some of them are located in our Galaxy and so are regarded as galactic sources, like e.g. supernova remnants (SNR). Among the extragalactic sources are radio galaxy jets, galactic clusters or intergalactic medium (IGM).

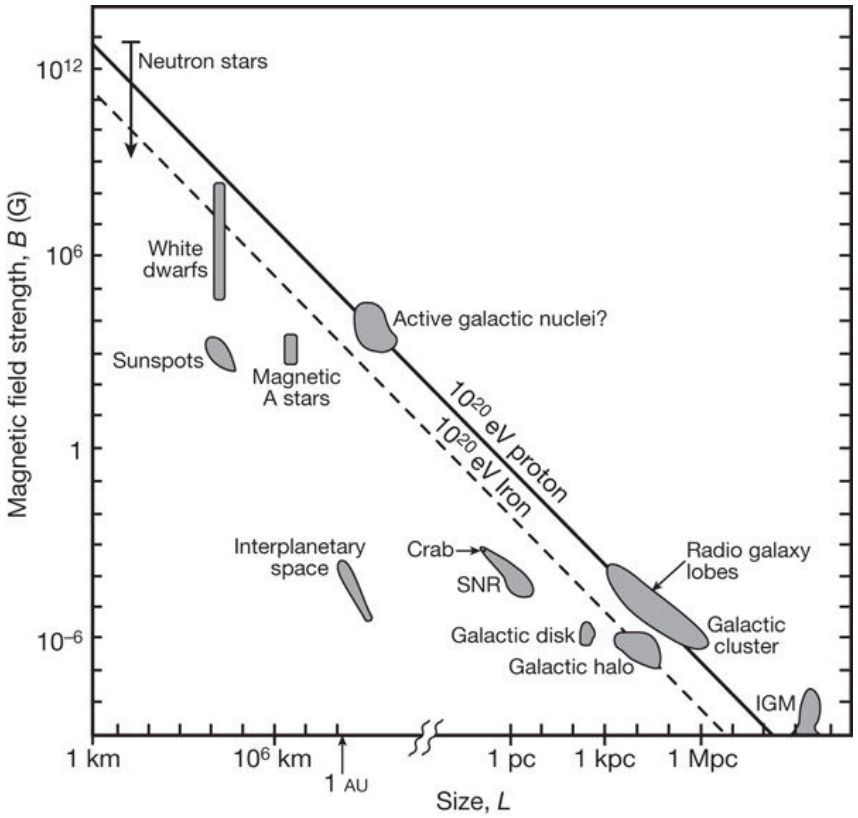

Figure 10: Hillas plot [44]. The possible sources for different energies are described on the plot.

A statistical bottom-up mechanism of particle acceleration linked with the supernova explosions was proposed by Fermi in 1949 (Fermi acceleration processes) [45]. This kind of acceleration processes provide the most promising explanation of particle acceleration. The Fermi processes of acceleration require existence of a plasma with magnetic field, like in supernova shock waves or medium in galaxy clusters. There are two scenarios of Fermi process considered, named as first and second order Fermi acceleration mechanisms. The second order Fermi acceleration mechanism takes place when a particle is moving through regions where inhomogeneous magnetic field is present. During propagation a particle reflects from these inhomogeneities moving with velocity v, like from a magnetic mirror. On average,the energy gain during each reflection is equal to:

$$\left\langle \frac{\Delta E}{E} \right\rangle = \frac{8}{3}\beta^2,$$

where $\beta = v/c$. The term "second order" comes from the value of the exponent of β. The simulated energy spectrum of particles accelerated in this process can be described by the power-law, like for the observed energy spectrum for cosmic rays. However, to begin acceleration this process requires injection of particles with energies exceeding thermal energy of the surrounding medium. Besides this mechanism is expected to have a slow energy gain due to a low density of interstellar matter.

In the first order Fermi acceleration process the particle gain higher energy during collisions with the shock waves produced e.g. in supernova explosions. The shock waves

can reach velocities 10^3 times larger than a mean velocity of the interstellar clouds in vicinity, becoming a supersonic interstellar shocks waves. Magnetic field before and after the shock wave is turbulent. A particle propagating in vicinity of shock wave has chaotic trajectory, passing through the shockwave many times. When a particle passes through the shock, it gets higher energy. In one round trip the particle mean energy gain is:

$$\left\langle \frac{\Delta E}{E} \right\rangle = \frac{4}{3}\beta.$$

The energy spectrum of many particles accelerated in this process has a spectral index close to the observed value, when particle leakage from Galaxy is taken into account. The first order Fermi process is regarded as the most promising explanation of the UHECR acceleration process.

If the bottom-up acceleration mechanism dominate, the observed energy density of cosmic rays ($\sim 1eV/cm^3$) requires the total power of Galactic UHECR sources of 10^{41} erg/s [4]. To fulfill this requirement, 10% of kinetic energy included in the most energetic Galactic sources (e.g. supernova explosions) is sufficient.

Among the bottom-up class acceleration mechanisms not only Fermi processes are regarded, but also the so-called single shot mechanism. This mechanism is of inductive kind, i.e. a charged particle may be accelerated by intense electric field during single passage of the particle through the field. So the charged particles could be accelerated by the electric field like in a linear accelerator. The maximum obtainable energy in this process is expected to be rather limited due to energy losses. However, it was shown that in specific configuration the accelerated particles can reach energies 10^{19} eV, even in a presence of the electric field with a low intensity.

Other mechanisms leading to production of UHECR, belonging to the top-down type, require existence of so far unknown unstable or meta-stable, super-massive particles [46]. The mass of decaying particles must be above the highest observed energy of UHECR ($\sim$320 EeV). Decaying super-massive particles will produce a cascade of energetic photons, neutrinos and light leptons with a small fraction of protons.

Two scenarios of UHECR production in top-down mechanism are considered. One is the radiation, interaction or collapse of topological defects producing super-massive particles which subsequently decay [47]. The topological defects are supposed to be traces of phase transitions in the early universe. Among the topological defects, which are regarded as possible sources of the UHECR, are cosmic strings. Another scenarios include decay of metastable super-massive relic particles from some primordial quantum field, produced after the inflationary stage of the Universe [48]. Their masses should be larger than 10^{21} eV, and they should decay at present to sustain observed flux of UHECR. Such massive particles should be pulled by gravitational force of galaxies and galaxy clusters, causing aggregation of massive particles in cluster's vicinity. Among decay products of super-massive particles leptons and other less numerous particles are expected. Also much more photons and neutrinos will be produced in super-massive particle decays, than by any other nucleus during interaction with matter. Therefore a significant photon fraction in

registered UHECR may confirm the validity of the top-down production scenarios, as well as a large cosmogenic neutrino flux. However, recent estimation of limits for UHE photon and neutrino fluxes exclude existence of such mechanisms with high confidence level (see section 2.2).

2.4 Identification of sources

As it was mentioned in section 2.1, the origins of the knees and ankle are still not certain. A direct identification of cosmic rays sources would give a strong evidence for existence of the bottom-up acceleration mechanisms. Additionally it may indirectly support the interpretation of the ankle as transition between the Galactic and the extragalactic cosmic rays.

Attempts of identification of the Galactic cosmic rays sources rely on detection of the UHE photons, UHE neutrons, UHE neutrinos and the most energetic charged components of UHECR. Two main techniques of sources identification are used. One is to search for excess in the UHECR incoming directions. It is done in a narrow solid angles on the celestial sphere (blind search). A location of possible excesses may coincide with positions of astrophysical objects, which may be subsequently regarded as possible sources. Another technique relies on analysis of correlation between the incoming directions of UHECR and the expected source positions taken e.g. from catalogues (targeting to sources).

Recent results of UHECR sources identification, relying on detection of the UHE photons, are as follows. The UHE photons are expected to be produced during interactions between the UHE protons and the matter located in vicinity of proton sources. UHE protons are considered, since they are at least a significant fraction of the cosmic rays near the ankle region. If interactions between UHE protons and matter have place, a production of neutral pions, subsequently decaying into photons, occurs. Also the inelastic nuclear collisions of accelerated protons with the matter in sources may result in the production of UHE photons. Photons, as neutral particles, will propagate directly from the sources to detector, without trajectory bend due to the magnetic field. However, UHE photons are probably attenuated during propagation to the Earth, with estimated attenuation length around 4.5 Mpc. So detectable volume for sources of UHE photons is small compared to the GZK sphere ($\sim$200 Mpc), but large enough to contain the Local Galaxy group [49].

Currently the upper limit of directional UHE photon flux from the point sources, estimated using the Auger data, indicate that no UHE photons from any single source located in the Galaxy, nor from the Local Group of galaxies was registered [50] (see fig. 11). Also the Telescope Array Collaboration did not report any photon point sources in the northern hemisphere [51].

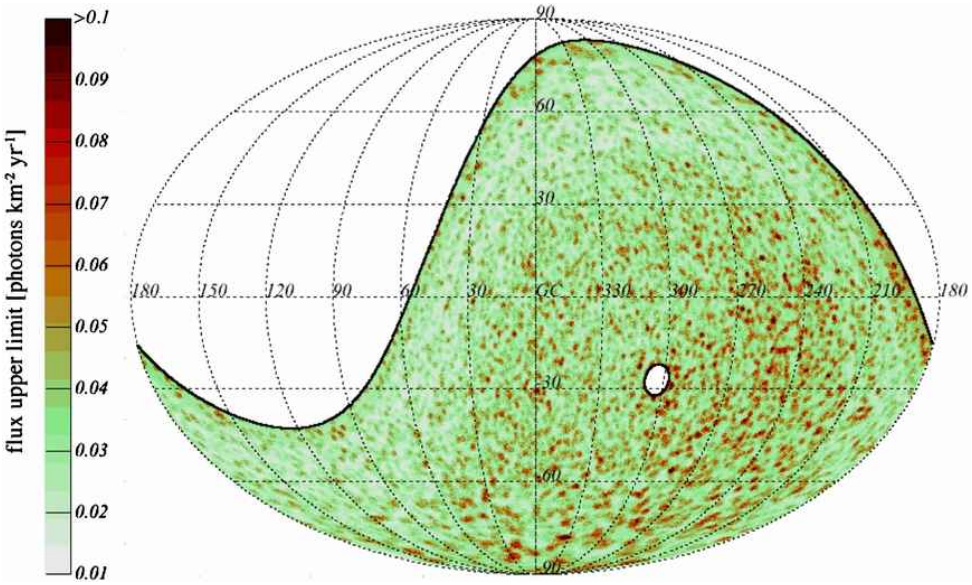

Figure 11: Photon flux limit sky map in galactic coordinates. Galaxy center is marked on the map as GC [50].

UHE neutrons, like UHE photons, propagate on straight lines. So neutrons also may be used for identification of cosmic ray sources [52]. The maximum distance of propagation for a free neutron, which in the rest frame has mean lifetime 886 seconds, depends on the neutron energy. The propagation distance of 1 EeV neutron is large enough for detection of a possible neutron source located in the Galactic disk, including the Galactic center. UHE proton sources are expected to produce some neutrons in charge-exchange interactions of proton.

Neutrons produce air showers identical to proton-induced showers. Since neutrons propagate from the source on straight lines, an excess of air showers arriving from some direction may indicate a position of the neutron source. However, the blind search for the UHE neutron sources did not confirm any point source on the sky. The upper limit of UHE neutron flux was estimated to not exceed 0.046 particle/km^2/year, which corresponds to the energy flux equal to $\sim$0.34 eV/cm^2/s (assuming energy spectrum E^{-2}, characteristic for the "first order" Fermi acceleration process). Such limit stands in opposition to expectations: there are known less energetic TeV gamma-ray sources (TeV sources) in Galaxy, which provide at least three times larger energy flux compared with the energy flux limit corresponding to the UHE neutrons. Assuming that Fermi acceleration processes in TeV sources are present, and that these sources are also possible EeV sources, the energy flux at EeV range should be at least similar to the detected energy flux coming from the TeV sources [52]. It was shown that if the Galactic sources produce protons with EeV energies continuously, the appropriate neutron flux should be detectable at Earth. Even targeting to sources did not confirm any statistically significant candidate neutron source on the sky [52].

The charged particles are also used for identification of the UHECR sources, despite that before hitting the Earth their trajectories are bent by the magnetic field of the Galaxy. Taking into account estimated intensity of the Galactic magnetic fields, the Larmor ra-

dius of most energetic registered UHECR is much larger than the diameter of Galaxy. Therefore the trajectories of UHECR light components are expected to be bent by less than 3°, when propagating from the source to the detector. This property is used for identification of sources by analysis of the incoming directions of the most energetic UHECR. The anisotropy in distribution of incoming directions on the celestial sphere, like localised excess or clustering, may coincide with the positions of possible sources [13]. However, the data suggest the more massive components are present in cosmic rays at energy above the first knee. This brings additional uncertainties in direct source identification, since these components have much more bent trajectories, compared to the proton. Besides, if the intensity of extragalactic magnetic fields is higher than expected, the trajectories will be also more bent, even for light elements. In case when the composition is heavier, or stronger intergalactic magnetic fields are present, it is expected that the incoming directions could show a directional excess on large angular scales on the celestial sphere, since the distribution of galaxies including potential sources is not isotropic. However, recent analysis of the incoming directions for the most energetic primaries ($E > 54$ EeV), taking into account large angular scales, does not confirm existence of significant correlation with sources (see fig. 12) [53].

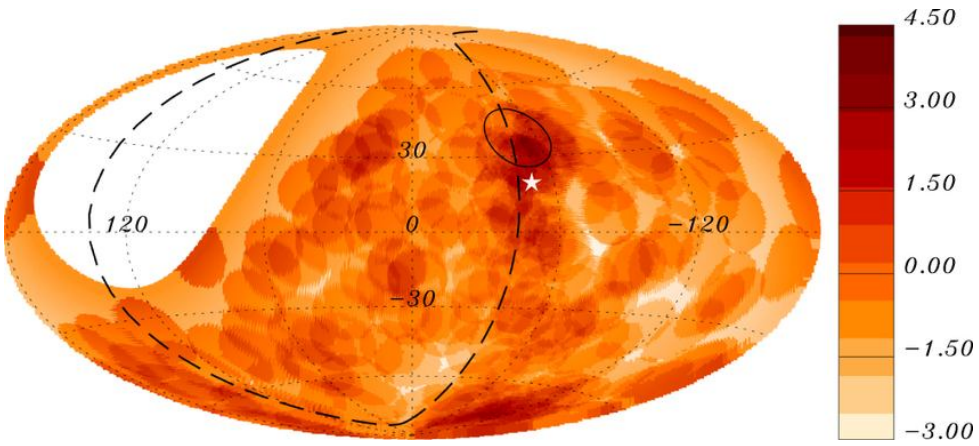

Figure 12: Significance of anisotropy excess map [13]. The white star indicates location of Cen A source. The highest significance found in data (4.3σ) is close to the Super-Galactic plane and Centaurus A region. However the simulations showed the data events have distribution compatible with isotropic distribution.

Cross-correlation between the arrival directions of charged UHECR with energies $E >$ 53 EeV, and the AGN positions in VCV catalog did not confirm AGN as a possible sources of UHECR. Data showed the correlation fraction is equal to $28.1^{+3.8}_{-3.6}\%$, which is only 2σ above the isotropic expectation (21%) [13]. Therefore the recent data set shows that the anisotropy in arrival directions is not significant. So no AGN as possible source of the UHECR was indicated as certain. In general the dipole anisotropy is visible for cosmic rays with energies above 8 EeV. The orientation of the dipole seems to prove the UHE cosmic rays are of extragalactic origin [54].

Current results do not confirm the existence of the Galactic sources. However, the attempts of sources identification presented above do not finally exclude the existence of the Galactic sources. Data collection is still in process, so attempts of galactic sources identification using data with larger statistics may provide confirmation of galactic sources.

2.5 Extensive air showers

When a high-energy cosmic ray propagates in the atmosphere, it interacts with air molecules. This leads to development of particle cascade in the atmosphere. Such cascade is called the extensive air shower. Development of the shower begins at high altitudes (15 - 35 km), where first interaction of the primary particle occurs, and is sustained through the whole thickness of atmosphere. All particles produced in the cascade (secondary particles), which have energy above production threshold, contribute to the shower development due to many subsequent interactions, in which production of other generations of particles occurs. Secondaries propagate towards the ground and spread over a large area. At highest energies they cover many square kilometers on the ground, depending on the inclination of the shower. The total number of secondary particles in a shower is large, for example a shower with energy 10^{19} eV has $\sim 10^{10}$ particles at stage of maximum development of shower.

Particles in a shower may be divided in three general types: muonic, hadronic and electromagnetic. A summary of shower components and interaction channels is presented in figure 13. The electromagnetic part of a shower is composed of photons, electrons and positrons. Other particles also contribute indirectly to this component, like the neutral pions which decay into photons, or muons through decays and interactions. For proton-induced air shower the electromagnetic part at sea level contains 99% of the total number of particles and carries 85% of the total shower energy, which makes the electromagnetic part the dominating component of the shower. The remaining part of shower is composed of muons (carrying $\sim$10% of energy), charged pions (carrying $\sim$4% of energy), neutrinos and less numerous baryons [46].

The shower has a large lateral distribution which arise from lateral spread of particles produced in hadronic interactions and Coulomb scattering of electrons, and is characterised by the air-density-dependent Molière radius (at sea level Molière radius $\sim$ 78 m). Lateral spread of muons is larger than the spread of electrons. On the other hand the muons reach the ground with small energy loss, and with very small scattering during propagation, while the electromagnetic part is attenuated much more intensive than the muonic part. Different propagation properties between the electromagnetic and the muonic parts of a shower result in a characteristic behaviour of muon-to-electron ratio on the ground. This ratio depends on the zenith angle of the shower, and tends to increase for more inclined showers which implies a general rule: the more inclined shower the larger relative muon content on the ground.

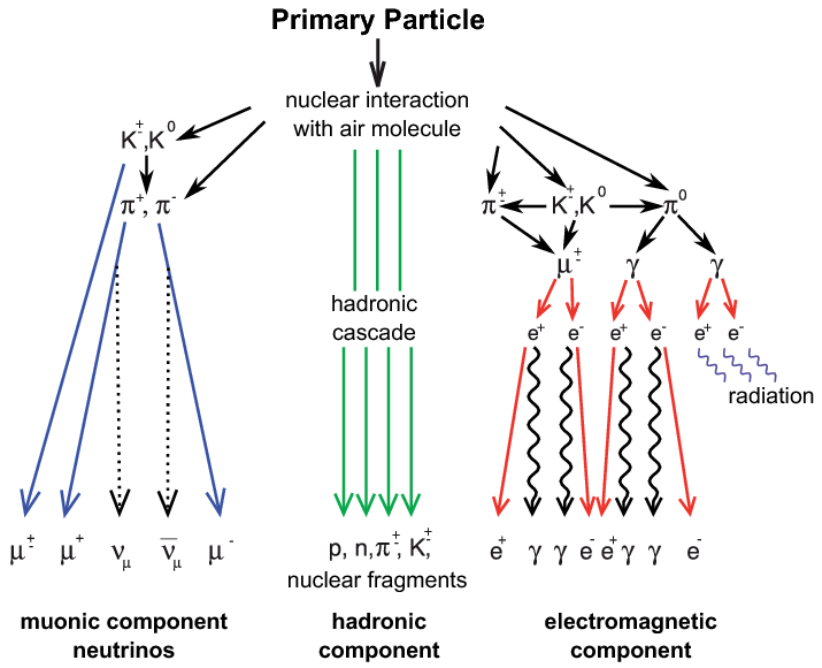

Figure 13: Particle shower components and interaction channels [55].

Since the number of the secondary particles increases with shower development, their energy is distributed among much more particles after each next interaction than in previous interaction. In effect, after some number of interactions, the energy of many secondaries start to drop below the interaction threshold where the particle production stops. When fraction of secondaries with energies below production threshold dominate, the shower reach the stage of maximum of its development. Shower maximum occurs at some height h above the ground, but atmospheric depth X instead of height is used as a parameter describing the matter measure in which the shower particles propagate and interact. By definition: for atmosphere described by density profile $\rho(h)$, the atmospheric depth at given height h is:

$$X = \int_h^\infty \rho(h')\mathrm{d}h',$$

and is expressed in g/cm^2. The atmospheric depth at which any shower has its maximum number of produced secondary particles is called the shower maximum and is denoted as $X_{\max}$.

Another parameter, called shower age s, is introduced to describe a universal properties of electromagnetic cascade in showers. Many properties like energy spectrum or distribution of secondaries may be described using this parameter. By definition

$$s = \frac{3X}{X + 2X_{\max}},$$

so the shower age has value $s = 0$ at the top of the atmosphere, and reaches value $s = 1$ at the shower maximum.

The shape of the shower is parameterised along its axis (longitudinal profile) and in a plane perpendicular to the shower axis (lateral distribution). The longitudinal profile

of a shower is parameterised by the following set of parameters: the number of particles at shower maximum $N_{\max}$, $X_{\max}$, the interaction length λ and atmospheric depth of first interaction X_1. Its functional form, known as Gaisser-Hillas (GH) function [56], is as follows:

$$N(X) = N_{\max}\left(\frac{X - X_1}{X_{\max} - X_1}\right)^{(X_{\max}-X_1)/\lambda} \exp\left(\frac{X_{\max} - X}{\lambda}\right).$$

Such parameterisation allows us to obtain values of $N_{\max}$ and $X_{\max}$ from a fit to data, which is used for determination of energy of the cosmic ray which induced a given shower. An example of shower visualisation with appropriate GH function is presented in figure 14.

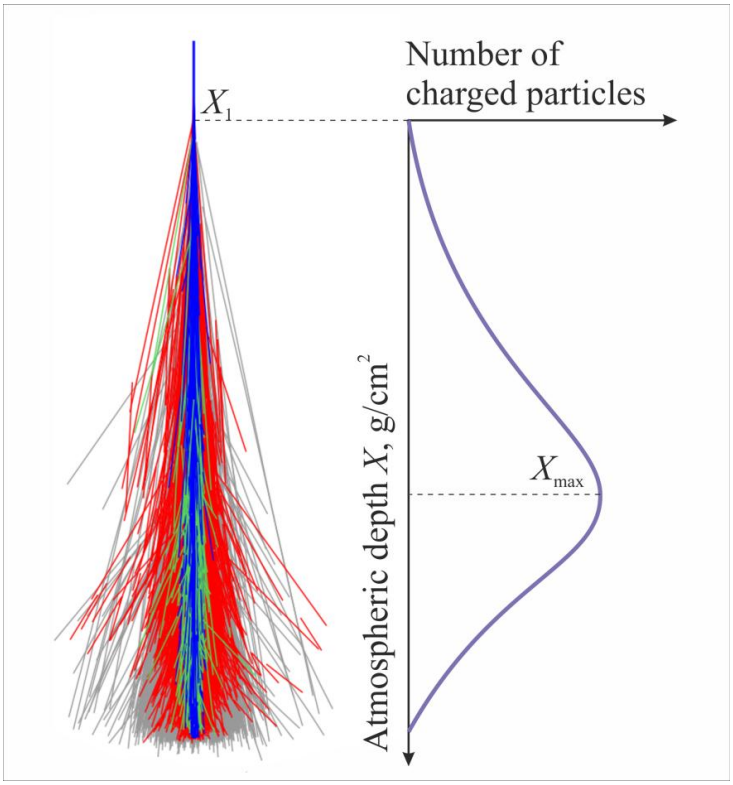

Figure 14: Development of extensive air shower and corresponding GH function. The primary enters the atmosphere from the top of the plot, having first interaction at altitude which corresponds to atmospheric depth of first interaction (denoted as X_1). The corresponding charged particle number profile is ploted with marked location of $X_{\max}$.

The longitudinal shower development can be also described analytically as a solution of the transport equation [57]. Such equation is used for fast calculation of the total energy deposited in atmosphere, and for calculation of the longitudinal distribution of particles in a shower in function of atmospheric depth.

The lateral distribution of particles in a shower is parameterised using the formula given by Nishimura, Kamata and Greisen (NKG) [58]. For electrons the NKG formula has a form:

$$\rho_{\mathrm{e}}(r, X) = N_{\mathrm{e}}(X)\frac{C(s)}{r r_1}\left(\frac{r}{r_1}\right)^{s-1}\left(1 + \frac{r}{r_1}\right)^{s-9/2},$$

where s - shower age, r_1 - Molière radius, $C(s)$ - a normalisation coefficient. The $C(s)$ is chosen to hold the expression:

$$\frac{2\pi}{N_{\mathrm{e}}(X)} \int_0^\infty r\rho(r)\mathrm{d}r = 1.$$

Example lateral distributions of different types of particles are presented in fig. 15.

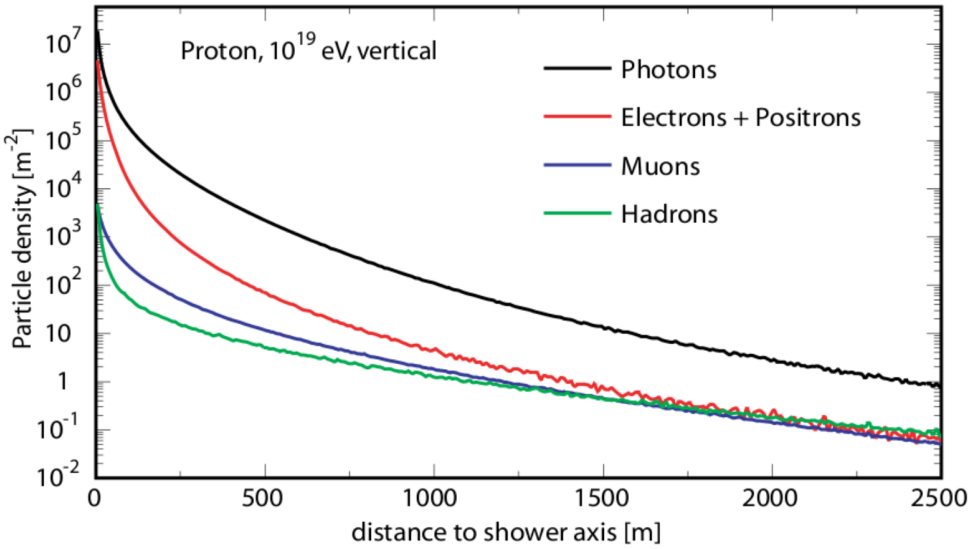

Figure 15: Lateral distributions of different shower components in a shower initiated by a 10^{19} eV proton.

Shower development has large shower-to-shower fluctuations, even when showers are induced by particles with the same mass and energy. Due to these fluctuations, analysis of a single shower provides information about properties of the primary particle with large uncertainty. Only a set of parameters describing the showers (like X_{max}, N_μ), which are measured with sufficient statistics, are used in data analyses.

Basic shower characteristics are related with the properties of the primary. A simple toy model of electromagnetic cascade evolution was proposed by Heitler (fig. 16, left) [59]. In the model it is assumed that primary particle with energy $E_\circ$ interacts with the air molecule, and produces two secondary particles with energies equal to $E_\circ/2$ each. Particle production occurs after traveling the distance equal to one interaction length. Each of the secondary, tertiary, etc. particles, which were produced in previous interaction, produce again a pair of particles. Each produced pair carry half of energy of the interacting particle. A production of the next generations of particles continues, until the energy of produced particles drops below the critical energy E_{c}, when no further particle multiplication is possible. Then the electromagnetic shower reaches its maximum of development. The number of particles after n steps will be then $N_n = 2^n$ and the energy of a given particle at step n will be $E_n = E_0/2^n$.

Such assumptions give the following properties of the Heitler model. The particle number at maximum of cascade development is $N_{\mathrm{max}} = E_\circ/E_{\mathrm{c}}$. The number of interactions needed to reach maximum of cascade development is $N_{\mathrm{max}} = \ln(N_{\mathrm{max}}/\ln(2))$. Since the

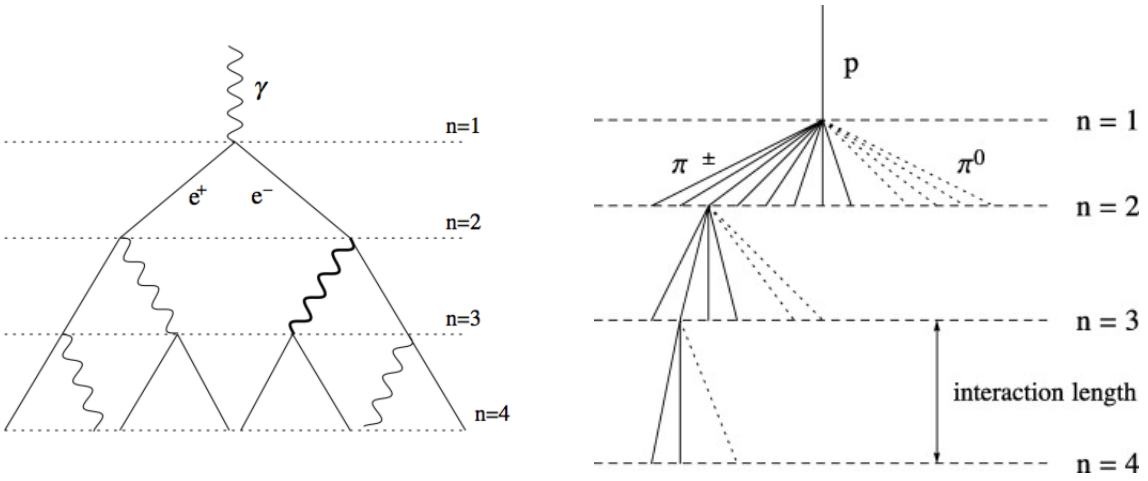

Figure 16: Left: Electromagnetic Heitler model [60]. Right: Hadronic Heitler-Matthews model [60].

depth of shower maximum must be $X_{\max} = X_1 + N_{\max} d$, where X_1 is a depth of the first interaction and interaction length $\mathrm{d} = \lambda_\mathrm{r} \ln(2)$, λ_r is a radiation length, the $X_{\max}$ can be expressed as:

$$X_{\max} = X_1 + \lambda_\mathrm{r} \ln(E_0/E_\mathrm{c}).$$

The elongation rate $D_{10} = \mathrm{d}X_{\max}/\mathrm{d}\log E_0$ in Heitler model is:

$$D_{10} = \lambda_\mathrm{r} \ln 10 \sim 85 \mathrm{g/cm}^2.$$

The Heitler model is simple, but correctly describes the electromagnetic showers up to the maximum of shower development. The model also shows the logarithmic energy dependence of $X_{\max}$.

The Heitler model was adopted and modified by Matthews to describe the hadronic showers [60]. The Matthews model shows analytical relations between energy of the primary, muon number, electron number and depth of the maximum. Despite its simplicity it helps to understand the basic properties of hadronic cascades, and shows a dependence between the basic parameters of hadronic interaction models and the air shower predictions. The Matthews extension introduces pion production and hadronic interaction length λ_I. In this model the hadronic shower development is as follows. After each step of thickness $\lambda_I \ln(2)$ a production of $2N_\pi$ charged pions (π^+, π^-) and N_π neutral pions π^0 occurs (fig. 16, right). The π^0 decay to photons and contribute to development of the electromagnetic cascade. Using the superposition approach (i.e. a nucleus with mass A and energy E can be described as superposition of A nucleons with energy E=E/A) the Matthews model describes:

- Energy of the primary E_0:

  $E_0 \approx 0.85$ GeV $(N_\mathrm{e}{+}24N_\mu)$,

  where N_e is number of electrons, N_μ is number of muons. E_0 is mass-independent and insensitive to fluctuations. This energy dependence is linear, and is used in data analysis for estimation of primary energy, when both, N_e and N_μ, are measured.

- $X^{\mathrm{p}}_{\mathrm{max}}$ for proton:

$$X^{\mathrm{p}}_{\mathrm{max}} = X_0 + \lambda_{\mathrm{r}} \ln(E_0/(3\mathrm{N_{ch}})).$$

- X_{max} is a function of the primary mass and shower maximum for proton $X^{\mathrm{p}}_{\mathrm{max}}$:

$$X^{\mathrm{A}}_{\mathrm{max}} = X^{\mathrm{p}}_{\mathrm{max}} - \lambda_{\mathrm{r}} \ln(\mathrm{A}),$$

  where λ_{r} - radiation length in the medium. This relation shows that the heavier nuclei the shallower shower induced, i.e. with smaller X_{max}.

- the number of muons:

$$N^{\mathrm{A}}_{\mu} = N^{\mathrm{p}}_{\mu} A^{1-\beta} = \mathrm{A}\left(\frac{\mathrm{E/A}}{E_{\mathrm{c}}}\right)^{\beta},$$

  where E_{c} is the critical energy at which charged pions decay into the muons, $\beta = \ln[N_{\mathrm{ch}}]/\ln[3N_{\mathrm{ch}}/2] \sim 0.9$ [4]. Therefore showers initiated by heavier primary with the same energy will contain larger number of muons.

- The elongation rate is described by

$$D^{\mathrm{A}}_{10} = \frac{\mathrm{d}X^{\mathrm{A}}_{\mathrm{max}}}{\mathrm{d}\log(E_0)} = \frac{\mathrm{d}X^{\mathrm{p}}_{\mathrm{max}}}{\mathrm{d}\log(E_0)} = D^{\mathrm{p}}_{10}.$$

  and is not mass-dependent.

- Number of muons in the shower N_{μ}:

$$N_{\mu} = (E_0/E_{\mathrm{c}})^{\beta_1},$$

  where $\beta_1 = \ln[1 + N_{\mathrm{ch}}]/ln[(1 + n_{\mathrm{ch}})/(1 - \kappa/3)]$, κ - inelasticity (a fraction of the total energy used for production of pions). This equation shows the N_{μ} is a model-dependent parameter. Introduction of inelasticity ($\kappa < 1$) does not change values of elongation rate nor primary energy. However, it decreases X_{max} by $\sim 17\mathrm{g/cm^2}$.

2.6 Detection methods of extensive air showers

There are two widely used complementary detector types used for detection of extensive air showers: surface detector arrays and fluorescence detectors. The surface detectors (SD) are usually scintillator detectors, like in Volcano Ranch Experiment [62] and AGASA [63], or water Cherenkov detectors like in Haverah Park [64] and Pierre Auger Observatory [17]. The SD are used to register the shower particle density on the ground. The individual detectors are installed over a large area, and they form a net of detectors. The total area of the SD stations is very small when compared to the area at which the shower particles spread (e.g. in the Pierre Auger Observatory the detector area ratio to area covered by particles on the ground is $\sim 10^{-6}$). Since the number of particles in shower is large (10^{10} and more), the small fraction of detection area is sufficient for proper reconstruction of lateral distribution of particles in a shower. Signals and their timings registered in surface

detector are used for reconstruction of shower geometry, lateral distribution of particle density in shower and energy of the primary.

Each detector station registers particles and their arrival times independently from others detectors. Arrival times are monitored in order to distinguish detection of shower particles from detection of accidental particles. When shower particles trigger a set of neighbor stations in characteristic time intervals, a detection of shower occurred. Then the signals in triggered stations may be used to fit the analytical formula for lateral distribution function (LDF). Station distances from shower core, measured in a plane perpendicular to the shower axis, are applied in a fit. Traditionally the integrated LDF is a measure of the total number of particles in shower, and so of the primary energy. The measured lateral distribution of particle density can be evaluated as:

$$S(\mathrm{r}) = S(\mathrm{r_{opt}}) f_{\mathrm{LDF}}(\mathrm{r}),$$

where $\mathrm{S}(r_{\mathrm{opt}})$ - a shower signal measured at some optimum distance from the shower axis, $f_{\mathrm{LDF}}(r)$ - parameterisation function, normalised to $\mathrm{S}(r_{\mathrm{opt}})$. For Auger observatory $r_{\mathrm{opt}} = 1000$ m [65]. The $f_{\mathrm{LDF}}(\mathrm{r})$ used in Auger observatory has a form:

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

where r_{700} means r=700 m, r_{1000} is r=1000 m, β-slope of the LDF dependent on the zenith angle. Having determined S(r) from the measurements, the total number of particles N can be computed by integrating:

$$n \sim \int 2\pi r S(r) \mathrm{d}r.$$

The energy estimator is a function of the expected signal of a shower at a reference zenith angle equal to 38° (denoted as S_{38}):

$$E = A S_{38}^{\mathrm{B}},$$

where A and B are parameters dependent on the characteristics of the detectors, and on properties of the particle shower, especially on the zenith angle [66]. Values of the parameters A and B need to be obtained from simulations. The parameter S_{38} answers for a question: for a shower developed under some zenith angle what the S_{1000} value would be, if this shower would have developed under reference zenith angle equal to 38°. The reference zenith angle (38°) was chosen to remove zenith dependence of the energy estimator. This dependence is removed using the CIC (Constant Intensity Cuts) method, in which S_{38} has a form:

$$S_{38} = \frac{S_{1000}}{CIC(\Theta)}.$$

The $CIC(\Theta)$ function has a form of second order polynomial: $CIC(\Theta) = 1 + \mathrm{a}x + \mathrm{b}x^2$, where $x = \cos^2(\Theta) - \cos^2(38°)$, and is obtained in a way described in [67]. It corrects the S_{1000} for different attenuation properties of inclined showers, and provides an unbiased energy-related parameter S_{38}.

More detailed description of geometry reconstruction is presented in the following chapter, where a description of the Pierre Auger Observatory is presented (section 3.6). Also more precise description of SD on example of detectors used in Pierre Auger Observatory is presented in section 3.6.

It was shown that during air shower development about 90% of the total air shower energy is used for ionization and excitation of the air molecules [68][69]. A spontaneous deexcitation of the nitrogen molecules effects in isotropic emission of fluorescence light, which wavelength is located mainly in the UV region (figure 17). A so-called fluorescence detector (FD) collects the fluorescence light, which is a base of another technique of extensive air showers detection. The registered fluorescence light is used for determination of the longitudinal profile of the shower and the calorimetric energy of the primary. The Gaisser-Hillas function is used as a functional parameterisation of the profile.

The number of photons emitted during deionization is proportional to the energy deposited in the atmosphere by a shower, i.e. $N_\gamma \sim dE/d\mathrm{X}$. The energy, determined using the fluorescence technique, is nearly the calorimetric energy E_{cal}, since:

$$E_{\mathrm{cal}} = \frac{E_{\mathrm{crit}}}{X_0} \int N_{\mathrm{ch}}(X)\mathrm{d}X,$$

where E_{crit} - critical energy, X_0 - radiation length in the air, N_{CH} - number of charged particles in a shower.

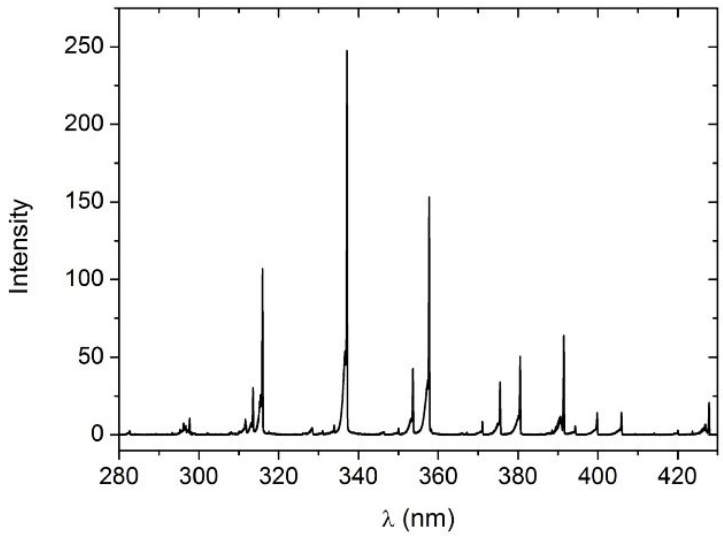

Figure 17: Spectrum of fluorescence light emited during deionization of air molecules previously excited by extensive air shower [70].

The shower geometry can be determined from the longitudinal profile. It is done using geometric dependencies between the location of FD eye and the timings of fluorescence light detection. More details of fluorescence reconstruction methodology is presented in section 3.5. A very efficient detection of showers is realised, when the showers are registered

simultaneously by both fluorescence and surface detectors. Such approach, called a hybrid registration, allows us for more precise determination of shower properties. A realization of the hybrid reconstruction idea is used e.g. in the Pierre Auger Observatory, which is described in next chapter.

3 Pierre Auger Observatory

3.1 General description

The Pierre Auger Observatory [17] was established to register extensive air showers induced by cosmic rays with energies above 10^{18} eV. Main goals of the observatory are: determination of flux, mass composition and distribution of arrival directions of cosmic rays.

The observatory is located in Argentina, Mendoza province, with the headquarter placed in Malargüe city. Malargüe is located at the south-west edge of the observatory site. This site was chosen as the location of the observatory, because it is characterised by a low light pollution, good weather conditions and existence of a large, relatively flat area, with a mean altitude 1420 meters above the sea level. These characteristics of observatory location are adequate for installation of cosmic rays detectors. The observatory was completed in year 2008, but data acquisition is running from year 2004, when the number of working detectors enabled for proper reconstruction of detected showers.

The layout of the observatory is presented in figure 18. Two main types of detectors are used: the fluorescence detector (FD) and the surface detector (SD). While the SD stations are spread over a round-shape area, all FD's are located on the perimeter of this area, having fields of view covering the whole SD site. Due to such configuration any shower registered by the SD will be also registered by at least one FD telescope (during FD acquisition). Since relative configuration of detectors allows to register simultaneously the same shower, a realization of hybrid detection of the extensive air showers is possible.

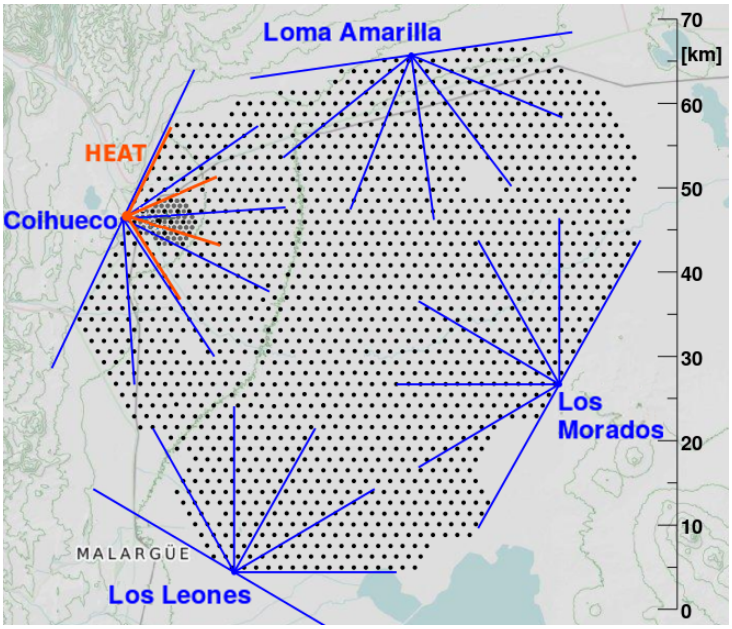

Figure 18: The Pierre Auger Observatory setup. Black dots indicate locations of the SD stations [70]. Four eyes of FD, named Coihueco, Los Leones, Los Morados and Loma Amarilla, and their fields of view are plotted (blue lines). Location of the headquarter is also marked as Central Campus.

3.2 Surface detector

To achieve a large statistics during detection of cosmic rays from the GZK cutoff region, where the flux is below 1 particle/km^2/century, the surface detector in Auger Observatory includes 1600 station units. They are uniformly spread over an area of 3000 square kilometers in equilateral triangle net. Distances between the individual SD stations are equal to 1500 m, which is sufficient for detection of showers at ultra high energies. The SD setup triggers about 10 stations for a vertical shower with energy 10^{20} eV, and typically over 20 stations for inclined shower with the same energy. Also, such configuration of SD stations allows us for precise determination of the shower geometry. For showers with zenith angles less than 60° the aperture of the SD is 7350 km^{2}sr. Almost 100% of showers with energies above 3*10^{18} eV should be registered by the SD. The SD array is also sensitive to showers arriving at large zenith angles.

The SD stations used in the Observatory are water Cherenkov detectors [71]. A schematic plot of single SD station is presented in figure 19. Each SD unit is a plastic cylinder-shape container with dimension 3.6 meters in diameter and 1.5 meters high. The inner side of the container is covered with a reflective material which reduces Cherenkov photon loses. Ultra pure water, used as the medium to produce Cherenkov radiation, is used in SD stations (thickness of water layer is 1.2m, volume about 12 000 liters).

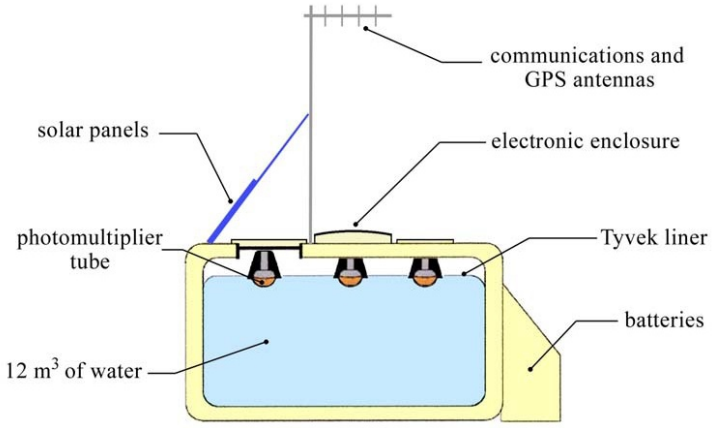

Figure 19: A schematic layout of a single SD detector [17].

Each detector is equipped with 3 photomultipliers (PMT) located at the top side of the SD station, which are used as Cherenkov signal receivers. The PMT signals are filtered, digitised and registered with 40 MHz sampling frequency. Example time distribution of the PMT signal, read from a single SD station, is presented in fig. 20.

To describe the SD signals, a Vertical Equivalent Muon (VEM) unit was introduced. The VEM units are used for a convenient description of registered SD signals. By definition: a single muon passing the SD station at its center, perpendicular to the ground, produces signals which integrated value is equal to 1 VEM. The total SD signals, as well as the PMT signal time distributions, are provided in VEM units, but the time distributions read from the PMT are also available.

The SD array operates constantly, under any weather conditions, even when some SD units are not functioning (typically more than 98.5% of the stations are operational at any time). The signals from SD stations are fed to the Central Data Acquisition System (CDAS) using radiocommunication. The CDAS, located in the headquarter of the Auger observatory, is a place where the data signals are stored in a raw form, and are accessible for Auger observatory collaborators in a raw form and as the reconstructed SD events.

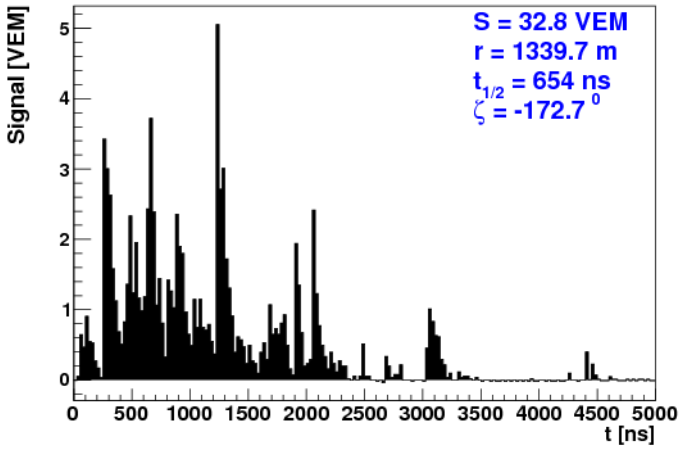

Figure 20: Example plot of signals in VEM units, read from a single PMT placed in example SD located at distance 1339.7 m from the shower core. The integrated (total) signal registered in this station is 32.8 VEM. The narrow peaks with high signal values are produced by muons. The wide background below 1 VEM are mainly from electrons.

A calibration of the SD stations is performed constantly, for each station separately, using the low energy atmospheric muons [72]. Such muons are uniformly distributed over the whole SD installation, and give a narrow and easily distinguishable peaks in the signal distribution. The calibration data are also stored in the CDAS.

From a single SD station the signal distributions from each of the three individual photomultipliers, calibration data, timing information and location on the ground are provided. These data are used to reconstruct the lateral distribution of particles in observed shower. Data SD reconstruction methodology is described in section 3.6.

3.3 Fluorescence detector

The fluorescence detector (FD) at Pierre Auger Observatory includes four eyes, which are called Los Leones (LL), Loma Amarilla (LA), Los Morados (LM) and Coihueco (CO). The FD is designed to guarantee registration of every shower with energy above 10^{19} eV, which occurs in the FD's field of view. Such shower will be registered by at least one FD eye. About 60% of such events will be registered by two or more FD eyes (stereo detection). Showers with energies above the energy of the GZK cutoff will be always registered as stereo events. Since the fluorescence light is very weak, and has the maximum of intensity in ultraviolet region, the FD operates during moonless nights only, which is 13% of the total operational time of Auger observatory. The aperture of FD

in the observatory changes from 900 km^{2}sr (for showers with energy $10^{17.5}$ eV), to 7400 km^{2}sr (for showers with energy 10^{19} eV).

A single fluorescence detector eye comprises six telescopes [17], which are placed permanently, without possibility of moving. The relative telescope configuration in an eye is shown in figure 21. Schmidt-type telescopes are used, which optics layout is shown in figure 22. They are designed in such a way to eliminate spherical and coma aberrations, but a modification of correction plate was introduced. The modification is that instead of the correction plate a segmented correction ring is installed at the outer edge of the telescope aperture [73]. Use of the correction ring enhances the telescope aperture by a factor of ~ 2. The UV filter is attached to reduce the sky background. The angular resolution of a single telescope is 0.5°, i.e. 1/3 of the pixel size of the PMT detector placed in telescope focus. This makes that a point source of light is registered mainly by a single PMT (depending on point spread function which is sensitive to atmospheric conditions).

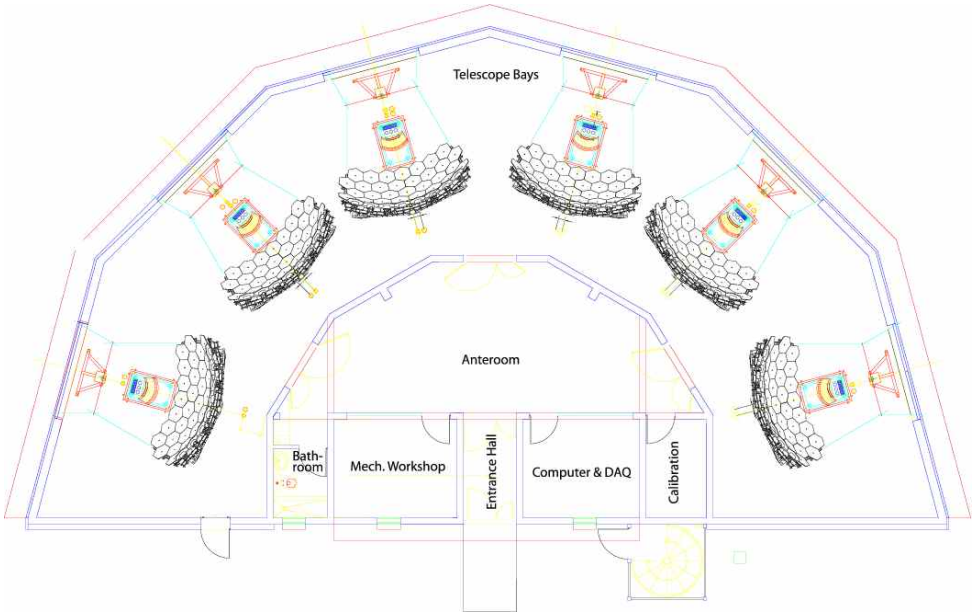

Figure 21: A single FD eye. The relative configuration of six telescopes is shown [73].

The field of view of a single FD eye covers the azimuth range of the SD site viewed from the eye location. A field of view of a single telescope is 30° in azimuth, and in elevation it covers a range 2-30.6° (in horizontal coordinate system centered in FD eye). So each FD eye has the total field of view equal to 180°x28.6°. In the focus of a single telescope a matrix of 22x20 photomultiplier tubes is installed (fig. 23). However, the areas between the PMT cathodes are inactive. So in order to reduce losses of light falling in the areas between cathodes, light collectors are applied. They reflect the light from the area between the PMT cathodes into the closest cathode.

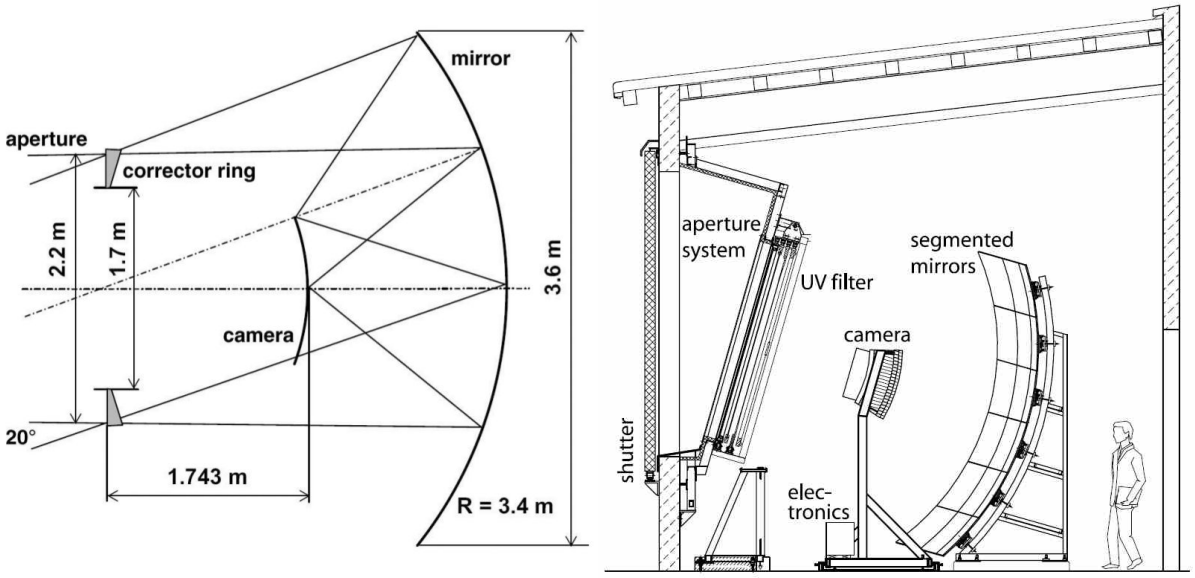

Figure 22: Left panel: schematics of optical layout [73]. Right panel: general configuration of a single telescope used in FD [73].

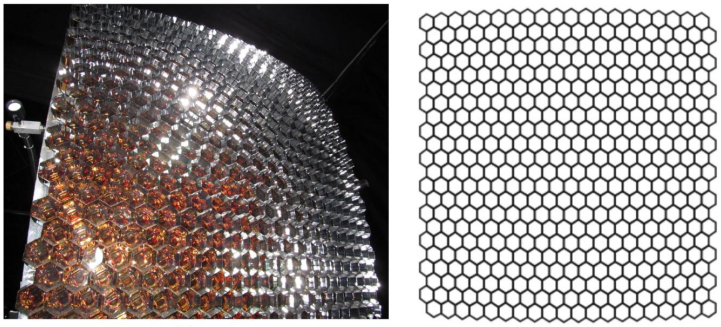

Figure 23: The PMT matrix. Left: a curvature of the focal plane is seen. Right: distribution of pixels in the total field of view of a single telescope in FD.

In order to take into account as many effects of light propagation through the telescope system as possible, and obtain the photon/ADC conversion, a calibration of FD is performed. The cumulative result of effects arising during light propagation in the optical system of FD is determined by a calibration with a drum, instead of testing the response of each telescope element separately. The drum is a 2.5 meter (in diameter) LED UV light source placed at the telescope aperture, which provides uniform light flux at the entrance pupil of the telescope. The drum allows us for absolute calibration of each PMT channel separately. Also additional calibration tools are used like relative PMT calibration before and after each data collection, relative FD spectral response in 5 different UV wavelength bands and independent absolute calibration tests performed using vertical laser shots made during data collection.

3.4 Atmospheric monitoring

A proper reconstruction of the longitudinal profile of a shower is crucial since the profile defines location of shower maximum, i.e. the directly measured mass-sensitive observable. During propagation from shower to detector the emitted photons are scattered on molecules (Reyleigh scattering) and on aerosols (Mie scattering) which are present in the air. The atmospheric conditions influence the location of reconstructed shower maximum [74]. These atmospheric effects must be taken into account during data reconstruction to reconstruct properly the shape of the longitudinal profile.

The SD array is also sensitive to the atmospheric conditions [75]. It was shown that the SD signals are modulated by the variations of the atmospheric conditions (temperature, pressure and density of air at ground level). This influences the SD energy estimator (SD signal at 1000 m from the core). Such dependence is caused mainly by air density-dependent Molière radius near ground, describing intensity of particle scattering. A decrease of air density increases Molière radius and thus broadens the lateral distribution of a shower, and vice versa. Also the development of a shower is pressure-dependent, since the air pressure determines the amount of matter in which the shower develops. Furthermore, the seasonal modulations may affect a search for large scale anisotropies of arrival directions.

During reconstruction of detected showers the atmospheric conditions, which were present during data collection, are taken into account. Thy are read from the Global Data Assimilation System database (GDAS), which is a weather conditions database provided by the National Center for Environmental Information [76]. Despite the GDAS, an additional monitoring of atmospheric conditions is performed locally in 5 weather stations, from which four are located at the four FD sites, and one in the center of the SD array (see fig. 24). The registration is performed in 5 minutes time intervals. In the early years of the Auger observatory, 331 balloon measurements of atmospheric conditions in different seasons were performed. The balloons were used to measure the vertical profiles of the atmospheric parameters above the observatory site. From these measurements the mean seasonal profiles were prepared to be available for shower reconstruction. The weather stations and the balloon measurements in the observatory were used to check the reliability of the GDAS data.

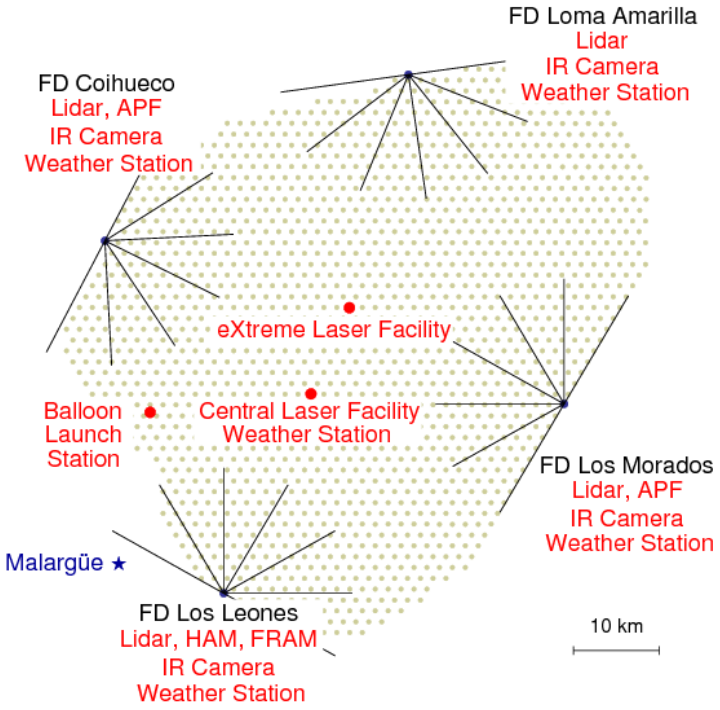

Figure 24: Atmospheric monitoring facility locations [77].

The balloon measurements were not continued when it turned out the GDAS data are appropriate, and no longer balloon data collection was necessary. Existing weather stations still perform measurements to supplement the GDAS data.

The constant direct measurements of aerosol density are necessary [78]. Currently 80% of registered hybrid events are registered with aerosol measurements. Hourly measurements of vertical aerosol optical depths are performed using two central lasers and four lidar stations. The aerosol monitoring is performed using calibrated laser beams shot in pulses at wavelength 355 nm. The laser facilities (named as Central Laser Facility and eXtreme Laser Facility) are located around the center of the Auger site. If the scattered laser light propagates in the same direction towards given FD, it is attenuated in the same way as fluorescence light coming from a shower [79]. This property allows us for adequate determination of aerosol concentration in the atmosphere during data acquisition. The power of laser beams is equivalent to the power of light emitted by a shower with energy 10^{20} eV [80]. The calibration shots occur every 15 minutes. The light from laser shots are registered by FD at specific GPS timings which is used to distinguish the laser shots from the light produced by the particle showers. The light registered in FD can be compared with the light registered in very clear nights (so-called reference nights). This comparison is used for determination of the scatter magnitude on aerosols, and subsequently allows us for determination of the optical depth of aerosols. Also the LIDAR installation is able to determine aerosol profile at many elevation angles.

The magnitude of aerosol scattering is measured by two devices placed 1 km from the CO and LM. The aerosol scattering measurements use a laser-based device which shoots horizontal beams in two wavelengths (350 and 390 nm). The beams pass through the field

of view of nearby FD. After registration of light scattered from such beam by FD, the phase function for aerosol scattering can be determined and used in shower reconstruction. A wavelength-dependent attenuation is also monitored by Horizontal Attenuation Monitors.

Cloud presence during observations deforms the registered fluorescence light from a shower. Such deformation is seen as attenuation if a part of a shower is developed in the clouds, or as amplification if the shower light is reflected from the clouds to the FD. The cloud correction is very difficult to calculate so the FD data gathered in cloudy nights are usually rejected. The cloud monitoring is performed using infrared cameras installed at each FD site. The cameras make one picture every 5 minutes for field of view the same as the FD field of view. Every 15 minutes all the sky is monitored. Besides, the cloud altitude and thickness is monitored by the LIDAR stations.

3.5 Reconstruction of hybrid events

Each shower registered by the FD causes a sequence of triggered pixels in the PMT matrix of a telescope. When a characteristic sequence of signals and timings indicate a possible event registration, the signals and timings are fed to the CDAS. An example of signal sequence triggered by a shower is presented in figure 25. Signals have a characteristic distribution in time as shown in the figure.

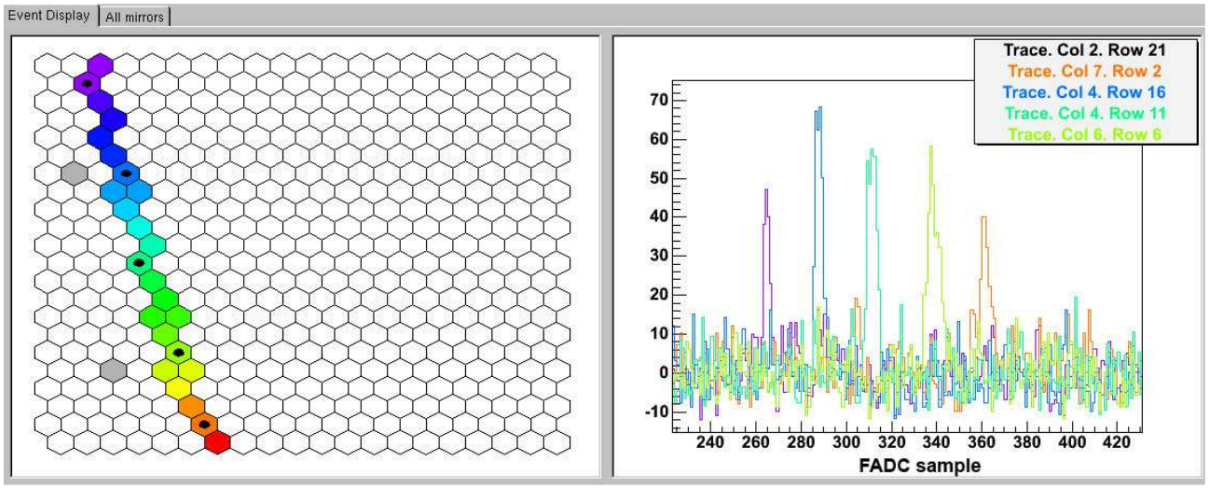

Figure 25: Visualisation of detected shower in FD [73]. The shower is seen in a form of track of the triggered pixels. A characteristic distribution of signals for 5 chosen pixels (denoted using black dots) is shown.

The reconstruction algorithm may be divided in the following steps. The first step is a determination of the shower-detector plane (SDP) using the observables from the FD. The SDP is a plane passing through the center of the FD telescope and the shower trajectory, as shown in figure 26. A shower track seen in the FD telescope lies on the SDP, when the track is projected on the sky seen from the FD site.

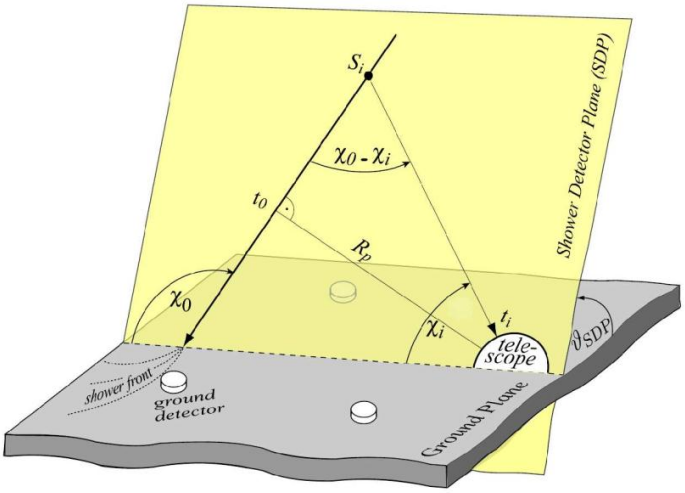

Figure 26: The geometrical configuration used during reconstruction [73].

Minimisation of the following chi-square quantity is used for determination of the SDP:

$$\chi^2_{\mathrm{SDP}} = \sum_{\mathrm{i}} \frac{(\vec{n_{\mathrm{i}}}\vec{r_{\mathrm{i}}})^2 w_{\mathrm{i}}}{\sigma_{\mathrm{SDP}}},$$

where the n_{i} is a unit normal vector of i-th pixel, r_{i} is a pointing direction at element of shower track seen in the i-th pixel, w_i is a weight for given individual pixel signal size, $\sigma_{\mathrm{SDP}} = 0.35^\circ$ is a constant of pointing accuracy. Summation index i is going on the individual triggered pixels only.

Next, a determination of the shower axis (lying on the SDP) is performed using the timing information from the triggered pixels in FD. The shower axis, lying on the SDP, can be described by two parameters: perpendicular distance (R_{p}), which is the closest distance between the shower axis and the telescope, and the angle between the shower axis and the horizontal line in the plane (χ_0). The direction of shower axis is obtained from the timings in individual pixels minimising the following quantity:

$$\chi^2_{\mathrm{Time}} = \sum_{\mathrm{i}} \frac{(t_{\mathrm{i}} - t_{\mathrm{meas,\,i}})^2}{\sigma_{\mathrm{i}}},$$

where σ_{i} is accuracy of i-th time measurement, t_{i} is a function of geometrical parameters dependent on the shower-FD configuration, and can be expressed as:

$$t_{\mathrm{i}} = T_0 + \frac{R_{\mathrm{p}}}{c} \tan\left(\frac{\chi_\circ - \chi_{\mathrm{i}}}{2}\right),$$

where c - the light velocity, T_0 - time when the shower front passes the point of closest approach R_{p}, χ_{i} - angle between the direction of i-th pixel and the horizontal line of the FD. When the measured angular speed $d\chi/dt$ does not change significantly during registration, the accuracy of parameters obtained using this equation is small. This occurs when the shower trace seen in FD is short. In such case the data points do not allow us for appropriate determination of the shower direction in SPD. This is why an additional timing information from the SD stations are taken into the fit, which improves significantly

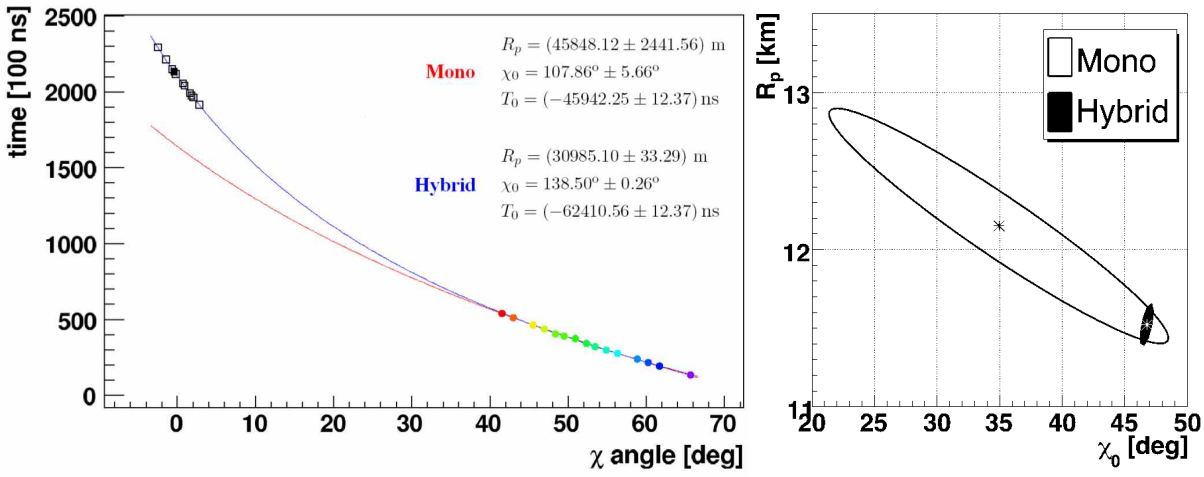

Figure 27: Left panel: Timings of registered signals used in the reconstruction [73]. The black squares correspond to the triggered SD stations. These timings are used in equation for t_i. Right panel: comparison of core location error ellipse in hybrid reconstruction and reconstruction without timings of SD (mono reconstruction).

the accuracy of geometry determination (fig. 27). This is a characteristic issue of the hybrid reconstruction and allows to reduce uncertainty of the shower axis determination.

Now the shower profile reconstruction is possible, which is the last stage of FD reconstruction. The shower distance determined in the previous steps is necessary to take into account atmospheric conditions, and to calculate properly the attenuation and scattering of the light arriving to the FD aperture. The profile reconstruction is performed using the signals registered in the PMT matrix. The signals are transformed into the flux of fluorescence photons. The total light registered in FD is composed of the fluorescence light coming from the shower, and additionally from direct and scattered Cherenkov light produced by high-energy charged particles in shower. The attenuation and multiple scattering of light is taken into account during reconstruction. The final result of this step is a longitudinal profile with errors (fig. 28).

The reconstructed energy of the primary is calculated from the longitudinal profile of energy deposited in the atmosphere. First, the calorimetric energy is calculated by integrating the profile:

$$E_{\mathrm{cal}} = \int_0^\infty \frac{dE}{dX}(X)dX.$$

The total primary energy is a function of calorimetric energy. Additionally the primary energy includes also so-called invisible energy factor f_{inv}, describing the part of the shower energy which is taken away by neutrinos and high-energy muons and hadrons which are not registered. This energy is taken into account by invisible energy correction f_{inv} in the following way:

$$E = (1 + f_{\mathrm{inv}})E_{\mathrm{cal}}.$$

f_{inv} value is about 10%. The overall accuracy of primary energy determination is about 14% due to the systematic uncertainty.

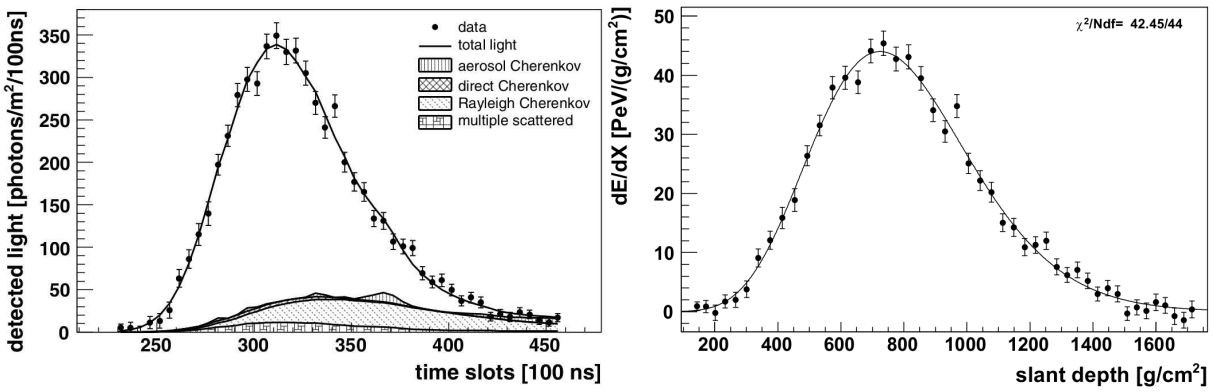

Figure 28: Left panel: example of a light at aperture measurement (dots) and reconstructed light sources (hatched areas) [73]. Right panel: Energy deposit profile reconstructed from the light at aperture shown on left side [73]. The line shows a Gaisser-Hillas fit of the profile. The reconstructed energy is 30 ± 2 EeV.

3.6 Reconstruction of SD events

The SD reconstruction of a shower is performed using an iterative fit algorithm. The main steps of this algorithm are the following. The SD reconstruction begins from initial determination of shower impact point on the ground and shower axis orientation. A signal-weighted barycenter of the stations involved in the fit is used as preliminary estimation of the impact point. This point is used as beginning of the coordinate system in iterative SD reconstruction procedure. A planar shower front, moving with the speed of light, is preliminary used as a real shower front approximation. In next steps of the iterative procedure the shower front is assumed to be curved. After determination of impact point, the analysis of timings in individual stations is used for determination of orientation of the shower axis.

Having established the SD geometry of the shower, the lateral distribution of the signals $S(r)$ is to be calculated. As it was mentioned in section 2.6, the signal in distance 1000 m from the core (S_{1000}) is used as the energy to determine the lateral distribution. So in Auger observatory the $f_{\mathrm{LDF}}(r)$ function has such form to be equal 1 at distance 1000 m:

$$f_{\mathrm{LDF}}(r) = \left(\frac{r}{1000m}\right)^{\beta}\left(\frac{r+700m}{1700m}\right)^{\gamma},$$

where β and γ - steepness parameters. The $f_{\mathrm{LDF}}(r)$ is a form of the NKG function [58]. Thus $S(r)$ is modeled as:

$$S(r) = S_{1000} f_{\mathrm{LDF}}(r).$$

The result of SD reconstruction can be visualised (see fig. 29). As was mentioned in section 2.6, the energy estimator of the primary is S_{38}.

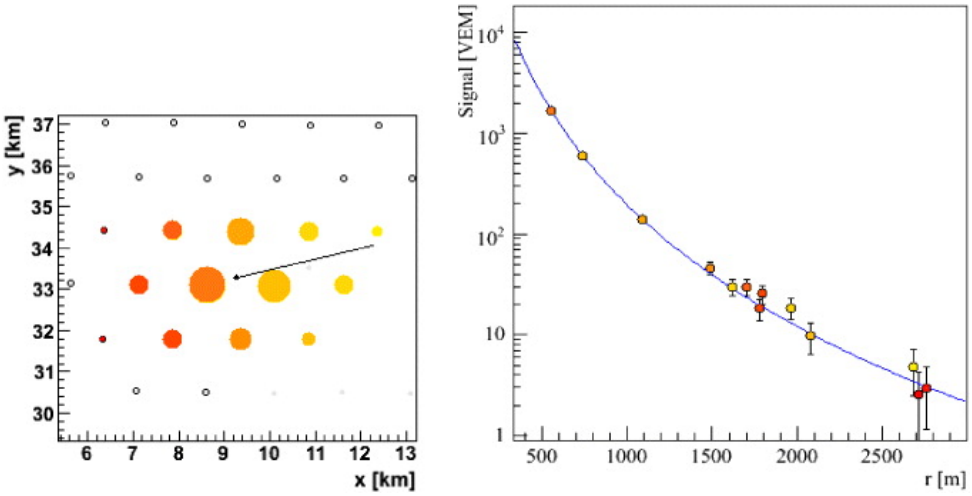

Figure 29: The visualisation of reconstructed SD event. Left: triggered stations are colored with yellow/orange. The color intensity depend on the trigger time (the lighter color the earlier arrival time of shower front). The line shows location of the impact on the ground and azimuth at he shower axis. Right: the lateral distribution of reconstructed SD event.

3.7 Offline software

Offline is a dedicated software framework, which is used to data analysis in the Pierre Auger Observatory [81]. The main purposes of this software are: simulation and reconstruction of collected data, simulation of calibration techniques. Investigation of air shower physics is also possible using the Offline. This software is written in C++ and allows to work with many data formats, including CERN Root and CORSIKA format. Also the GDAS database is available for the Offline software. It is also possible to include user procedures during compilation. The Offline is one of the basic tools used during this study.

The Offline consists of three main parts: processing modules, data modules and detector description. The structure of the modules and their dependencies are presented in figure 30. Sequencing of the modules is performed by instructions provided in an XML file. Such approach is used to separate the data from the algorithms operating on the data. The three main parts of Offline are complemented by a set of foundation classes and utilities for error logging, physics and mathematical manipulation. Also a package for manipulation of abstract geometrical objects is available. The Offline framework parts have the following properties. The detector description includes read-only detector characteristics, i.e. the configuration, size and status of the detectors (FD and SD). Also the data from atmosphere monitoring and periods indicating when the FD was operated successfully (uptime information) are included.

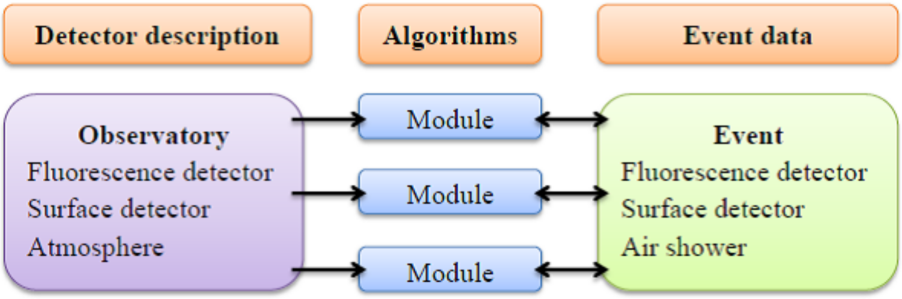

Figure 30: The Offline structure.

The processing modules, stored in XML-files, are used to provide object-oriented data analysis. During the analysis a special file (ModuleSequence.xml) includes a sequence of instructions which call the desired analysis module. A user can arrange a sequence of instructions for his own purposes, also additional new analysis sequences may be written and called for, during the Offline run. This provides flexibility of the Offline software in data analysis run. Standard settings in modules can be easily changed by using booststrap.xml file, where the standard values of all parameters may be overridden.

The data modules organise the event structure during data analysis. The event includes data from detectors and air shower. The data module allows to prepare any kind of data on output (raw, calibrated or Monte Carlo). The output has ROOT file format called Advanced Data Summary Tree (ADST). Such format was designed for purposes of the Auger observatory for storage of data analysis results. The ADST format allows us for writing of FD and SD variables which can be used for further physics-oriented or instrumental analysis.

4 Extensive air shower simulations

4.1 Introduction

An interpretation of air showers induced by the cosmic rays relies on simulations of shower development. Therefore the ability of extensive air shower simulations is the integral part of the data analysis. Currently no analytical model describing detailed shower development is available, so to perform such simulations the Monte Carlo approach is used. The simulations are based on hadronic interaction models and include calculation of particle transport and modelling of interactions between the shower particles and the air molecules. The used hadronic interaction models were tested and tuned using accelerator data at energy 7 TeV in the center of mass (which is equivalent to cosmic ray energy 10^{17} eV). On the other side the energy of interactions, which occur during air shower development, reach 10^{21} eV. Since the interaction models were tested at energies much below the energies required in the air shower simulations, the extrapolations of interaction models are used in air shower simulations.

For electromagnetic and weak interactions the model predictions at energies higher than energy accessible in accelerator measurements are consistent with air shower data. A use of the hadronic interaction models at higher energies is problematic, since the models are a mixture of fundamental theoretical ideas and empirical parameterisations based on accelerator data. So the extrapolations of the accelerator data are used to provide hadronic interaction models at higher energies. In effect both, the empirical parameterisations and their extrapolations beyond the energy region accessible in acceleration measurements, bring uncertainty in interaction models. Such uncertainties may impact on the air shower simulations and subsequently on the data interpretation [82].

The impact of the model uncertainties on the observables of simulated showers were studied numerically e.g. in [83]. Such studies were performed in the following way. For given values of observables (cross section, elasticity or multiplicity) a set of simulations was performed. Next, the set of simulations was repeated from the beginning, but each time a different relative value of cross section was used, while the other properties remained unchanged. The same strategy was applied for another sets of simulations where the elasticity only, and subsequently multiplicity only was changed. Due to such studies it was shown that the shower observables are sensitive to the model uncertainties with the same magnitude as they are sensitive to the primary mass [83]. A short summary, showing shower observables sensitivity to model properties changes, is shown in fig. 31. It is seen that change of a given interaction model property (cross-section, elasticity or multiplicity) influences the shower characteristics. For example, when the cross-section increases (while other properties are unchanged), the $X_{\max}$ and its dispersion decreases. On the other side when only the elasticity increases, the $X_{\max}$ and its dispersion also increases. So changes of elasticity induce reverse behaviour of $X_{\max}$ and its dispersion when compared to the changes induced by the cross-section. Similarly, the N_{e}, as well as the RMS of $X_{\max}$, are under influence of cross section changes. The sensitivity of N_{e} and RMS $X_{\max}$ differs, since

the N_e and its dispersion is much more sensitive for multiplicity than X_{max} (dispersion of X_{max} is almost insensitive for multiplicity). The uncertainty of multiplicity influences also on N_μ, which is positively correlated, while the N_e is negatively correlated with the multiplicity changes. Such dependencies between the shower observables and changes of interaction model properties may be used for estimation of model characteristics, but external knowledge on the primary mass is needed.

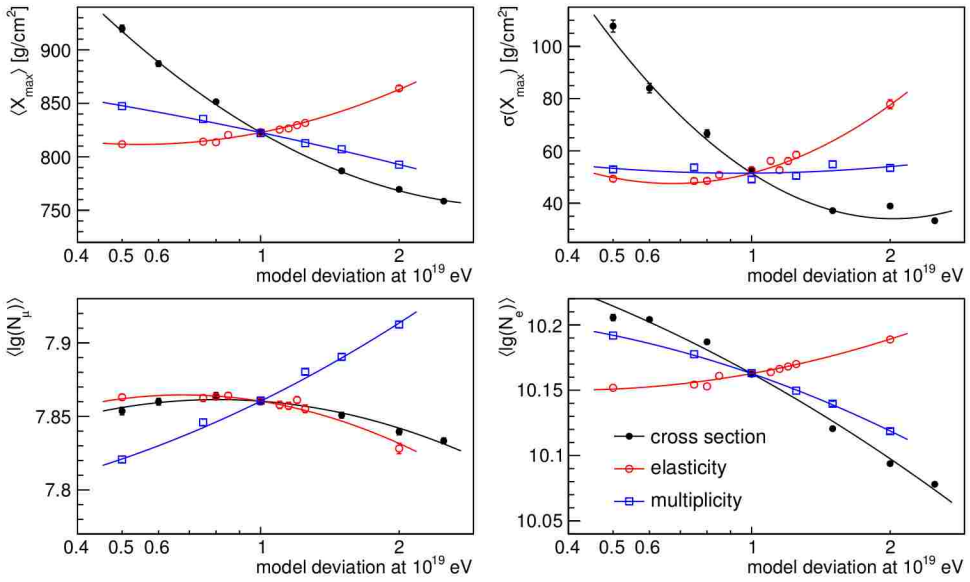

Figure 31: The plots show the sensitivity of shower parameters (N_μ, X_{max} and their dispersions) to changes of interaction model properties (cross section, elasticity and multiplicity) [84]. The characteristics were obtained through independent simulations, performed with assumed different values of model properties. For example, the sensitivity of the muon number and electron number on the cross section (black line) was obtained from simulations in which only the relative value of cross section was changed (elasticity and multiplicity were kept constant with value corresponding to model deviation equal to 1). The sensitivity of the muon number and electron number on the elasticity and multiplicity was obtained in a similar way as for the cross section.

4.2 Interaction models

The most recent versions of EPOS and QGSJet hadronic interaction models are retuned to achieve consistency with the LHC accelerator data measured at energy 7 TeV in the center of mass system. These models are denoted as EPOS LHC and QGSJetII-04, and from now will be called as LHC-tuned models. Their status in air shower physics is as follows. It was shown that the LHC-tuned models give more consistent results in shower development simulations when compared with the pre-LHC models [85]. As an example, the elongation rate obtained in shower simulations using the LHC-tuned models, is compared with the results obtained with the pre-LHC model versions (figure 32).

It is seen in the plot that the pre-LHC models (black and green lines) have different elongation rate. In effect the mean X_{max} difference between the EPOS LHC and QGSJetII-

04 models changes for proton from 0 g/cm^2 at energy 10^{17} eV to about 50 g/cm^2 at 10^{20} eV. Such difference may influence the mass composition studies when the pre-LHC models are used. On the other hand the LHC-tuned models (blue and red lines) provide more similar elongation rate to each other. Only a small difference of $X_{\max}$ ($\Delta X_{\max} \sim 20$ g/cm^2) between the LHC-tuned models is present. Such $\Delta X_{\max}$ is close to the experimental systematic error in Auger observatory. For iron the elongation rate is also similar for the LHC-tuned models.

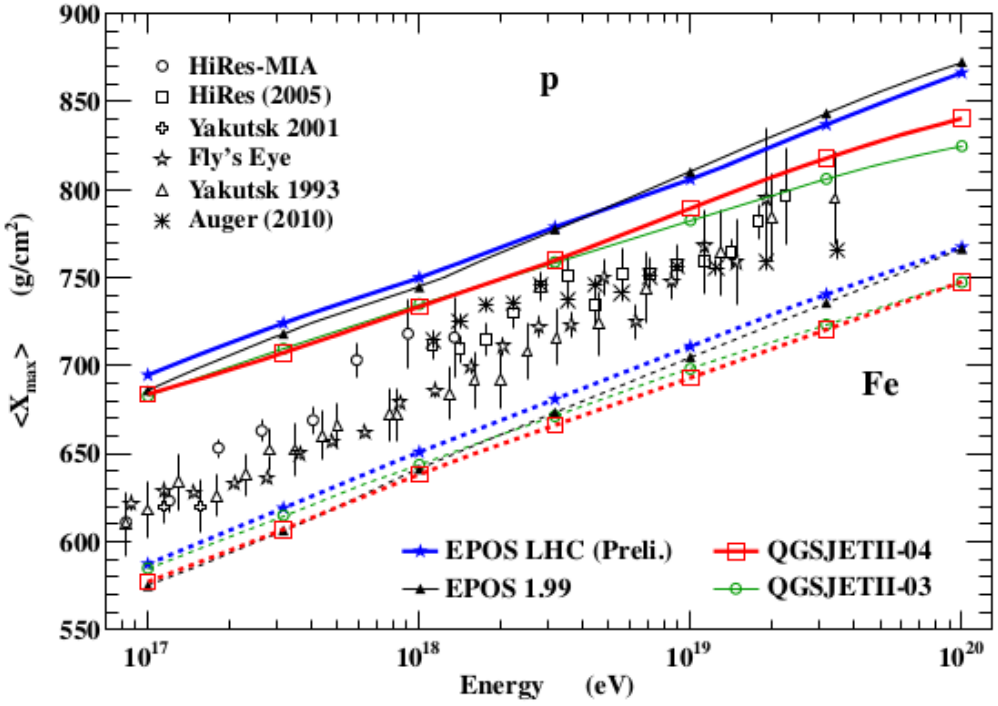

Figure 32: Mean Xmax for proton- and iron-induced simulated showers in function of the primary energy [86]. Solid lines indicate proton and dashed lines indicate iron as primary. Full triangles - EPOS 1.99, open squares - QGSJETII-04, open circles - QGSJETII-03, full stars - EPOS LHC. Real data are also presented.

The muon number predicted by the LHC-tuned models shows 30% to 80% deficit at 10^{19} eV. Such disagreement in muon number between the data and simulations influences on the interpretation of measurements. It is clearly seen in analysis in which the composition of cosmic rays was derived using muon number as a mass-sensitive quantity [61]. In the analysis the showers with zenith angle larger than 60° were used, since they provide almost a direct measurement of the muon number (electromagnetic part is largely attenuated). The composition derived using estimated muon number seems to be dominated by iron or more massive elements for all energy ranges (see fig. 33, left panel). This stands in opposition to the other composition measurements based on the $X_{\max}$. The main source of the discrepancy above is that the simulations provide lower muon number than detected.

Since the more massive element the larger muon number, the data compared with simulations suggests that the cosmic rays are composed of more massive elements. This

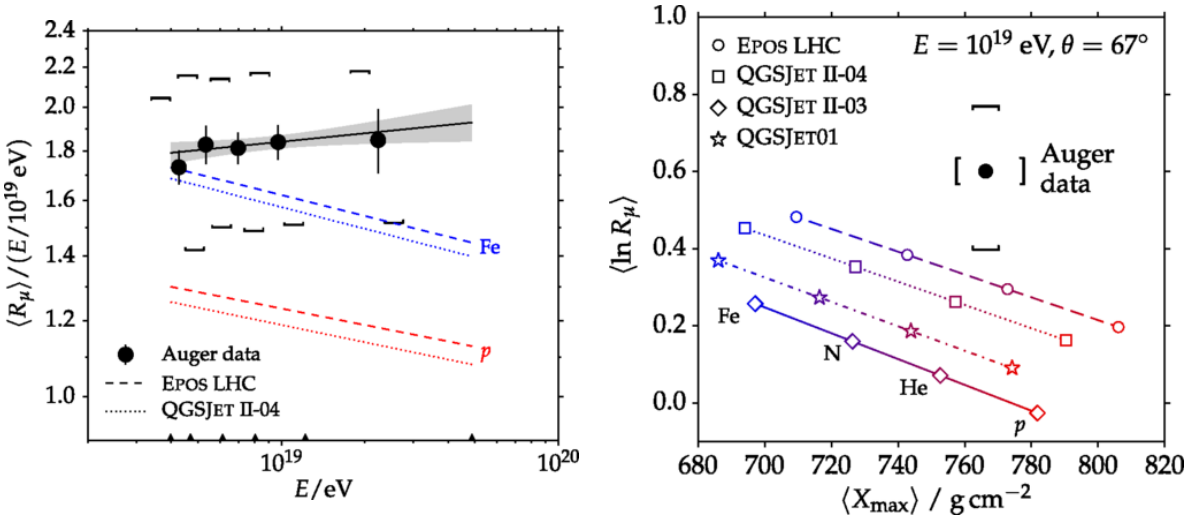

Figure 33: Left panel: Average muon number in air showers scaled to reference muon number derived from simulations of showers at energy 10^{19} eV [61]. The dependence of the muon number on the energy suggests heavy composition for all energy ranges. Right panel: Average logarithmic muon number in function of the average shower depth for used data set [61].

is clearly seen when the simulated muon numbers and X_{max} are compared with the data (fig. 33, right): the used data sample lies outside any model used. Such disagreement can be removed by increasing muon number by 30-80%. Some aspects of hadronic interactions uncertainties, and implications on the shower observables, are described in section 4.1.

4.3 Thinning

A precise simulation of air shower development would require tracing of all particles in the shower. Since at shower maximum the number of particles induced by UHECR scales with energy (appriximate number of particles is roughly equal to E in GeV), the simulation time of all particles in a single shower would take weeks. Besides, the amount of output data after simulation of all particles would require a terabyte disk space for one shower. Therefore a statistical thinning algorithm (thinning) was proposed by Hillas [87] to increase effectiveness of CPU and disk space usage in Monte Carlo simulations of shower development.

When the thinning is applied, not all particles are simulated, but only some randomly chosen representative part of the total number of particles. The thinning is used for all secondary particles which have energy smaller than a chosen energy level, described by the thinning level $\varepsilon_{\rm th} = E/E_0$. During simulations this condition is applied every time when particle production occurs. If the sum of the energies of simulated particles falls below the thinning energy:

$$\sum_{\rm j} E_{\rm j} < \varepsilon_{\rm th} E_0,$$

only one of the secondary particles is followed in simulations. This particle is selected randomly, with the probability:

$$p_{\rm i} = \frac{E_{\rm i}}{\sum_{\rm j} E_{\rm j}}.$$

All other particles are discarded. For the tracked particle a weight $w_{\rm i} = 1/p_{\rm i}$ is attributed, which allows us for energy conservation. In case when only some part of secondaries falls below the thinning level, the particles are selected with probability:

$$p_{\rm i} = \frac{E_{\rm i}}{\varepsilon_{\rm th} E_0},$$

and are attributed with weight $w_{\rm i} = 1/p_{\rm i}$. The gain of thinning is that a rather constant number of particles is followed in simulations, instead of their exponentially growing number.

The thinning artificially increases fluctuations, which depend on the thinning energy. To reduce thinning fluctuations the particle selection rule was introduced. This rule states that particles with weights larger than some chosen $w_{\rm lim}$ are not thinned, but all their secondaries are tracked. Such modification increases CPU usage and size of output data. However, this thinning trick can be optimised such that the artificial fluctuations will be minimal for a given CPU usage [88]. The optimum $w_{\rm lim}$ can be calculated from expression:

$$w_{\rm lim} = \log(E_\circ)\varepsilon_{\rm th}.$$

Example fluctuations in lateral distribution for chosen particle types, and for optimum and not optimum thinning values, are shown in figure 34.

The increased amount of output data, caused by use of the optimal thinning method, can be reduced if additional weighting of particles is provided. It is so-called radial thinning, which is used in regions close to the core of the air shower, where the number of particles is so large that the detectors are saturated. When the radial thinning is used, only a small fraction of the total number of particles, which are located in the region close to the core, will be written to the output file. The probability of particle retain in the radial thinning is proportional to $(r/r_0)^k$, where $r_\circ \sim 150$ m, k=2,...,6. The data amount is reduced by a factor of 2 to 5, without a significant increase of statistical fluctuations in the regions closer than r_0 from the core.

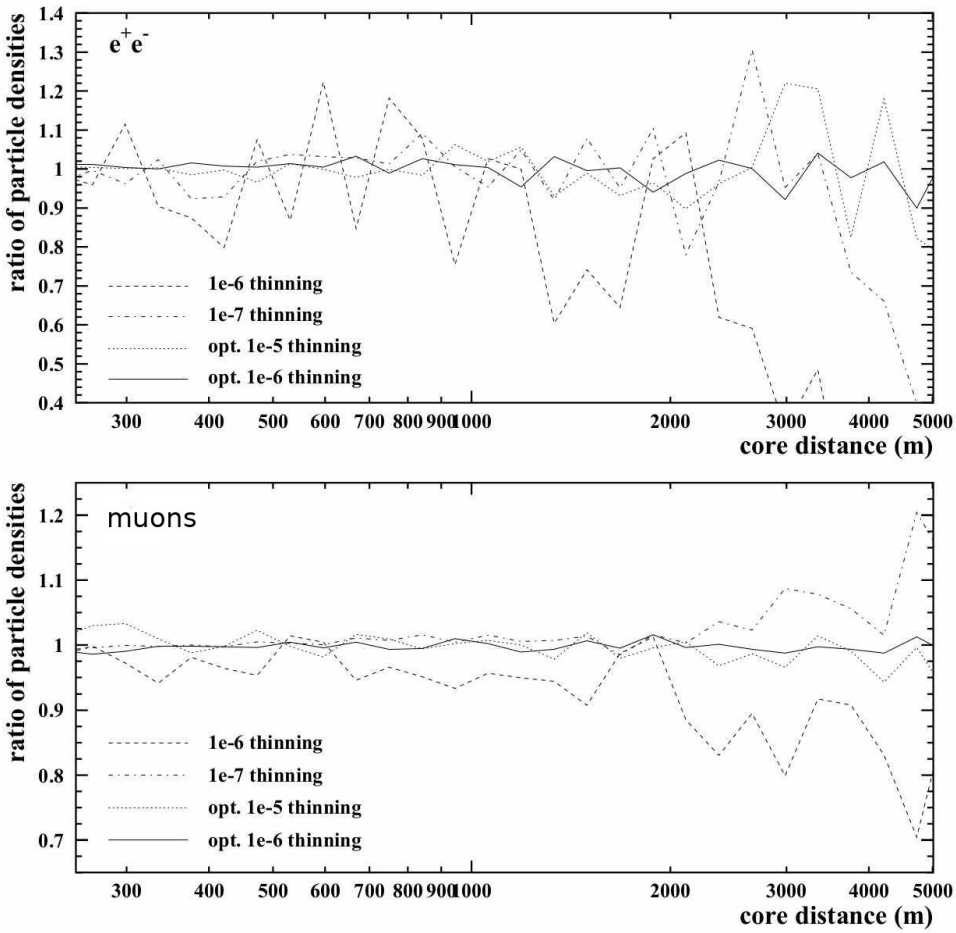

Figure 34: Comparison of fluctuations between the optimal (solid lines) and not optimal thinning (dashed lines) [88]. Eight vertical proton showers per each thinning level were averaged and their mean lateral distributions were divided by the reference lateral distribution.

4.4 CORSIKA software

CORSIKA is a Monte Carlo simulator of extensive air shower development [89]. The simulations of particle interactions in CORSIKA are performed with high accuracy, which was achieved by taking into account as many processes as possible, including transport

mechanisms of particles through the atmosphere and interactions with the air molecules. The simulation of interactions with the air molecules includes calculation of ionization energy losses as well as the Coulomb multiple scattering. Also the decays of produced secondary particles are taken into account with all known modes of decay, taking into account its branching ratio down to 1%. Additionally the deflection in the Earth's magnetic field is calculated. The simulator also includes the thinning algorithm.

The atmosphere simulated in CORSIKA is composed of nitrogen (78.2%), oxygen (21.0%) and argon (0.9%). The density of atmosphere is parameterised in 5 layers. It is possible to set any desired atmosphere profile at input. Also a curved atmosphere is available in simulations.

During CORSIKA run the particles selected by the thinning algorithm are tracked, therefore many information are accessible for further simulations or analysis. These data are stored in a file for chosen observation level. Such approach allows us for detailed study of shower features. CORSIKA uses the EGS4 (Electron Gamma Shower system version 4) code [90], which is dedicated for detailed MC simulations of all electromagnetic processes, including the processes with small cross-section like direct $\mu^+\mu^-$-pair production and the photo-nuclear reactions with photons and neutrons of nuclei of the atmosphere. Many high-energy interaction models are available in CORSIKA (e.g. EPOS, QGSJet, SIBYLL, DPMJET), as well as low energy interaction models (GHEISHA, UrQMD).

A run of CORSIKA is steered by keywords, in which many parameters may be set by the user. Among them are parameters related with shower geometry, energy cutoffs, atmosphere profiles, magnetic field, algorithms of particle tracking, format of simulation output, etc. The keywords are stored in a text file which can be easily manipulated by the user.

4.5 CONEX

The CONEX software [91] [92] was developed as a separate package, which now is implemented in CORSIKA. The program is used for a fast, one-dimensional simulations of shower profile. Two main stages of simulation are included during a CONEX run. In the first stage an explicit MC simulation of particle cascade is performed at energies above some energy threshold. This stage is continued as long as all hadrons have energies above the threshold, which usually is 1% of the energy of primary particle. The characteristics of all sub-threshold particles are written and subsequently used as the source terms for numerical solution of the cascade equation. In the second stage a solution of nuclear-electro-magnetic cascade equations is performed. The solution is obtained for all sub-cascades in smaller regions, calculated along the direction of the shower axis.

CONEX and CORSIKA use the same run keyword parameters. Therefore the same steering files can be used for simulations in both programs, with appropriate keyword switching between the two programs. This makes CORSIKA a very suitable tool for cosmic ray research.

5 The top-down reconstruction method

5.1 Introduction

The simulations of extensive air showers predict too small amount of muons, which is known as muon number problem (see section 4.2). Since data interpretation rely on simulations, the problem with muons induces deep implications in data interpretation, suggesting much heavier composition of cosmic rays than composition derived with $X_{\max}$ (see chapter 4.2). Even for the interaction models lately tuned to LHC measurements the muon problem is still present, despite the tuned models provide consistent results with each other. The reason for it is a lack of accelerator data to tune the models at energies which occur in shower development. While the used interaction models were tuned to reproduce accelerator data at 7 TeV in center of mass, the shower simulations are performed at the energies even 100 times more. So to perform simulations the extrapolations of interaction model properties are used. Simulations performed with such models provide data interpretation suggesting much heavier composition of cosmic rays, which stands in clear opposition to the other cosmic rays studies (see chapter 4.2).

In order to reduce the muon number problem, a top-down reconstruction method was proposed [93]. The main aim of the method is to quantify the discrepancies in simulations and to provide useful hints which will help to improve the interaction models and minimise the muon number problem. The top-down method rely on simulations which use hadronic interaction models. For a given data shower the main aim of the top-down reconstruction is to predict signals in detectors on simulation basis. Therefore the method looks for a simulated shower which has a particle distribution of electromagnetic origin along the shower axis (longitudinal profile) similar to the longitudinal profile of data shower (a reference profile). The reference longitudinal profile is linked with the electromagnetic part of shower, so the method relies on the fact that this part is properly simulated. As data shower, which is the input to the top-down, the air shower detected and reconstructed in a standard way is used. Such detected and reconstructed air shower (an event) includes values of signals in detectors from real data shower. As an output, the top-down provides a reconstructed event, but here the signals in detectors are determined using Monte Carlo simulations. The simulated SD signals in output shower, which are the fingerprints of interaction models, may be then compared with the data. Thanks to such approach the method links the properties of interaction models with the properties of particle showers.

Shower development is characterised by large fluctuations. So in order to generate shower, which have shape of longitudinal profile close to the profile of data shower, many simulations are required. The simulations are performed by CONEX or CORSIKA and provide Monte Carlo showers (MC showers). Among many simulated showers, one or few of them will have the shape of longitudinal profile similar to the data profile. So the method reproduces the set of fluctuations which lead to shape of the longitudinal profile similar to the profile in data shower. For the MC shower with a profile similar to the data shower the observatory response is simulated, providing the SD signals. The SD signals include

also the muon signals which are traces of hadronic interaction properties. Comparison of simulated SD signals with the corresponding signals in data shower brings opportunity to investigate the lateral distributions of simulated showers. Since the distributions are sensitive to the hadronic interaction models, analysis of the distributions provides an opportunity to investigate indirectly the interaction models at energies above the energy at which the accelerator data are available. So it is expected that the top-down method will allow us to calibrate the interaction models, and to reduce discrepancy between the data and simulations.

The simulation procedure in top-down method takes into account the available information about the fluorescence detector, state of the atmosphere and the detector characteristics, to exclude possible sources of systematic errors in results. In such simulations the axis orientation of MC shower, seen from the observatory site, has the same orientation as the data shower axis (axes of both showers have the same zenith angle and azimuth). Additionally the location of MC shower axis on the ground covers the location of data shower axis, so they are equally distant from the FD. The primary type is assumed, and a narrow energy range is set, centered at the energy of data shower ($E_0^{\mathrm{sim}} = E_0^{\mathrm{data}} \pm \sigma_{E_0^{\mathrm{data}}}$). For the MC shower, which is most similar to data shower, the observatory response is simulated, providing as output so-called MC event. The MC event includes the simulated SD signals, which can be compared with the data SD signals to study details of shower development. An example comparison of SD signals obtained in the top-down reconstruction is shown in figure 35. Since current hadronic interaction models do not reproduce properly the SD signals, the attempt of overall rescaling of hadronic shower was done to provide more consistent description of the simulations with the data [93]. It was shown that energy rescaling is not necessary for showers with energy 10^{19} eV, if mixed composition is assumed. Only the muonic signals are higher in data by 30 to 60% depending on the model. A source of this discrepancy is not certain.

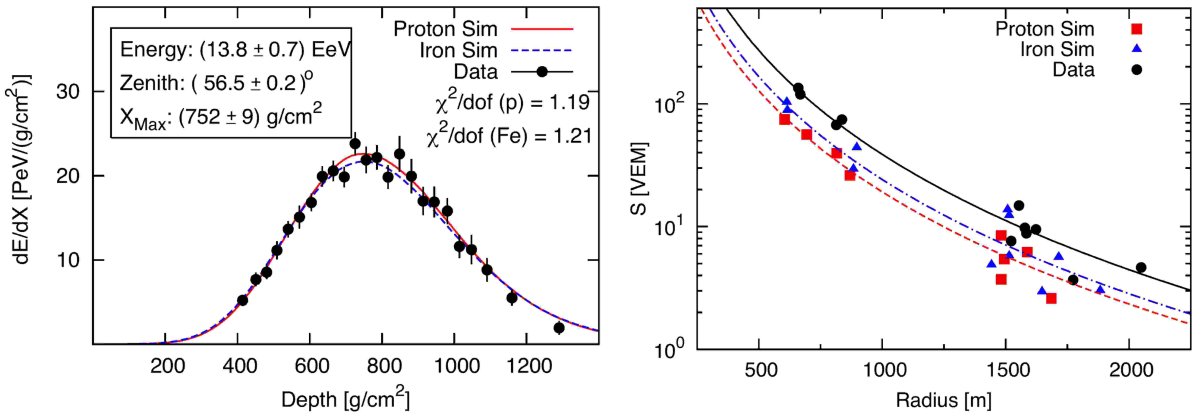

Figure 35: For a given shower the simulations performed with proton and iron as primary reproduce the longitudinal profile correctly (left panel). On the other side the same simulations produce too low number of particles in hadronic part of shower, hence the simulations give systematically lower signals in detectors (right panel). Figure taken from [93].

5.2 Scheme description

The top-down reconstruction chain used in this thesis is an improved version of the reconstruction chain described e.g. in [93]. The main improvement is the use of CONEX and CORSIKA as the main simulation tools, instead of SENECA software. The CORSIKA was developed intensively during last years, so when compared with the SENECA, CORSIKA currently provides more reliable and accurate simulation results. Other minor improvements include the use of real atmosphere models based on fit to GDAS database, applying SD core asymmetry removing, use of core location on the ground which is consistent with core location of data shower, and a comparison of SD signals in results. All of the improvements are described in detail below.

The simulation chain is divided into two main parts. The main goal of the first part is to provide a MC shower with its longitudinal profile most similar to the data shower. The second part of the chain includes a simulation of the surface detector response. A schematic of the algorithm used is presented in figure 36, from which individual parts will be described in more detail in next sections. Each chain element is used independently in an organised sequence. When calculations in some chain element are completed, an intermediate output is provided i.e. a simulated shower, event or just steering parameters. The output is written to an appropriate file, and subsequently used by the next chain element invoked in the sequence. The intermediate outputs are used to control the individual chain elements, among which are the searching procedure, SD core location asymmetry removal and transition from CONEX to CORSIKA. Also the input atmosphere profiles are investigated, despite this is one of the initial input.

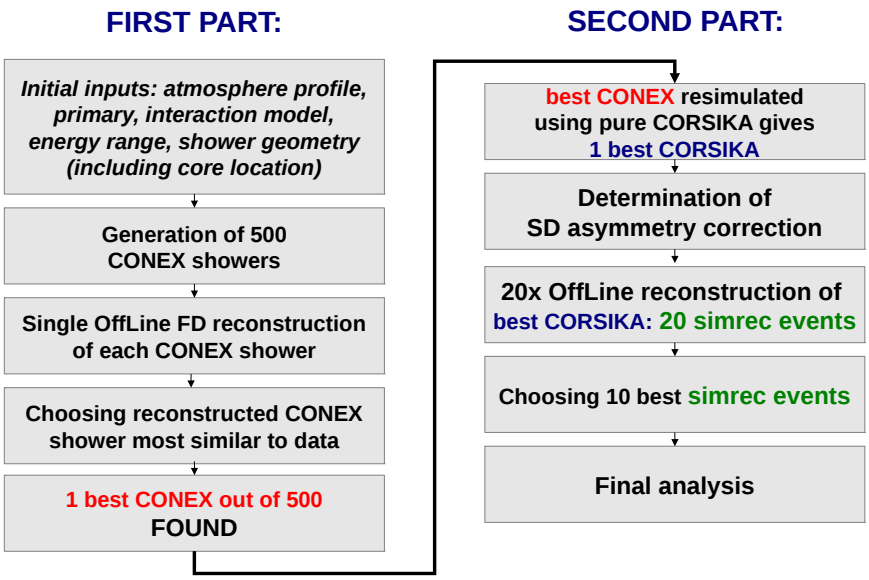

Figure 36: A scheme of used top-down algorithm.

At the beginning the following estimates of input parameters, required by CONEX, are provided. From the data shower: zenith angle, azimuth, and also the observation level on the ground are taken as fixed. The primary energy is drawn from a range $E_0^{\mathrm{sim}} = E_0^{\mathrm{data}} \pm \sigma_{E_0^{\mathrm{data}}}$. The environmental parameters are: magnetic field at Auger site, and atmosphere profile for time of data shower detection. Also the primary type must be assumed. In this study proton and nitrogen are used in two independent, separate runs.

5.3 Atmosphere profiles

The atmosphere profiles in CONEX and CORSIKA are numerically represented by a set of coefficients which approximate the real atmosphere profiles from GDAS database. A dedicated fitting procedure was used to find the coefficients. In order to assure a good quality of atmosphere approximation, the following quality cut is set into procedure: the maximum difference between the approximation and the GDAS profile must be less than 4 g/cm^2. The accuracy of atmosphere profiles, used in CORSIKA and CONEX, is investigated by comparing the atmosphere profiles evaluated from the used parameterisation with the GDAS atmosphere profiles. Since real atmosphere profile can change significantly even during one night, each data event has its own set of profile coefficients, prepared for exact time corresponding to the time of shower detection. So, for each data event, an individual atmosphere profile is read from the GDAS database, for the time closest to event time.

Since the major part of extensive air shower development occurs within an altitude below 20 km above the sea level, the accuracy of profile evaluations are investigated up to this altitude. Example plot of differences between the complete GDAS profile and the profile evaluated from CORSIKA atmosphere coefficients, plotted for altitude range from ground to 100 km (for one chosen event), is shown in figure 37. It is seen that the differences for this particular profile are not larger than $\sim$1 g/cm^2, within the altitude range from the ground up to 10 km which is the most important for shower development.

To see if all evaluations of the atmosphere profiles hold the desirable characteristics, the histograms presenting differences between profiles of atmospheric mass (GDAS profile - fit) are computed for all used atmosphere profiles. The 1 to 20 km altitude range is divided in four sub-ranges with limits at 5,10 and 15 km, in order to see whether any significant tendency of profile differences occur for specific altitude range. Profile evaluations are calculated in equidistant steps (0.2 km), at altitude values equal to altitudes of atmosphere profiles read from the GDAS. So no interpolation is used in calculation of differences. The results for chosen altitude sub-ranges are presented in figure 38. The histograms show that the differences in all used profile evaluations do not exceed 4 g/cm^2 within all altitude ranges. This pass the required assumed quality (value of 4 g/cm^2 is a maximum allowed difference in fitting algorithm). Up to 10 km the absolute value of the differences are less than 2 g/cm^2, which, taking into account the Offline accuracy of X_{max} determination ($\sim$15 g/cm^2), proves a sufficient accuracy of the profile evaluations used in CORSIKA. Above 10 km the differences become more significant, but still within $\pm$4 g/cm^2 range, as assumed.

The histograms show the atmosphere fits for all data events are sufficiently accurate and should not introduce any significant noise nor bias in the top-down reconstruction.

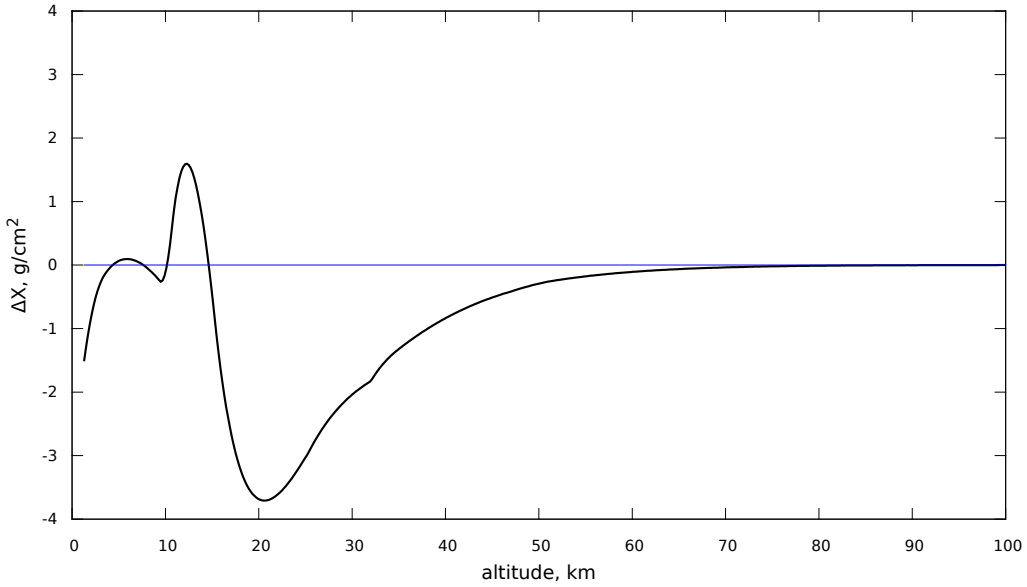

Figure 37: Example differences between the GDAS profile and its approximation used in CORSIKA. Event date: Jan. 19, 2005.

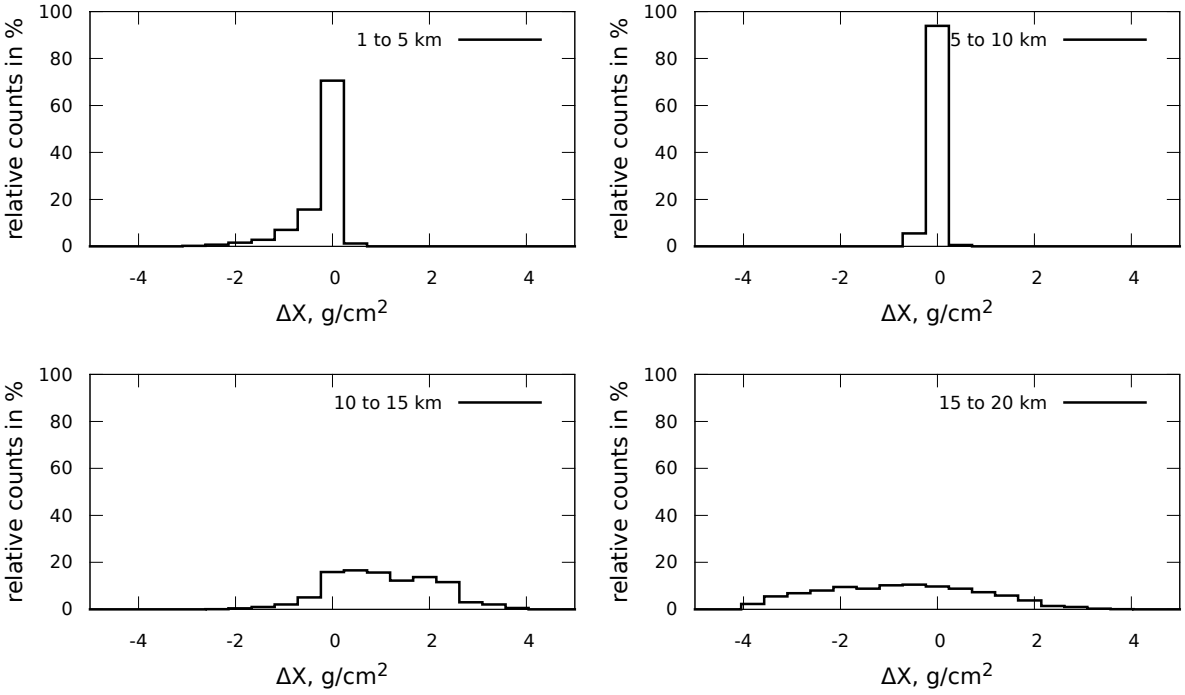

Figure 38: Histograms of differences between the fitted atmosphere profiles and profiles taken from the GDAS database. Four subranges of altitude are shown.

5.4 Generation of 500 CONEX showers

For a given data shower, using the input parameters listed above, a generation of 500 CONEX showers is perform. Due to the Monte Carlo approach, each generated shower has a different fluctuation set which leads to different shape of each longitudinal profile. The random fluctuation set, which occur in simulations, is steered by keywords (SEED 1, SEED 2, SEED 3), of which values are stored in order to be used again in the second part of the algorithm. In simulation with more massive primary the energy threshold for cascade equation is scaled with primary mass. The CORSIKA and CONEX enable such scaling with preservation of shape and fluctuations, to achieve optimum CPU usage.

5.5 Single Offline FD reconstruction

For each of the generated CONEX profiles a comparison with the data longitudinal profile is to be done. The data provide only longitudinal profile of energy deposited in the atmosphere as seen in the FD. Therefore, before comparison of profiles, the FD response for each CONEX profile is simulated using the Offline software, providing a CONEX event. Since the CONEX provides only the longitudinal profile of simulated shower, its reconstruction is performed only for FD (i.e. no SD are used in reconstruction).

To perform a simulation of Auger observatory response, the Offline requires additional information about core location in the Auger site. Shower core location on the ground is described by two coordinates called northing (N) and easting (E). The northing and easting are given in meters in direction to the north and east respectively. For reconstruction of CONEX showers their core locations in Auger site are set to coincide with the core location of the data shower. Also the date of simulation is set as the date of data shower detection. Therefore not only the same FD-shower distance and hence the same scattering and attenuation of light is present, but also the same SD stations will be included in the subsequent simulation chains.

A randomization of the core position is performed to average out possible reconstruction-dependent effects, and to imitate "natural noise" of the shower reconstruction. The values of northing and easting are randomly chosen with uniform distribution, in a square-shaped area centered at the reconstructed SD core position (figure 39). Such randomization is implemented in the Offline software, with a uniform distribution of initial core position as the only available mode. Dimensions of the square-shaped area are taken as $\pm 1\sigma_n$ and $\pm 1\sigma_E$, where σ_n and σ_E are taken from data shower. Since the inaccuracy of the SD core position is small, usually having values of $\sigma_{\mathrm{N,E}} < 50$ m, the size of the square-shaped area is negligible relative to distance between the SD units (1500 m).

To see performance of FD reconstruction a comparison of CONEX profiles and its FD reconstructions (fig. 40, table 1.) is performed. CONEX profiles used in this comparison are so-called the CONEX profiles with the lowest value of χ^2-measure described in next paragraph (5.6). Here and in subsequent comparisons the following parameters of the Gaisser Hillas function, fitted to longitudinal profiles included in the comparison, are

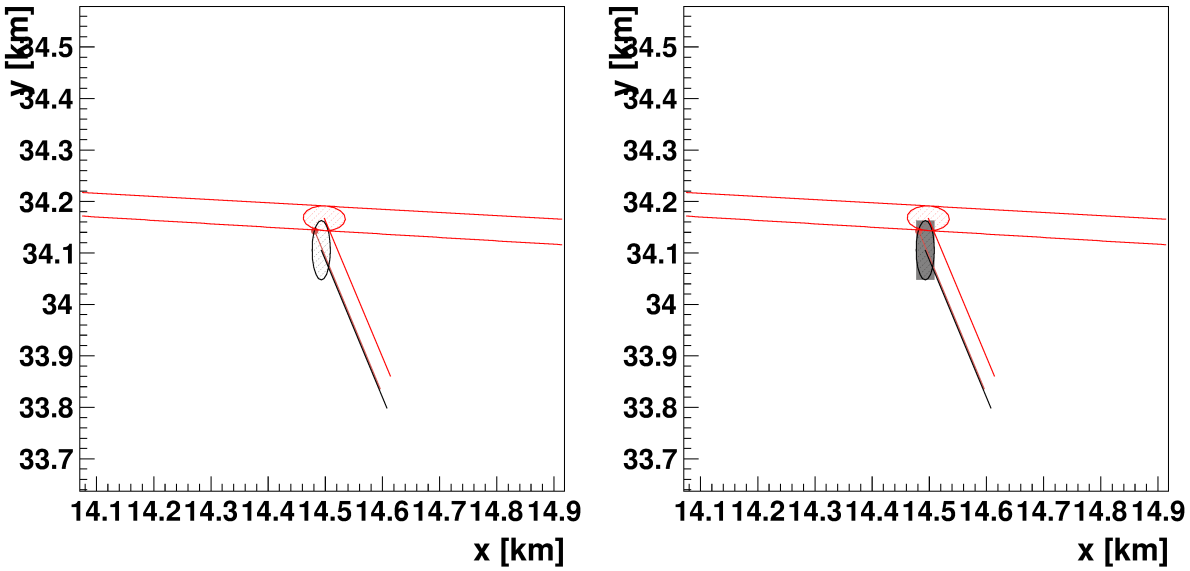

Figure 39: Left: core location in example data shower obtained in standard Offline reconstruction. 1σ error ellipses of reconstructed core location are shown (black from SD reconstruction, red from Hybrid reconstruction). The red lines indicate how the error ellipse is seen from the FD telescope placed at the left side of the plot (not included in figure). Northing and easting is denoted as x and y respectively, but in this plot the distances are measured from chosen reference point in Malargüe city. The data SD core position (center of the black ellipse) is used to set the initial position in Offline, to perform CONEX shower reconstruction. Right: the shaded square-shaped area, centered in the black ellipse, is used as allowed area for randomization of core position.

used: $X_{\max}$, $N_{\max}$ and additionally $E_{\text{cal}}^{\text{ratio}}$ as the integral of the GH function. The scaling factor $N_{\max}$ is replaced by $dEdX_{\max}^{\text{ratio}}$ in case of energy deposit profile, indicating the value of energy deposited at shower maximum. The $X_{\max}$ differences are shown and comparison of the remaining parameters ($E_{\text{cal}}^{\text{ratio}}$ and $dEdX_{\max}^{\text{ratio}}$) are in form of ratios.

While the $\Delta X_{\max} \sim 0$, the distributions of calorimetric energy ratios and $dEdX_{\max}^{\text{ratio}}$ ratios are evidently shifted by about 1σ to lower values. The reason for that it may be due to attenuation of light caused by the dust on the UV filter placed at the FD entrance pupil, which is currently under intensive investigations by other members of Auger observatory. However, due to use of the same detector setup in generation and in reconstruction of MC events, the bias is reduced, which will be proven in next paragraph (5.6). So this effect is taken into account during the top-down reconstruction and should not influence the results.

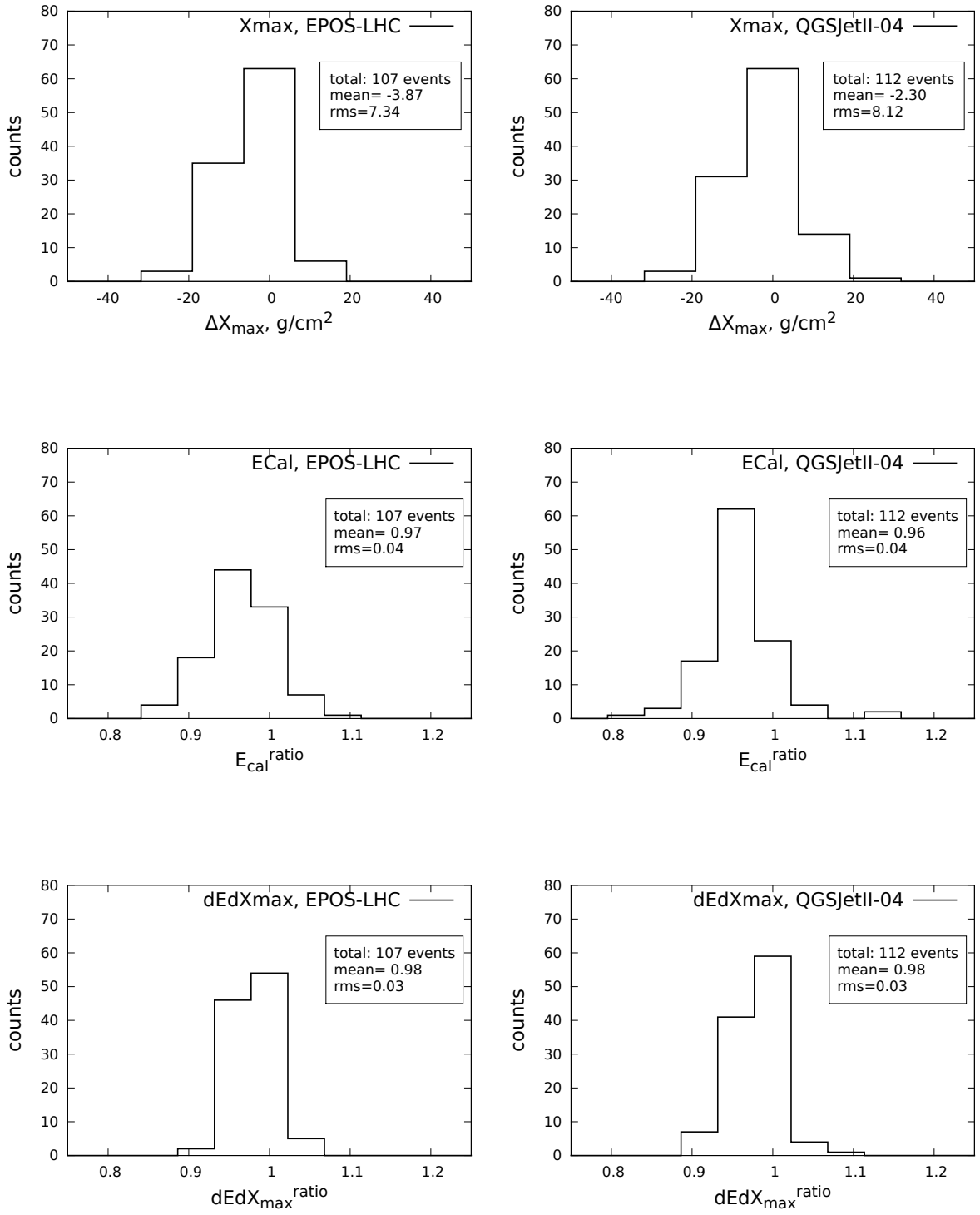

Figure 40: Reproduction of the longitudinal profile parameters after FD reconstruction of CONEX profile, i.e. the difference between the true value generated by CONEX and the one after detector simulation (Offline). The parameters are $\Delta X_{\max} = X_{\max}^{\text{FD simrec}} - X_{\max}^{\text{CONEX}}$, $dEdX_{\max}^{\text{ratio}} = dEdX_{\max}^{\text{FD simrec}}/dEdX_{\max}^{\text{CONEX}}$, $E_{\text{cal}}^{\text{ratio}} = E_{\text{cal}}^{\text{FD simrec}}/E_{\text{cal}}^{\text{CONEX}}$

model	$\Delta X_{\max}$, g/cm^2	$dEdX^{\text{ratio}}_{\max}$	$E^{\text{ratio}}_{\text{cal}}$
EPOS	-3.87 (7.34)	0.98 (0.03)	0.97 (0.04)
QGS	-2.30 (8.12)	0.98 (0.03)	0.96 (0.04)

Table 1: Summary of figure 40: mean values of $\Delta X_{\max} = X^{\text{FD simrec}}_{\max} - X^{\text{CONEX}}_{\max}$, $dEdX^{\text{ratio}}_{\max} = dEdX^{\text{FD simrec}}_{\max}/dEdX^{\text{CONEX}}_{\max}$, $E^{\text{ratio}}_{\text{cal}} = E^{\text{FD simrec}}_{\text{cal}}/E^{\text{CONEX}}_{\text{cal}}$ and their standard deviations (in brackets).

5.6 Choosing reconstructed CONEX shower most similar to data.

In this step all 500 CONEX profiles are compared with the data profile which is treated as a reference. This is done to find the best shower, called here the "best CONEX" (see fig. 36, last step in the first part). The profiles used in comparisons are represented by measurements of energy deposited in the atmosphere. Due to fluctuations each profile has deposit values measured at different atmospheric depths, so it is not possible to compare the measured values directly. Therefore the following χ^2-type measure is applied. Let the number of measurements of deposited energy in data be n, and in the CONEX event $n2$. Also let the simulated measurements of deposited energy from CONEX event is denoted as $LP^{\text{sim}}(X)$, and their errors respectively $\sigma^2(LP^{\text{sim}}(X))$. Similarly, let the data profiles be denoted as: $LP^{\text{dat}}(X)$ and $\sigma^2(LP^{\text{dat}}(X))$. Also the Gaisser-Hillas fits to the longitudinal profiles are available. For the simulated profile the GH fit is denoted as $GH^{\text{sim}}(X)$, and for the GH fit to data profile as $GH^{\text{dat}}(X)$. The GH function will be used to calculate the χ^2. The calculation is performed in the following way: first a calculation of χ^2 for the LP^{dat} relative to the GH^{sim} fit is done:

$$\chi^2_1 = \sum_{\text{i}=1}^{n} \frac{(GH^{\text{sim}}(X^{\text{dat}}_{\text{i}}) - LP^{\text{dat}}(X^{\text{dat}}_{\text{i}}))^2}{\sigma^2(LP^{\text{dat}}(X^{\text{dat}}_{\text{i}}))}$$

Next a calculation of χ^2 for the simrec LP^{sim} with to the GH^{dat} fit is done:

$$\chi^2_2 = \sum_{\text{i}=1}^{\text{n2}} \frac{(GH^{\text{dat}}(X^{\text{sim}}_{\text{i}}) - LP^{\text{sim}}(X^{\text{sim}}_{\text{i}}))^2}{\sigma^2(LP^{\text{sim}}(X^{\text{sim}}_{\text{i}}))}$$

The effective χ^2 has a form:

$$\chi^2 = \frac{1}{\text{n+n2}} \left(\chi^2_1 + \chi^2_2\right).$$

An example best CONEX profile and best CONEX event, selected using this χ^2 measure, are presented in figure 41. Also a visualisation of behaviour of the χ^2 measure for one example event is shown on figure 42.

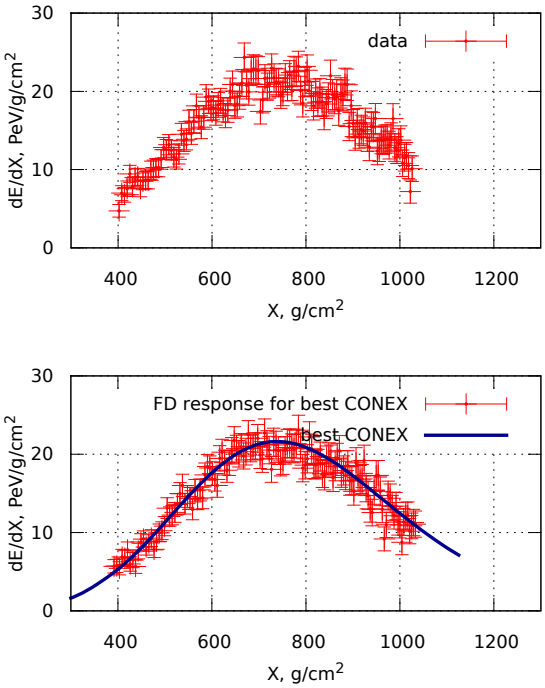

Figure 41: Top: example of a FD profile data event. This profile was used as reference profile to find the best CONEX (see lower plot). Bottom: best CONEX profile (line) and best CONEX event seen in FD (red points). The FD response for best CONEX profile indeed have the same shape as the data profile plotted above.

Performance of the procedure is investigated now by comparing the "best" longitudinal profiles for all MC data events, which is done similarly like in the section 5.5. So the test includes comparisons of longitudinal profiles between the best CONEX events (so CONEX profiles after FD reconstruction) and the corresponding MC events. The comparison will allow us for estimation of the accuracy of χ^2 measure, and so the quality of the selected best CONEX event. The results are presented in fig. 43 with summary in table 2. It is seen that on average the $X_{\max}$, $dEdX^{\text{ratio}}_{\max}$ and FD calorimetric energy of the CONEX events are reproduced well. The accuracy obtained for the $X_{\max}$ is $\sim$5 g/cm^2, which is better than the accuracy of real data reconstructed in Auger Observatory (15 g/cm^2). It indicates that the assumed number of 500 CONEX profiles used in searching procedure is sufficient to provide satisfactory accuracy of selected best CONEX event. Also, the calorimetric energy and deposit maximum are reproduced well, with $\sim$3% accuracy, which is satisfactory. It is notable that a small bias, seen in comparison presented in previous paragraph (5.5), is indeed irrelevant. The accuracy of used parameters can be treated as the mean accuracy of profiles provided by the searching procedure. So this test proves that the searching procedure provides a reliable simulated representation of the MC events profiles.

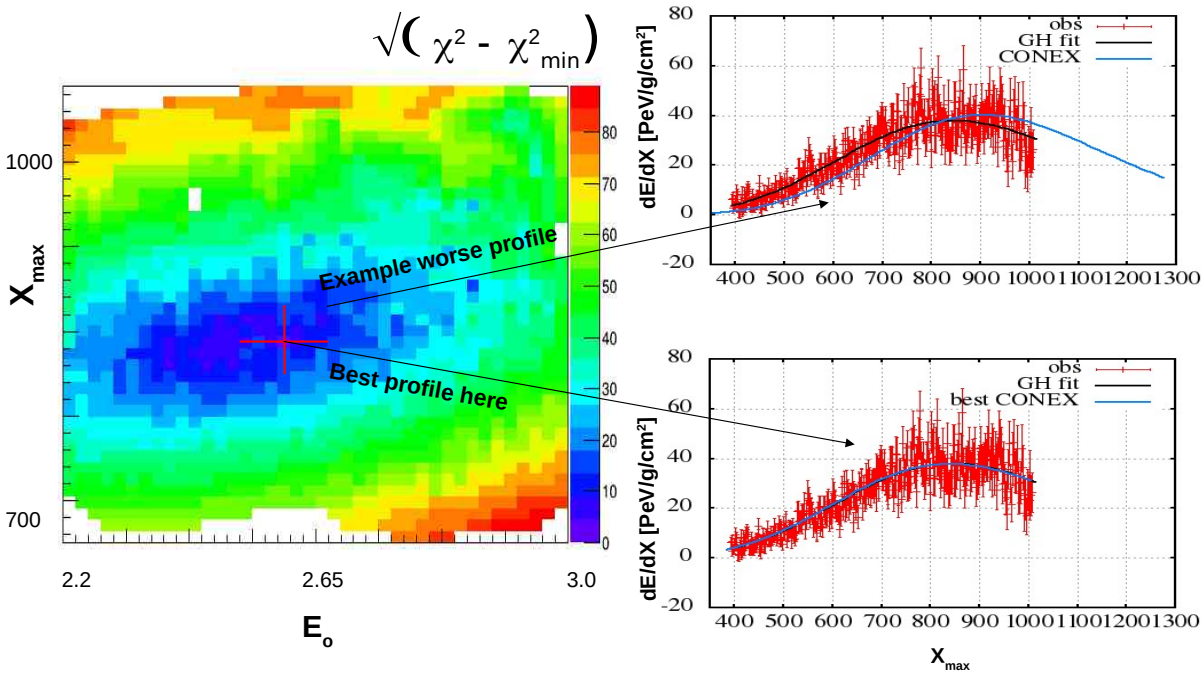

Figure 42: A two-dimensional histogram of $\sqrt{(\chi^2 - \chi^2_{\rm min})}$ in energy-$X_{\rm max}$ parameter space for one example event (left side). A region with minimum of χ^2 is visible. The red cross indicates location where the CONEX simulation with the lowest χ^2 is located (best CONEX). The profile of best CONEX is shown on the plot of observed data profile (reference, denoted as "obs", right side, lower panel). It is seen the profile seem to have shape very similar to the shape of reference profile. Since there are 500 CONEX simulations, each simulation lying outside the minimum regions has worse shape when compared to the reference profile. The larger distance from the minimum the worse shape when compared with the reference. An example profile, which is located outside the minimum region is plotted on the right side, upper panel. Such profile, with χ^2 larger than minimum, has visually different shape when compared to the reference profile.

model	$\Delta X_{\rm max}$, g/cm^2	$dEdX^{\rm ratio}_{\rm max}$	$E^{\rm ratio}_{\rm cal}$
EPOS	0.42 (5.31)	1.00 (0.03)	0.99 (0.04)
QGS	1.50 (4.71)	1.00 (0.04)	0.99 (0.03)

Table 2: Summary of figure 43: mean values of $\Delta X_{\rm max} = X^{\rm FD\ simrec}_{\rm max} - X^{\rm obs}_{\rm max}$, $dEdX^{\rm ratio}_{\rm max} = dEdX^{\rm FD\ simrec}_{\rm max}/dEdX^{\rm obs}_{\rm max}$, $E^{\rm ratio}_{\rm cal} = E^{\rm FD\ simrec}_{\rm cal}/E^{\rm obs}_{\rm cal}$ and their standard deviations (in brackets).

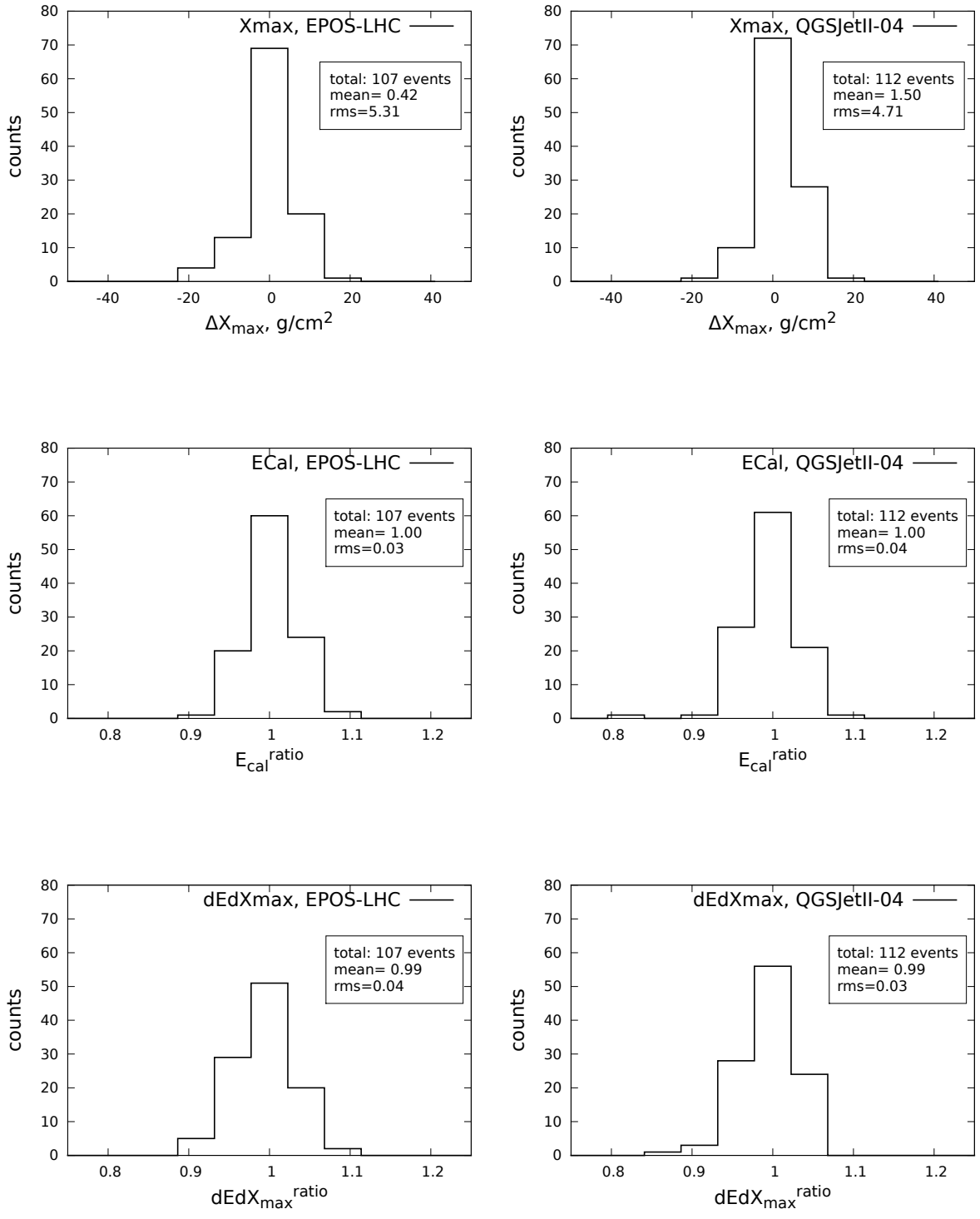

Figure 43: Reproduction of the longitudinal profile parameters in the "best CONEX" profile (i.e. CONEX profile after Offline FD reconstruction) selected using the χ^2 measure. The parameters are $\Delta X_{\max} = X_{\max}^{\text{FD simrec}} - X_{\max}^{\text{obs}}$, $dEdX_{\max}^{\text{ratio}} = dEdX_{\max}^{\text{FD simrec}}/dEdX_{\max}^{\text{obs}}$, $E_{\text{cal}}^{\text{ratio}} = E_{\text{cal}}^{\text{FD simrec}}/E_{\text{cal}}^{\text{obs}}$

5.7 Transition from CONEX to CORSIKA.

The top-down reconstruction chain includes a precise resimulation of the best CONEX shower. The optimal thinning is used, with $\varepsilon_{\rm th} = 10^{-6}$. Almost all values of steering parameters used in previous part are the same, including the keywords SEED 1, SEED2, SEED3, which steer the set of random fluctuations which occur during simulation of shower development. So the resimulation of the CONEX profile in CORSIKA is expected to preserve its shape.

As a result the raw CORSIKA shower is obtained, which is used in remaining steps of the top-down chain. During the resimulation the shapes of the longitudinal profiles in raw CORSIKA and CONEX showers should be similar. However, it is not excluded that sometimes the same values of SEED keywords in CONEX and CORSIKA will not provide the same shape of shower profile. Therefore many additional tests of shower shape preservation were performed in the tests of algorithm used. The tests proved that the shower profile shape is preserved during resimulation. Despite the other tests of shape preservation a control of shape preservation of the used simulations in this analysis is performed here to exclude a possibility of rare shape disagreement. The control includes a comparison of profile shapes between a given CONEX shower and its resimulation to corresponding CORSIKA shower.

Such control, performed with the simulations used in reconstruction of the MC events, is shown in fig. 44 with summary in table 3. The histograms show that during the transition from CONEX to CORSIKA the calorimetric energy is preserved. It is also seen that there is mentioned before ∼-3 g/cm^2 disagreement in $X_{\rm max}$, which is a small value compared with the accuracy of $X_{\rm max}$ determination in Auger observatory, so it is acceptable. Since the energy is conserved, the shift in $X_{\rm max}$ induces a change of $dEdX^{\rm ratio}_{\rm max}$, to sustain the $E^{\rm ratio}_{\rm cal}$ (area below the profile). After transition from CONEX to CORSIKA, the value of $dEdX^{\rm ratio}_{\rm max}$ changes by ∼1%. Its impact on the results of the entire reconstruction chain is investigated in the chapter 6.2.

parameter	EPOS	QGSJetII-04
$\Delta X_{\rm max}$, g/cm^2	-2.35 (3.10)	-2.92 (2.97)
$E^{\rm ratio}_{\rm cal}$	1.00 (0.01)	1.000 (0.01)
$dEdX^{\rm ratio}_{\rm max}$	1.01 (0.01)	1.01 (0.01)

Table 3: Summary of figure 44: mean values of $\Delta X_{\rm max} = X^{\rm CORSIKA}_{\rm max} - X^{\rm CONEX}_{\rm max}$, $dEdX^{\rm ratio}_{\rm max} = dEdX^{\rm CORSIKA}_{\rm max}/dEdX^{\rm CONEX}_{\rm max}$, $E^{\rm ratio}_{\rm cal} = E^{\rm CORSIKA}_{\rm cal}/E^{\rm CONEX}_{\rm cal}$ and their standard deviations (in brackets).

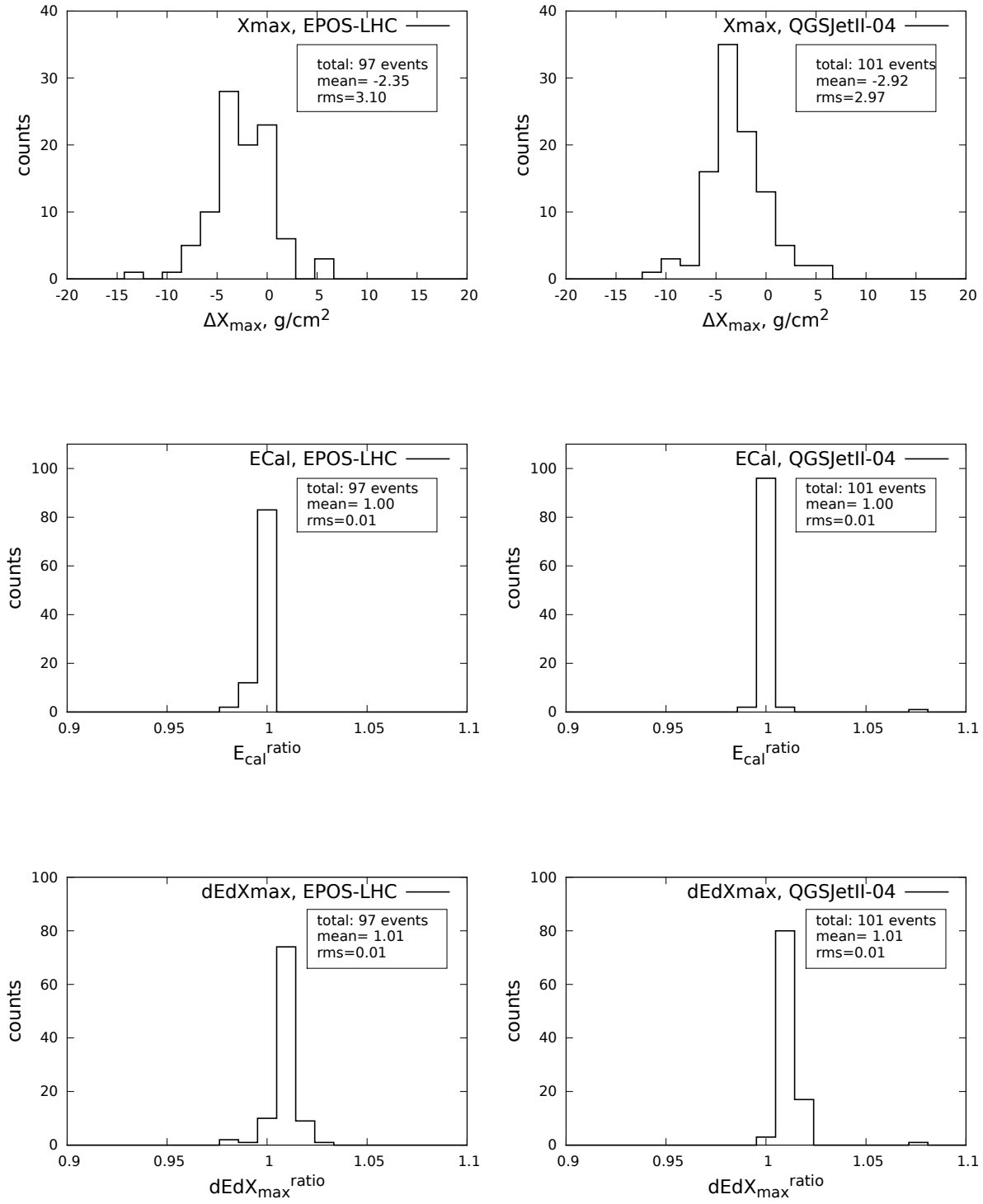

Figure 44: Control of longitudinal profile parameters changes during CONEX to CORSIKA transition. The parameters are $\Delta X_{\max} = X_{\max}^{\mathrm{CORSIKA}} - X_{\max}^{\mathrm{CONEX}}$, $dEdX_{\max}^{\mathrm{ratio}} = dEdX_{\max}^{\mathrm{CORSIKA}}/dEdX_{\max}^{\mathrm{CONEX}}$, $E_{\mathrm{cal}}^{\mathrm{ratio}} = E_{\mathrm{cal}}^{\mathrm{CORSIKA}}/E_{\mathrm{cal}}^{\mathrm{CONEX}}$

5.8 Determination of the SD asymmetry correction

The result from the previous step is called a "simrec shower". Now the Auger observatory response for the simrec shower is to be simulated in the hybrid mode, using the Offline software, which will provide a "simrec event". Before the Offline simulation is performed, the initial core position is corrected for the following effect.

When an inclined shower hits the SD site, the stations located upstream register earlier stage of shower development than stations located downstream. So the upstream stations register larger signals than downstream stations, which results in so-called SD asymmetry of core position, since the station signals are used as weights in the barycenter fit during SD reconstruction. The asymmetry is seen as if the reconstructed position of shower core on the ground is dragged in upstream direction. A schematic of SD-shower configuration leading to this effect is shown in figure 45.

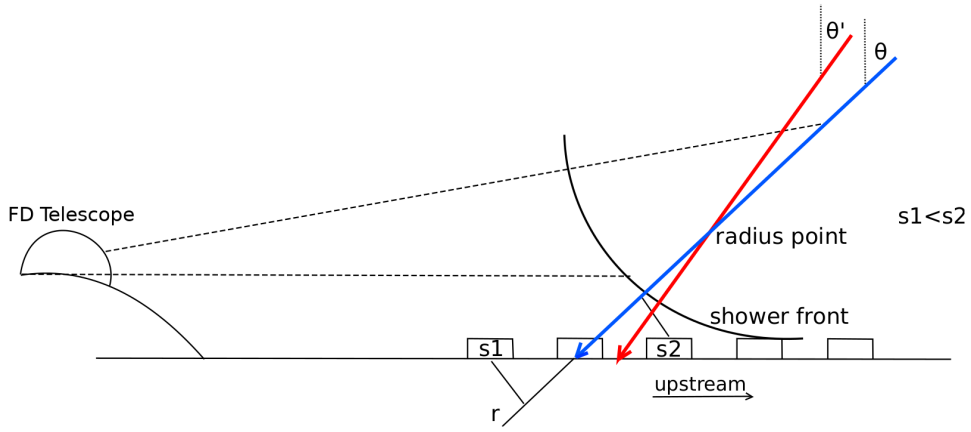

Figure 45: Schematics of geometry leading to core asymmetry mechanism. The dashed lines indicate field of FD telescope view. The blue line indicates the initial shower axis (before reconstruction), the red line indicates the dragged shower axis as seen after reconstruction. It is seen that, due to shower inclination, the station assigned as s2 detects shower at earlier stage of development that station s1. In effect the density of particles is larger in the s2 station, and so the s2 station signal. In result the reconstructed shower axis is dragged towards the upstream direction during reconstruction (red line).

This effect causes that triggered SD stations detect different parts of the shower front in the simulations than in the data. Since the final analysis will use ratios of signals in individual SD stations, a given SD station must detect the same part of the shower relative to shower axis as in the data shower. The core shift on the ground caused by the asymmetry effect can exceed 100 meters, so is larger than the core position accuracy of most data events.

The core location asymmetry correction is determined for each data event individually through simulations. For each CORSIKA shower 49 Offline SD response simulations are performed with fixed core position on the ground. Using fixed core position during this simulation causes that the mean reconstructed position will be affected by the asymmetry effect. So such reconstruction enables determination of the shift due to SD asymmetry for each data shower. The SD responses for shower are calculated using a fast and simplified

algorithm (Offline module FastTankSimulator). This is sufficient for core position correction determination, because only the estimated total SD signal is necessary to reconstruct the core location.

The reconstructions provide new 49 positions of core on the ground which are scattered according to reconstruction accuracy, and systematically shifted due to the SD core asymmetry effect (fig. 46). The mean core position on the ground is shifted by the asymmetry effect in azimuth direction of the shower, which angular value depends mainly on the configuration of the closest SD station in respect to the shower azimuth. This shift is regarded as the correction value, and is applied during the final Offline reconstruction in the opposite direction, along calculated azimuth of the shift.

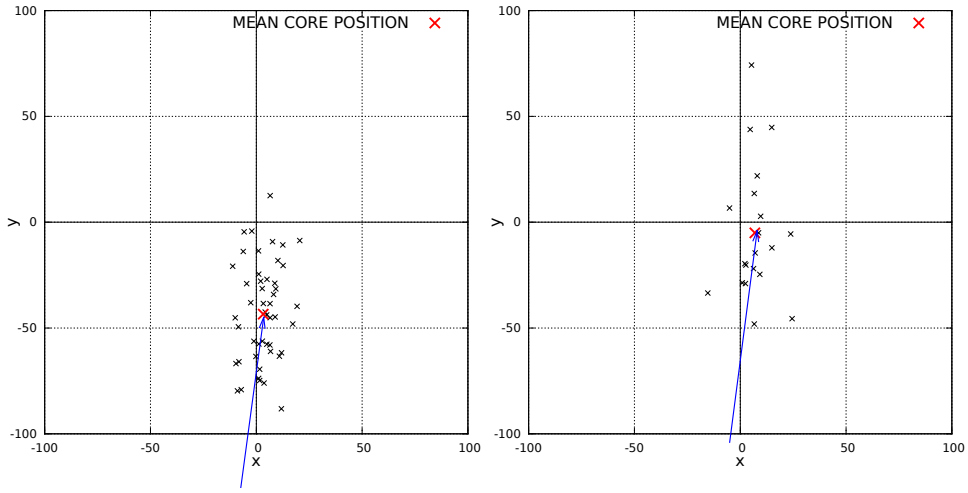

Figure 46: Left: Plot of 49 reconstructed core locations relatively to initial position used in SD simulation. The blue arrows shows estimated direction of the shower azimuth. We see the points are scattered in ellipse corresponding to the error ellipse of data shower shown in figure 37, but the mean position of this ellipse is shifted in the upstream direction. This shift is caused by the asymmetry effect. Removing of this effect is based on application of seen shift of the mean position. Right: The relative core positions from final 20 Offline reconstructions, after applying the core correction. The mean core position is now much closer to the data core position placed at coordinate 0,0.

The properties of the algorithm for SD core location asymmetry removing (in short: core correction) will be investigated now. The test includes a comparison of core locations of the simrec events with corresponding core locations of the MC events. So the core positions in MC events are used as the reference positions, for which the Northing and Easting differences ($\Delta N = N_{simrec} - N_{MCdata}$ and $\Delta E = E_{simrec} - E_{MCdata}$) are calculated. Northing and Easting are used since these coordinates are set as the initial core positions in Offline simulation of Auger observatory response. Also impact of the core correction on the shower axis geometry and SD station distances from core measured in shower plane are investigated.

5.8.1 Core position improvement

The SD core positions were reconstructed in two ways: with and without application of the core correction. It is done to see how big improvement the core correction gives. Resulting histograms of the corrected and not corrected Northing and Easting differences (denoted as $\Delta N_{\rm corr}$,$\Delta E_{\rm corr}$, $\Delta N_{\rm nocorr}$, $\Delta E_{\rm nocorr}$) are presented in fig. 47. The histograms of coordinate differences for not corrected core positions show that the maximum value of shift due to core asymmetry exceed 100 meters. The corrected core positions show a significant improvement. The improvement is seen not only in smaller differences, (summarised in table 4.), but also in the shapes of their distribution. On average the corrected and not corrected differences have zero mean values. Taking into account the $\sigma_{\Delta N_{\rm corr}}$ and $\sigma_{\Delta E_{\rm corr}}$, the estimated value of position accuracy is ∼25 m after application of the core correction, which is two times better than accuracy of not corrected position. Additionally the control histograms of differences of the core location altitude between the MC data events and the corrected simrec events are shown in fig. 48. It is seen that the SD core location correction does not change the observation level on the ground. Only a small scatter, caused most likely due to the Offline reconstruction accuracy, is visible ($\sigma_{\Delta height}$ ∼1 m).

	model	$\mu_{nocorr.}$, meters	$\sigma_{nocorr.}$, meters	$\mu_{corr.}$, meters	$\sigma_{corr.}$, meters
ΔN	EPOS	2.5	47.9	-0.7	16.2
	QGSJetII-04	49.7	7.7	-1.2	20.0
ΔE	EPOS	2.0	55.9	3.0	21.2
	QGSJetII-04	4.8	53.5	2.0	15.7

Table 4: Summary of figure 47: mean (μ) and standard deviation (σ) values of northing $\Delta N = N_{simrec} - N_{MCdata}$ and easting $\Delta E = E_{simrec} - E_{MCdata}$ differences for not corrected and corrected simrec events. Due to application of the SD core location asymmetry removing the reconstructed core positions are in better agreement with the MC data events.

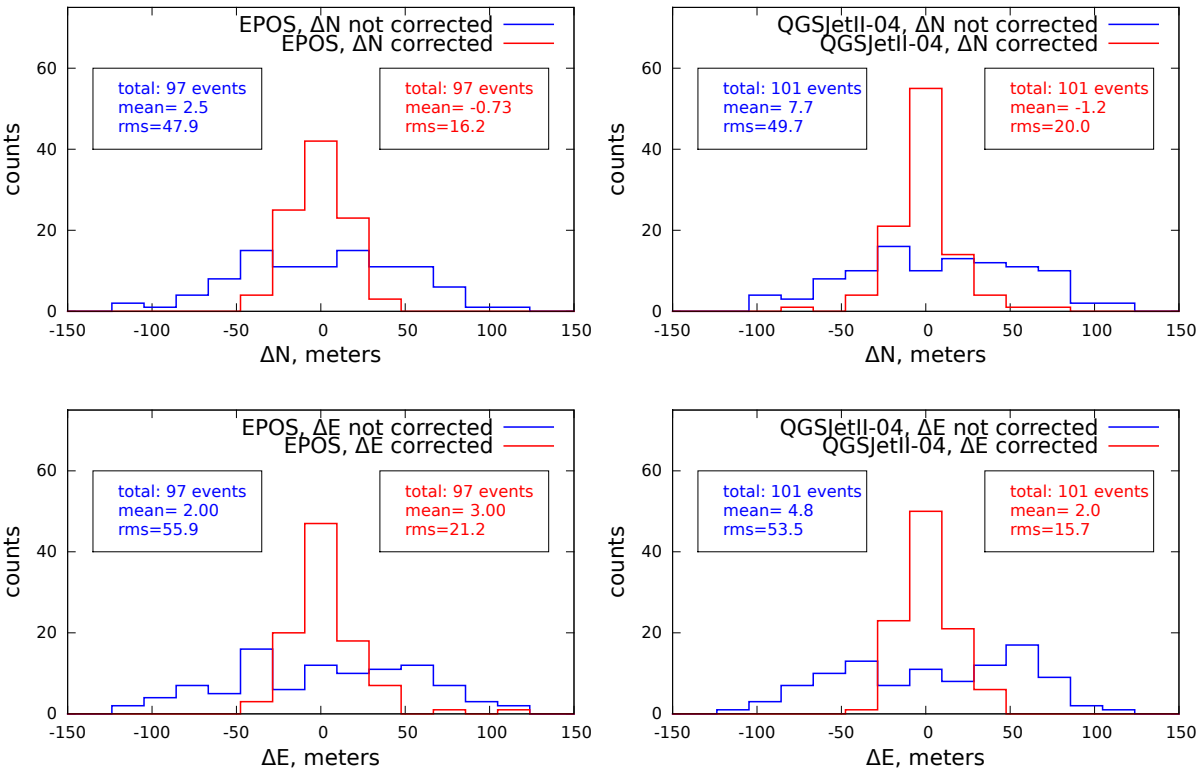

Figure 47: Northing ($\Delta N = N_{simrec} - N_{MCdata}$) and Easting differences ($\Delta E = E_{simrec} - E_{MCdata}$). Distribution of reconstructed core positions in respect to the MC data without core correction (blue line). Due to application of the SD core location asymmetry removing the reconstructed core positions are in better agreement with the MC data events (red line). The mean values and standard deviations are summarised in table 4.

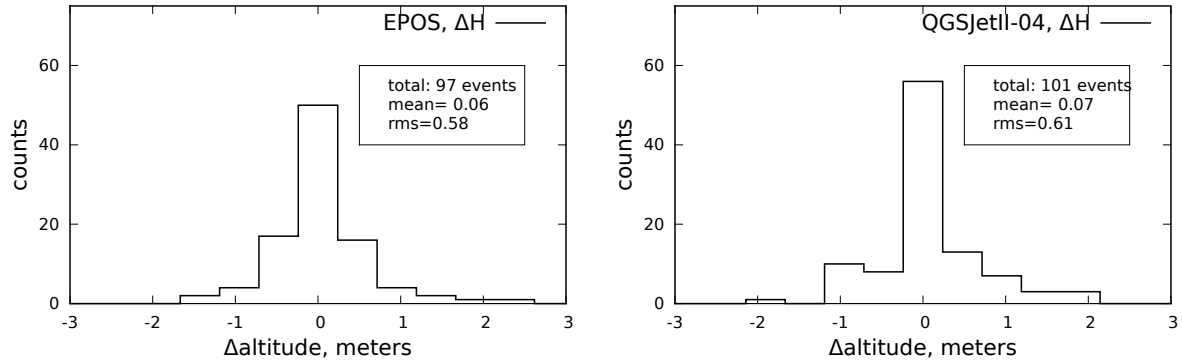

Figure 48: Control plot of altitude ($\Delta Z = Z_{simrec} - Z_{MCdata}$) differences for simrec events after application of the SD core location asymmetry removal. The histograms show that the mean altitude values of reconstructed core positions are not changed significantly due to application of the core correction.

5.8.2 SD station-shower core distances

The SD station-shower core distances ($\rho = \rho_{\text{simrec}} - \rho_{\text{MC data}}$) measured in shower plane (SP) are used in analysis of the SD detector signals. The SP is a plane perpendicular to the shower axis, so is also inclined to the ground. The SD station azimuth in shower plane is irrelevant, since the showers are assumed to be symmetrical in the coordinate system related with the shower axis.

All efforts linked with improvement of the core location on the ground are done to provide the core distances in simrec events as close to the corresponding core distances in MC data as possible. So, the differences of SD-shower core distances ($\Delta\rho$) are compared for corrected and non-corrected to core correction simrec events. The SD-shower core distances in corresponding MC events are used as the references. The histograms of $\Delta\rho$ are presented in fig. 49 with summary in table 6. It is seen that the core correction reduced disagreement between the SD station distances for factor ∼3 (see table 5). The distribution of the corrected SD station distance differences has gaussian-like shape. This suggest that introduction of the core correction refines the SD station distances in reconstructed simrec events, leaving only normally distributed reconstruction error. Taking the standard deviation of corrected $\Delta\rho$ distribution after core correction as the accuracy of the SD station distances, the accuracy is $\sim$ 15 meters in shower plane. Such accuracy of ρ is only 1% of distance between SD stations in Auger site.

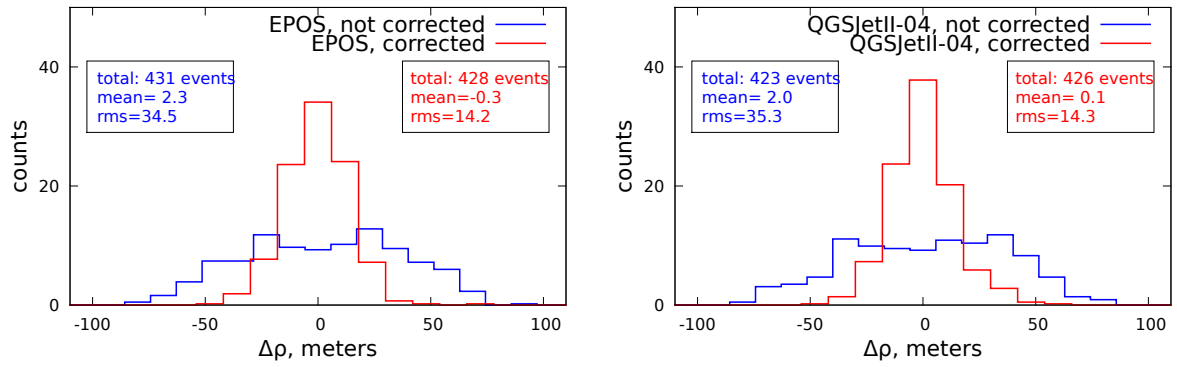

Figure 49: Histogram of SD station-shower core distance differences ($\rho = \rho_{\text{simrec}} - \rho_{\text{MC data}}$) measured in shower plane. The reference distances are taken from the MC data. Both, the corrected on the core correction SD station distances (red line) and non-corrected on the core correction SD station distances (blue line) are plotted. It is seen the application of the core correction reduces the differences of core to SD station distances between MC data and its reconstructions.

model	$\rho_{\text{nocorr.}}$, meters	$\sigma_{\rho_{\text{nocorr.}}}$, meters	$\rho_{\text{corr.}}$, meters	$\sigma_{\rho_{\text{corr.}}}$, meters
EPOS	2.3	34.5	-0.3	14.2
QGSJetII-04	2.0	35.3	0.1	14.3

Table 5: Summary of figure 49: mean SD staion to core distances ($\rho = \rho_{\text{simrec}} - \rho_{\text{MC data}}$) and its standard deviation values (σ) for not corrected and corrected simrec events. Due to application of the SD core location asymmetry removing the reconstructed SD staion to core distances are in better agreement with the MC data events.

5.8.3 Impact of the core correction on shower axis orientation

In section 5.8.1 it was shown that the SD core locations on ground are improved. However, orientation of the FD shower axis may suffer from the applied core correction [94]. From the geometrical point of view the core correction is a simple spatial parallel shift of the initial shower axis location on the ground. The shifted core position is set as the initial position in Offline. So the location of the axis in FD field of view (FD geometry) may change.

A possible impact of the FD core asymmetry removing on the FD geometry of reconstructed shower axis will manifest itself as the zenith and azimuth changes. The changes of direction of the reconstructed shower axis in hybrid mode, after use of the core correction, are investigated. It is done by testing behaviour of the azimuth differences ($\Delta\phi$) and zenith angle differences ($\Delta\Theta$) of FD shower axis between MC data events and simrec events. The FD geometries are investigated for the entire top-down reconstruction.

The histograms in fig. 50 show the impact of the core asymmetry removing on the FD axis geometry after Offline simulation, with summary in table 4. The Offline simulation of detector response introduces some reconstruction noise, but the noise has rather gaussian-like distribution. The obtained shape of distributions indicates that the core correction impacts the final shower axis orientation. For FD geometry the azimuth suffers most due to core correction. Also some changes of the zenith angle are seen, but impact of the core correction on zenith angle is smaller than on azimuth. The mean values of $\Delta\phi$ and $\Delta\Theta$ distributions are equal to $\sim$0, so no bias is introduced. Taking into account the accuracy of FD shower axis reconstruction, which is $\sim$0.6°, the mean light in FD should not be affected significantly due to shower axis change caused by the core correction. Only a possible larger scatter of the values of deposited energy in shower profile may occur.

	model	FD geometry
$\Delta\Theta$, degree	EPOS	0.04 (0.20)
	QGSJetII-04	0.01 (0.24)
$\Delta\phi$, degree	EPOS	0.00 (0.29)
	QGSJetII-04	-0.02 (0.40)

Table 6: Summary of figure 50: mean zenith ($\Delta\Theta = \Theta_{\rm simrec} - \Theta_{\rm obs}$) and azimuth ($\Delta\phi = \phi_{\rm simrec} - \phi_{\rm obs}$) changes and their standard deviations (in brackets) of reconstructed FD shower axis orientations due to the core correction.

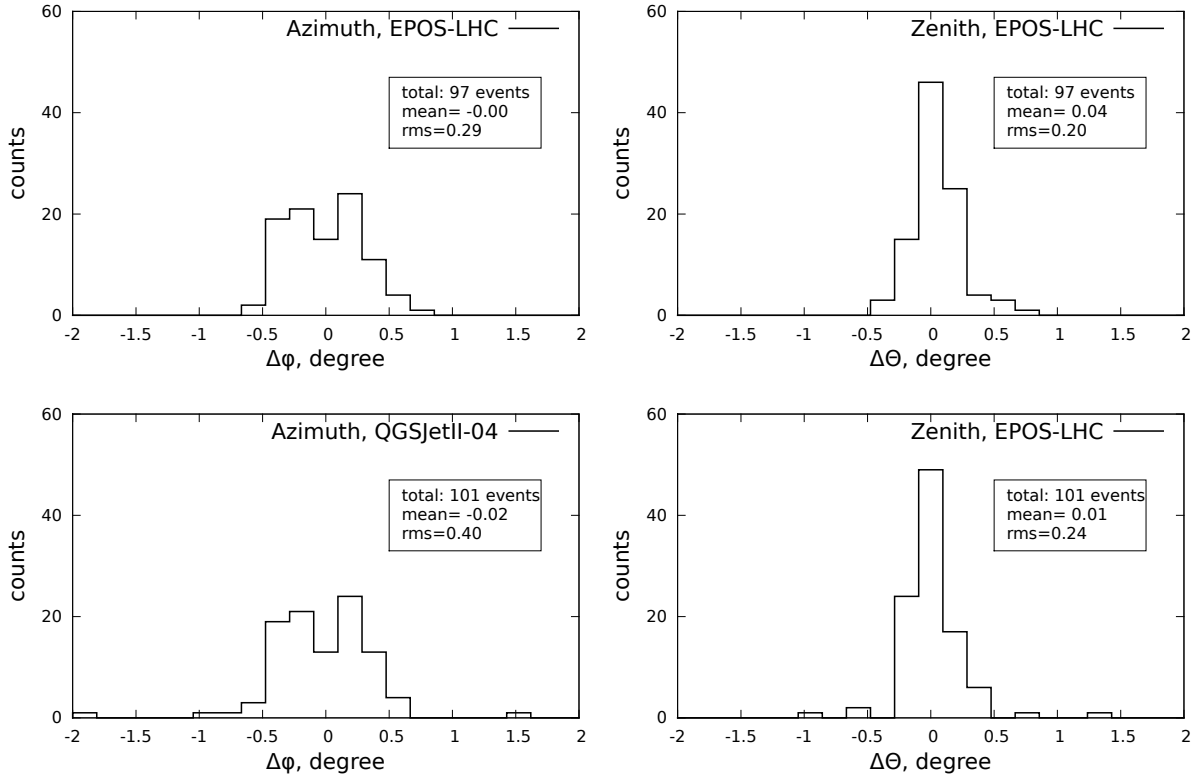

Figure 50: Histograms of the zenith ($\Delta\Theta$) and azimuth ($\Delta\phi$) changes of reconstructed FD shower axis orientations due to the core correction. The changes are measured as $\Delta\Theta = \Theta_{\rm simrec} - \Theta_{\rm obs}$ and $\Delta\phi = \phi_{\rm simrec} - \phi_{\rm obs}$ calculated for final simrec events (i.e. after the whole top-down reconstruction run).

5.9 20 Offline reconstructions and choosing 10 best simrec events

In the final step the Auger observatory response is simulated and Offline hybrid reconstruction of the CORSIKA shower is done, which provides the simrec event. The initial core position is randomised like in section 5.5. The top-down reconstruction is performed on identical detector setup as data shower, which enables additional elimination of possible detector-dependent effects in individual stations. For example, if some detector station has some source of a bias, which is not taken into account, the results of both, simulated and data showers, are biased in the same direction. Considering the fact that the results of top-down are analysed using ratios of detector signals, the bias of signal in a given SD station will be reduced, since it will be the same station in data and in simulations. Also, it may happen that some SD station is not triggered in the data shower due to maintenance, despite this unused SD station was exposed on the stream of shower particles and would be triggered. In such case, by applying the identical detector setup as in data shower, such not working station will be excluded from the simulation of the detector response. This is important since the analysis includes also a comparison of S_{1000}, which are estimated from LDF fit to SD signals. Such approach enables a reduction of possible biases which may be sensitive to the number of triggered stations used in the fit.

The SD signals are simulated using GEANT4 algorithm, which is implemented in Offline. 20 Offline runs are performed for one CORSIKA shower, providing 20 simrec events. From them 10 events are selected for comparison of the station signals with the data event. The selection criterion is the χ^2 measure described above. So 10 final simrec events, with the lowest values of the χ^2, are used in subsequent SD signal analysis. Such selection is done to reduce natural noise which is present in reconstruction. After last step of the top-down reconstruction we have 10 simrec events, which are the final result of the top-down reconstruction chain applied for one data shower. All 10 simrec events are used for calculation of the mean SD signal ratios using convention data/simrec. These ratios are the final result of the entire top-down reconstruction method.

6 Validation of the method

6.1 Generation of validation events and analysis strategy

The top-down method is validated using Monte Carlo test events, which were generated using LHC-tuned interaction models. The validation events will be reconstructed with the top-down method using the same models as used for generation of MC events. The MC events used for validation are the CORSIKA showers for which the observatory response is simulated in Offline. So the same hadronic interaction properties are used for generation and top-down reconstruction of the events. In such approach it is expected that, if the method is consistent, the validation will show that the mean signals in detectors are similar to the signals in MC events. The method was prepared to investigate a set of real data events registered in Auger observatory. So it was assumed that the MC events, which will be used in the validation of the top-down chain, should imitate the set of real data events described in next chapter. Imitation of the real events means that the MC events should have the same geometrical properties (zenith angles, azimuths, core locations in the PAO site), times of detection, and the energy. So the same attenuation properties of light seen in the FD, as well as the same SD setup in Auger observatory site, will be present in the MC events and real data events.

To generate the desirable MC events, which will be a simulated representation of the real events, the top-down reconstruction chain is used. The main property of the chain is used, i.e. that the chain provides a simulated event which is similar to the data event (similar are the incoming directions, longitudinal profiles, energies of primary, locations in the detector site, atmospheric conditions, etc...). So, the real events, which are analysed in the thesis, are used as the prototypes of MC events. However, the top-down chain provides 10 final simrec events after reconstruction of one data event. To sustain the same statistics in MC events as in the real data, the MC event which have the lowest χ^2_{eff} is chosen. A simplified schematic presenting generation of a single MC event, and the entire validation chain, is shown in figure 51. The basic properties of generated MC data events ($X_{\max}$, zenith angles and azimuths) are summarised in the histograms in figures 52, 53, 54.

The MC events are generated using two hadronic interaction models tuned to LHC data at 7 TeV: EPOS LHC and QGSJetII-04 (used separately), and proton as primary. Only proton is used during generation of MC events. The proton showers are characterised by largest fluctuations, when compared with the showers induced by more massive particles. So, it is assumed that if the validation of the top-down chain, performed using proton MC showers, proves its reliability, the reconstruction of real data events, performed assuming more massive primary at initial inputs of the reconstruction chain, will be correct. However, the validation includes a reconstruction of the proton MC events using proton and nitrogen as primaries in two separate runs. The top-down reconstruction with proton as primary is performed to prove reliability of the reconstruction chain. Use of nitrogen

in the validation is done to see whether the top-down chain is mass-sensitive.

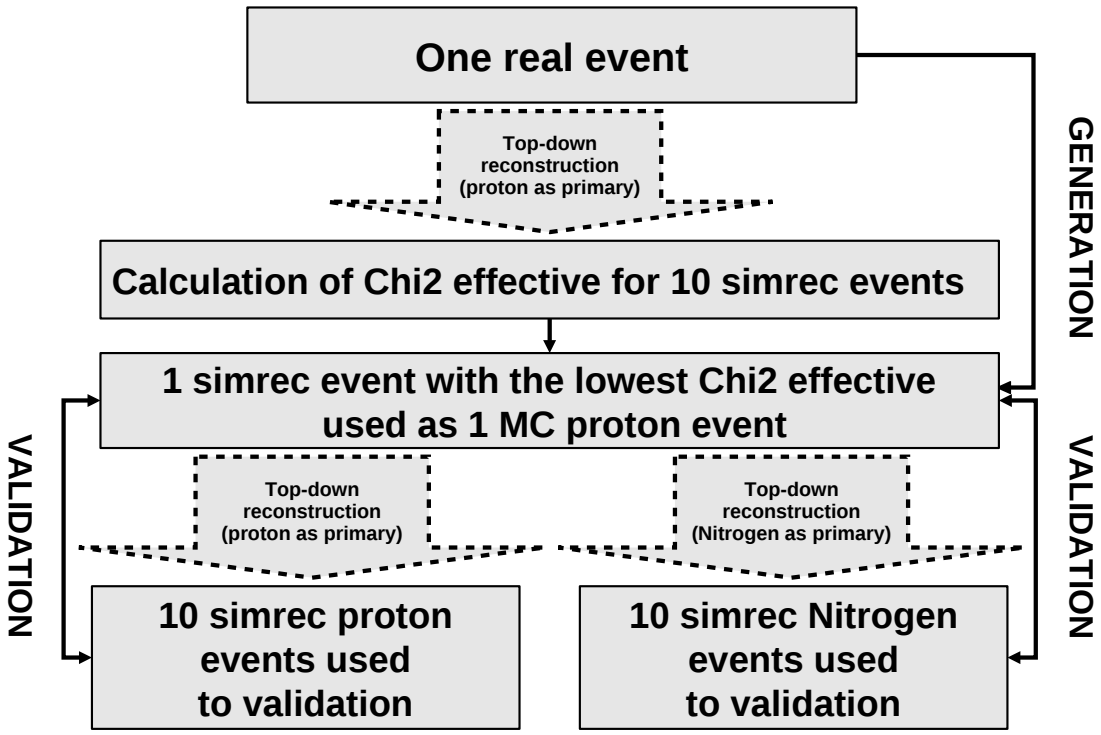

Figure 51: A simplified schematics of preparation and validation procedure of the entire top-down reconstruction chain, applied for one data event. This procedure is used independently for all events and for two interaction models (EPOS LHC and QGSJetII-04). The comparison between one MC event and the 10 simrec events are used in validation for calculation of one mean difference/ratio of parameter and its dispersion. Use of proton and Nitrogen is performed in two separate runs.

During validation the whole top-down method is treated as a black box. The accuracy of the method is determined and compared with accuracy of the individual chain element which can be a source of possible bias, like transition from CONEX to CORSIKA. During the transition a small shift of the $X_{\max}$ and $dEdX_{\max}$ occur. For validation of the entire chain the outputs are compared with corresponding inputs, and the estimation of longitudinal profiles reproduction accuracy as well as quality of the reproduced SD signals are performed. The tests of shower parameters related with shower geometry, like incoming directions, core location on the ground, include analysis of the differences (MC events - simrec). In case of signals in detectors the ratios (MC events / simrec) are used. The differences and ratios are calculated in the following way. First the mean value of a given parameter is calculated from 10 output simrec events. Next, appropriate subtraction or division is performed.

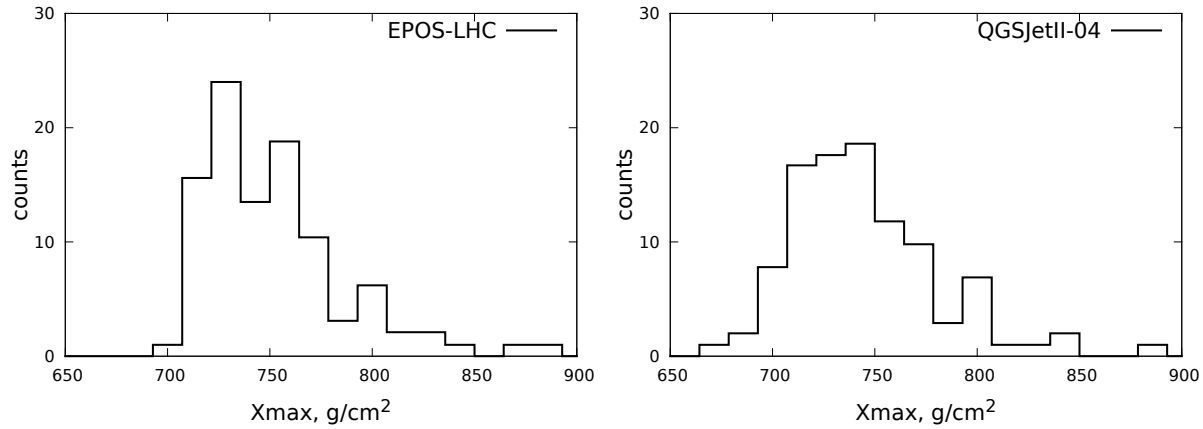

Figure 52: Review of the $X_{\max}$ distributions of MC data events used for validation. The MC events are generated using EPOS and separately QGSJetII-04. Values of $X_{\max}$ for both interaction models are plotted separately.

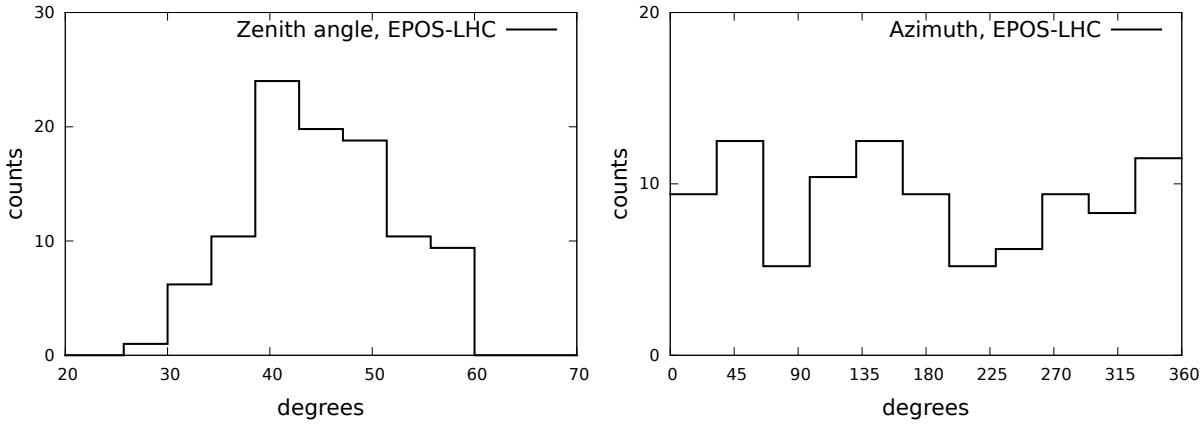

Figure 53: Review of the incoming directions (zenith and azimuth angles) of the generated MC data events used for validation (model EPOS).

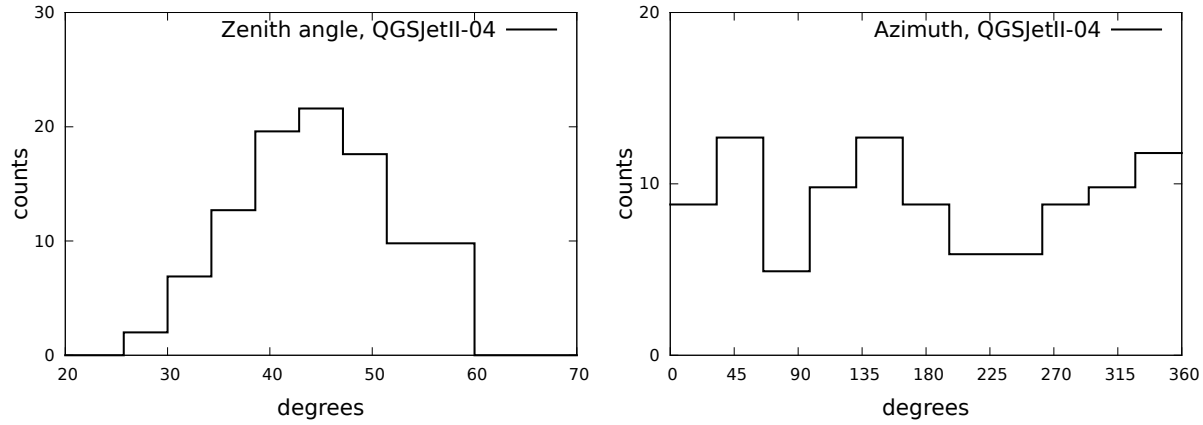

Figure 54: Review of the incoming directions (zenith and azimuth angles) of the generated MC data events used for validation (model QGSJetII-04).

6.2 Top-down reconstruction: longitudinal profile

In previous chapter it was shown that the parameters of longitudinal profiles are preserved with satisfactory accuracy in almost all chain elements. Only after the chain element, in which transition from CONEX to CORSIKA is performed, the profile parameters have a small bias ($\sim$-3 g/cm^2 disagreement in $X_{\max}$, see section 5.7). This may be regarded as a possible source of bias for the entire chain. Impact of this disagreement on the entire top-down reconstruction chain is now investigated. It is done by comparing the profile parameters after the entire reconstruction.

The comparison (fig. 55, with summary in table 7), shows that after use of the entire chain, the mean $\Delta X_{\max}$ is systematically shifted by about -2 g/cm^2 for both interaction models. This shift may be a result of the previously mentioned shift which appears during CONEX to CORSIKA transition (section 5.7). So the shift seems to remain through the subsequent steps of the method, but is not statistically significant when compared to the standard deviation of the $\Delta X_{\max}$ distribution ($\sim$ 8 g/cm^2). Since the Offline $X_{\max}$ accuracy for Auger events is $\sim$ 15 g/cm^2, the systematic shift in $X_{\max}$ obtained for simrec events should not affect significantly the results of the top-down reconstruction.

The $dEdX_{\max}$ and E_{cal} are reproduced with 4% accuracy in the whole reconstruction chain. No shift of the deposit maximum is visible. So it appears that the shift of the energy deposit at shower maximum, which appears during CONEX to CORSIKA transition, does not impact the $dEdX_{\max}$. The whole chain provides 4% accuracy of energy and $dEdX_{\max}$ reproduction. Before the CONEX to CORSIKA transition (i.e. in step choosing reconstructed shower most similar to data) the $dEdX_{\max}^{\mathrm{ratio}}$ accuracy is $\sim$3% (see section 5.6). Taking the above into account it appears that the shift of the energy deposit at shower maximum, causes only reduction of the accuracy of reconstructed $dEdX_{\max}$ by increasing standard deviation of reproduced energy deposit by about 1%. This is probably due to selection of final 10 simrec events, performed in the last step, which reduces the $dEdX_{\max}^{\mathrm{COR}}$ shift as a possible source of bias. It also suggests that the $dEdX_{\max}^{COR}$ shift is small, compared with the fluctuations of maximum of energy deposit which arise during simulation of light in FD aperture. It is not excluded, that this is an effect of the larger spread of final axis orientations, which is not correlated with CONEX to CORSIKA transition. So, for the whole analysis chain the two effects slightly increase the noise of reconstruction, for which the $dEdX_{\max}^{COR}$ shift is not significant. Despite it the bias seems to not amplify, so the analysis chain appears to be stable.

Taking into account the results above, it seems that the entire chain reproduces the longitudinal profile of real data with better accuracy than the accuracy provided by the standard Offline reconstruction. Even despite the possible bias sources like the $dEdX_{\max}^{\mathrm{ratio}}$ shift, or change of axis orientation due to core correction. The biases appearing during transition from CONEX to CORSIKA do not impact the final results.

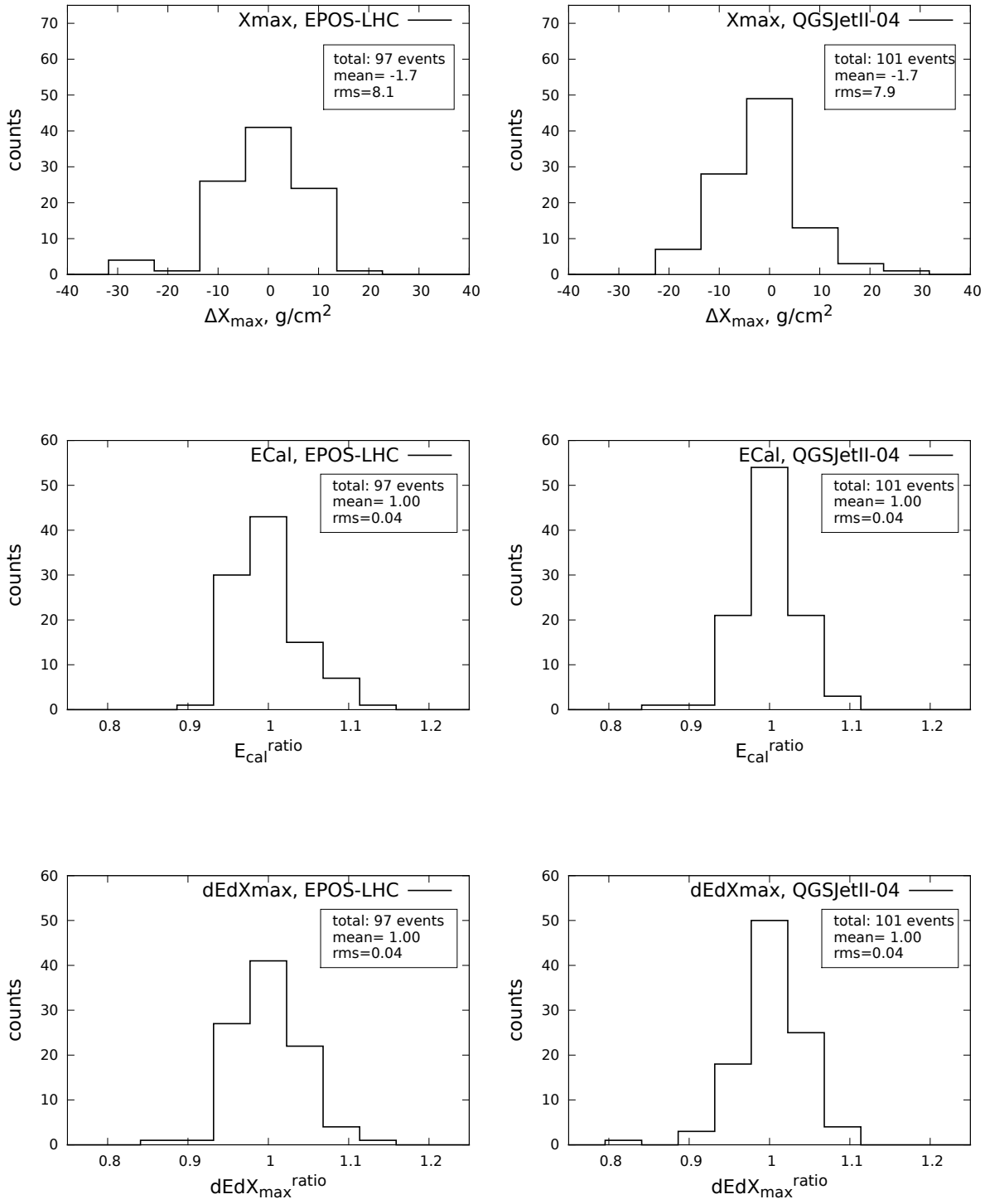

Figure 55: Reproduction of the longitudinal profile parameters after top-down reconstruction of the MC data. The parameters are $\Delta X_{\max} = X_{\max}^{simrec} - X_{\max}^{\text{obs}}$, $dEdX_{\max}^{\text{ratio}} = dEdX_{\max}^{simrec}/dEdX_{\max}^{\text{obs}}$, $E_{\text{cal}}^{\text{ratio}} = E_{\text{cal}}^{simrec}/E_{\text{cal}}^{\text{obs}}$

	model	mean	RMS
$\Delta X_{\max}$, g/cm^2	EPOS	-1.7	8.1
	QGSJetII-04	-1.7	7.9
$E_{\text{cal}}^{\text{ratio}}$	EPOS	1.00	0.04
	QGSJetII-04	1.00	0.04
$dEdX_{\max}^{\text{ratio}}$	EPOS	1.00	0.04
	QGSJetII-04	1.00	0.04

Table 7: Summary of figure 55: mean values of $\Delta X_{\max} = X_{\max}^{simrec} - X_{\max}^{\text{obs}}$, $dEdX_{\max}^{\text{ratio}} = dEdX_{\max}^{simrec}/dEdX_{\max}^{\text{obs}}$, $E_{\text{cal}}^{\text{ratio}} = E_{\text{cal}}^{\text{simrec}}/E_{\text{cal}}^{\text{obs}}$ and their standard deviations (in brackets).

The histograms of $E_{\text{cal}}^{\text{ratio}}$ and $\Delta X_{\max}$ may be regarded as a projection of energy-$\Delta X_{\max}$ parameter space on one of the coordinates. For all MC events set such parameter space is plotted for both used interaction models (see figure 56). Such presentation of the whole data set allows us to check if there is any correlation between outliers in the histograms. In such plot it is seen that the EPOS model has all points in energy-$X_{\max}$ parameter space distributed symmetrically around center, in an oval-shape cloud. There is no visible correlation, the outliers are distributed around the cloud of points. This suggest that the SD station signal comparison, shown in the next chapter, will not be affected by FD effects. For the QGSJetII-04 model the points are also distributed around the center of the parameter space, however, the outliers are located in such way that suggest small negative correlation. Since the number of outliers is small (∼3% of the total number of points), their presence may be an effect of rare fluctuations, which occurred during simulation of shower development. To discriminate the reason for such behaviour a more numerous data sample is needed.

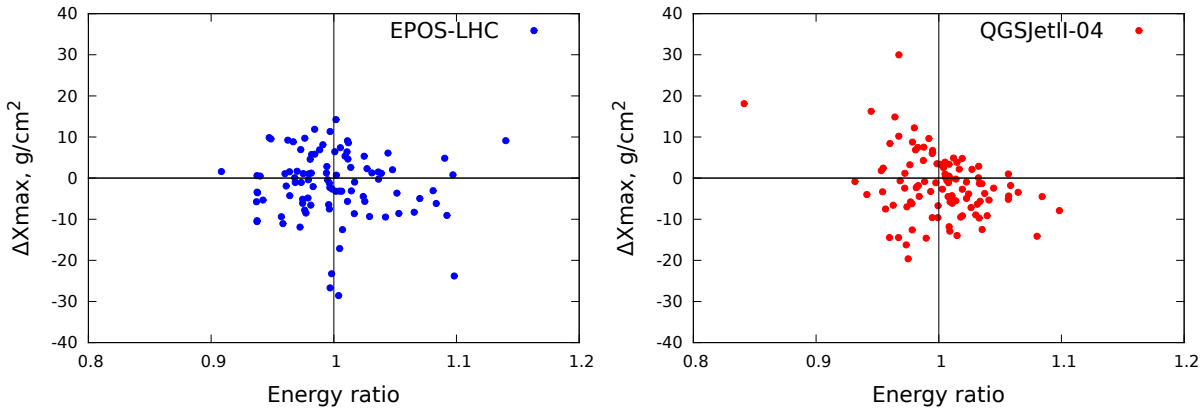

Figure 56: Plot of location of the reconstructed events on energy-$X_{\max}$ plane. No significant correlation or shift of the whole sample is visible. That seems to confirm that no bias in top-down reconstruction of the MC events is present.

The results above suggest that no statistically significant bias is visible in longitudinal profile parameters reproduced using the entire top-down reconstruction. So, since the MC events set seems to be reproduced well, a reproduction of the SD station signals are to be investigated.

6.3 Top-down reconstruction: SD signals

Determination of SD station signal ratios is the essential point of the top-down reconstruction. Not all triggered stations are used in signal ratios determination. A "10 VEM" cut for expected signal is applied to avoid a trigger bias. This bias arises mostly for stations located far from core, and is due to the SD station signal fluctuations. For stations with lower signals the fluctuations may cause that a given station sometimes is not triggered, when the fluctuation cause a lower signal in station. In such case the SD with higher signals will be triggered (upward fluctuation), which may lead to overestimation of the mean signal, since the stations with lower signals will not be included in calculation of the mean signal. The value of 10 VEM is commonly used in Auger data analysis to remove bias due to upward fluctuations. Additionally the cut on station distance in shower plane is used (minimum allowed shower-plane distance is 400 meters), to avoid effects due to signal saturation. This analysis is focused on a test of the total SD station signal reconstruction, so no disentangling of shower components is performed. So all particles passing through the SD are treated collectively in the results shown below (total SD station signals are used).

The accuracy of SD station signal reconstruction is analysed as follows. Only the SD stations which passed the 10 VEM cut are taken into account. For used energy of showers (10^{19} eV) this cut limits the maximum station distances to around 1500 m. The differences of SD station signal values are calculated (SD station signals in MC events - reproduced SD station signals). Signal differences are then binned in bins with equal width, and normalised to the standard deviation in bin. The results are plotted in function of core distance in shower-plane (fig. 57). Also a distribution of the entire set of signal differences is shown (fig. 58). Next, the SD station signal ratios for all triggered SD stations are calculated and binned in function of core distance in shower-plane (fig. 59, 60). Finally, the properties of S_{1000} signal ratios are investigated in function of shower-plane core distance. The values of S_{1000} are obtained as standard result in Offline SD reconstruction from fit of NKG function to the reconstructed signals (fig. 61).

It can be seen that on the average the relative signal differences lie within $\pm 1\sigma$ region, where σ is the standard deviation used as a measure of fluctuations in bin (fig. 57). So, for all MC events, no statistically significant bias in reconstructed SD station signals is visible for all used core distances. Both interaction models give similar results, showing only small fluctuations of binned values around the zero value. The histogram of relative signal differences for all stations (fig. 58) suggest their normal distribution. Such behaviour of reconstructed signals allows us to calculate the signal ratios in SD stations. The convention used in SD station signal ratios calculations is (MC events SD signal/simrec SD signal).

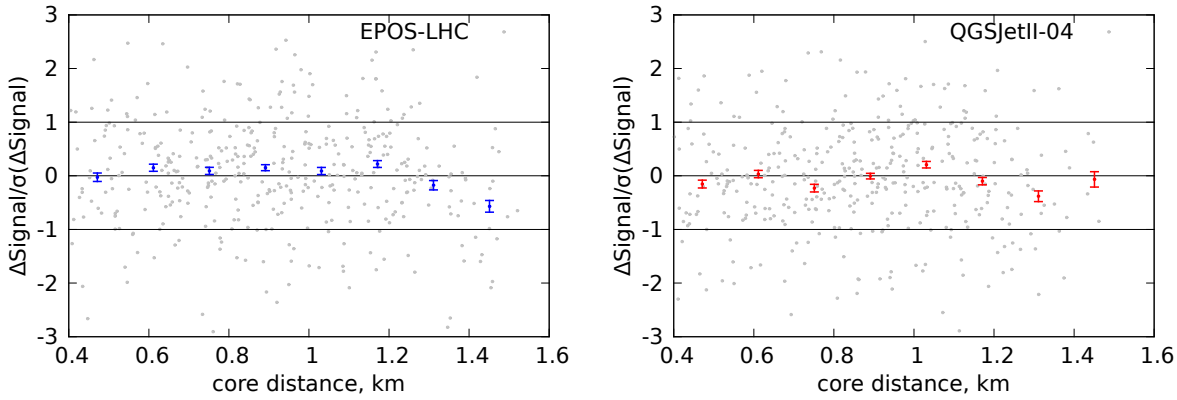

Figure 57: Reconstructed SD station signals relative to standard deviation in bin in function of shower plane core distance.

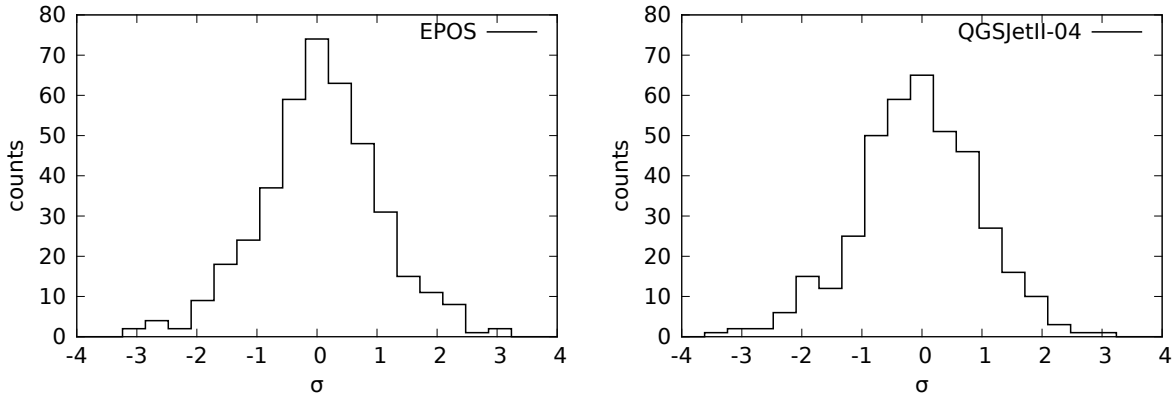

Figure 58: Histogram of reconstructed SD station signals relative to standard deviation in bin (σ).

The total signal ratios are plotted in a function of core distance (fig. 59). The mean value for all signal ratios is 1.01±0.01 (EPOS), and 1.02±0.01 (QGSJetII-04). The behaviour of the binned values of signal ratios does not show any statistically significant drift up to distance $\sim$ 1200 m. Above that distance a small bias seems to occur. A closer look at the signal ratios (fig. 60) show the EPOS interaction model has the binned ratios equal to 1 within 1σ for almost all bins. No statistically significant trend is visible. Only in the last bin, centered around distance 1.4 km from the core, the binned value of SD station signal ratio shows tendency to be higher than 1. For the QGSJetII-04 model the binned SD station signal ratios differ from 1 1 within 2σ. The overall mean of reconstructed SD station signal ratios for QGSJetII-04 is 1.02±0.01, which may suggest a small bias, but more numerous data set is necessary to diagnose it. Currently the systematic shift of SD station signal ratios between the two models (seen on fig. 60) is not statistically significant. The error bars in the binned SD station signal ratios for QGSJetII-04 model overlap with bars of the EPOS-LHC in almost all bins, therefore statistically the eventual bias is small. The fact, that also the $X_{\max}$-energy plot shows not uniform distribution of

reconstructed events for QGSJetII-04 model (fig. 56) indicate that the method applied for this model must be used carefully.

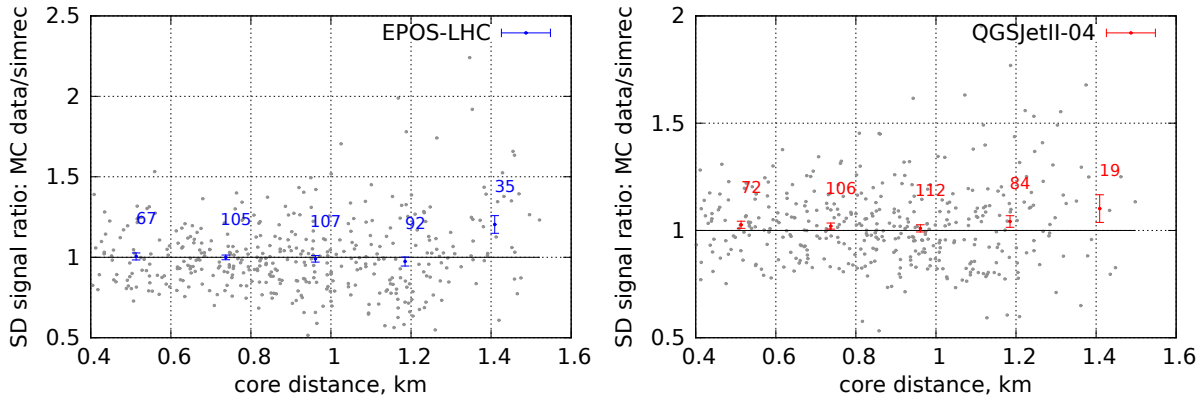

Figure 59: SD station signal ratios for individual stations (simrec/MC data). Grey points present values for all triggered stations. Binned values are coloured. Number of stations in bin are on the plot. The binned values are compared in figure 60.

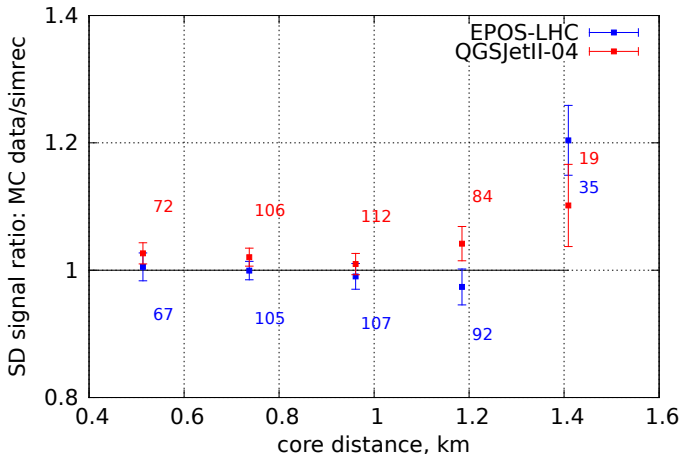

Figure 60: A comparison of binned SD station signal ratios from plots in figure 59 (simrec/MC data).

S_{1000} ratios are determined from the fit of LDF function to the SD signals. Obtained S_{1000} ratio characteristics is plotted in figure 61. It has a flat characteristics for all zenith angles. The overall mean of S_{1000} ratios is: 0.98±0.02 (EPOS), 1.02±0.02 (QGSJetII-04).

It is seen that the SD station signals, represented either by signals in individual stations and by the S_{1000}, are reproduced properly for EPOS model and with sufficient accuracy for QGSJetII-04 model. The problem of bias existence for QGSJetII-04 model is to be investigated using much larger amount of data, but from statistical point of view the overall mean lies within 2σ. Taking it into account the further analysis wit hthe real data will be carried out for both models.

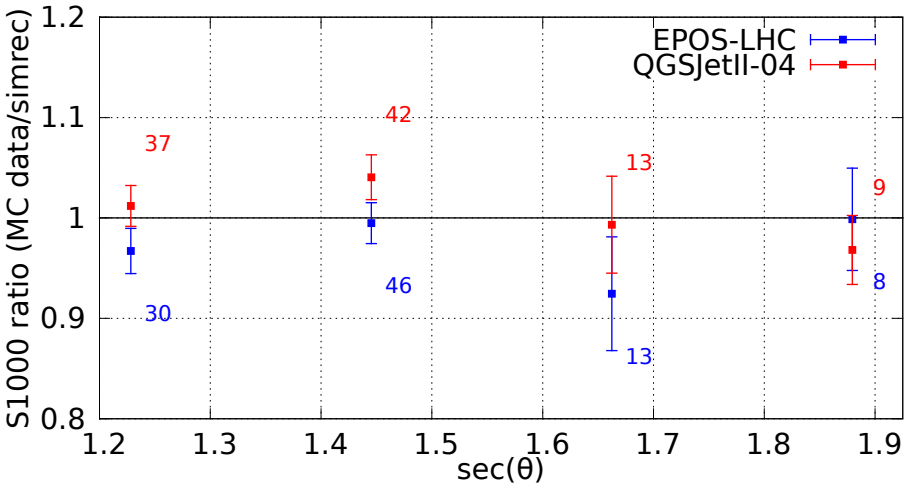

Figure 61: Binned S_{1000} signal ratios in function of sec(Θ) for two interaction models (EPOS-LHC and QGSJetII-04, simrec/MC data). The numbers of events in bin are shown on the plot. No significant correlation of S_{1000} signal ratios for reconstructed MC data is visible.

6.4 Top-down reconstruction: mass sensitivity

Primary mass is a parameter which must be assumed to perform the top-down reconstruction. In order to determine how much the method is mass-sensitive, the top-down reconstruction of proton MC events is performed assuming nitrogen as the primary. The choice of nitrogen as primary comes from interpretation of mass determination for real data described in section 2.2. This is why the nitrogen will be also used as a primary in reconstruction of real data events (see next chapter).

The mass discrimination test was performed by an analysis of the S_{1000} ratios calculated as $S_{1000}(\text{proton})/S_{1000}(\text{nitrogen})$ (fig. 62, left panel). No statistically significant dependence on the shower inclination is visible for the S_{1000} ratios. Taking into account the fact, that nitrogen have more interacting protons, the number of muons produced in interactions is expected to be larger, hence the S_{1000} ratio should be smaller than 1, which is confirmed in this test. This confirms mass sensitivity property of top-down reconstruction method.

Also the SD station signal ratios (figure 62, right panel) have values lower than one, but their dependence on the core distance is not flat within all ranges of core distance. Starting from a distance $\sim$ 1000 m the plot shows following tendency: the more distant stations, the lower signals coming from the nitrogen simrec events. The reason for it is as follows. The air showers induced by more massive nuclei develop faster than proton shower, which is due to larger number of particles in primary. In the mass sensitivity test the proton shower is to be reconstructed with nitrogen. So it means that the nitrogen shower should be characterised by $X_{\max}$ close to proton shower to provide successful reconstruction. Since the nitrogen shower develops faster, the simulation with nitrogen which will have $X_{\max}$ close to the $X_{\max}$ of proton shower needs to start deeper in the

atmosphere. In effect the lateral distribution of the nitrogen shower on ground will be steeper. Since at regions close to the core the lateral distribution of nitrogen shower is larger than proton shower, the ratio is lower than 1. At larger distances the steeper lateral distribution of nitrogen shower starts to be lower than the distribution of proton shower, hence the proton shower distribution starts to dominate. In effect at larger distances the ratios of the distributions, calculated as proton distribution/nitrogen distribution, will have values above 1. It is also notable that both models are very consistent with each other. These properties of interaction models for MC data may be also useful in study of the interaction models using top-down method.

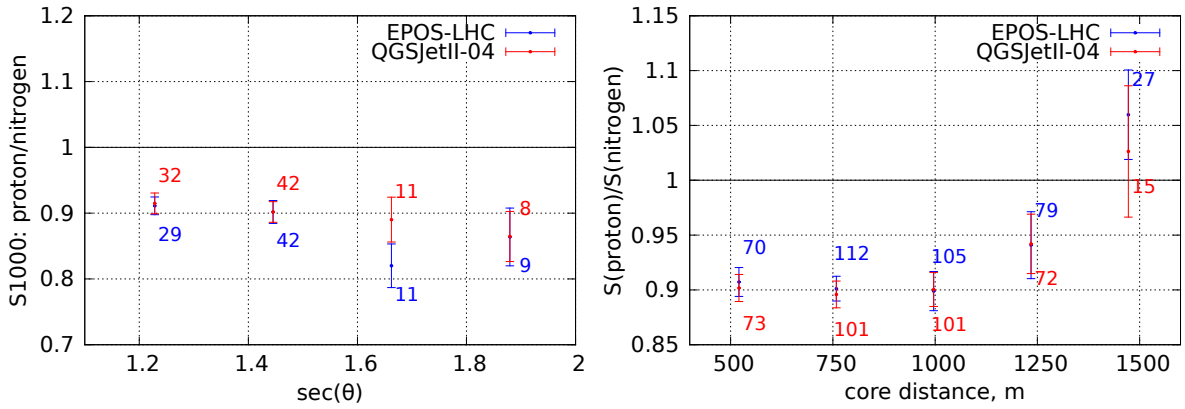

Figure 62: Binned values of SD station signal ratios (MC proton event/simrec nitrogen event) for individual stations. The numbers of stations in bin are shown on the plot.

6.5 Summary for validation of the top-down method

The validation of the entire top-down reconstruction chain seems to confirm, that the results obtained with the top-down simulation chain are consistent. It was shown that the top-down reconstruction method reproduces the geometrical and physical characteristics of reconstructed Monte Carlo showers with satisfactory accuracy for both EPOS-LHC and QGSJetII-04 models. Also it was shown that the updated simulation chain does not introduce any significant bias, even despite that in some inner steps of the chain small distortions of the longitudinal shape appear (CONEX to CORSIKA transition). The impact of these distortions on the final results seems to be negligible, they do not propagate through the remaining chain elements, influencing on the results only by decreasing of the accuracy of results. For QGSJetII-04 model a possible small bias in reproduced SD signals exist, but larger data sample is required to confirm it. For the used sample the overall SD signal ratios are equal to 1 with 95% confidence level for both models. Taking into account that both MC event sets were generated in similar way, and the EPOS showers have no bias, a possible bias for QGSJetII-04 may be linked directly to the intrinsic properties of the model. However, since for the QGSJetII-04 the possible bias is small ($\sim$2% in signals), and not statistically significant, this model will be also used in reconstruction of the real data in next chapter.

7 Application to the real data

7.1 Data set

The top-down reconstruction chain, described and tested in previous chapters, has been applied to the real data hybrid events. They are selected from the events registered in the Auger observatory in years from 2005 to 2012. Only the high quality events, i.e. events registered during good weather conditions and having high signal to noise ratio are taken into account. The selection is performed using quality cuts described below.

The cuts used are similar to the cuts in [95], which include cuts linked with the weather conditions and cloud presence, as well as the cuts limiting minimal quality of longitudinal profile, and are as follows. To discriminate events detected during bad weather conditions a vertical atmospheric optical depth (VAOD) is used as the main selection criterion. When the integrated value of the VAOD from ground to altitude 3 km is larger than 0.1 at the time of shower detection, such event is omitted. Another used selection criterion linked with the weather conditions is the cloud presence during detection of shower. The events are rejected if a cloud was detected along direction of the shower axis. In case when the clouds were detected in other directions than direction of the shower axis, it was required that the cloud base must be located above altitude for which the corresponding atmospheric depth is 400 g/cm^2. Such value of cloud base altitude is linked with geometry of the FD telescope field of view. The main purpose of this cut is to eliminate events detected with additional fluorescence light coming from the reflections on the clouds. In case of a lack of the cloud data, the corresponding cut limits allowed average cloud fraction during detection of shower to be below 25% of the whole sky.

For the events which passed the criteria above, another set of selection criteria linked with quality of the registered shower profiles are applied. First the quality of FD calibration is taken into account during data selection. It is done by approving the events registered during periods with high quality calibration of PMT gains in FD PMT matrix. The profile quality cuts exclude also events characterised by gaps in the longitudinal profile. If the gap is larger than 20% of the total observed track length, measured in atmospheric depth, the event is rejected. Such gap occurs when e.g. shower track passes through the fields of view of two or more telescopes in the same FD eye. It was required that the registered track length should be larger than 300 g/cm^2. Such a cut was applied to assure a proper determination of $X_{\max}$ from Gaisser-Hillas function fit. Since the longer shower track the closer shower core, the cut for observed track length eliminates also too close events. A summary histograms presenting χ^2/ndf of Gaisser-Hillas fit and FD-core distances are shown in figure 64.

An additional quality cut, introduced in this analysis, is that the shower track seen in the FD must trigger at least 12 pixels in the PMT matrix. It is done to assure that the signal-to-noise ratios in FD signals are high. Due to such cut the minimum FD-shower distance (R_{p}) of events is limited, since this cut additionally limits minimum angular size of a shower as seen in an FD telescope, and so the minimum distance from FD eye. The

distances between a shower and FD do not exceed 25 km (figure 64, right panel). On figure 64 it is seen that most of events have R_{p} from 10 to 20 km. More distant events have angular length in FD field of view too small to trigger 12 pixels. Closer events have too short track lengths in the FD field of view, so they are rejected due to cuts described above. The example two profiles from the selected real data set are shown in figure 65. The GH fit (red line) describes well the observed longitudinal profile, the maximum of the GH fit is determined well. So all the profiles in the entire data set have visual characteristics between the two presented profiles. A narrow energy range is chosen: $10^{18.8}$ to $10^{19.2}$ eV, since in such range the composition of cosmic rays does not evolve significantly. Such energy range is identical to the energy range used in a similar analysis [93], performed with SENECA as the simulator of shower development. The histograms of X_{max} and $\sigma_{X_{\mathrm{max}}}$ distributions in the selected data sample is shown in figure 63 (lower panel). It is seen that the majority of events have values of $\sigma_{X_{\mathrm{max}}}$ below 15 g/cm^2. Most of the selected events have X_{max} located between 700 and 770 g/cm^2. The deepest data shower has X_{max} around 900 g/cm^2. The final set of the used real data includes 113 events.

Before the top-down reconstruction is performed, a reconstruction of the raw real data is performed in the standard way using the hybrid mode (see section 3.5). The same Offline version (v2r9p1 Valentine) is used in the standard reconstruction and in subsequent detector response simulation performed during the top-down reconstruction. The real showers are reconstructed with taking into account the so-called halo effect, which is one of the effects arising during propagation of light to the FD pixels. This effect is not simulated in Offline, only an empirical estimation of its contribution to the total FD light is taken into account during standard reconstruction of the real data. Therefore the real data used in the top-down reconstruction are previously reconstructed in a standard way with taking into account the halo effect. For MC events the halo was not simulated and so no correction was necessary during standard reconstruction of these events.

A subsequent top-down reconstruction of the real data was performed independently for two interaction models (EPOS LHC and QGSJetII-04), and for two types of primary particle (proton and nitrogen). Therefore each event was reconstructed four times, each time with different model and primary assumed in the following combinations: proton with EPOS, proton with QGSJetII-04, nitrogen with EPOS, nitrogen with QGSJetII-04. The choice of used primaries (proton and nitrogen) is based on other composition studies, described in section 2.4. In the cited studies data suggest a large nitrogen fraction in cosmic rays at energy $\sim 10^{19}$ eV. Both types of primary are used in separate runs.

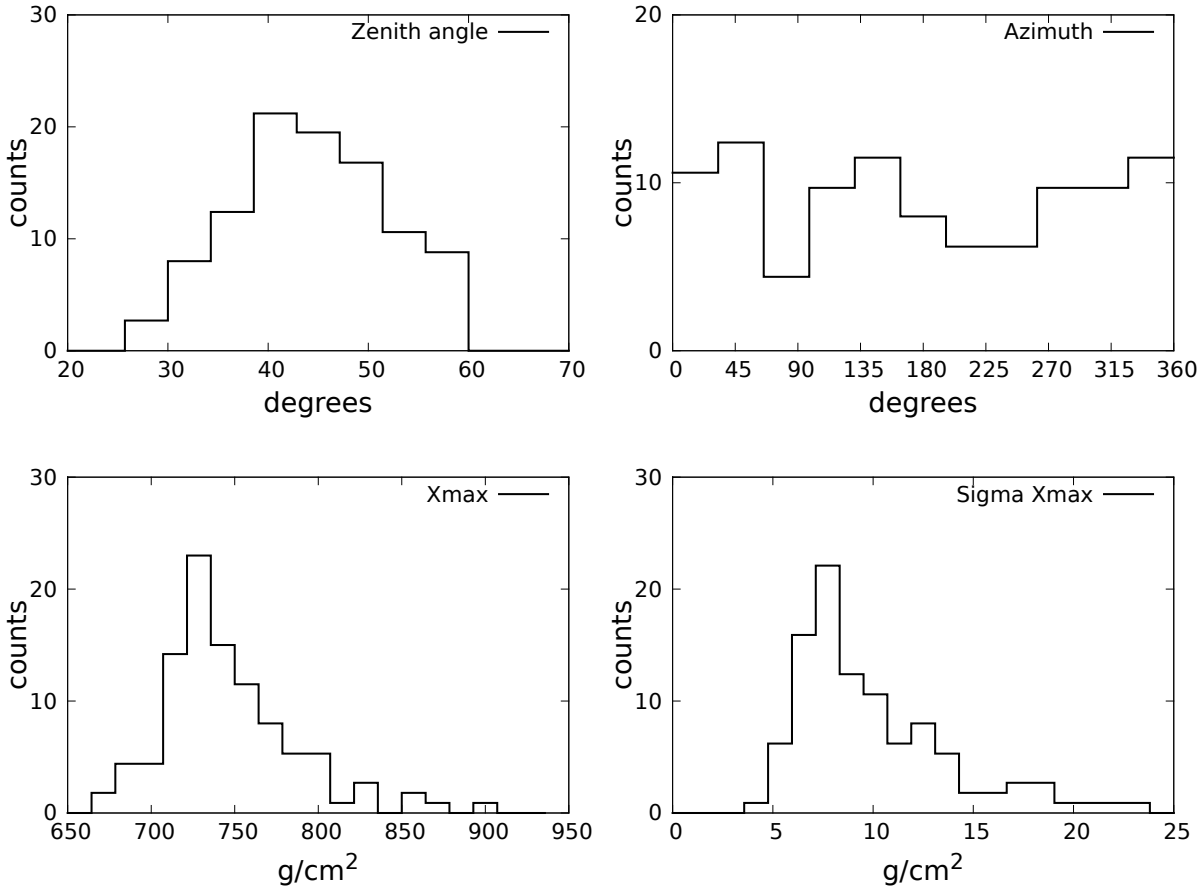

Figure 63: Review of basic characteristics of 113 real data events used in top-down reconstruction. The plotted characteristics include incoming directions of primary in local coordinate system (zenith and azimuth), and values of $X_{\max}$ and $\sigma_{X_{\max}}$.

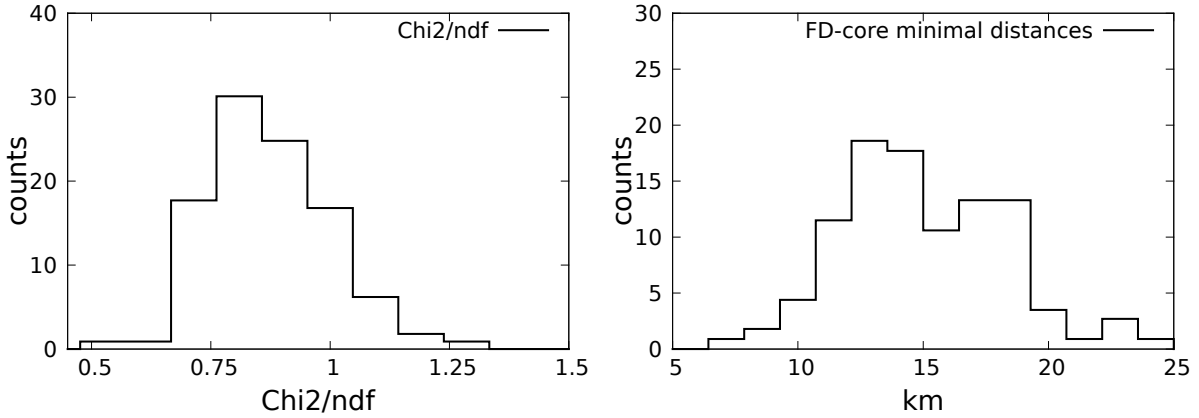

Figure 64: Left panel: profile accuracy measured as Chi2/ndf for 113 real data events used in top-down reconstruction. Right panel: distribution of FD-shower minimum distance of 113 real data events.

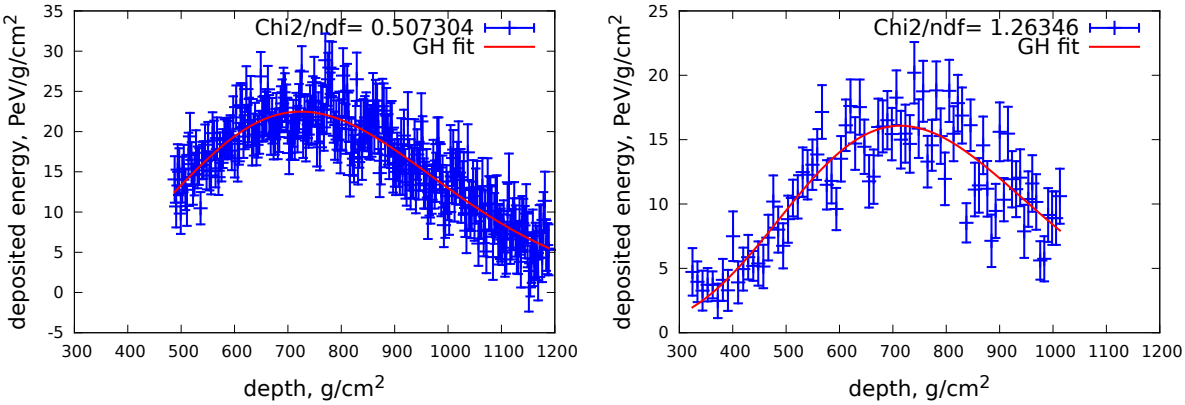

Figure 65: Longitudinal profiles of two real data events: with smallest and largest χ^2/ndf. Red line is a fitted Gaisser-Hillas function.

The true composition of cosmic rays in the data showers is unknown. Therefore if some real data events were induced by more massive primary, the top-down reconstruction of such events with proton as primary may not succeed, since the showers induced by more massive primary have shallower $X_{\max}$. In such case the real data event, which cannot be reconstructed using the top-down method, is just omitted from the further analysis. In effect, the number of finished reconstructions for proton is 98 using EPOS-LHC and 105 using QGSJetII-04, while the initial number of events is 113. For nitrogen as primary the number of reconstructions is 99 using EPOS-LHC and 95 using QGSJetII-04.

7.2 Cross checks after top-down reconstruction

Cross-checks are done to control the quality of the top-down reconstruction. As it was mentioned before, the reconstruction relies on a comparison of the data longitudinal profile with simulated longitudinal profile. So the cross-checks performed here include comparison between the longitudinal profiles at input (from the real data events) and the corresponding reconstructed profiles at output (final stage of the chain). Such checks are necessary to control quality of the real data reconstruction. It will be also helpful to exclude possibility that some rare fluctuation occurred. Such fluctuation may occur e.g. during transition from CONEX to CORSIKA (see section 5.7).

For comparison of the simulated longitudinal profiles with the data profile the following parameters are used (similarly to the validation in section 6.2): $X_{\max}$, $E_{\text{cal}}^{\text{ratio}}$, $dEdX_{\max}^{\text{ratio}}$. The appropriate histograms are presented in figures 66, 67, 68, with summary in table 8. Each primary and interaction model is plotted separately. Also, in order to see if any correlation between the reproduced energy and $X_{\max}$ occurred, a plot of these parameters on energy-$X_{\max}$ plane is shown in fig. 69.

It is seen that on average the basic characteristics of the longitudinal profiles are reproduced well by the top-down reconstruction. The profile parameters are preserved within 1σ. It indicates that the top-down reconstruction does not introduce any statistically significant bias for the chosen narrow energy range for the real data events. The mean uncertainty of $X_{\max}$ reproduction is around 10 g/cm^2 for both models and both primaries. This uncertainty is close to the uncertainty of the input real data events, for which the mean $\sigma_{X_{\max}}$ is $\sim$ 10 g/cm^2 (see $\sigma_{X_{\max}}$, fig. 63 lower right panel). A reproduction of E_{cal} and $dEdX_{\max}$ of the profile is also satisfactory, with uncertainty around 4%-5%. It is similar to the uncertainty obtained in reconstruction of the MC events ($\sim$4%). So the quality of top-down reconstruction of the real data is similar to the quality obtained in the validation. Also the correlation plot in energy-$X_{\max}$ parameter space, for each reconstruction separately (fig. 69), does not show significant correlation nor shift. A small number of outliers are present for both primaries and interaction models. The presence of outliers may indicate that the assumption about the primary type is not correct. Despite it the cross-checks allow us to expect that results of the SD signal reproduction obtained for the whole data sample (presented in next section) should not be affected by statistically significant effects.

	model (primary type)	mean	st. dev.
$\Delta X_{\max}$, g/cm^2	EPOS (proton)	1.1	9.1
(fig. 66)	QGSJetII-04 (proton)	-1.5	9.4
	EPOS (nitrogen)	-1.2	10.7
	QGSJetII-04 (nitrogen)	-3.8	9.6
$dEdX_{\max}^{\text{ratio}}$	EPOS (proton)	1.00	0.05
(fig. 67)	QGSJetII-04 (proton)	1.00	0.05
	EPOS (nitrogen)	1.0	0.04
	QGSJetII-04 (nitrogen)	1.00	0.05
$E_{\text{cal}}^{\text{ratio}}$	EPOS (proton)	0.99	0.04
(fig. 68)	QGSJetII-04 (proton)	1.00	0.04
	EPOS (nitrogen)	0.99	0.04
	QGSJetII-04 (nitrogen)	1.00	0.04

Table 8: Summary of the cross checks shown in figures 66, 67, 68: mean values of $\Delta X_{\max} = X_{\max}^{simrec} - X_{\max}^{\text{obs}}$, $dEdX_{\max}^{\text{ratio}} = dEdX_{\max}^{simrec}/dEdX_{\max}^{\text{obs}}$ $E_{\text{cal}}^{\text{ratio}} = E_{\text{cal}}^{simrec}/E_{\text{cal}}^{\text{obs}}$ and their standard deviations.

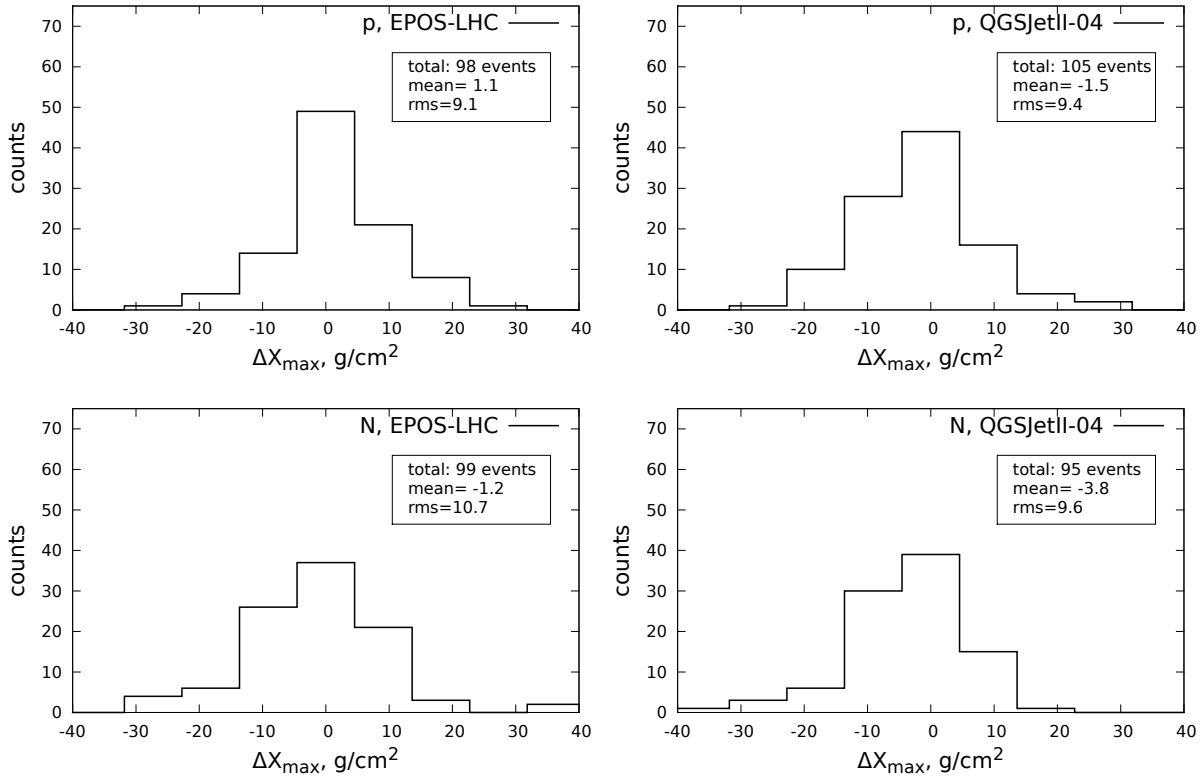

Figure 66: Cross checks of Xmax reproduced during top-down reconstruction ($\Delta X_{\max} = X_{\max}^{simrec} - X_{\max}^{\text{obs}}$).

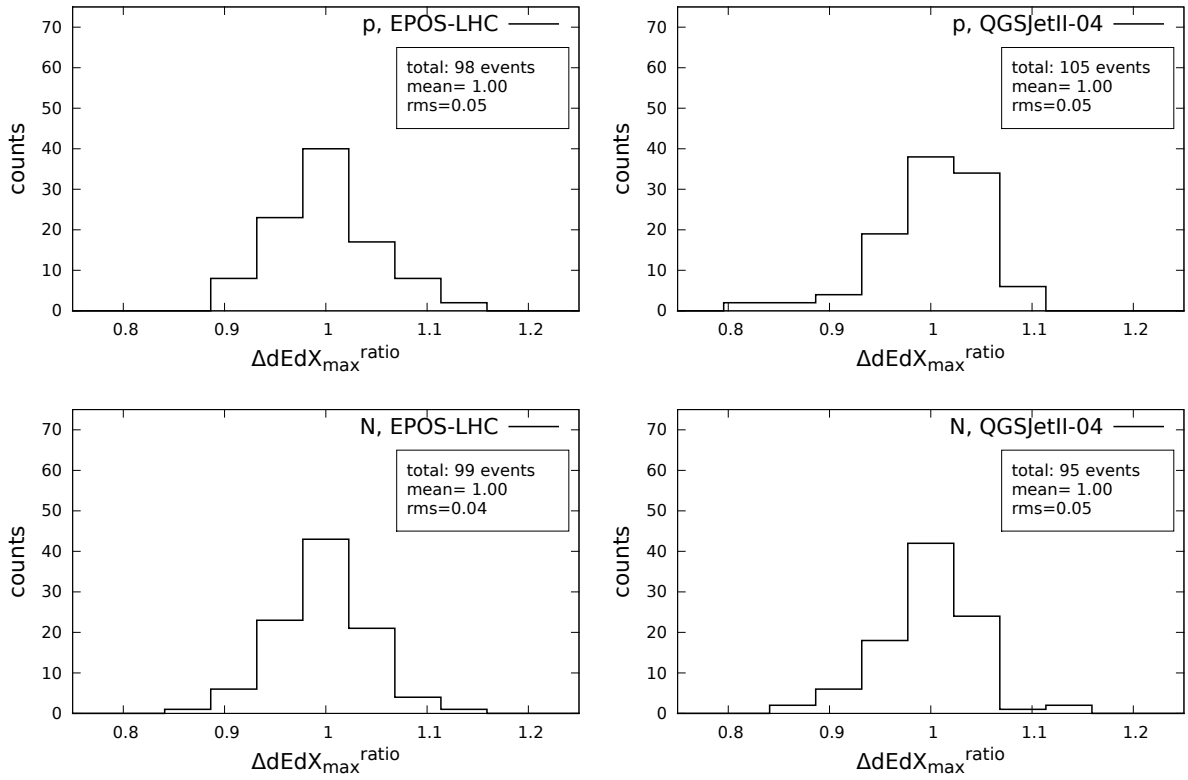

Figure 67: Cross checks of dEXmax reproduced during top-down reconstruction ($dEdX^{\rm ratio}_{\rm max} = dEdX^{simrec}_{\rm max}/dEdX^{\rm obs}_{\rm max}$).

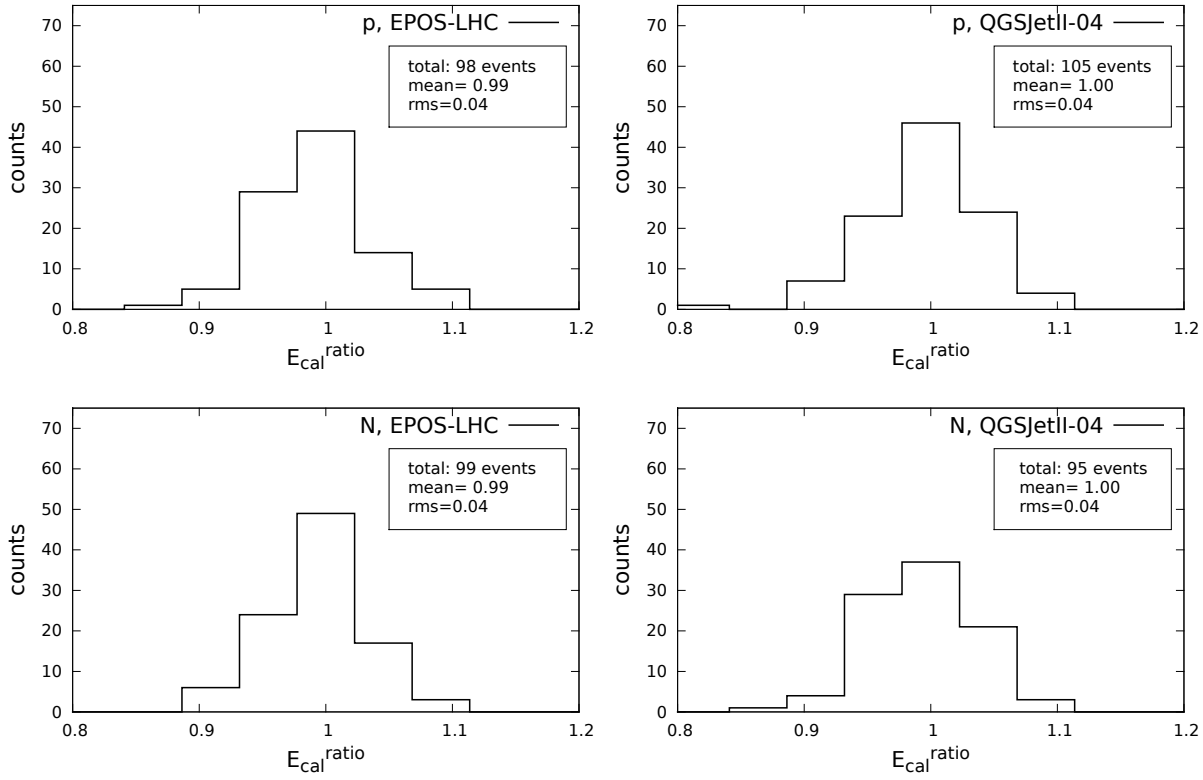

Figure 68: Cross checks of Ecal reproduced during top-down reconstruction ($E^{\rm ratio}_{\rm cal} = E^{simrec}_{\rm cal}/E^{\rm obs}_{\rm cal}$).

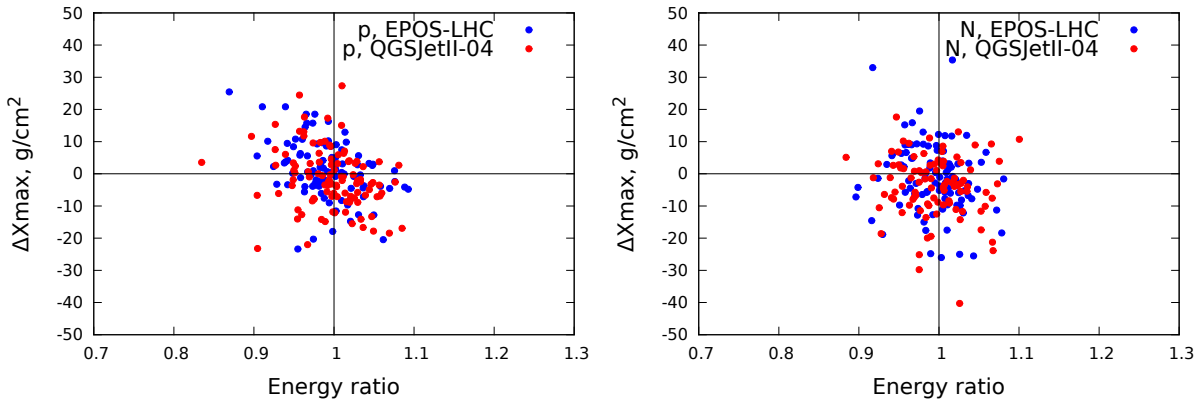

Figure 69: Plot of location of the reconstructed events on energy-$X_{\max}$ plane. No significant correlation or shift of the whole sample is visible. That seems to confirm that no bias in top-down results is present.

7.3 SD signals reconstruction

The SD station signals reconstructed in the top-down are compared with the signals from the real data. To compare the signals, the ratios of S_{1000} values are calculated (data/simrec), as well as ratios of individual SD station signals (data/simrec). All ratios of S_{1000} are binned in equidistant steps of sec(Θ), and similarly in case of individual SD station signals. The binned S_{1000} ratios and SD signal ratios for the whole reconstructed data set are presented in figures 70 and 71 respectively, with summary in table 9. Similarly like in the validation, the 10 VEM cut is applied to avoid the trigger bias in SD signals. The SD-core distances of stations included in the ratio calculation lie within range 400-1200 meters. The upper bound of allowed SD-core distance is due to the possible bias in SD signal ratios at larger distances, which occurred in the validation (see section 6.7).

For both models the reproduced SD signals are smaller than in the real data. The ratios, calculated as: data signal/simrec signal, have values larger than 1. The S_{1000} ratios tend to increase with increasing zenith angle for both primaries. For proton showers this dependence was also obtained in a similar analysis in [93]. This dependence is linked with smaller contribution of the electromagnetic component to SD signals in more inclined showers. It is notable that for nitrogen as primary the discrepancy between simulations and data S_{1000} ratios is reduced, but still is present. Since the mass composition derived from the $X_{\max}$ suggests large contribution of components with mass close to the mass of nitrogen (see section 2.2), it is not surprising that the top-down reconstruction results with nitrogen as primary are in better agreement with the data. Assuming that the models are correct, the better agreement of signal ratios obtained for nitrogen could be tentatively interpreted as an independent confirmation of a heavier mean mass composition of cosmic rays at energy $\sim 10^{19}$ eV. The additional simulations for more massive primaries must be performed to quantify the dependence. However, such behaviour of the results suggest that the method may be feasible in mass-oriented analyses.

For both models the individual SD station signal ratios (fig. 71) are characterised by an absence of significant dependence on core distance. It is seen that there is a systematic difference between the models in all plots of signal ratios. The QGSJetII-04 model always tends to predict smaller SD signals than the EPOS-LHC model, so the data/simrec signal ratio is higher in the convention used. This property is seen in the entire range of core distance for SD signal ratios. The difference between the models is statistically significant. In the validation it was shown that for the MC events the QGSJetII-04 reproduces the overall SD signal ratio about 2% lower in simrec events. Assuming that the difference between the model simulations is a result of a simple 2% scaling of the signals reproduced by QGSJetII-04 model, the mean SD signal ratio would be around 1.22$\pm$.02, but the data provide 1.33$\pm$.02 (for proton). Taking it into account the difference is expected to be linked with the properties of the models. Another fact supporting the conjecture that the difference is linked with the model properties is the muon number in showers simulated using the two models. It is shown e.g. in [86], where muon number tends to be higher for EPOS. Hence the simulated SD signals (and S_{1000}) are expected to be systematically

higher for EPOS than for QGSJetII-04 (so the ratios are lower for EPOS). This indicates that the top-down reconstruction method gives consistent results when compared with other analyses performed with the models used. It also proves that the method is reliable for analysis of interaction model testing, and may enable further improvement of the models.

	model (primary type)	mean	st. dev.
S_{1000}^{ratio}	EPOS (proton)	1.18	0.02
	QGSJetII-04 (proton)	1.30	0.02
	EPOS (nitrogen)	1.05	0.02
	QGSJetII-04 (nitrogen)	1.15	0.02
SD station signal ratios	EPOS (proton)	1.18	0.02
	QGSJetII-04 (proton)	1.32	0.02
	EPOS (nitrogen)	1.06	0.01
	QGSJetII-04 (nitrogen)	1.15	0.01

Table 9: Summary of the SD station signals reproduction for the real data events: mean of S_{1000} staton signal ratios and SD station signal ratios (the used convention for ratio calculation is data/simrec). The values of ratios are systematically larger than 1, which means that the models reproduce too small signals. Table refers to figures: 70, 71.

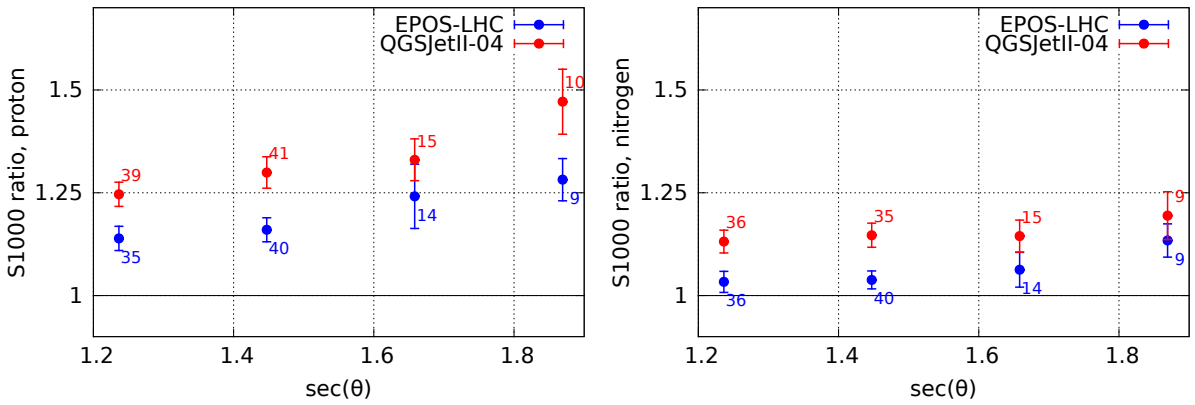

Figure 70: S_{1000} ratios in function of sec(Θ) (simrec/real). The numbers located close to the points indicates number of entries in bin.

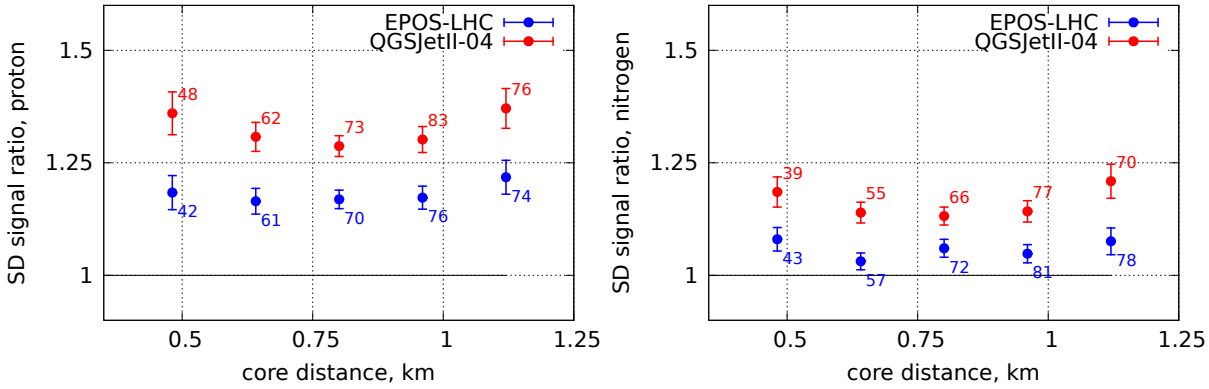

Figure 71: SD signal ratios in function of core distance (simrec/real data). Numbers of stations in bin are shown on the plot.

8 Conclusions and outlook

Properties of the top-down method were investigated in this work to prove usefulness of the method in testing of hadronic interaction models. The investigation included a test of the top-down method using generated Monte Carlo events. The main goal of the test was to show that the top-down method reproduces properly the signals in Surface Detector (SD) stations.

Another intention was to show that when the top-down method is used to reconstruct the real data, a discrepancy in Surface Detector station signals between the data and simulations is obtained. This discrepancy is expected to be similar to that obtained in the other analysis performed which different simulation tools. A detailed analysis of the discrepancy will be helpful to investigate properties of hadronic interaction models. Therefore the output signals obtained in the top-down method should not be affected by a bias or a noise, of which a source might be the top-down reconstruction itself, or other factors. It is also important to have in mind that the air showers in general (and so also the simulated showers) are characterised by large fluctuations. So the following factors are investigated, regarded as possible sources of bias or errors. Among them are the atmosphere quality and a shift of core position on ground. Also a widely used in simulations mean detector level may be a source of bias, since in the real world the detectors are placed at different altitudes. So instead of the mean detector level a precise altitude equal to altitude of the detector is used in simulations. In this thesis these sources of possible bias are taken into account and their impact on the quality of reconstruction are estimated.

A proper choice of atmosphere model used in simulations seems to be important, since the real atmosphere profile tends to vary significantly even during a single day. This is why a use of mean seasonal atmosphere model in simulations may be regarded as a possible source of noise. Therefore, instead of using a mean seasonal atmosphere model, each event is reconstructed with its unique atmosphere model assigned to given real data event. Such atmosphere model is obtained from a fit to the measured atmosphere profile at the time closest to the time of detection of the shower. The uncertainty of the fit turned out not to exceed 4 g/cm^2, and is much smaller than the $X_{\max}$ uncertainty of the Pierre Auger Observatory (15 g/cm^2). Such uncertainty is believed to be sufficient to provide unbiased results due to seasonal variations of atmosphere profile.

An improvement of the initial core location on the ground (core correction) is another investigated item. The measured core position is used during the top-down simulation run as an initial location of simulated shower on the Auger site. The idea is that the locations of the real data shower and its top-down reconstruction should coincide within the data shower core location uncertainty (usually around 10-40 meters). Since the reconstruction procedure tends to provide core location shifted up to 120 meters, the core correction had to be introduced. This correction is especially important since in this study the locations

of Surface Detector stations relative to simulated shower core position should be as similar as possible to the real data core positions. So the point is to sustain the same positions of the Surface Detector stations relative to the simulated and real data core position, so that a given Surface Detector station has the same fixed position during the top-down simulation as in the real event. Otherwise, when the core shift is not taken into account, a given Surface Detector station will detect a different part of the simulated shower. To see if the improved core location on ground works fine, for each triggered Surface Detector station a comparison of SD station-core distance differences is done. It is shown that the SD station-core distances are improved after use of the core correction. It turned out that for used hadronic interaction models (EPOS LHC and QGSJetII-04) the uncertainty of the core position without correction is $\sim$50 m, while with the core correction is $\sim$20 m. In effect the SD station-core distances measured in shower plane are also improved (without correction the accuracy is $\sim$35 m, with correction $\sim$14 m). So, with the core correction the uncertainty of core position is decreased by factor $\sim$2. When the core correction is applied, the achieved core position uncertainty in shower plane is only 1% of the distance between the SD stations.

However, the improvement of the initial core location may influence the shower axis. If so, the total light in FD may suffer from the shower axis change, since the amount of air in which the fluorescence light will travel through will be different. This may be important in the last step, when selection of final 10 Offline reconstructions is performed. The influence of the core correction on the shower axis orientation was analysed by comparing reconstructed shower axis in two separate reconstructions: with and without the core location correction. It was shown that the core correction does not change significantly the orientation of the shower axis (the change is less than 0.1°). Such systematic change of shower axis is negligible, since the uncertainty of axis orientation in Pierre Auger Observatory is $\sim$0.6°. Therefore the applied core correction is useful.

A comparison of the longitudinal profiles was performed for individual chain element denoted as "single FD reconstruction". For such comparison the profile parameters ($X_{\max}$, $dEdX_{\max}$ and E_{cal}) are used. A single FD reconstruction is performed during the first stage of the method, when the initial shower profile calculations are in progress. The initial simulation provides a raw simulated shower (represented by the longitudinal profile only) which is the input in this step. The output of this step is the FD response which is subsequently compared with the input. It was shown that in this step the reproduction of the longitudinal profile suffers a small bias most likely due to a dust accumulation on FD. The bias is that in the real event the light entering to the FD is reduced which causes the profiles are seen as lower at shower maximum by about 3%. In effect the reconstructed calorimetric energy is smaller by almost 3%. This effect is taken into account since the same detector setup is used in generation and reconstruction of simulated events, as well as in reconstruction of the real data. So all events are biased in the same direction, which result in proper reproduction of the whole top-down method as shown below.

Next the accuracy of the longitudinal profile reproduction for the whole top-down

reconstruction chain is investigated. It is done by comparing the profile parameters ($X_{\max}$, $dEdX_{\max}$ and E_{cal}) between the input data shower and the top-down output. It came out that for the whole top-down reconstruction the $X_{\max}$ is reconstructed with uncertainty 8 g/cm^2. The remaining profile parameters ($dEdX_{\max}$ and E_{cal}), for which ratios are used, are reconstructed with 4% uncertainty. Such uncertainty is smaller than uncertainty of the standard reconstruction in the Pierre Auger Observatory, for which $X_{\max}$ uncertainty is 15 g/cm^2 and energy determination uncertainty is about 8%. Therefore it was shown that the top-down method, as used in this work, reproduces the longitudinal profile well. Since the uncertainty of the reconstruction is smaller than the uncertainty of the real data, it seems that the top-down method does not introduce any additional bias or noise in the longitudinal profile reproduction. Also no significant correlation between the $X_{\max}$ and the energy was found. This indicates that the longitudinal profile is reproduced well, and so it enables an investigation of the Surface Detector signals.

To perform successful top-down reconstruction, one must find a simulated shower most similar to data. To achieve it, a generation of many longitudinal profiles is performed during a top-down run. A generation of simulated profiles is time consuming. Therefore, to reduce CPU usage, a simplified calculations of such profiles are performed. When the most similar profile is found, it is resimulated more precisely in next steps of the method. However, sometimes it may occur that the resimulated profile has a shape which differs from the initial profile. In such case this profile should be omitted. To assure that such possibility did not occur during top-down reconstruction, a comparison of initial profiles and their resimulations were performed. The comparison proved that all used profiles were properly resimulated, with a small, negligible shift in $X_{\max}$ by about 2-3 g/cm^2. The remaining parameters (calorimetric energy, $dEdX_{\max}$) are reproduced with low uncertainty (better than 1%).

The top-down reconstruction relies on the comparison of longitudinal profiles (simulated profile is compared with the data profile). The properties of selection process were also studied. The selection procedure is performed using a χ^2-like measure. The uncertainty of the comparison may be treated as the uncertainty of the measure for assumed 500 initial simulation used to find one best profile which match the data profile. Such comparison, performed similarly like described above for the longitudinal profile, showed that the searching procedure works with a satisfactory uncertainty ($\Delta X_{\max} \sim 0 \pm 5$ g/cm2, the $E_{\mathrm{cal}}^{\mathrm{ratio}} \sim 1 \pm 3\%$, $dEdX_{\max}^{\mathrm{ratio}} \sim 1 \pm 4\%$ between the data profile and the most similar profile). It seems the assumed 500 initial simulations are sufficient to find a profile which is similar to data, and which will be used in subsequent steps.

Since the longitudinal profile is accurately reproduced, the next step was to compare the Surface Detector station signals. The individual Surface Detector station signals are reproduced well, without any visible drift up to 1200 m from the shower core. For larger distances, limited to 1500 m from the shower core due to "10 VEM cut", large fluctuations have place. The fluctuations cause that the Surface Detector station signal ratio has value around 1.15. The overall mean ratio of Surface Detector station signals for EPOS is

1.01±0.01, and for QGSJetII-04 is 1.02±0.01. So for EPOS reconstructed Surface Detector signals are equal to 1 within 1 sigma. The 2 sigma shift of the mean Surface Detector signal ratio for QGSJetII-04 may suggest existence of a bias, but further studies are necessary to quantify it. For a given shower the Surface Detector signal at 1000 m from the shower core is used for parameterisation of lateral distribution of shower particles. Value of the Detector signal at 1000 m is obtained from a fit to Surface Detector station signals and assigned as S_{1000}. The properties of S_{1000} reproduction were also investigated. It turned out that the S_{1000} is reproduced with a satisfactory uncertainty (the S_{1000} ratios are: 0.98±0.02 for EPOS and 1.02±0.02 for QGSJetII04).

The mass sensitivity of the method is preliminarily investigated by performing the top-down reconstruction of proton events generated using Monte Carlo method, assuming nitrogen as the primary. The results indicate that the top-down method gives different values of reconstructed Surface Detector signals when different primary is used, which is according to expectations. The Surface Detector signals for nitrogen as the primary are larger than for signals obtained with assumed proton primary. Also different distributions of particles on ground for the two primaries are obtained, which is seen in decreasing value of Surface Detector signals for nitrogen in stations more distant than ∼1000 m.

Finally, the method was applied to 113 real data events. Again, proton and nitrogen as primaries were used during reconstruction performed in two separate runs (each primary was reconstructed with EPOS and separately with QGSJetII-04 interaction models). The main results of the reconstruction are as follows. It was shown that for the real data the top-down method provides smaller Surface Detector signals, which is a result of the muon number problem. Such discrepancy is obtained for both used primaries (proton and nitrogen). Only the stations located at distance up to 1200 m from the core are taken into account, since for larger distances the fluctuations are larger and the statistics starts to be too low (which was shown in the test with MC events).

For the real data the top-down reconstruction with nitrogen as primary fits better to the data. It is in agreement with the mean composition derived from the $X_{\max}$, which suggests the real data composition is closer to nitrogen than to proton. Additionally it is notable that the S_{1000} and Surface Detector signal ratios for QGSJetII-04 are systematically higher than for EPOS. This difference is a result of different models properties. Now it is clearly seen that a proper reproduction of Surface Detector signals, which also includes reduction of possible noise sources which may affect the results, is crucial for further studies. The difference between the Surface Detector signals may be used e.g. to study the properties of the interaction models with the top-down as a tool. The obtained differences are thus in agreement in other analyses of the used interaction model properties, and is in agreement with expectations. On the other side, since the Surface Detector signal ratios change with primary mass, this dependence may be used as another source of valuable clues about mass composition of cosmic rays. So the top-down method may be used in composition studies of the cosmic rays. The method feasibility, either in hadronic interaction properties tests or in composition studies, may provide a significant

improvement of UHECR knowledge.

Summarising, it can be stated that the method, as presented in the thesis, can be used in studies of properties of the cosmic rays. Also, the method is sensitive to the properties of different interaction models, and therefore it is feasible to study their properties. In a further study the reconstruction should be extended to higher energies and different types of primaries to quantify the discrepancies and to provide other hints about cosmic rays composition.

Bibliography

[1] D.H. Perkins, *Particle Astrophysics, second edition,* Oxford University Press, Oxford, 2009.

[2] J. Beringer et al., *Cosmic rays,* Physics Review D 86, 010001 (2012)

[3] S.P. Swordy, *The energy spectra and anisotropies of cosmic rays,* Space Science Reviews 99, 85 (2001).

[4] J. Blümer, R. Engel, J.H. Hörandel, *Cosmic rays from the knee to the highest energies* Progress in Particle and Nuclear Physics 63, 293 (2009).

[5] Pierre Auger Collaboration, *Observation of the suppression of the flux of cosmic rays above* 4×10^{19} *eV,* Physics Letters B 685, 239 (2010).

[6] G. Matthiae, *The cosmic ray energy spectrum as measured using the Pierre Auger Observatory,* New Journal of Physics 12, 075009 (2010).

[7] J.R. Horandel, *On the knee in the energy spectrum of cosmic rays,* Astroparticle Physics 19, 193 (2003).

[8] V.S. Ptuskin et al, *Diffusion and drift of very high energy cosmic rays in galactic magnetic fields,* Astronomy & Astrophysics 268, 726 (1993)

[9] A.M. Hillas, *Can diffusive shock acceleration in supernova remnants account for high-energy galactic cosmic rays?* Journal of Physics G: Nuclear and Particle Physics 31, R95 (2005).

[10] R. Aloisio, V. Berezinsky, A. Gazizov, *Transition from galactic to extragalactic cosmic rays,* Astroparticle Physics 39-40, 129 (2012).

[11] K. Greisen, *End to the Cosmic-Ray Spectrum?,* Physics Review Letters 16, 748 (1966)

[12] R. Aloisio, V. Berezinsky, and A. Gazizov, *Ultra high-energy cosmic rays: The disappointing model,* Astroparticle Physics 34, 620 (2011)

[13] Pierre Auger Collaboration, *Searches for Anisotropies in the Arrival Directions of the Highest Energy Cosmic Rays Detected by the Pierre Auger Observatory,* Astrophysical Journal 804, 15 (2015)

[14] C. Blacksley, *Photodetection Aspects of JEM-EUSO and Studies of the Ultra-High Energy Cosmic Ray Sky,* Ph.D. Thesis, University Paris Diderot (Paris 7); Defended 19th November, 2013 (arXiv:1406.5818).

[15] M.S. Longair, *High-energy Astrophysics, Volume 1: Particles, photons and their detection, second edition,* Cambridge University Press, Cambridge, 1992.

[16] A. Schulz, *The measurement of the energy spectrum of cosmic rays above* 3×10^{17} *eV with the Pierre Auger Observatory,* Proceedings of 33^{rd} Inernational Cosmic Ray Conference, Rio de Janeiro 2009.

[17] Pierre Auger Collaboration, *Properties and performance of the prototype instrument for the Pierre Auger Observatory,* Nuclear Instruments and Methods in Physics Research A 523, 50 (2004)

[18] Antoni et al., *The Cosmic-Ray Experiment KASCADE,* Nuclear Instruments and Methods in Physics Research 513, 490-510 (2003).

[19] H. Kawaii et al., *Telescope Array Experiment,* Nuclear Physics B (Proc. Suppl.) 175, 221 (2008).

[20] T. Abu-Zayyad et al., *The prototype high-resolution Fly's Eye cosmic ray detector,* Nuclear Instruments and Methods in Physics Research A 450, 253 (2000)

[21] H. Ulrich et al., *Energy spectrum and elemental composition of cosmic rays in the PeV region,* The European Physical Journal C 33, s01, 944 946(2004).

[22] Piere Auger Collaboration, *Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Measurements at Energies above* $10^{17.8}$ *eV,* Physical Review D 90, 122005 (2014).

[23] Piere Auger Collaboration, *Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Composition Implications,* Physical Review D 90, 122006 (2014).

[24] Pierre Auger Collaboration, *Search for photons with energies above* 10^{18} *eV using the hybrid detector of the Pierre Auger Observatory,* arXiv:1612.01517 (2016).

[25] V.S. Berezinsky, G.T. Zatsepin, *Cosmic rays at ultra high energies (neutrino?),* Physics Letters B 28, 423 (1969)

[26] M. Ahlers, *High-Energy Cosmogenic Neutrinos,* Physics Procedia 61, 392 (2015).

[27] E. Waxman, J. N. Bahcall, *high-energy astrophysical neutrinos: The upper bound is robust,* Physical Review D 64, 023002 (2001).

[28] R.J. Protheroe, *High-energy Neutrinos from Blazars,* arXiv:astro-ph/9607165 (1996).

[29] P. W. Gorham et al. *Observational constraints on the ultrahigh-energy cosmic neutrino flux from the second flight of the ANITA experiment,* Physical Review D 85, 049901(E) (2012).

[30] M.G. Aartsen et al. *Probing the origin of cosmic rays with extremely high-energy neutrinos using the IceCube Observatory,* Physical Review D 88, 112008 (2013).

[31] Pierre Auger Collaboration, *An improved limit to the diffuse flux of ultra-high-energy neutrinos from the Pierre Auger Observatory,* Physics Review D 91, 092008 (2015)

[32] O.E. Kalashev et al., *Ultrahigh-energy neutrino fluxes and their constraints,* Physical Review D 66, 063004 (2002).

[33] The Pierre Auger Collaboration, *Search for photons above* 10^{18} *eV with the hybrid detector of the Pierre Auger Observatory,* Journal of Cosmology and Astroparticle Physics 04, 009 (2017).

[34] T.W. Kephart, T.J. Weiler, *Magnetic Monopoles as the Highest Energy Cosmic Ray Primaries,* arXiv:astro-ph/9505134 (1995).

[35] C.O. Escobar, R.A. Vazquez, *Are high-energy Cosmic Rays Magnetic Monopoles?,* Astroparticle Physics 10, 197 (1999).

[36] Pierre Auger Collaboration, *Search for ultrarelativistic magnetic monopoles with the Pierre Auger Observatory,* Physical Review D 94, 082002 (2016).

[37] E.N. Parker, *The Origin of Magnetic Fields,* Astrophysics Journal 160, 383 (1970).

[38] M. Ambrosio et al., *Final results of magnetic monopole searches with the MACRO experiment,* European Physical Journal C25, 511 (2002).

[39] R. Abbasi et al., *Search for Relativistic Magnetic Monopoles with IceCube,* Physical Review D Vol 87 022001 (2013).

[40] M. Detrixhe et al., *Ultrarelativistic magnetic monopole search with the ANITA-II balloon-borne radio interferometer,* Physical Review D 83, 023513 (2011).

[41] D. Hogan, et al., *Relativistic Magnetic Monopole Flux Constraints from RICE,* Physical Review D 78, 075031 (2008).

[42] M. Lemoine, G.Sigl, *Physics and Astrophysics of Ultra high-energy Cosmic Rays,* Springer-Verlag 2001.

[43] A.M. Hillas, *The origins of ultra-high-energy cosmic rays,* Annual Review of Astronomy and Astrophysics 22, 425 (1984).

[44] P.M. Bauleo, J.R. Martino, *The dawn of the particle astronomy era in ultra-high-energy cosmic rays,* Nature volume 458, pages 847-851 (2009).

[45] E. Fermi, *On the Origin of the Cosmic Radiation,* Physical Review 75, 1169 (1949).

[46] T. Stanev, A. Letessier-Selvon, *Ultrahigh-energy Cosmic Rays,* Reviews of Modern Physics 83, 907 (2011).

[47] T. Vachaspati, *Formation, Interaction and Observation of Topological Defects* arXiv:astro-ph/9903362 (1999)

[48] V. Berezinsky, *Ultra high-energy cosmic rays,* Nuclear Physics B - Proceedings Supplements 70, 419 (1999).

[49] M. Risse, P. Homola, *Search for ultra-high-energy photons using air showers,* Modern Physics Letters A 22, 749 (2007).

[50] Pierre Auger Collaboration, *A search for point sources of EeV photons,* Astrophysical Journal 789, 160 (2014).

[51] R.U. Abbasi et al., *A Northern Sky Survey for Point-Like Sources of EeV Neutral Particles with the Telescope Array Experiment,* Astrophysical Journal 804, 133 (2015).

[52] Pierre Auger Collaboration, *A Search for Point Sources of EeV Neutrons,* Astrophysical Journal 789, 160, (2014).

[53] Pierre Auger Collaboration, *Searches for Anisotropies in the Arrival Directions of the Highest Energy Cosmic Rays Detected by the Pierre Auger Observatory,* Astrophysical Journal 804, 15 (2015).

[54] The Pierre Auger Collaboration, *Observation of a large-scale anisotropy in the arrival directions of cosmic rays above* 8×10^{18} *eV,* Science 357, Issue 6357, p. 1266 (2017)

[55] A. Haungs, *KCDC - The KASCADE Cosmic-ray Data Centre,* arXiv:1504.06696

[56] T.K. Gaisser, A.M. Hillas, Proc. of 15^{th} ICRC, Plovdiv, Bulgaria, 1977, vol. 8, pag. 353.

[57] T. Stanev, *high-energy cosmic rays (second edition),* Springer Verlag, 2010.

[58] K. Greisen, *Cosmic rays showers,* Annual Review of Nuclear Science 10, 63 (1960).

[59] W. Heitler, *The Quantum Theory of Radiation,* Oxford University Press, 3 ed. (1954).

[60] J. Matthews, *A Heitler model of extensive air showers,* Astroparticle Physics 22, 387 (2005).

[61] Pierre Auger Collaboration, *Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events,* Physical Review D 91, 032003 (2015); ERRATA: Phys. Rev. D 91, 059901 (2015).

[62] J. Linsley, L. Scarsi, B. Rossi, Journal of the Physical Society of Japan Suppl. C 17, 91 (1962).

[63] N. Chiba et al., Nuclear Instruments and Methods in Physics A 311, 338 (1992).

[64] D.M. Edge et al., *The cosmic ray spectrum at energies above* 10^{17} *eV,* Journal of Physics A 6, 1612 (1973).

[65] D.Newton et al., *The optimum distance at which to determine the size of a giant air shower,* Astroparticle Physics 26, Issue 6, 414 (2007).

[66] S. J. Sciutto, *Physics of Astroparticles,* Brazilian Journal of Physics 37, Issue 2B, 494 (2007).

[67] J. Hersil et al., *Observations of extensive air showers near the maximum of their longitudinal development,* Physical Review Letters 6, 22 (1961).

[68] M. Risse, D. Heck, *Energy Release in Air Showers,* Astroparticle Physics 20, 661 (2004).

[69] H.M.J. Barbosa et al., *Determination of the calorimetric energy in extensive air showers,* Astroparticle Physics 22, 159 (2004).

[70] M. Ave et al., *Spectrally resolved pressure dependence measurements of air fluorescence emission with AIRFLY,* Nuclear Instruments and Methods in Physics Research Section A, Vol. 597, Issue 1, p. 41 (2007).

[71] Pierre Auger Collaboration, *The Surface Detector System of the Pierre Auger Observatory,* Nuclear Instruments and Methods in Physics A 586, 409 (2008).

[72] Pierre Auger Collaboration, *Performance of the Pierre Auger Observatory Surface Detector,* arXiv:astro-ph/0709.1823 (2007)

[73] Pierre Auger Collaboration, *The Fluorescence Detector of the Pierre Auger Observatory,* Nuclear Instruments and Methods in Physics Research A 620, 227 (2010).

[74] J. Blümer et al., *Atmospheric Profiles at the Southern Pierre Auger Observatory and their Relevance to Air Shower Measurement,* Proceedings of 29^{th} ICRC, Pune 2005.

[75] Pierre Auger Collaboration, *Atmospheric effects on extensive air showers observed with the Surface Detector of the Pierre Auger Observatory,* Astroparticle Physics 32, 89 (2009); Erratum-ibid.33, 65 (2010).

[76] https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-data-assimilation-system-gdas

[77] Pierre Auger Collaboration, *Atmospheric aerosols at the Pierre Auger Observatory and environmental implications,* The European Physical Journal Plus 127, 97 (2012).

[78] Pierre Auger Collaboration, *Atmospheric Monitoring and its Use in Air Shower Analysis at the Pierre Auger Observatory,* Proceedings of 31^{st} ICRC, Łódź 2009.

[79] Pierre Auger Collaboration, *Measuring Atmospheric Aerosol Attenuation at the Pierre Auger Observatory,* Proceedings of 33^{rd} ICRC, Rio de Janeiro 2013.

[80] Pierre Auger Collaboration, *Atmospheric Aerosol Measurements at the Pierre Auger Observatory,* Proceedings of 31^{st} ICRC, Łódź 2009.

[81] S. Argiro et al., *The Offline Software Framework of the Pierre Auger Observatory,* Nuclear Instruments and Methods in Physics A 580, 1485 (2007).

[82] J. Knapp et al., *Extensive Air Shower Simulations at the Highest Energies,* Astroparticle Physics 19, 77 (2003).

[83] R. Ulrich, R. Engel, M. Unger, *Hadronic Multiparticle Production at Ultra-High Energies and Extensive Air Showers,* Physical Review D 83, 054026 (2011).

[84] K.H. Kampert, M. Unger, *Measurements of the Cosmic Ray Composition with Air Shower Experiments,* arXiv:1201.0018 (2012).

[85] D. d'Enterria et al., *Constraints from the first LHC data on hadronic event generators for ultra-high-energy cosmic-ray physics,* arXiv:1101.5596 (2011).

[86] T. Pierog, *Modelling hadronic interactions in cosmic ray Monte Carlo generators,* EPJ Web of Conferences 99, 09002 (2015).

[87] A.M. Hillas, Proceedings of 17^{th} ICRC, Paris 1981, 193.

[88] Pierre Auger Collaboration, *A thinning method using weight limitation for air-shower simulations,* Astroparticle Physics 15, 259 (2001).

[89] D. Heck et al., Report FZKA 6019, Forschungszentrum Karlsruhe (1998).

[90] W.R. Nelson, H. Hirayama and D.W.O. Rogers, Report 265, Stanford Linear Accelerator Center (1985).

[91] T. Bergman et al., *One-dimensional Hybrid Approach to Extensive Air Shower Simulation,* Astroparticle Physics 26, 420 (2007).

[92] T. Pierog et al., Proceedings of 32^{nd} ICRC, Beijing 2011, contr. 1170

[93] Pierre Auger Collaboration, *Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory,* Physical Review Letters 117, 192001 (2016)

[94] S. Falk, R. Engel, R. Ulrich, M. Unger - private communication.

[95] Pierre Auger Collaboration, *Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: measurements at Energies above* $10^{17.8}$ *eV,* Physical Review D 90, 122005 (2014).

Acknowledgments

This research was partially supported by the National Science Centre grant No. 2016/23/B/ST9/01635.

I also would like to express a great gratitude to my colleagues from the pierre Auger Observatory project for their support which has been essential for this thesis preparation. I am especially grateful to my PhD supervisor, prof. Henryk Wilczyński, and auxiliary supervisor, dr Tanguy Pierog, for their guidance, enormous help, patience and support during the time of the research and the thesis preparation. I also owe a debt of gratitude to Ralph Engel, Darko Veberič, Markus Roth and Michael Unger from Karlsruhe Institute of Technology for their valuable advices and help during my research.