Energy Spectrum and Mass Composition of Cosmic Rays and How to Publish Air-Shower Data

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

von der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

Dipl. Phys. Sven Schoo

aus Rheine

Tag der mündlichen Prüfung:	17. Juni 2016
Referent:	Prof. Dr. Johannes Blümer
Korreferent:	Prof. Dr. Ulrich Husemann

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

Karlsruhe, 23.05.2016

.....

(**Sven Schoo**)

Abstract

Energy Spectrum and Mass Composition of Cosmic Rays and How to Publish Air-Shower Data

The energy spectrum and mass composition of cosmic rays is the main source of information used to constrain astrophysical models on the origin, acceleration, and propagation of cosmic rays. The present analysis aims to reconstruct a consistent energy spectrum and mass composition in the primary energy range of 10^{15} eV to 10^{18} eV by a combined reconstruction of the data of the KASCADE and KASCADE-Grande detectors. The new reconstruction delivers a significantly improved accuracy compared to the standalone analyses in particular for the determination of the number of electrons and muons. Nonetheless, a consistent reconstruction of the energy spectrum and composition for events located in KASCADE and KASCADE-Grande is still not possible due to differences between measurements and simulated data, which have been revealed only because of the better reconstruction of the air-showers. Where the features of the spectrum and composition found earlier (i.e. the knee as a cutoff of the contribution of light primaries, a "heavy knee" at about 10^{17} eV, as well as an ankle-like feature of the light component well below 10^{18} eV) are confirmed, the achieved accuracy is so high that now even remaining deficiencies of the hadronic interaction models to describe the development of the air showers are unveiled. In addition, it is demonstrated that a consistent reconstruction will be possible once a hadronic interaction model describes the measurements sufficiently well. A comparison of the results based on the two hadronic interaction models QGSJetII4 and EposLHC has revealed that a consistent result for light primaries is achieved. It is motivated that the remaining difference between the two models probably stems from a different extrapolation of the proton-proton interactions to proton-nucleus and nucleus-nucleus interactions. It is also found that the general differences between simulated data and measurements for those models are larger than the remaining differences between the two models.

Since KASCADE has finished its data taking and has meanwhile been dismantled, the availability of its high-quality data to the community and the general public is achieved by setting up the KASCADE Cosmic ray Data Centre (KCDC). Its main task is to publish the data together with the necessary documentation on the detector and reconstructed quantities. In the future, also the entire software of the web portal is going to be released as open source software. Having this in mind, its design is based on plugins to ensure that the available functionality can be easily extended or modified in order to accommodate possibly different requirements of other experiments. The current version of KCDC was developed as a part of this thesis.

Zusammenfassung

Energiespektrum und Massenkomposition der kosmischen Strahlung und wie man Luftschauer-Daten veröffentlicht

Das Energiespektrum und die Massenzusammensetzung der kosmischen Strahlung sind wichtige Informationen zur Unterscheidung astrophysikalischer Modelle des Ursprungs, der Beschleunigung und der Propagation der kosmischen Strahlung. Die vorgestellte Analyse hat die konsistente Rekonstruktion des Energiespektrums und der Komposition im Bereich von 10^{15} eV bis 10^{18} eV als Ziel. Dabei soll, verglichen mit den ursprünglichen Analysen, eine deutlich verbesserte Genauigkeit erreicht werden. Diese neu entwickelte Rekonstruktion ermöglicht in der Tat eine signifikant verbesserte Genauigkeit, insbesondere in der Bestimmung der Anzahl der Elektronen und Myonen im Luftschauer. Trotzdem ist noch immer keine konsistente Rekonstruktion des Energiespektrums und der Massenzusammensetzung für Ereignisse, die in KASCADE beziehungsweise in KASCADE-Grande gemessen wurden, möglich. Die Ursache liegt darin, dass die verwendeten hadronischen Wechselwirkungsmodelle die Natur nicht ausreichend genau beschreiben, was erst aufgrund der verbesserten Rekonstruktion der Luftschauer deutlich wurde. Die bereits vorher entdeckten Strukturen im Energiespektrum und in der Komposition (d.h. das Knie als Abbruch in dem Fluss leichter Elemente, das „schwere Knie“ bei etwa 10^{17} eV und eine knöchelartige Struktur im Spektrum der leichten Elemente weit unterhalb von 10^{18} eV) konnten bestätigt werden. Zusätzlich konnten nun aber, aufgrund der sehr hohen Genauigkeit der rekonstruierten Schauerparameter, verbleibende Ungenauigkeiten in der Beschreibung der Schauerentwicklung durch die hadronischen Wechselwirkungsmodelle aufgedeckt werden. Es wurde jedoch auch gezeigt, dass eine konsistente Rekonstruktion prinzipiell möglich ist, sobald ein Wechselwirkungsmodell zur Verfügung steht, welches die Messungen gut genug wiederspiegelt. Ein Vergleich der Resultate, die auf den Wechselwirkungsmodellen QGSJetII4 und EposLHC beruhen, hat ergeben, dass diese für leichte Teilchen kompatible Ergebnisse liefern. Demnach dürften die Unterschiede zwischen den beiden Modellen hauptsächlich auf eine unterschiedliche Extrapolation von Proton-Proton-Wechselwirkungen zu Proton-Kern- und Kern-Kern-Wechselwirkungen zurückzuführen sein. Es konnte ebenfalls gezeigt werden, dass die generellen Unterschiede zwischen simulierten und gemessenen Daten mittlerweile größer sind als die noch vorhandenen Unterschiede zwischen den beiden Modellen.

Da das KASCADE Experiment die Datennahme abgeschlossen hat und mittlerweile abgebaut ist, möchte die KASCADE-Grande Kollaboration die Verfügbarkeit der qualitativ hochwertigen Daten sicherstellen. Dazu wurde das KASCADE Cosmic ray Data Centre (KCDC) entwickelt, welches, neben den Daten, die benötigten Dokumentationen zum Detektor und zu den Daten zur Verfügung stellt. Da die Software in Zukunft frei verfügbar gemacht werden soll, basiert diese auf einem modularen Design. Dadurch soll eine einfache Modifikation oder Erweiterung der momentanen Funktionalität KCDCs möglich gemacht werden, um KCDC an die Anforderungen anderer Experimente anpassen zu können. Die aktuelle Version von KCDC wurde im Rahmen dieser Arbeit entwickelt.

Contents

Preface

The air-shower experiments KASCADE and its extension KASCADE-Grande can look back on a very successful past, having made important contributions to the current knowledge on the energy spectrum and mass composition of cosmic rays in the primary energy range of 10^{15} eV to 10^{18} eV. This was achieved by comparing the measured properties of extensive air showers induced by the primary particle to detailed simulations of these particle cascades. This dependence on simulations is useful and a common way to interpret air showers, but it is also a hindrance. On one hand, this dependence can be employed to test various aspects of hadronic interactions. This has been done to great success using KASCADE and KASCADE-Grande measurements, which cover energies and kinematic space regions not accessible by man made accelerators. For this kind of tests, KASCADE was especially well suited, as the hadronic, electromagnetic, and muonic component of an air shower have been measured simultaneously, the latter being even evaluated at four different muon energy thresholds. On the other hand, this dependency introduces limitations on how accurately the energy and mass of a primary particle can be reconstructed. Therefore, the more accurate the hadronic interaction models describe nature, the more accurate the estimate of the energy spectrum and mass composition will be.

The experimental installations of the KASCADE facility have been dismantled by now and the personpower dedicated to the analysis of more than 20 years of data is continuously decreasing. However, the hadronic interaction models have not evolved as far as to allow the statement that the high-quality data accumulated over all these years could not be used any more to help further improving them. And, by doing so, to improve also the determination of the energy spectrum and mass composition of cosmic rays. In fact, a combined analysis of the KASCADE and KASCADE-Grande measurements – presented in part 1 of this work – has led to a significantly more accurate reconstruction of the shower observables allowing to test hadronic interactions in greater detail than before. As will be shown, the reconstruction has improved so much that the hadronic interaction models are now the dominant source of systematic uncertainty and the main factor that prevents KASCADE and KASCADE-Grande from providing a consistent reconstruction of the composition from 10^{15} eV to 10^{18} eV. This is a pity as the analysis presented in this thesis provides a degree of detail and accuracy not reached up to now in this energy range.

Hence, to ensure that this valuable data is not lost to the community, the KASCADE-Grande collaboration has started a pioneering project for the publication of cosmic ray data, namely the KASCADE Cosmic ray Data Centre (KCDC). The newest and most sophisticated version of KCDC was developed as a part of this work and is presented in part 2. Having the release of its software (as open-source software) in mind, it is based on a modular design making it easy to be extended or modified to

match the requirements of other experiments as well.

It is a main motivation of KCDC that this large amount of high-quality data can be used also in the future to improve substantially our knowledge on the nature of cosmic rays and hadronic interactions.

Part I.

The Energy Spectrum and Mass Composition of Cosmic Rays

1. Introduction

The first part of this work is focused on the determination of the energy spectrum and mass composition of cosmic rays. They have been discovered more than 100 years ago by Victor Hess [1], who studied the development of the ionization in the atmosphere as a function of the height above sea level. The amount of radiation was expected to decrease with increasing height as it was believed at that time that the radiation originates from radioactive elements in the soil. Hess, however, measured an increase of the radiation level, which he interpreted to be caused by what he called "Höhenstrahlung", i.e. by radiation originating from outside the Earth's atmosphere. The true nature of this radiation was discovered several years later in 1929 by Bothe and Kolhörster [2] and in the 1930s by Rossi [3], and Auger [4]. They registered multiple particles in coincidence in separate detectors. It was concluded that at least part of the "Höhenstrahlung" is caused by particles (the cosmic rays) interacting with the atmosphere inducing extended particle cascades - called air showers. As will be shown in chapter 2, the mass composition and energy spectrum of cosmic rays are crucial for understanding the origin of the primary cosmic rays, as well as the acceleration and propagation mechanisms. While up to about 10^{14} eV, cosmic rays can be studied directly by satellites or balloon-borne experiments, ground based experiments are needed to measure cosmic rays at higher energies via the detection of the induced air showers (known to reach at least up to an energy of several 10^{20} eV as first measured by Fly's Eye in Utah, USA in 1991 [5]).

The air-shower experiments KASCADE and KASCADE-Grande have measured cosmic rays from $\approx 10^{15}$ eV to $\approx 10^{18}$ eV. KASCADE has recorded more than 1.735 billion events in the time from 25/10/1996 to 15/1/2013. KASCADE-Grande, which was an extension of the KASCADE experiment, started to take data on 20/12/2003 and measured about 894 million events until it has been dismantled at 31/10/2012. The energy range covered by the two detectors is of particular interest because changes in astrophysical processes are expected to occur, which would introduce prominent features to a otherwise structure-less spectrum. The first imprint would be related to the maximum energy that can be reached at supernova remnants, which are assumed to be the source of the bulk of galactic cosmic rays, i.e. cosmic rays originating from within our galaxy (see e.g. [6]). Such a source cutoff would result in a charge dependent drop in the contributions of the various elements to

the all-particle energy spectrum. Whether these cutoffs are visible individually and if the first one is the source of the "knee", depends on their relative abundances. Even the most powerful supernova remnants are not capable of accelerating particles up to the highest energies observed. Therefore, there has to be a transition to at least one additional population. The most common assumption is that cosmic rays at the highest energies are of extragalactic origin (see e.g. [7, 8]). Depending on the astrophysical modeling of galactic and extragalactic contributions, two or three populations are needed to explain the measured all-particle energy spectrum. Hence, an additional set of cutoffs, shifted towards slightly higher energies, could be present. However, since this second galactic component would be sub-dominant, these cutoffs would likely be visible only in the fluxes of individual elements and not in the all-particle spectrum. Finally, a transition to the extragalactic component is expected as there is experimental evidence by anisotropy studies [9] that the highest energetic particles are of extragalactic origin. Assuming a charge dependent leakage from their home galaxy, such a transition should become visible for light primaries first.

The various models can only be distinguished by measuring very accurately the energy spectrum and energy-dependent mass composition. In case of KASCADE and KASCADE-Grande, the mass sensitivity was achieved by simultaneously measuring the number of electrons and muons at observation level. Because a detailed modeling of air showers is needed for the indirect observation to relate the measured number of particles to the energy and mass of the primary particle, a simplified model of air showers will be discussed in chapter 3.

The separate analyses of the KASCADE and KASCADE-Grande data already gave important results regarding the energy spectrum and mass composition of cosmic rays, which are discussed in chapter 4. However, having employed two different reconstruction strategies, their results are not directly comparable as they intrinsically suffer from different systematic uncertainties. Therefore, a combined analysis, treating both arrays as one single detector, has been devised. This is the subject of chapter 5. The general idea was to derive one single estimate of the energy spectrum and mass composition covering the entire energy range accessible for the two detectors, namely from 10^{15} eV to 10^{18} eV. In addition, a more detailed estimate should be possible due to a more accurate reconstruction of the shower observables. As will be shown in section 5.2, the latter is indeed achieved. However, the accuracies of the reconstruction of N_e and N_μ are not exactly the same for KASCADE and KASCADE-Grande giving rise to systematic differences between showers located in the respective other detector despite the combined shower reconstruction procedure. It will be demonstrated in section 5.3 that this is not a problem in itself, as these systematic uncertainties are well understood and the flux [1] can, in principle, be corrected. However, it will also be shown that even slight differences between simulation and measurement cause discrepancies between the results obtained for measured events located in the respective other array. Unfortunately, this limitation cannot be removed at present with the existing hadronic interaction models and one will have to wait for the definite result of the combined analysis until these hadronic interaction models experience a further improvement. The current limitations of the

[1] The flux is the number of particles arriving at Earth within a certain area, time, solid angular range, and energy range. Its unit is given by $[\mathrm{m}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}\,\mathrm{eV}^{-1}]$.

analysis are summarized at the end of the section, so that a discussion of the results can take place in chapter 6, which will conclude part 1 of this work.

2. Cosmic Rays

Even more than 100 years after the discovery of cosmic rays it is not fully understood where they come from, how they are accelerated, and how they reach Earth. In the following a short overview of the main features observed so far in the energy spectrum of cosmic rays and how these features might be related to acceleration and propagation mechanisms is given. More detailed reviews can be found e.g. in [10, 11].

While the bulk of cosmic rays is expected to originate from within our galaxy (the currently favored acceleration mechanism being discussed in section 2.2), it is commonly assumed that the magnetic fields permeating our galaxy are not strong enough to keep the highest energetic particles from leaving the Milky Way. Therefore, at the highest energies the energy spectrum of cosmic rays should be dominated by an extragalactic component. Models have been developed that try to deduce at which energy a transition from galactic to extragalactic cosmic rays is likely to occur taking into account the observed features of the energy spectrum of cosmic rays. Different types of models are presented in section 2.3.

2.1. The Energy Spectrum of Cosmic Rays

The energy spectrum of cosmic rays spans more than 13 decades in energy. It is shown in the range of $10^8\,\mathrm{eV}$ to $10^{21}\,\mathrm{eV}$ in figure 2.1. Except for the lowest energies, at which an influence of solar magnetic fields cannot be neglected, the spectrum approximately follows three power-laws, defined as $\frac{\mathrm{dN}}{\mathrm{dE}} \propto \mathrm{E}^{-\gamma}$. Up to about $4 \times 10^{15}\,\mathrm{eV}$ the index of the underlying power-law is given by $\gamma \approx 2.7$. The change of slope at this energy is called the "Knee" of cosmic rays, which marks a transition to a steeper power-law with index $\gamma \approx 3.1$. The currently favored explanation of this structure is based on the acceleration and propagation of cosmic rays involving magnetic fields, as explained in the next sections. In this frame, the "Knee" marks the energy at which the dominant contributors to the total cosmic ray population can no longer effectively accelerate protons or other light primaries to higher energies. Alternatively, the galactic magnetic fields can no longer keep particles with small charge numbers in our vicinity and a charge dependent leakage from our galaxy

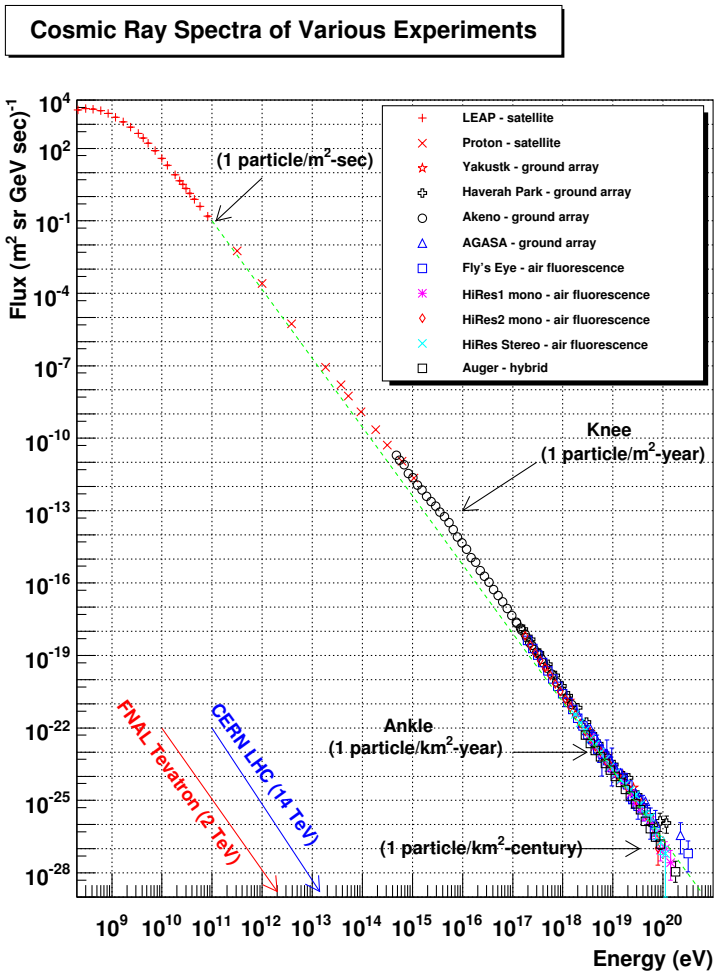

Figure 2.1.: Energy spectrum of primary cosmic rays. The green line corresponds to a single power law and illustrates how the measured spectrum deviates from a featureless spectrum. Picture taken from [16].

would reduce the flux towards higher energies. In fact, it has been shown that the "Knee" is caused by a decreasing flux of light primaries [12–14] and a "Heavy Knee" has been found at roughly a factor of 26 times higher energy [15], supporting the picture of charge dependent knee positions and, therefore, a connection to the acceleration and/or propagation mechanisms.

Another change of the power-law index at about 4×10^{18} eV to $\gamma \approx 2.7$ is generally associated with the final transition to a purely extragalactic population [5]. In reference [17], the onset of a new light component at a few 10^{17} eV has been found, which might mark the beginning of this transition. Some of the various models trying to explain the observed flux and composition are presented in section 2.3.

The reason for the cutoff of the spectrum at the highest observed energies is still under debate. The favored theories are the "GZK-cutoff" [18, 19] named after Greisen, Zatsepin, and Kuz'min and, alternatively or in addition that the accelerators have finally reached their maximum possible energy. While the latter is compatible with a mixed composition up to the highest energies, the former approach is based on the assumption that the extragalactic component consists mainly of protons. The cutoff is then caused by interactions of these protons with the photons of the cosmic

microwave background radiation [1] (CMB). Each of these reactions lowers the energy of the proton by about 20 %, therefore, the probability for a proton to reach the Earth with an energy above 6×10^{19} eV depends on its initial energy and the length of the path taken from its source to Earth.

The most recent measurements of the composition at the highest observed energies have been taken by the air shower experiments TA (Telescope Array, Utah, USA) and the Pierre Auger Observatory (Argentina). Both experiments cover a large area with detectors measuring samples of the particles produced in the air shower that reach the ground level. In addition, both setups are complemented by telescopes measuring the fluorescence light emitted isotropically during the development of the shower. Therefore, both experiments can combine the measurements of the longitudinal and lateral shower profiles.

While the results of TA are interpreted as compatible with a pure proton composition [20], Auger results seem to favor a transition to a heavier mixture of primaries [21].

2.2. The Origin of Cosmic Rays

Once the energy spectrum and the elemental composition are reconstructed, models regarding the acceleration and propagation of cosmic rays can be tested. A common assumption is that the bulk of cosmic rays are accelerated in supernova remnants. As shown in reference [22], a simple calculation can be performed to test the general viability of this assumption. It is stated that the energy density of cosmic rays is approximately $1\,\mathrm{eV/cm^3}$ and assuming the thickness of our galaxy to be 300 pc and its radius to be 15 kpc, the needed cosmic ray luminosity of our galaxy would be $L = \frac{V\rho}{\tau} \approx 5 \times 10^{40}\,\mathrm{erg/s}$. The time that the particles are kept in our galaxy is assumed to be $\tau = 6 \times 10^6$ years. The rate of supernova explosions in the galaxy is about three per 100 years corresponding to a total energy output of about $10^{42}\,\mathrm{erg/s}$, therefore, supernovae could provide enough energy. However, supernovae are not expected to be able to accelerate particles up to the highest energies. While the sources of cosmic rays are in principle unknown, pulsars and active galactic nuclei are among the possible candidates [22] that could be able to accelerate the particles to the highest observed energies. Usually, in addition to the aforementioned, gamma-ray bursts have been viewed as possible source candidates too. If a significant amount of protons are accelerated at gamma-ray bursts, some of them would interact with the gamma rays present in the burst via e.g. $\mathrm{p} + \gamma \rightarrow \Delta^+ \rightarrow \mathrm{n} + \pi^+$. The pions would then decay ($\pi^+ \rightarrow \mu^+ + \nu_\mu$ followed by $\mu^+ \rightarrow \mathrm{e}^+ + \nu_\mathrm{e} + \overline{\nu}_\mu$) and some of the neutrinos produced should be measured by e.g. the large neutrino detector IceCube. However, the results of the IceCube collaboration disfavor models using gamma-ray bursts as a significant source for the highest energies, because they have not measured enough neutrinos in coincidence with more than 500 observed bursts [23].

Several possible acceleration mechanisms are discussed. Among these are the acceleration of charged particles at the poles of pulsars or even the decay of heavy particles produced during the periods of high temperatures just after the Big Bang. Especially when considering supernovae as possible sources, the first order Fermi

[1] $\gamma_\mathrm{CMB} + \mathrm{p} \rightarrow \Delta^+ \rightarrow \mathrm{p} + \pi^0$, $\gamma_\mathrm{CMB} + \mathrm{p} \rightarrow \Delta^+ \rightarrow \mathrm{n} + \pi^+$

acceleration is discussed. As derived in reference [24], particles can be accelerated at the shockfront of e.g. supernovae. For each cycle (passing the shockfront twice), the particles gain a fraction of their initial energy as defined by $< \frac{\Delta E}{E} >= \frac{4}{3}\frac{V}{c}$, V being the velocity of the shockfront, which can be of the order of 10^4 km/s in case of young supernovae [24]. After n cycles, the energy of the particle has reached $E_\mathrm{n} = E_0(1 + \frac{\Delta E}{E})^\mathrm{n}$. After each cycle, the particle escapes the accelerator with a probability of P_e, the number of particles reaching an energy of at least E is, therefore, given by $N = N_0(\frac{E}{E_0})^{\log(1-P_\mathrm{e})/\log(\beta)}$ ($\beta = \frac{\Delta E}{E}$). Hence, the number of particles escaping with a certain energy is given by $N(E)\mathrm{d}E = const \times E^{-1+\log(1-P_\mathrm{e})/\log(\beta)}\mathrm{d}E$. For the acceleration at shockfronts, this results in $N(E)\mathrm{d}E \propto E^{-2}\mathrm{d}E$, which is not the observed power-law as was shown in the previous section, but the difference can be assigned to propagation processes discussed in the next section.

If the acceleration at supernova remnants is expected to be the dominant mechanism for the bulk of cosmic rays, a modification of the original energy spectrum is needed. This may take place during the propagation of cosmic rays. These modifications can be caused by general energy losses e.g. due to ionisation. Also, particles of a certain element might get lost in interactions with the interstellar medium (ISM) or due to decay, however, the products may contribute to the number of particles at a lower energy or to the numbers of a different element. In addition, a re-acceleration e.g. at magnetic gas-clouds is considered as a possibility. This mechanism is known as the "second order Fermi acceleration" [24, 25]. Depending on the direction in which the cloud itself is moving, the particle may gain or lose energy, however, assuming a statistically uniform distribution of velocity vectors, the movement of the particle itself will result in a positive, mean energy gain of $< \frac{\Delta E}{E} >= \frac{8}{3}(\frac{V}{c})^2$. This mechanism cannot be considered as the primary acceleration mechanism, though, because the speed of the clouds is far below the speed of supernova shockfronts and it is its squared value that enters into the equation. Therefore, the "injection problem" is of an even greater concern. This problem is based on the acceleration rate for low energetic particles being too low compared to the ionisation losses. Therefore, the particles would require a sufficiently high energy - already at injection. It is not clear where this "pre-acceleration" takes place yet.

Some of the major models regarding the sources and the propagation of cosmic rays are discussed in the following.

2.3. The Propagation of Cosmic Rays

2.3.1. Ankle, Dip or Mixed

When discussing possible astrophysical models to describe the all-particle energy spectrum in the light of source distributions, particle propagation and the origin of the ankle, three types of models are frequently considered. These are labeled "Ankle", "Dip", and "Mixed" type models. In the following, a brief characterization of these models is given, mainly based on [8], where the plots shown below have been taken from.

All three models focus on the simulation of the propagation of extragalactic cosmic rays. They are based on assumptions regarding the distribution of possible sources, the energy spectra and absolute abundances for the various primaries at

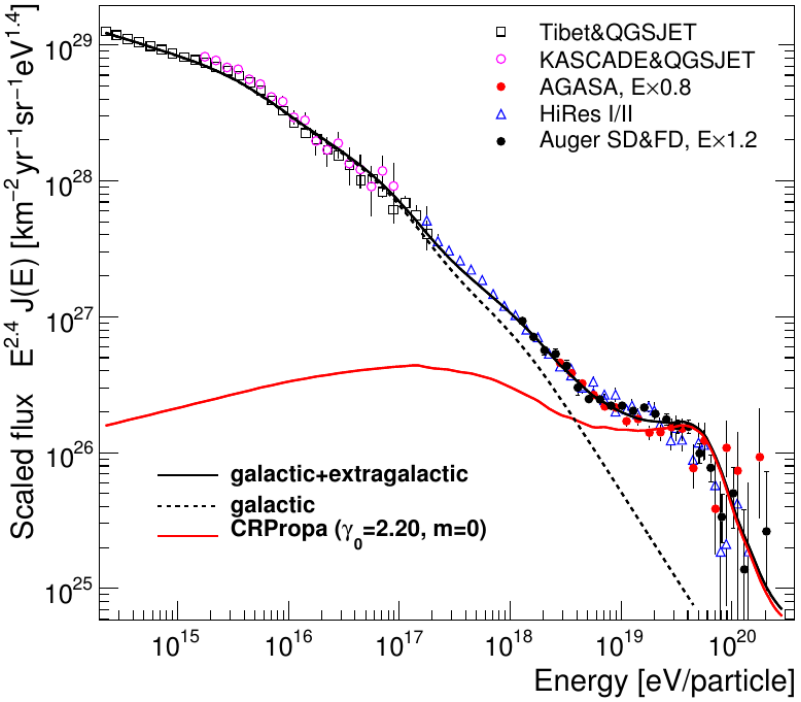

Figure 2.2.: All-particle energy spectrum of cosmic rays together with predictions of an "Ankle"-model obtained using the CRPropa package [26] (taken from [8]).

the sources, the structure of galactic and extragalactic magnetic fields, and possible interactions of the cosmic rays with the interstellar medium and the cosmic microwave background radiation. The simulations have been implemented using the CRPropa package [26] [2], which is a publicly available tool designed to simulate the propagation of ultra-high energy cosmic rays through the galactic and extragalactic environment.

The output of the models is a prediction of the energy spectrum of extragalactic cosmic rays. The contribution of galactic cosmic rays is assumed to be the difference to the all-particle energy spectrum.

An "Ankle" model (shown in Fig. 2.2) is called so because in this kind of models the ankle in the energy spectrum of cosmic rays is the result of the onset of an extragalactic proton component on the total galactic flux. The assumed extragalactic flux is relatively flat in its development with energy, therefore, in addition to the "standard" galactic component defined by charge dependent cutoffs (the proton cutoff energy being at the knee) another galactic contribution is needed to fill the "gap" between both components. Since the extragalactic component is assumed to consist (almost) purely of protons, the end of the spectrum may be modeled by the GZK-cutoff.

In figure 2.3, a "Dip" model is shown. These models try to explain the ankle with the interaction of protons with photons of the cosmic microwave background. In these interactions, the protons lose energy by the production of electron-positron pairs. The ankle is, therefore, not caused by an onset of an additional component,

[2] The exact version used is not stated in [8].

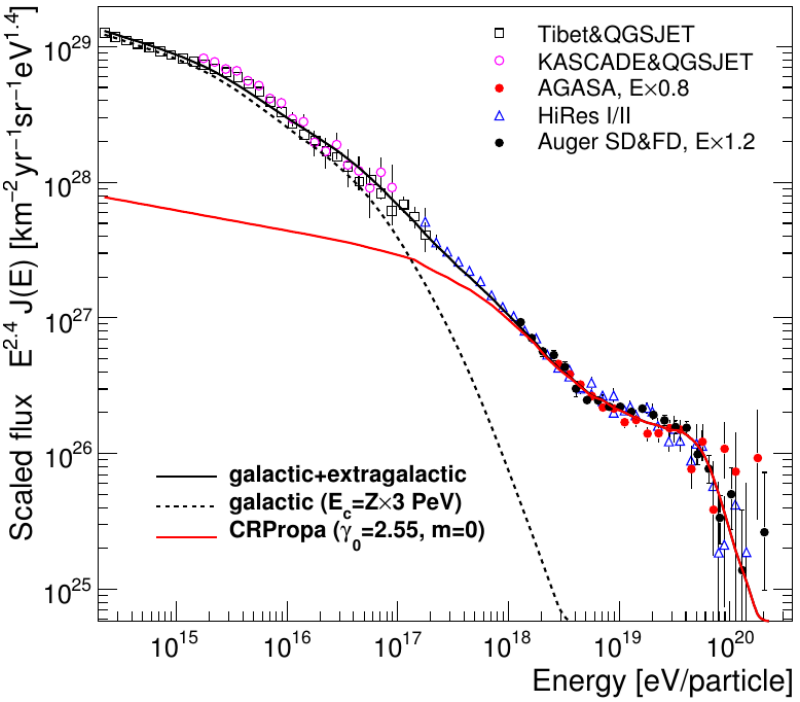

Figure 2.3.: All-particle energy spectrum of cosmic rays together with predictions of a "Dip"-model obtained using the CRPropa package [26] (taken from [8]).

but by a modification of the extragalactic flux. In addition, the transition occurs at much lower energies, just above 10^{17} eV, hence, the needed galactic flux can be explained using the "standard" component only. While for the "Ankle" model a pure proton composition is not strictly necessary, the "Dip" model could be invalidated by a composition measurement that results in a proton contribution of less than about 90 %. The reason is the origin of the dip, as these interactions are shifted towards higher energies for increasing number of nucleons, therefore, the shape of the dip would be different for a mixed composition.

A "Mixed" model is shown in figure 2.4. The assumptions for the exact composition vary among the models, however, often a composition similar to the one measured at lower energies is assumed. The transition takes place at a lower energy compared to the "Ankle" model, but still at an energy too high to explain the galactic component by the "standard" component only. The shape of the extragalactic component towards the end of the spectrum is governed by Lorentz-factor dependent photo-disintegration processes [27]. The suppression of the individual components starts at an energy of $E \geq A \times 10^{18}$ eV. The disintegration of light and medium-mass primaries should result in a lighter composition towards higher energies. The end of the spectrum may be shaped by the GZK-cutoff, which suppresses the proton flux above about 4×10^{19} eV. However, the minimum energy for the photo-disintegration of iron primaries is slightly above this energy. Therefore, if the mixed composition consists of a significant fraction of iron primaries, the end cannot be explained by the GZK-cutoff alone.

The flux at the highest energies is commonly assumed to be of extragalactic origin, however, some models try to explain the entire spectrum assuming galactic contributions, only [28].

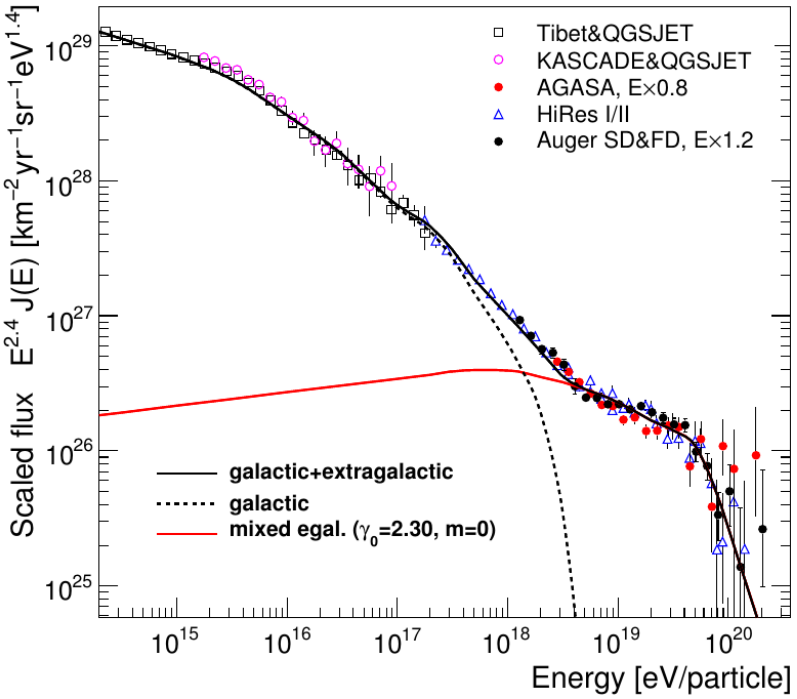

Figure 2.4.: All-particle energy spectrum of cosmic rays together with predictions of a "Mixed"-model. The red line (mixed egal.) corresponds to the estimate of the extragalactic contribution to the all-particle energy spectrum. (taken from [8]).

Measuring the energy spectrum and mass composition in great detail is necessary to test and restrict the above mentioned models and their various parameterizations. Additionally, the investigation of anisotropies in the arrival direction of cosmic rays can be used to restrict these models and maybe even to identify single sources.

The gyroradius of a charged particle is

$$r_{\mathrm{g}} = \frac{pc}{Z\mathrm{e}} \frac{\sin(\theta)}{Bc}$$

where p is the particle's momentum, Z the charge number of the particle, B the strength of the magnetic field, and θ the angle between the particle's trajectory and the magnetic field.

As discussed in e.g. [24], for $B \approx 3 \times 10^{-10}\,\mathrm{T}$, the gyroradius of a proton and iron primary with an energy of $10^{15}\,\mathrm{eV}$ is of the order of 0.36 pc and 0.014 pc, respectively. For energies of $10^{19}\,\mathrm{eV}$ the gyroradii of proton and iron primaries increase to 3.6 kpc and 140 pc, respectively. Hence, while at lower energies the scale of the irregularities of the magnetic field is large enough to effectively isotropize (by the scattering of the cosmic rays) the arrival directions of cosmic rays, the gyroradius increases significantly towards large energies and the degree of isotropy should decrease. Therefore, nearby sources or luminous sources of high-energy particles could be seen as small scale anisotropies in the arrival direction of cosmic rays and may be correlated with known source candidates. Such a search for correlations of ultra-high energy events with source candidates from catalogs and the galactic center or plane has been performed e.g. by the Pierre Auger Observatory [9]. No significant deviations from an

isotropic distribution were found in either case, strongly suggesting that the cosmic rays at the highest energies are probably of extragalactic origin.

The lack of anisotropies is quite interesting; its implications depend to some extent on the composition of the highest energetic cosmic rays. For a very light composition, one would expect the particles to point at least in the general direction of their source region. So either the composition is heavier, resulting in larger deflections due to the involved magnetic fields, or the assumption that there should be only a fairly low number of sources within the GZK-horizon is inaccurate.

In any case a detailed measurement of the composition also at the highest energies will help to shed light on that problem. The present work concentrates on the composition at energies below the Ankle, but this has direct impact on these models as a presumably light extragalactic component should become visible in the spectrum of light primaries already below the ankle, i.e. in the energy range covered by KASCADE-Grande. Also, the models include predictions of the extragalactic flux below 10^{18} eV, which can be compared with the measured results together with the corresponding predictions of the flux of galactic cosmic rays.

2.3.2. Peters Cycles

Throughout the analysis presented in this work, two composition assumptions will be used to test the reconstruction procedures. The models representing these assumptions are the H4a [29] and the GST3 [30] [3] model.

Both models are mainly based on two concepts. One is the idea of Hillas [31], which assumes up to three populations of cosmic rays, shown in figure 2.5.

The first population is assumed to stem from the acceleration at supernova remnants and its spectrum is assumed to lie within the hatched area in figure 2.5. Another component is a contribution of extragalactic origin. Denoted as "EGAL p" is an estimated flux of a pure proton contribution. The required flux from within the galaxy is labeled as "Total-E=egalP", which lies within the shaded area, therefore, this model would not need a third component. However, two other assumed extragalactic components require a galactic contribution outside of the shaded area. Therefore, a third component, often denoted as the galactic component B, is needed. This idea of different populations contributing to the total flux within different energy ranges, was adopted in [29, 30] to define the two models.

Another assumption used in the construction of the H4a and GST3 models is based on the idea of Peters [32]. Since the acceleration and propagation mechanisms involve the collision-less diffusion in magnetic fields, the cutoff energies of different elements within one population are expected to be shifted towards higher energies according to the charge of the nucleus. This is the idea of the charge dependent knee-positions mentioned in section 2.1.

Summarizing, the H4a and GST3 models consist of several populations of cosmic rays. A population is defined by contributions from different primaries. The flux for each primary is expressed as a single power-law with an exponential cutoff. The energies of the latter are based on a common cutoff energy, which is multiplied by the charge of the corresponding primary.

[3] This is not an official name, however, the authors are Gaisser, Stanev, and Tilav and the model assumes three different populations of cosmic rays, hence it is referred to as GST3.

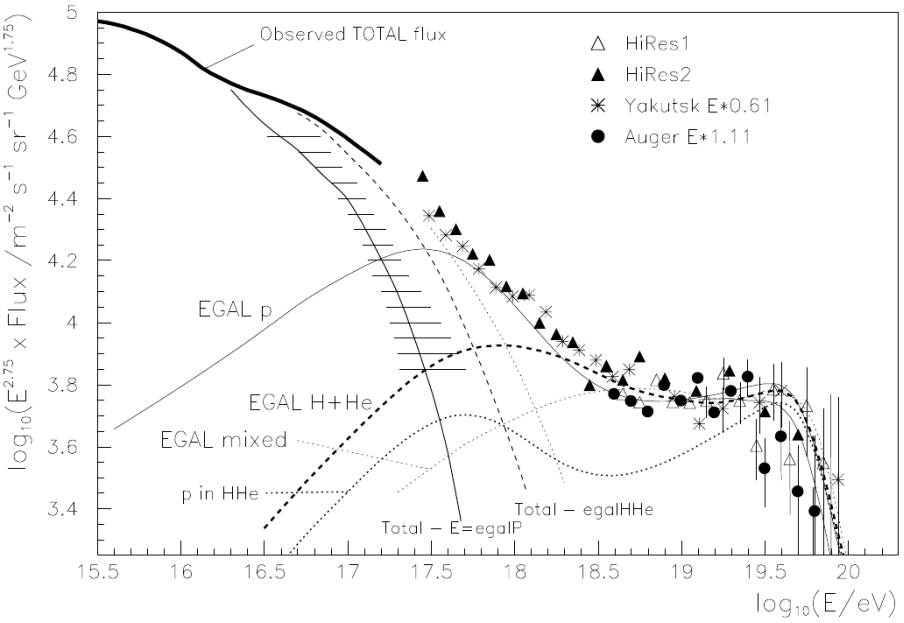

Figure 2.5.: The energy spectrum of cosmic rays and theoretical prediction for the galactic and extragalactic components. The spectra from the Pierre Auger Observatory and Yakutsk have been normalized to the HiRes spectrum (taken from [31]).

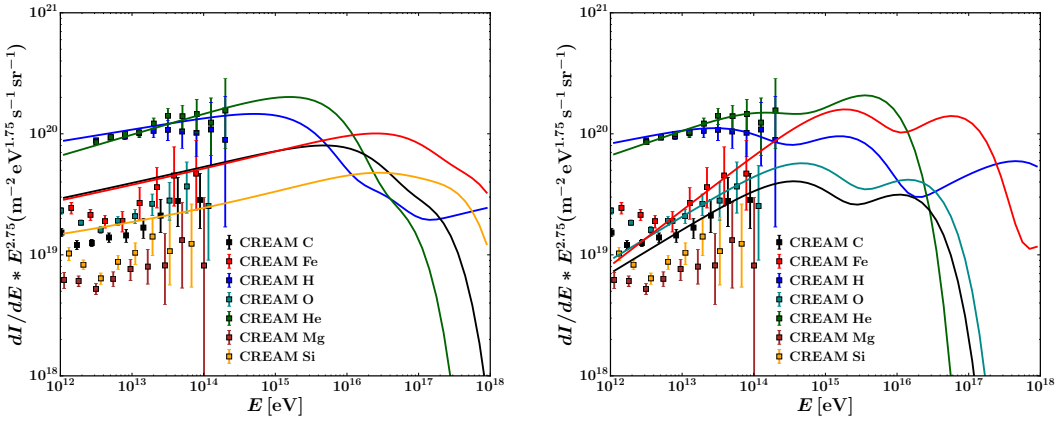

Figure 2.6.: CREAM data (markers) together with the predictions of the H4a (l.h.s.) and the GST3 (r.h.s.) models, represented as lines.

The model predictions are shown in figure 2.6 together with the results of the Cosmic Ray Energetics and Mass (CREAM) experiment (markers), which was a balloon-borne experiment directly measuring the flux and mass of the primary particles. The results of CREAM are believed to be based on the most accurate direct measurements above TeV energies.

In case of the H4a model (l.h.s.), the lines corresponding to the predicted flux of C and Si primaries are actually the combined fluxes of C+O and Mg+Si, respectively. For the sake of section 5.3.3, it is worth to note already at this point that the predictions at lower energies seem to ignore the change of slopes visible in the spectra. In fact, the indices of the power-laws of the first population are quite close to those given by CREAM in [33], which have been obtained by fits to the total energy range. The model consists of three populations. The first one has a cutoff at 4×10^{15} eV. The second population is considered to correspond to Hillas' galactic component B. Having a cutoff at 30×10^{15} eV, it bridges the gap to an extragalactic proton component with its cutoff at 60×10^{18} eV.

The predictions of the GST3 model (r.h.s.) closely follow the CREAM fluxes above the kinks, however, the model does not contain contributions from Mg and Si. Since the air-showers induced by Si and Fe primaries are not very different from each other, the GST3 model can be used to test how well the reconstruction procedure can distinguish between the two primaries. The original GST3 model includes contributions of nuclei heavier than iron. These are not included here, therefore, the presented GST3 model is actually a subset of the original GST3 model as it is published in [30]. Compared to the H4a model, an additional population has been added with a cutoff at 120×10^{12} eV. It has been introduced to take into account that the CREAM data exhibits ankle-like features in the spectra of the various masses. Using the corresponding slope above the kinks for a population with a cutoff at 4×10^{15} eV would result in a too large contribution of heavy primaries at larger energies. There is no additional population between the galactic component with a cutoff at 4×10^{15} eV and the extragalactic component with a cutoff at 1.3×10^{18} eV, which has also contributions of iron primaries below 10^{18} eV as can be seen in the figure.

Being fit to the data of several experiments (including their estimate of the mean logarithmic mass (lnA) [4] – or equivalents expressed in terms of lnA – where given) these models represent the current understanding of the energy spectrum and mass composition of cosmic rays. Therefore, they are used to test the reconstruction procedures of the combined analysis to assert the capability of the combined analysis of KASCADE and KASCADE-Grande to successfully reconstruct not only the all-particle spectrum, but also a realistic mass composition [5].

[4] lnA is the mean of the logarithm of the mass of the primary particles. It is 0 for protons (log(1)) and about 4 for iron (log(56)).

[5] It is worth noting that in [30] a modified version of the GST3 model is available, which adds a fourth component consisting of protons. Doing so results in a better agreement of the predicted mean lnA beyond 10^{18} eV.

3. Extensive Air-Showers

Extensive air showers (EAS) are cascades of secondary particles induced by high-energy particles hitting the Earth's atmosphere. The products of the first interaction will interact with the atmosphere themselves producing another generation of particles. As long as the produced particles have enough energy to effectively produce another generation, the cascade will go on. At some point this is no longer the case and the shower will start to die out.

Having at least a general idea of how air showers develop in the atmosphere and how they depend on the energy and mass of the primary particle is crucial for understanding the basic strategy of the presented analysis. This chapter has been written for this purpose and is based on simplified models for electromagnetic and hadronic cascades as derived in [34]. Additional information can be found therein and in e.g. references [35] and [36].

3.1. An Air-Shower Toy Model

The schematic view of a photon induced air shower is shown on the l.h.s. of figure 3.1. The incoming photon produces an electron positron pair via pair-production. After one radiation length χ_0 (approx. 36.7g/cm^2 in air), each of them produces a photon via bremsstrahlung [1].

It is assumed that each particle of a generation has the same energy, i.e. after each interaction length, the total number of particles is doubled and the energy of each particle is half of the energy of its parent particle. The maximum number of particles is, therefore, given by $N_{\max} = E_0/E_c$ with E_c being the critical energy. For electrons, this is roughly 84 MeV. Below this energy the energy loss due to ionization is dominant. For photons, the critical energy is about 24 MeV. Photons with an energy below this value lose their energy primarily by Compton-scattering.

The depth of the shower maximum, i.e. the depth at which the number of particles is at its maximum, can be inferred using the mean traversed atmosphere between two

[1] The mean free path for the pair-production is approximately $\frac{9}{7}\chi_0$ (see e.g. Ref. [35])

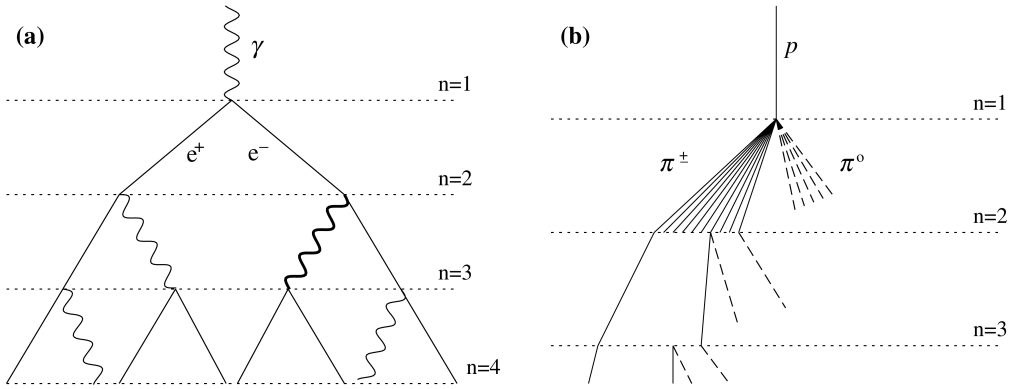

Figure 3.1.: A simplified model of an electromagnetic (l.h.s.) and a hadronic (r.h.s.) shower. Picture taken from [34].

interactions, which is $\lambda = \chi_0 \ln(2)$, i.e. the length after which the electron has lost half of its energy, and the number of particles after n interactions is $N_n = 2^n = e^{X/\chi_0}$, with $X = n\lambda$. From $N_{\max} = E_0/E_c = e^{X_{\max}/\chi_0}$ follows $X_{\max} = \log(E_0/E_c)\chi_0$.

Although this model is simplified, it explains the linear rise of $N_{\max}$ and the logarithmic rise of $X_{\max}$ with the primary energy.

A similar model can be derived for hadron induced showers. A schematic view of a proton induced shower is given on the r.h.s. of figure 3.1. It is a simple model assuming that the only particles being directly produced in the interactions are pions with 2/3 being charged pions and 1/3 being neutral. In each interaction n_{mult} pions are produced. After n interactions, $N^{\pi}_{\text{n}} = (n_{\text{ch}})^{\text{n}}$ charged pions have been produced. Each pion has an energy of $E^{\pi}_{\text{n}} = \frac{E_0}{(\frac{3}{2}n_{\text{ch}})^{\text{n}}}$ [2]. Once this energy has reached a critical value, no new pion generations are produced any more and the existing ones decay into muons and neutrinos. The number of muons is, therefore, $(n_{\text{ch}})^{\text{n}_c}$ with $\text{n}_c = \frac{\log(\frac{E_0}{E_c})}{\log(\frac{3}{2}n_{\text{ch}})}$ being the total number of pion generations. This can be written as $N_\mu = (\frac{E_0}{E_c})^{\beta}$ with $\beta = \frac{\log(n_{\text{ch}})}{\log(\frac{3}{2}n_{\text{ch}})}$. Depending on the assumptions for n_{ch}, the value of β is between 0.85 and 0.92.

Within this model the energy remaining in the hadronic component after n interactions is $E_{\text{hadr}} = \frac{2}{3}E_0$. Since the neutral pions decay almost immediately into two photons, producing an electromagnetic shower, the energy transferred to the electromagnetic component is 56 % of the primary energy after the second interaction and already 90 % after the first six interactions. Hence, it is assumed that the shower maximum is reached with the maximum of the electromagnetic component. Considering only the first interaction, which takes place in a depth corresponding to the mean interaction length λ_{I}, $X_{\max}$ can be estimated by $X_{\max} = \lambda_{\text{I}} + \log(\frac{E_0/n_{\text{mult}}}{E_c})\chi_0$ with E_0/n_{mult} being the energy of one neutral pion and assuming that the $\frac{1}{3}n_{\text{mult}}$ sub-showers do not influence each other.

The proton case can be generalized for heavier nuclei in the following way. It is assumed that the energy of the nucleus is evenly shared among the A nucleons. The shower maximum is given by:

[2] $n_{\text{mult}} = n_{\text{ch}} + n_{\text{neutral}} = \frac{3}{2}n_{\text{ch}}$

$$X^{\mathrm{A}}_{\mathrm{max}} \approx X_{\mathrm{max}}(E_0/A).$$

In contrast to the shower maximum, the number of muons of each sub-shower are added to the total number of produced muons:

$$N^{\mathrm{A}}_{\mu} \approx A \cdot (\tfrac{E_0/A}{E_{\mathrm{c}}})^{\beta} = A^{1-\beta} \cdot N_{\mu}.$$

Interestingly, the total number of particles at the shower maximum is independent of the mass of the primary particle:

$$N^{\mathrm{A}}_{\mathrm{max}} \approx A \cdot \tfrac{E_0/A}{E_{\mathrm{c}}} = \tfrac{E_0}{E_{\mathrm{c}}} = N_{\mathrm{max}}.$$

Using this simplified view of a hadronic shower, it is possible to explain the basic concept of the energy and mass reconstruction of the primary particle. Measuring the longitudinal shower profile, the number of particles at the shower maximum is accessible, giving the primary energy of the particle. Comparing the depth of the maximum with detailed Monte-Carlo simulations, an estimate of the mass can be constructed.

For KASCADE, the lateral distributions of electrons and muons at ground level have been measured. With the number of electrons alone, neither the energy nor the mass can be reconstructed without some assumption on the composition. This is due to the fact that KASCADE, being located at 100m above sea level, does not measure the total number of particles produced, but the number of particles that reach the ground. Therefore, although the number of electrons at the shower maximum is independent of the mass and directly linked to the primary energy, the mass dependent depth of the shower maximum results in showers induced by heavy particles being more attenuated compared to proton induced showers, because the shower maximum of the former is reached at higher altitudes. A second observable is needed to distinguish between low-energy protons and high-energy heavy primaries. In case of KASCADE this is the number of muons which exhibits a mass dependence in itself and has a different attenuation length.

In practice, this over-simplified model is too simple for a meaningful interpretation of the properties of measured air showers. The following section will give a brief introduction to the hadronic interaction models used in the present analysis.

3.2. Monte-Carlo Simulations

Having measured the total number of muons and electrons reaching the detector plane, a comparison to detailed simulations of the air showers is needed to infer the properties of the primary particle, such as its mass and energy. For KASCADE and KASCADE-Grande, the CORSIKA (COsmic Ray SImulations for KAscade) [37] air-shower simulation tool has been used. The modeling of low-energy interactions (up to 200 GeV lab energy) is based on the FLUKA model [38]. Hadronic interactions at higher energies are described by different models. Used in this analysis are QGSJetII2 [39], QGSJetII4 [40, 41], and EposLHC [42], where the latter two are new versions of the models and have already been tuned to the 7 TeV LHC data.

For the simulation of extensive air showers two main components are needed. One is needed for the bookkeeping (where are the particles, what are their momenta, what are their types etc.), defining the target (i.e. what is the composition and density of the atmosphere at a given height), determining how the particle's trajectories are influenced by e.g. Earth's magnetic field etc. The second component is the hadronic interaction model, which is given the information on the projectile and the target. It has to take into account numerous aspects of hadronic interactions such as the probability for the two particles to interact with each other, how many particles are produced, what their types and momenta are, what possible fragments of the primary particle are and how the energy is distributed among these, and many others. The various parameters of these models have to be tuned by comparing their predictions with the measurements of experiments. The latter could be detectors at particle accelerators, however, also air-shower measurements are used. The former have the important advantage that the properties of the projectiles can be controlled, while for the latter the maximum energy is higher, although the unknown energy and mass of the primary particle is a significant hindrance.

Having simulated the entire shower, the trajectories and energies of the produced particles are known and are fed into a Geant3 [43] based detector simulation named CRES (Cosmic Ray Event Simulation). The output format is identical to the format in which the measured data is stored. Therefore, the shower reconstruction is the same for measured and simulated data. The reconstruction is done with KRETA (KASCADE Reconstruction of ExTensive Airshowers). The procedure will be briefly explained in chapter 4.

Several differences within the high-energy hadronic interaction models lead to different numbers of muons and electrons and their energy and lateral distributions. This changes the interpretation of the measurements regarding the energy spectra of cosmic rays as well as the mass composition. In the following, the most important aspects are going to be briefly discussed.

Within the simplified picture described in the previous section, two dependencies of N_e and N_μ at ground level emerge. One is the interaction cross section, the other is the multiplicity. Both determine how fast the shower develops and, hence, the depth of the shower-maximum, which is directly related to the number of particles that reach the observation level. This is oversimplified, of course. As pointed out in e.g. [34] and [44], considering the multiplicity alone is not sufficient. What is needed in addition is the energy transferred to the secondaries, i.e. the inelasticity, which is the energy not carried away by a leading particle $(1 - E_{\text{leading}}/E_0)$. The lower the inelasticity, the deeper the shower penetrates into the atmosphere and the larger the number of measured particles gets.

The simple model assumed a production of pions only, and a charged to neutral pion ratio of 2 to 1. However, since the neutral pions feed the electromagnetic component, their fraction of all produced particles is of great importance. For example, the number of muons is strongly connected to the amount of energy transferred to the electromagnetic component. As discussed in [45], an enhanced baryon/antibaryon production can result in a larger muon content, because fewer neutral pions are produced. In addition, the production of e.g. ρ^0, which decays predominantly into two charged pions, could enhance the number of muons. It is worth noting that

recent NA61/SHINE [3] measurements [46] of the $\pi^- + carbon$ interaction suggest that the current models underestimate the fraction of ρ^0, which could result in a too small muon content predicted by the simulations. Interesting to note is that the longitudinal momentum fraction distributions for neutral pions and rho-mesons in the interactions $\pi^+ + p \rightarrow \pi^0$ and $\pi^+ + p \rightarrow \rho^0$ seem to be sufficiently well estimated [47].

As mentioned earlier, air-shower measurements can be used to test hadronic interaction models. In fact, doing so has always been one of the main scientific goals for the KASCADE and KASCADE-Grande experiments. Especially KASCADE was even built for the task, as it contained detectors for simultaneously measuring the hadronic, electromagnetic, and muonic component. For the latter, the measurement at different threshold energies was possible and, using the muon tracking detector, studies of the muon production height were performed [48].

In [49, 50] the predictions of various models regarding the number of hadrons, their maximum energy, and the sum of the hadron energies have been compared with KASCADE measurements. Also the hadron lateral distributions [51] and observables related to it [52] have been studied. To complement this, the attenuation length of hadrons in extensive air showers was studied as well [53].

Naturally, also the lateral distribution of muons at observation level can be compared to simulations [54]. However, using the local muon density distributions at various ranges of distances of the core to the detector and at different muon energy thresholds [55] it is possible to compare not only the lateral distributions, but also the energy distribution of muons with the model predictions.

Similar to the attenuation length of hadrons, the attenuation length of muons can be used to test hadronic interaction models. This was recently performed using KASCADE-Grande measurements [56] revealing a discrepancy for all current hadronic interaction models, when compared to measurements.

All these tests have been performed several years ago. Unfortunately, there is presently not enough personpower within the KASCADE collaboration to repeat these analyses testing the up-to-date models. However, the number of hadrons and the sum of their energies are included in the dataset published by KCDC, which is presented in Part 2 of this thesis. Therefore, the information is not lost and the analyses can be repeated by the community also for future models.

This is only a short summary of the different kind of tests performed using the data of the KASCADE and KASCADE-Grande experiments. A more complete list can be found at `https://web.ikp.kit.edu/KASCADE/`.

[3] NA61/SHINE is a fixed target experiment at the CERN Super Proton Synchrotron (SPS), which operates at different energies in the GeV range.

4. KASCADE and KASCADE-Grande

The original analyses of KASCADE were based on the KASCADE detectors only. Although KASCADE-Grande utilized the muon detectors of the former KASCADE array, the number of charged particles was reconstructed using the KASCADE-Grande detectors only. Since the arrival direction and the position of the shower core have been obtained without using the muon detectors, both standalone analyses obtained completely independent results. The combined analysis (this work), however, is based on the simultaneous use of the KASCADE and KASCADE-Grande detectors, i.e. it is the analysis that combines both detector setups. The reconstruction of the number of electrons and muons follows the procedure developed for the standalone analyses of KASCADE. Therefore, this section is dedicated to an overview of the detector setups, reconstruction procedures, and main results of the standalone analyses. The comparison of the results at the end of this chapter will provide additional motivation to use the combined reconstruction, apart from other reasons such as an improved accuracy of the reconstruction of N_e and N_μ.

4.1. KASCADE

4.1.1. Experimental Setup

KASCADE was an experiment designed to measure several components of extensive air-showers. It was equipped with a central detector used to study mainly the hadronic and muonic component near the shower core and a muon tracking detector. The relevant component for the studies presented here was the detector array shown schematically on the left hand side of figure 4.1.

The detector array had been equipped with 252 detector stations of which the outer 192 stations had $3.24\,\mathrm{m}^2$ of shielded and $1.57\,\mathrm{m}^2$ of non-shielded detectors. The non-shielded detectors alone could not provide a separation between electrons and muons, however, the mass sensitivity of the experiment was based on the simultaneous measurement of both kinds of particles. Therefore, an lead-iron (10 cm and 4 cm thick, respectively) layer was installed below the non-shielded detector component.

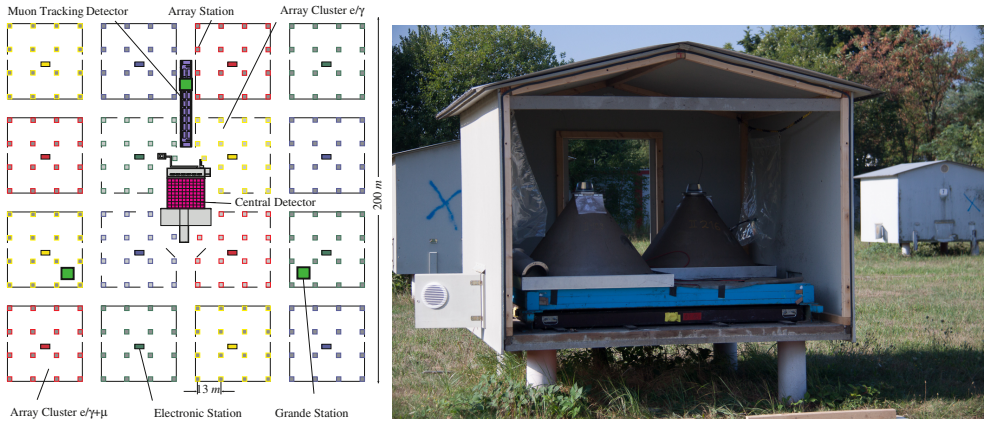

Figure 4.1.: Left: Layout of the KASCADE experiment [57]. Right: Picture of one of the 192 stations equipped with both shielded and non-shielded scintillation detectors.

It was built to shield against the electromagnetic component of the air shower allowing mainly muons with an energy above 250 MeV to pass the shielding. A second detector below this shielding was installed to register the remaining muons. The remaining 60 stations near the center of the array had twice the number of non-shielded detectors (i.e. 3.14 m^2 each) but did not include shielded ones. The reason to omit the shielded detectors for these stations was the increasing probability of high-energy photons and electrons to pass the shielding for stations close to the shower core.

With this setup it was possible to reconstruct the number of muons as well as the number of electrons.

4.1.2. Shower Reconstruction

The reconstruction of the shower is organized in three steps. Step one is used to get a first estimate of the particle numbers, arrival direction, and the position of the core. The latter is achieved by using the center of gravity of the signals of the non-shielded scintillation detectors. Assuming a plane shower front, the arrival direction is estimated, while the geometrically weighted signals of the non-shielded and shielded detectors are used to retrieve the starting values for N_{ch} and N_{μ}, respectively.

For steps two and three, the energy deposits have to be transformed into particle numbers. This is done using a lateral energy correction function [58], which describes the average energy deposited per charged particle taking into account the size of the shower obtained in the respectively previous step and the distance of the station to the shower core. The lateral energy correction function used for the electromagnetic component is shown in figure 4.2 for proton-initiated showers of three different sizes ($\log_{10}(N_{\mathrm{e}}) = 5$ (dashed), 6 (solid), and 7 (dotted dashed)). Its value is based on the energy deposited per electron ($E^{\mathrm{e}}_{\mathrm{dep}}$) and photon ($E^{\gamma}_{\mathrm{dep}}$) as well as the ratio of photons to electrons.

In step two, the arrival direction is improved by fitting a conically shaped shower front to the arrival times of the first particles measured with the non-shielded detectors. The core position, N_{ch} and the so called shower-age, i.e. the slope parameter

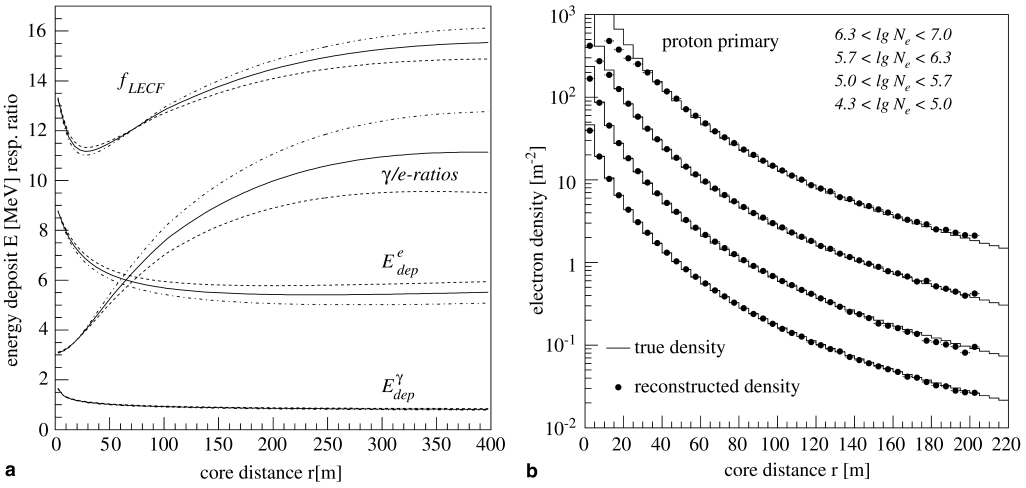

Figure 4.2.: Left: lateral energy correction function used to convert energy deposits to numbers of particles. This is shown for $\log_{10}(N_\mathrm{e}) = 5$ (dashed lines), $\log_{10}(N_\mathrm{e}) = 6$ (full lines), and $\log_{10}(N_\mathrm{e}) = 7$ (dashed dotted lines) [58]. Right: Resulting mean lateral distribution of electron densities for simulated proton showers [58].

of the function are then obtained by fitting an NKG-like [1] lateral density function (LDF) [59, 60] (Eq. (4.1)) to the particle densities measured at the non-shielded detectors. Such a lateral density function is shown on the right hand side of figure 4.2. The original NKG function has been derived analytically for e/γ induced showers describing the lateral distribution of the electromagnetic particles. However, the original function cannot describe the electromagnetic component of hadron induced showers well. This is commonly attributed to the fact that the electromagnetic component of a hadron induced shower is a superposition of a large number of electromagnetic sub-showers. Therefore, a different parameterization of the lateral density function has to be used.

$$\rho = N_\mathrm{e} \cdot c(s) \cdot \left(\frac{r}{r_0}\right)^{s-a} \cdot \left(1 + \frac{r}{r_0}\right)^{s-b} \tag{4.1}$$

with

$$c(s) = \frac{\Gamma(b-s)}{2\pi r_0^2 \Gamma(s-a+2)\Gamma(a+b-2s-2)}. \tag{4.2}$$

In case of KASCADE, the parametrization used is $a = 1.5$, $b = 3.6$, and $r_0 = 40\,\mathrm{m}$. Γ is the Gamma function.

The reconstruction of N_μ is performed simultaneously, therefore, the muon lateral density function is known at the beginning of step three and its information is included in the fitting procedure for the non-shielded detectors resulting in the total number of electrons at observation level.

The number of muons is obtained in a similar way. The transformation of energy deposits takes into account an N_e dependent probability of electrons, photons or

[1] The NKG function is named after Nishimura, Kamata, and Greisen.

hadrons passing the shielding. These particles falsely counted as muons dominate the signal for stations with a distance within 40 m of the core, therefore, these stations are ignored for the fit. In addition, the NKG-like function (Eq. (4.1) with $a = 1.5$, $b = 3.7$, and $r_0 = 420\,\mathrm{m}$) is known to deviate from the true muon distribution towards larger distances to the shower core, hence, instead of using the total number of muons, the truncated number of muons is used (N_μ^{tr}), which is the result of the integration of the fitted lateral density function in the range from $40 - 200\,\mathrm{m}$, i.e. where the KASCADE detectors provide sampling points for the fit. In addition, because of the low muon densities, the shape parameter, i.e. the slope of the lateral density function is fixed to a value derived using simulations. This value changes depending on the shower size. The only remaining free parameter is, therefore, the total number of muons.

In [58] a more detailed description of the reconstruction procedure can be found.

4.1.3. Analysis

The reconstruction of the energy spectrum is based on the two-dimensional shower-size spectrum ($\log_{10}(N_\mathrm{e})$ vs $\log_{10}(N_\mu)$) shown in figure 4.3. Each incident particle populates a certain region of the two-dimensional distribution depending on its mass and energy. The extent of these regions depends on the intrinsic shower-to-shower fluctuations, which themselves depend on the energy and mass of the primary particle. In addition, these regions are modified and smeared out by inaccuracies in the reconstruction of N_e and N_μ and the trigger and reconstruction probabilities for showers of the corresponding size.

The problem of extracting the energy spectrum and mass composition from the $\log_{10}(N_\mathrm{e})$ vs $\log_{10}(N_\mu)$ distribution can be expressed as

$$\vec{Y} = \mathbf{R}\vec{X} \tag{4.3}$$

with $\vec{Y}$ containing the cell contents as shown in figure 4.3, $\vec{X}$ being the corresponding mass and energy of the primary particle and $\mathbf{R}$ being the *response matrix* of the detector. The unfolding technique used to solve this equation is described in [12], the resulting spectra for protons, helium, and carbon using QGSJet01 [61] as the hadronic interaction model are shown in figure 4.3.

An important conclusion was that the knee-like structure around 6 PeV in the all-particle energy spectrum might be caused primarily by a decreasing flux of helium primaries instead of protons as it was commonly believed. However, the choice of the hadronic interaction model has a certain influence on the resulting composition. For example, a model that predicts significantly more muons will result in a lighter overall composition, which would enhance the proton component.

Assuming that the "proton-knee" is indeed located at an energy of about 2-3 PeV, the "helium-knee" would be expected to be at an energy of 4-6 PeV or 8-12 PeV depending on whether it depends on the charge or on the mass of the primary particle. Unfortunately, the spectra shown in figure 4.3 do not allow to distinguish between these two scenarios, therefore, there was a need to look at the position of the knee for heavier elements, e.g. iron, which is expected to be just below 100 PeV

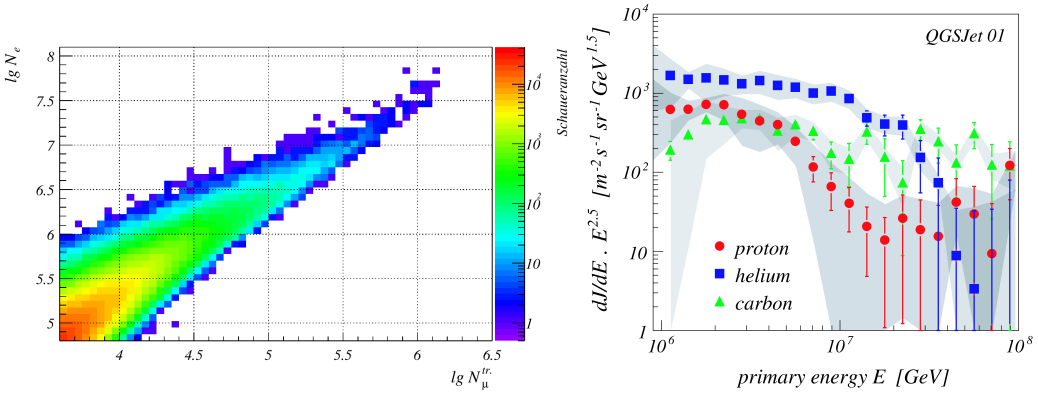

Figure 4.3.: Left: The $\log_{10}(N_\mathrm{e})$ vs $\log_{10}(N_\mu)$ distribution as measured by KASCADE [12]. Right: Resulting energy spectra for protons, helium and carbon primaries [12].

(in case of a charge dependence). This is the upper limit of the energy range for KASCADE, however, its extension KASCADE-Grande was designed to be able to reconstruct events with energies up to 1 EeV .

4.2. KASCADE-Grande

4.2.1. Experimental Setup

KASCADE-Grande was designed as an extension of the original KASCADE-array. The main goal was to extend the energy range accessible with KASCADE in order to cover the estimated energy range of a possible knee-like structure in the energy spectrum of iron primaries. This structure was expected to be visible at an energy of about 10^{17} eV in case of charge dependent knee-positions (see chapter 2). To achieve this goal and to increase the upper energy limit to 10^{18} eV, 37 additional stations containing 10 m^2 of non-shielded scintillation detectors, each were added to the original KASCADE array. One of these stations and their locations relative to the KASCADE array are shown in figure 4.4. Since the 37 stations did not contain shielded detectors, the reconstruction of N_μ was only achieved by using the shielded detectors of KASCADE.

The reconstruction procedure used is very similar to the one used for KASCADE. It will be briefly explained in the next section. A detailed discussion can be found in [63].

4.2.2. Shower Reconstruction

The reconstruction procedure is again split into several separate steps. The first step results in estimates for the core position, arrival direction, N_{ch}, and N_μ. It is identical to the first step of the KASCADE reconstruction. The lateral density function used for estimating N_{ch} is also a modified NKG-function, however, N_μ is estimated using a function proposed by Lagutin and Raikin [64].

The transformation of energy deposits into particle densities is again achieved by the application of an lateral energy correction function shown in figure 4.5. Its shape is mainly defined by the distance-dependent e/γ ratio and the energy distributions of electrons and photons. Above around 400 m only muons contribute to the energy deposit and above 450 m the lateral energy correction function describes the energy deposited by a single vertically incident muon. Different from the lateral energy correction function of KASCADE, a dependence on the shower size has been found to be negligible.

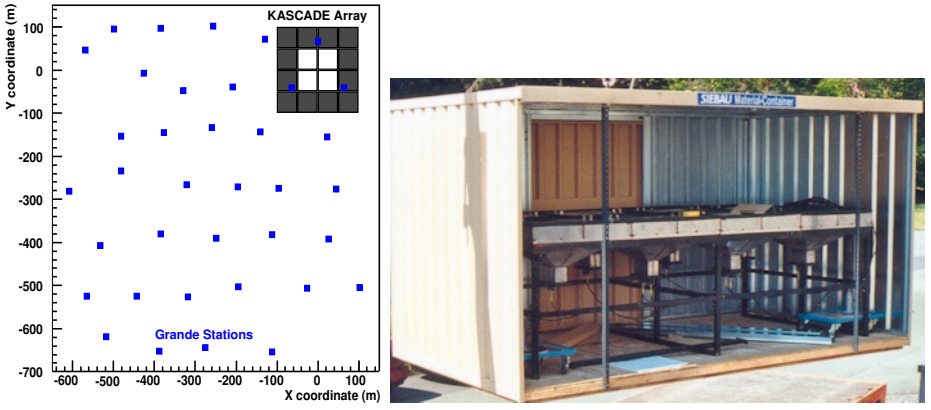

Figure 4.4.: Left: Layout of the KASCADE-Grande array. Right: Picture of a KASCADE-Grande station [62].

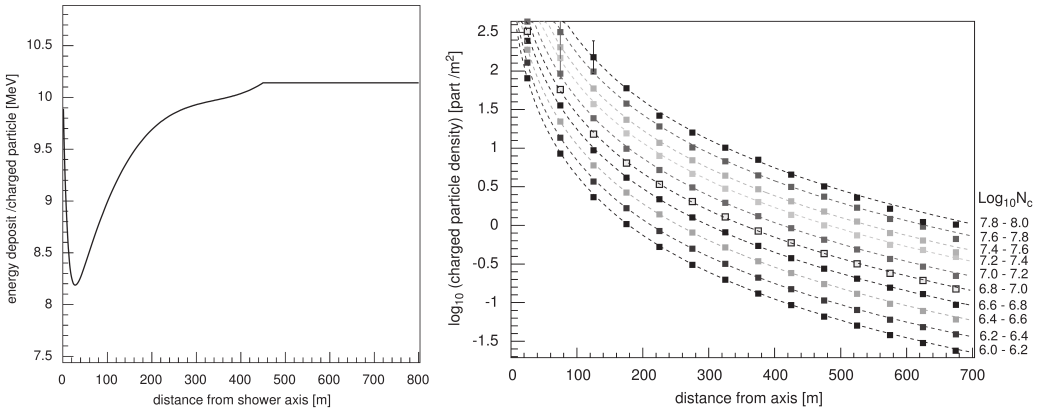

Figure 4.5.: Left: The lateral energy correction function used to convert energy deposits to particle numbers [63]. Right: Resulting mean lateral distribution of particle densities [63].

As a second step, an improved estimate of the core position is obtained. This is achieved by shifting the core position found in step one on a 7×7 grid with an 8 m spacing. Each of these positions is used as a fixed parameter for the fit of the lateral density function, the shape and N_{ch} being the only free parameters left. The position that results in the smallest χ^2 is used in the next two steps, which are used to obtain a more precise arrival direction by fitting the measured arrival times using a theoretical shower front derived from simulations. After that the shape and N_{ch} are fitted again using the current estimate of the core position, which is improved for the last time by fitting the lateral density function again, keeping the shape and N_{ch} fixed. This final position is now used, while the arrival direction, the shape of the lateral density function and N_{ch} are estimated for the last time.

Using the core-position and arrival direction obtained during the reconstruction of N_{ch}, N_{μ} is reconstructed by fitting the muon densities measured by KASCADE with N_{μ} being the only free parameter.

With the information now available on the number of charged particles and the number of muons, the energy spectrum and mass composition can be estimated.

4.2.3. Analysis

The reconstruction of the energy is again based on the two-dimensional shower size spectrum shown in figure 4.6. Because of larger uncertainties in N_{ch} and N_{μ} and low statistics at the highest energies, the unfolding analysis used for the KASCADE analysis cannot be used for KASCADE-Grande up to 10^{18} eV. Therefore, in order to really reach 1 EeV, a different approach was used.

The reconstruction is based primarily on N_{ch}, however, the ratio of N_{ch} to N_{μ} is used - as a function of N_{ch} - to take into account that the energy corresponding to a certain measured N_{ch} depends on the mass of the primary particle. As explained e.g. in [15], the following formulas are used:

$$\log_{10}(E) = [a_{\mathrm{H}} + (a_{\mathrm{Fe}} - a_{\mathrm{H}}) \cdot k] \cdot \log_{10}(N_{\mathrm{ch}}) + b_{\mathrm{H}} + (b_{\mathrm{Fe}} - b_{\mathrm{H}}) \cdot k \qquad (4.4)$$

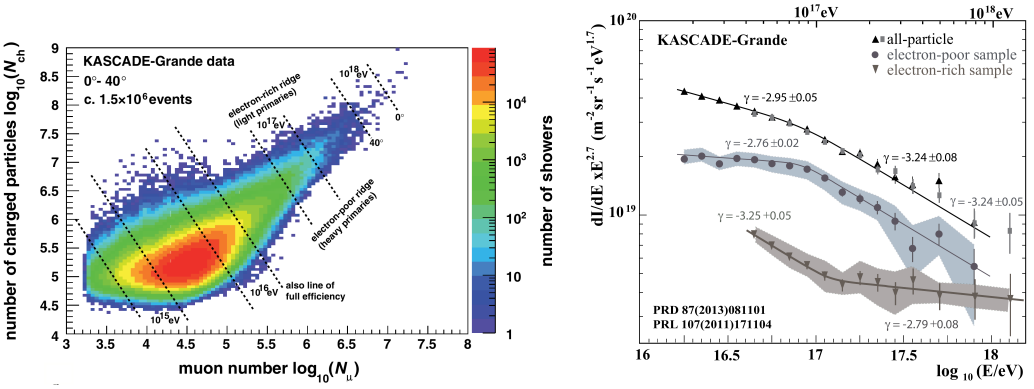

Figure 4.6.: Left: The $\log_{10}(N_{\mathrm{ch}})$ vs $\log_{10}(N_{\mu})$ distribution as measured by KASCADE-Grande [15]. Right: Resulting energy spectra for all particles together with the spectra of an electron-rich and an electron-poor sample. The all-particle spectrum with the lower threshold and the spectrum of heavy primaries (electron-poor) have been published in [15]. The other spectra are taken from [17].

with

$$k = \frac{\log_{10}(N_{\mathrm{ch}}/N_{\mu}) - \log_{10}(N_{\mathrm{ch}}/N_{\mu})_{\mathrm{H}}}{\log_{10}(N_{\mathrm{ch}}/N_{\mu})_{\mathrm{Fe}} - \log_{10}(N_{\mathrm{ch}}/N_{\mu})_{\mathrm{H}}} \tag{4.5}$$

and

$$\log_{10}(N_{\mathrm{ch}}/N_{\mu})_{\mathrm{H,Fe}} = c_{\mathrm{H,Fe}} \cdot \log_{10}(N_{\mathrm{ch}}) + d_{\mathrm{H,Fe}} \tag{4.6}$$

The coefficients are derived from simulated proton and iron induced showers. The parameter k is defined to be 0 for the mean $\log_{10}(N_{\mathrm{ch}})$-$\log_{10}(N_{\mu})$ values for protons and 1 for the mean $\log_{10}(N_{\mathrm{ch}})$-$\log_{10}(N_{\mu})$ values for iron assuming that the intermediate-mass primaries yield $\log_{10}(N_{\mathrm{ch}})$-$\log_{10}(N_{\mu})$ values in between proton and iron. Therefore, the value of $\log_{10}(N_{\mathrm{ch}})$-$\log_{10}(N_{\mu})$ is sensitive to the mass of the primary particle and can be used to separate the events into mass groups. This is done by comparing the measured value of k with the mean, energy-dependent values of k for simulations of showers induced by particles of five different masses. This procedure is explained in detail in [15], but will also be discussed further in the next chapter, as it is used also for the combined analysis.

The results - from two analyses [15, 17] which used QGSJetII2 based simulations - are shown on the r.h.s. of figure 4.6. The spectrum of heavy primaries (electron-poor) exhibits a knee-like structure at just below 10^{17} eV. This is the energy at which an "iron-knee" is expected in case of charge dependent knee positions. The heavy component is defined to contain events with a k-value above the mean between carbon and silicon simulations and it depends, therefore, on the hadronic interaction model used for the simulations. The mass group for light primaries (electron-rich) is defined in the same manner, using the mean k values between helium and carbon. The spectrum of this mass group shows an ankle-like structure at about $10^{17.1}$ eV. This structure is caused by an onset of another component, which is in general considered to be of extragalactic origin.

The dependence of the all-particle energy spectrum and mass composition on the hadronic interaction model used is discussed in [65]. While the absolute flux and

relative abundances of the various primaries depend on the model, the same features are obtained independently of the hadronic interaction model employed for the analysis.

Due to the differently defined separation of the events into two mass groups, the two spectra do not add up to the all-particle spectrum, i.e. a medium mass group is not included in either of the two mass groups.

4.3. Comparison of the Results

There have been several standalone analyses performed for both KASCADE and KASCADE-Grande. The analysis presented for KASCADE so far was based on QGSJet1 as it was the original analysis that revealed the cause of the knee in the all-particle energy spectrum, however, the results for KASCADE-Grande were mainly based on QGSJetII2, because at the time this result was obtained, the newer version of the QGSJet model was available. For the comparison of both arrays, a more up to date result for KASCADE was chosen, which is also based on QGSJetII2 and the unfolding approach. This way systematic differences due to the hadronic interaction model are avoided.

In principle, there is a KASCADE-Grande analysis that is also based on an unfolding procedure [66], however, that analysis does not reach up to 10^{18} eV. Therefore, the results using the standard, k based approach are used, which will lead to the same conclusion, anyway.

The results are shown together in figure 4.7. The spectrum for iron primaries of KASCADE connects nicely to the spectrum for heavy primaries of KASCADE-Grande. Therefore, at least in the overlapping energy region, the heavy mass group (defined as Si + Fe) should consist mainly of iron. The medium mass group of KASCADE is defined to contain He + C + Si. The mass group below iron, represented here by silicon (Si) should not be present in a significant amount, otherwise the heavy mass group of KASCADE-Grande should have a larger flux. The light mass group of KASCADE-Grande should consist mainly of H + He, however, KASCADE predicts only a very small and decreasing contribution of protons, hence, the light mass group should consist mainly of helium, at least up to the ankle-like feature in its spectrum. The medium mass group of KASCADE, therefore, consists of He + C, the latter not being shown for KASCADE-Grande. However, it is included in the all-particle spectrum.

Distinct inconsistencies arise, when taking a closer look at the results. The iron flux of KASCADE gets larger than the flux of heavy primaries as it is reconstructed with KASCADE-Grande. The all-particle flux reconstructed by KASCADE is larger than the one obtained using KASCADE-Grande. This seems to be due to the medium-mass component first, and later due to the iron component. The easiest explanation is an absolute shift of the mass scale, resulting in a higher energy reconstructed for events in KASCADE. This cannot be proven, however, as it is unknown if the estimated contribution of protons to the light mass group of KASCADE-Grande is larger than the proton flux of KASCADE. Hence, it is not clear whether there is a constant shift of the reconstructed mass or whether the mass scale of KASCADE-Grande is “stretched” compared to KASCADE. It is also not clear if such a different scale would be due to the slightly different reconstruction procedures employed to

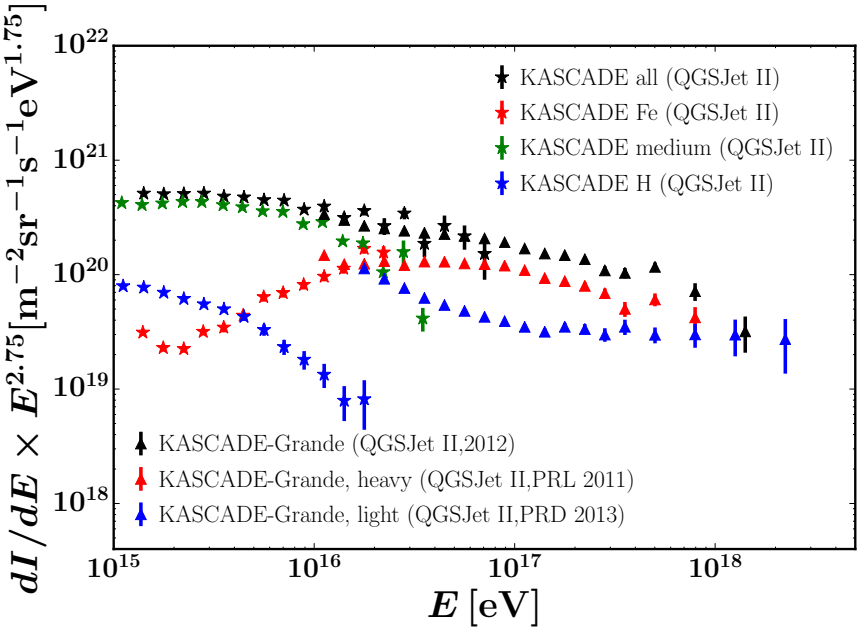

Figure 4.7.: The all-particle energy spectra for the standalone KASCADE and KASCADE-Grande analyses are shown together with the corresponding separation into various mass groups. KASCADE results have been taken from [67]. KASCADE-Grande: All-particle spectrum [68], heavy mass group [15], light mass group [17].

get the number of muons and electrons or a discrepancy between simulations and measurements, as the muonic component is sampled at different ranges from the shower core depending on the distance of the core to the KASCADE array. While a certain influence of the procedure used to reconstruct the energy is likely, it cannot account for the entire difference shown here, as the results for KASCADE-Grande based on the same unfolding approach [66] also show a lighter composition compared to KASCADE.

No systematic comparison of the results obtained using the respective other detector has been performed up to now. The combined analysis that employs a consistent reconstruction strategy to all events is presented in the next chapter. It will either get a consistent composition estimate or it will reveal a deeper inconsistency that is not directly related to the reconstruction procedures used.

5. Data Analysis

5.1. Motivation for a Combined Analysis

Up until now data taken by KASCADE and its extension KASCADE-Grande have been analyzed independently of each other, although KASCADE-Grande used the shielded detectors of KASCADE for the reconstruction of the total number of muons.

Treating both detectors as one has several advantages over the standalone analyses. For events located in the former KASCADE array the main advantage is the increased distance range in which sampling points are available. Because of this, the reconstruction of showers induced by primaries with higher primary energy is possible.

For KASCADE-Grande the main benefit is the availability of 252 additional detector stations among which 3 Grande stations are located. This should result in a more accurate reconstruction of the shower observables.

The aim of the combined analysis was to utilize this improved reconstruction to get one single, consistent spectrum in the energy range from 10^{15} eV to 10^{18} eV. The focus is on the mass composition, which is one of the most important sources of information needed to restrict astrophysical models on the origin and propagation of cosmic rays. With the improved reconstruction of the shower observables, a study of the elemental composition of high-energy cosmic rays - more detailed than before - is possible.

5.2. Shower Reconstruction

The number of electrons and muons have to be extracted from the energy deposited in the KASCADE and KASCADE-Grande detectors. The procedure follows the one developed for the KASCADE standalone analyses (see section 4.1.2). This section is focused on the quality of the reconstructed shower properties.

Since the inclusion of additional detectors is expected to result in a more accurate reconstruction of N_e and N_μ, the accuracies reached by the combined analysis

are compared to the accuracies of the standalone analyses. This is presented in section 5.2.2, but before the general data quality requirements are discussed in section 5.2.1.

The data set contains only a subset of the true number of air showers that hit the detector. This is partly due to the trigger efficiency, which means that not all low-energetic showers trigger the data taking. In addition, the applied selection criteria remove events from the total data set based on the quality of their reconstruction. Hence, the combined trigger and reconstruction threshold has to be inferred. The threshold is defined in terms of the primary energy of the cosmic ray and corresponds to the energy above which all events of the original data set are still present in the final event sample. The efficiency is discussed in section 5.2.3 and the final event sample is presented in section 5.2.4.

The reconstruction is performed separately for events located in KASCADE and KASCADE-Grande, although in both cases all the detectors of the two arrays are used and the same procedure is applied to all events. The reason to distinguish between events located in KASCADE and KASCADE-Grande is that the accuracy of the reconstruction of N_e and N_μ is different for events located in the respective other array. In addition, the muonic component is only measured by the KASCADE detectors, hence, a different part of the muonic component is sampled for events located in KASCADE compared to events located in KASCADE-Grande. If the reconstruction procedure would be tuned independently of the location, the differences in the reconstruction would be effectively ignored for events located in KASCADE, because the fiducial area covered by KASCADE is ten times smaller than the area selected for KASCADE-Grande. Nonetheless, by combining the two detectors, the quality of the reconstruction is improved for both sets of events and the corresponding results can still be merged once the composition of cosmic rays has been reconstructed, which will make a correction for the known reconstruction effects possible.

5.2.1. Quality Cuts

To ensure good quality of the reconstructed quantities, several selection criteria have been applied to exclude events with a bad or failing reconstruction.

In order to avoid border effects at the edges of the detector arrays, a smaller fiducial area was chosen compared to what is covered by the detectors. The chosen area is shown in figure 5.1 in comparison to the areas used in the standalone analyses.

In the following the selection criteria applied will be discussed.

Run: Measurement periods during which problems with either the KASCADE or KASCADE-Grande array occurred are excluded in addition to those where at least one of them was inactive, due to e.g. maintenance.

ANKA: ANKA [69], a synchroton radiation facility at KIT, was inducing fake events during beam injection or beam stops. The procedure to identify these events is discussed in [70].

The above mentioned criteria are only applied to measured data. The following selection criteria are applied to both, measured and simulated events. However, these criteria are applied prior to the actual combined processing of the data, in

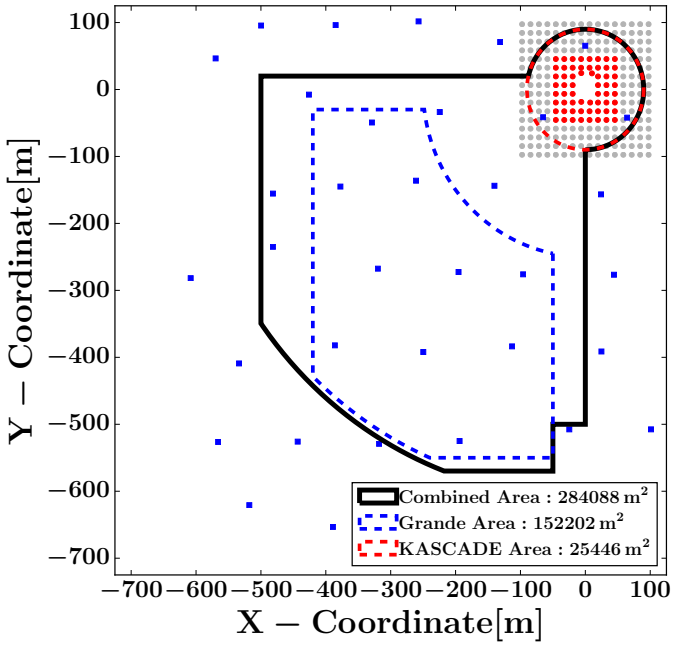

Figure 5.1.: The fiducial area chosen for the combined analysis compared to the ones used for the standalone analyses. Squares mark the locations of the 37 KASCADE-Grande stations. The KASCADE stations are shown as circles. The outer 192 stations, shown in gray, have been equipped with shielded scintillators in addition to non-shielded ones.

order to reduce the size of the dataset. Refined selection criteria are specified during the analysis and are discussed later.

N_e, N_μ: In order to reduce the number of events, for KASCADE a minimum N_e of $10^{3.2}$ electrons is used, which is well below the threshold of full efficiency. For KASCADE-Grande a minimum N_e of $10^{4.8}$ electrons is used. For the same reason a loose cut on N_μ is defined as $10^{3.0}$ muons for both KASCADE and KASCADE-Grande. In addition, non-physical N_e to N_μ ratios are prohibited, as can be seen in figure 5.2, which will be discussed later.

Zenith: The maximum angle relative to the zenith considered in this analysis is 30°.

Age: The age parameter, which is the slope of the lateral density function used for the reconstruction of N_e, is connected to the shower age and can be used to identify fits with an unphysically flat or steep lateral density function. For events located in KASCADE, the age parameter was chosen to be within the range from 0.2 to 1.48. For events within KASCADE-Grande, the range was chosen to be 0.15 to 1.48.

Stations: For events located in KASCADE, at least six stations have to have been triggered. For KASCADE-Grande the minimum number of stations with valid time information [1] is 12. In addition, the stations are organized in overlapping trigger

[1] This is the time of the arrival of the first particle at the individual detector station. These time stamps are used to decide whether a station is considered as a part of the event or not. In addition, the time information is used to infer the arrival direction of the primary particle, hence a minimum number of time stamps is needed to accurately reconstruct this direction.

hexagons. At least for one of these hexagons defined by seven stations, all stations have to be triggered.

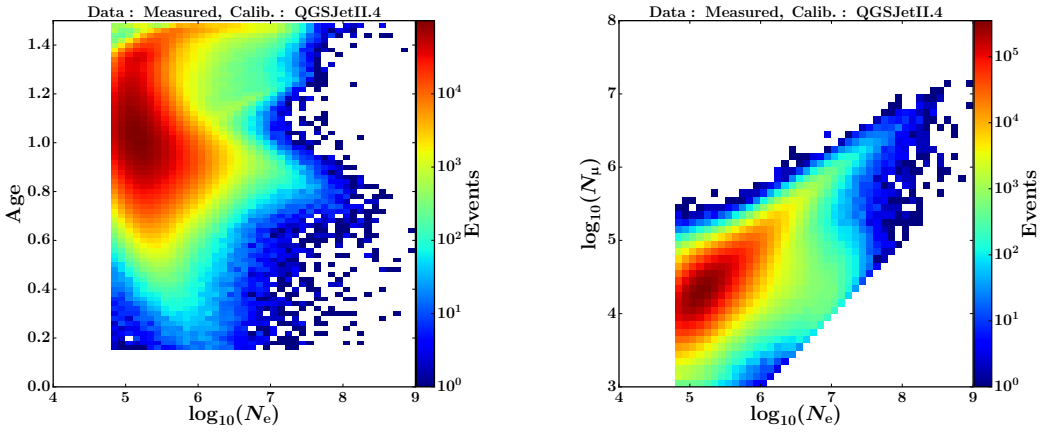

Figure 5.2.: Left: The Age vs N_e distribution for events located inside KASCADE-Grande, having only applied the "pre-processing cuts" described in the text. Right: The N_μ vs N_e distribution for events located inside KASCADE-Grande, for the same cuts.

The resulting distributions of Age vs N_e and N_μ vs N_e are shown in figure 5.2 [2] for events located in KASCADE-Grande. They contain a lot of non-physical structures, such as the blob emerging at the electron-rich edge of the N_μ vs N_e distribution. In principle, one would expect a Age vs N_e distribution that has a decreasing width of the Age values with the number of electrons, because the energy is proportional to N_e and, therefore, the shower-to-shower fluctuations of the shower maximum decreases. Hence, one would not expect the "arms", which are visible on the l.h.s. of figure 5.2. The general increase of the spread between $10^{4.8}$ and $10^{5.8}$ electrons can be explained by the trigger and reconstruction efficiency.

The selection criteria employed to remove these non-physical structures will be explained using events located in KASCADE-Grande, because the event sample for KASCADE is already much cleaner without having applied the corresponding criteria and the effect is, therefore, better visible for KASCADE-Grande. The result of the application of the criteria will be shown for both areas.

On the l.h.s. of figure 5.3 the mean difference of $\log_{10}(N_{ch}^{level1})$ to $\log_{10}(N_{ch}^{level3})$ is shown as a function of Age and N_e. $\log_{10}(N_{ch}^{level1})$ corresponds to the number of charged particles as it is estimated in the first step of the reconstruction (level1) and is based on the energy deposited in the detector stations. The $\log_{10}(N_{ch}^{level3})$ is the number of charged particles as it is obtained from the fitted lateral distribution of the particle densities. A large difference between these two quantities indicates a failed reconstruction of N_e, because a certain number of particles should deposit a corresponding amount of energy in the detectors.

It was argued that the decreasing shower-to-shower fluctuations of the depth of the shower maximum should result in a decreasing spread of the Age as a function of

[2] Many plots shown in this work contain a title which specifies which model was used to interpret the data ("Calib.: X") and whether the data that is interpreted has been simulated or measured ("Data: X"). In the case that it has been simulated, the hadronic interaction model is given.

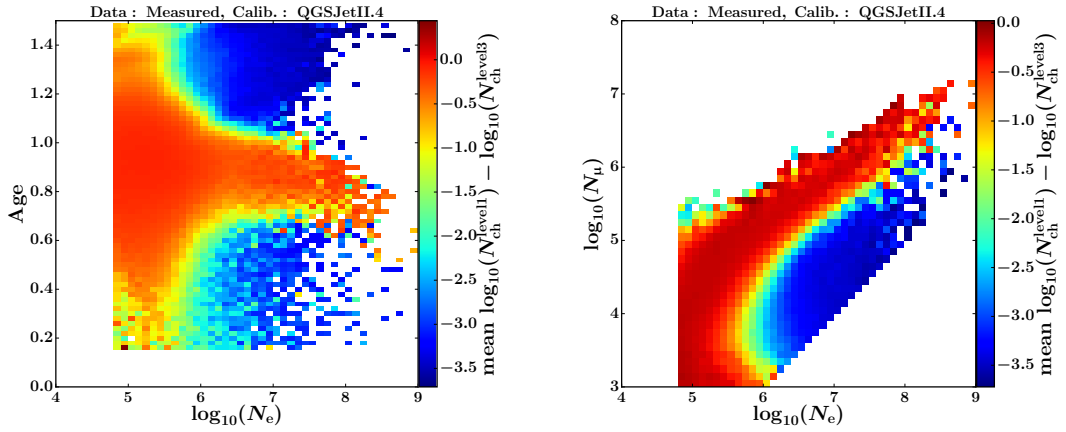

Figure 5.3.: The mean difference of $\log_{10}(N_{\mathrm{ch}}^{\mathrm{level1}})$ as estimated from the energy deposited in the stations (level1) to the final, reconstructed $\log_{10}(N_{\mathrm{ch}}^{\mathrm{level3}})$ (level3) is shown for events located inside KASCADE-Grande, having only applied the "pre-processing cuts" described in the text. Left: Dependence on the Age and N_{e}. Right: Dependence on N_{μ} and N_{e}.

$\log_{10}(N_{\mathrm{e}})$. This is what is visible in figure 5.3. The development of the Age would be as expected, if the "arms" were missing. The events populating these exhibit a large difference between $\log_{10}(N_{\mathrm{ch}}^{\mathrm{level1}})$ and $\log_{10}(N_{\mathrm{ch}}^{\mathrm{level3}})$. They correspond to small, low-energetic showers with an overestimated number of charged particles. As shown on the r.h.s., these events are the same that populate the blob in the N_{μ} vs N_{e} distribution.

It is important to know that at the first step of the reconstruction, the core of the shower is defined to be within KASCADE or KASCADE-Grande based on the energy deposits in the stations. $\log_{10}(N_{\mathrm{ch}}^{\mathrm{level1}})$ is then only estimated from the energy deposits of the corresponding array. For very small showers, this procedure does not always work properly. The problem is that the lateral energy correction function is parameterized in terms of the distance of the station to the core and the current estimate of $\log_{10}(N_{\mathrm{ch}})$. For small distances, the change of the lateral energy correction function with $\log_{10}(N_{\mathrm{ch}})$ is small, however, at large distances, the energy deposited by a single charged particle increases notably with $\log_{10}(N_{\mathrm{ch}})$.

If a low-energetic shower hits and triggers KASCADE, a small $\log_{10}(N_{\mathrm{ch}})$ will be reconstructed. If a second much smaller shower develops simultaneously and hits a KASCADE-Grande station, its associated particle density will be overestimated by far. The reason is that the lateral energy correction function is much smaller at far distances for low values of $\log_{10}(N_{\mathrm{ch}})$ than it is for large values of $\log_{10}(N_{\mathrm{ch}})$ and, hence, even small energy deposits can result in a large particle density. Of course, for large showers hitting KASCADE, this is not a problem, because the value of the lateral energy correction function will be large and the density in the Grande stations will be low (if they are not hit by the large shower, but really get a small energy deposit unrelated to the large shower). The signal of the single station with a large uncertainty will not significantly influence the fit of the lateral density function compared to the 252 KASCADE stations with a large density and, hence, smaller uncertainties. It was also no issue for the standalone analyses, because the

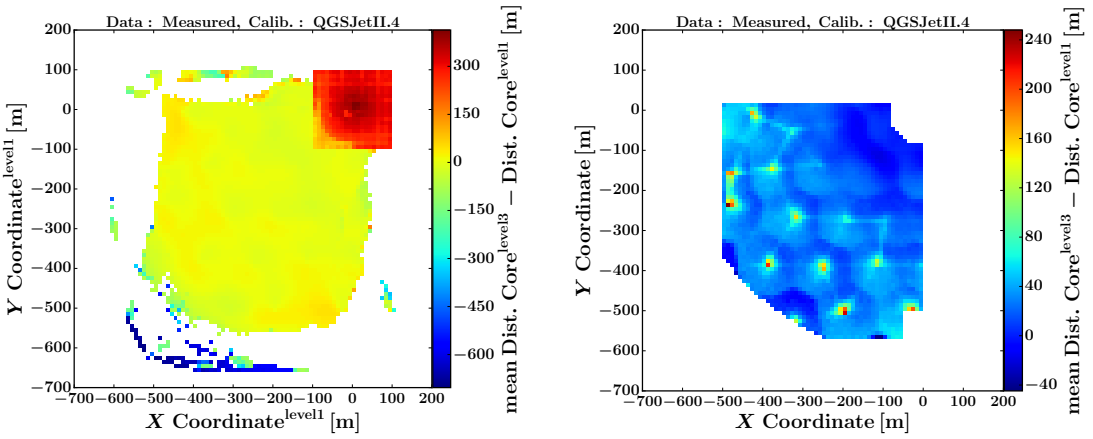

Figure 5.4.: The mean difference of the distance of the shower core to the center of the KASCADE array as estimated from the energy deposited in the stations (level1) to the final, reconstructed core obtained by the fit of the lateral density function (level3) is shown for events located inside KASCADE-Grande, having only applied the "pre-processing cuts" described in the text. Left: Dependence on the core position as reconstructed at level1. Right: Dependence on the final core position reconstructed at level3.

KASCADE-Grande stations were not included in the reconstruction of the lateral density function.

What actually happens in case of small showers is that the core will be pulled to the KASCADE-Grande station with the overestimated particle density. This is shown in figure 5.4, which displays the mean difference of the distance of the shower core to the center of the KASCADE array as estimated from the energy deposited in the stations (level1) to the final, reconstructed core obtained by the fit of the lateral density function (level3) as a function of the core position. The selection of the events is based on the final position of the core as it is reconstructed at level3. On the l.h.s. the core position at level1 is used. All events that have not been reconstructed to be located in KASCADE (at level1) show a fairly low mean absolute difference between the distances at level1 and level3. These are the larger showers that have already triggered KASCADE-Grande and, hence, have correctly been reconstructed to be located in KASCADE-Grande already at level1 with $\log_{10}(N_{\mathrm{ch}})$ being, therefore, correctly estimated by the energy deposited in the KASCADE-Grande stations. The largest shift away from KASCADE is observed for events that have been correctly reconstructed to have hit KASCADE at level1. On the r.h.s. of figure 5.4 the mean values are shown as they depend on the final core position (level3). It is clearly visible that those events have been dragged to the KASCADE-Grande stations. The shift in the distance to KASCADE being smaller than the distance of the respective KASCADE-Grande station to KASCADE is due to the mean shift in each cell being shown, which is dominated by the bulk of regular events that exhibit only small shifts.

The larger the shift is, the larger the Age, and the larger the difference of $\log_{10}(N_{\mathrm{ch}}^{\mathrm{level1}})$ to $\log_{10}(N_{\mathrm{ch}}^{\mathrm{level3}})$ gets. The reason is that the farther away the core is from KASCADE, the larger the particle density will be estimated by the lateral energy correction func-

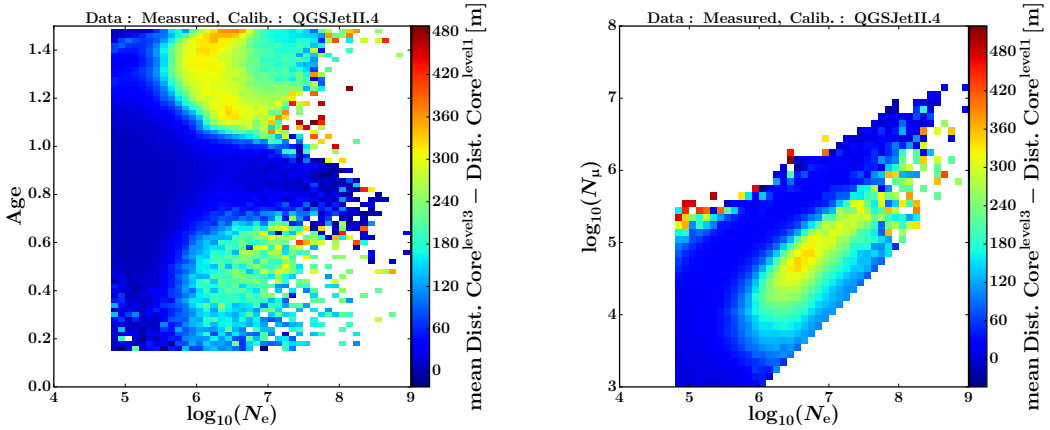

Figure 5.5.: The mean difference of the shower core to the center of the KASCADE array as estimated from the energy deposited in the stations (level1) to the final, reconstructed core obtained by the fit of the lateral density function (level3) is shown for events located inside KASCADE-Grande, having only applied the "pre-processing cuts" described in the text. Left: Dependence on the Age and N_e. Right: Dependence on N_μ and N_e.

tion. Hence, the densities rise for KASCADE, which is now far away from the new core position (and at a KASCADE-Grande station) and the lateral density function gets flatter, resulting in a larger Age. These are the events that populate the "arm" at large values of the Age and populate the blob in the two-dimensional shower-size spectrum. Within about 20 to 70 m to the core, the lateral energy correction function will return larger values for small $\log_{10}(N_{ch})$ than for large values of $\log_{10}(N_{ch})$. Therefore, if a KASCADE-Grande station at the border to KASCADE is hit, the calculated particle densities for KASCADE stations within that range will be underestimated, while the densities for the remaining stations, which are still close, will not change much. Hence, the reconstructed lateral density function will have a tendency to be too steep and the "arms" at low values of the Age are populated.

That this is really the case can be seen in figure 5.5. Shown is the mean difference of the distance of the core to the center of KASCADE between level3 and level1. For $\log_{10}(N_e)$ below 6, i.e. below full reconstruction efficiency, there is a tendency for the core to be pulled towards a station of KASCADE-Grande. The shift is small, however, it has similar effects and was the main reason for "smaller arms" observed in the standalone analyses of KASCADE-Grande. There is a clear difference between the values below an Age of 0.8 and above it. For values below 0.8, the mean difference is much smaller than it is for values above 0.8.

The strategy to remove the non-physical structures is to use the correlation between the shift of the core and the difference between $\log_{10}(N_{ch}^{level1})$ and $\log_{10}(N_{ch}^{level3})$ to define selection criteria on the data. The criteria are defined as $\log_{10}(N_{ch}^{level1})$-$\log_{10}(N_{ch}^{level3})> 0.5$ and abs(Dist. Core^{level3} − Dist. Core^{level1}) < 70 m. The correlation between the shift of the core and the difference between $\log_{10}(N_{ch}^{level1})$ and $\log_{10}(N_{ch}^{level3})$ is shown in figure 5.6 before the criteria (l.h.s.) and after the criteria (r.h.s) have been applied. The distribution behaves as expected from the above discussion.

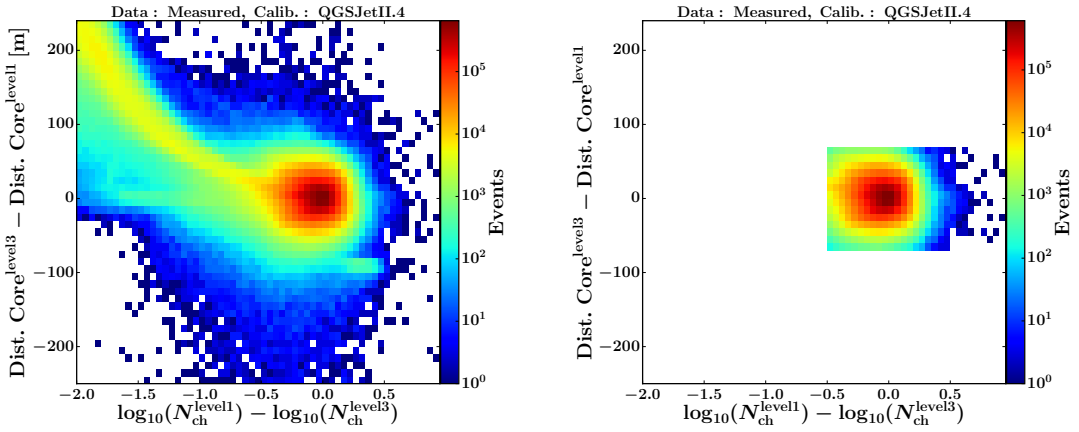

Figure 5.6.: The mean difference of the shower core to the center of the KASCADE array as estimated from the energy deposited in the stations (level1) to the final, reconstructed core obtained by the fit of the lateral density function (level3) is shown for events located inside KASCADE-Grande as a function of the difference between $\log_{10}(N_{\mathrm{ch}}^{\mathrm{level1}})$ and $\log_{10}(N_{\mathrm{ch}}^{\mathrm{level3}})$. Left: Before the level1-level3 cut has been applied. Right: After the cut has been applied.

The final distributions of Age vs N_{e} and N_{μ} vs N_{e} are displayed in figures 5.7 and 5.8, respectively, for events located in KASCADE (l.h.s) and for events with core positions inside KASCADE-Grande (r.h.s.). Almost the entire distortions visible in figure 5.2 are removed by the level1-level3 cut. Only a small change is introduced by the visible linear Age vs N_{e} cut, which further limits the maximum Age as a function of N_{e}.

Note that an additional, efficiency-dependent cut in the N_{e}-N_{μ} plane is applied, which will be discussed later. Also, the above mentioned selection criteria have been defined using measured data, simulations are almost unaffected by the level1-level3 criteria. This can be seen in figure 5.9, which displays the difference in the distance of the core positions as reconstructed in level1 and level3 against the difference in the reconstructed number of electrons (l.h.s.) and the N_{μ} vs N_{e} distribution (r.h.s) for events located inside KASCADE-Grande for QGSJetII4 based simulations, before the level1-level3 cuts have been applied.

The reason is quite simple. For simulated data, only a single shower hits the arrays, i.e. there never is a second, smaller shower developing simultaneously. The remaining effect is a tendency for small showers to be pulled to a nearby station. The corresponding difference in the distance of the core to the individual stations is much smaller, compared to a shower being pulled by a few hundred meters as it is observed in the measured data. This also explains why there is no tendency for an event to be mainly pulled to or to be mainly pushed away from KASCADE.

KASCADE is almost unaffected by the influence of a simultaneously developing small shower, because the probability for a small shower to trigger KASCADE-Grande is very small. Therefore, KASCADE is dominated by the small showers that have triggered KASCADE. This involves small changes to the core position which are also visible in the simulated data and the cut on the change of the distance is not

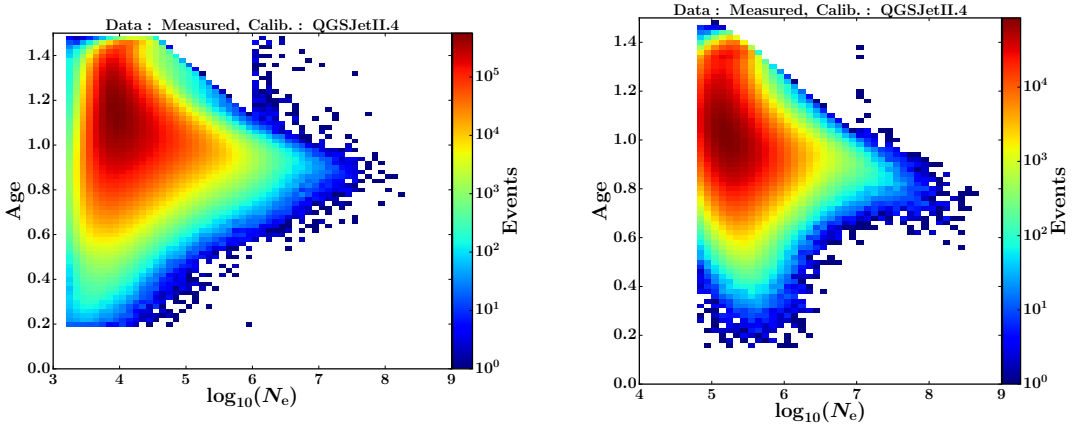

Figure 5.7.: The Age vs N_e distribution for events located inside KASCADE (l.h.s) and KASCADE-Grande (r.h.s) after the level1-level3 cuts have been applied.

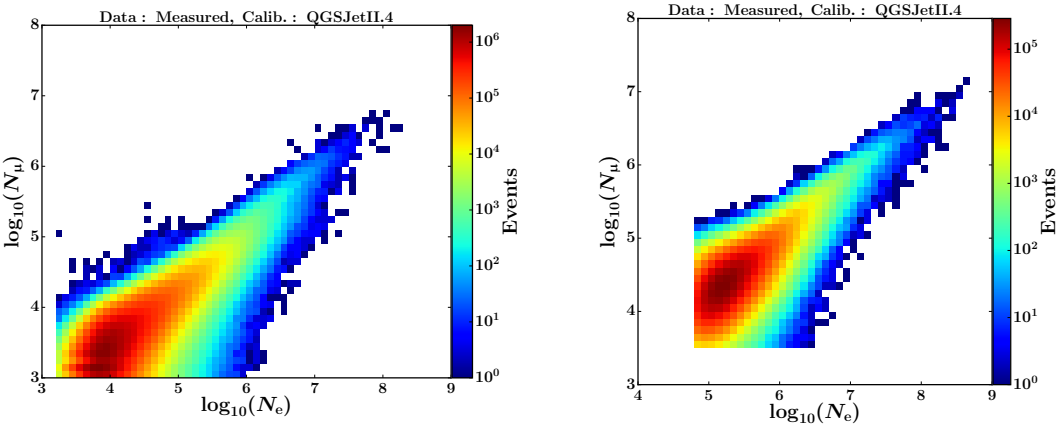

Figure 5.8.: The N_μ vs N_e distributions for events located inside KASCADE (l.h.s) and KASCADE-Grande (r.h.s) after the level1-level3 cuts have been applied. The distributions are shown for measured events with zenith angles between 0° and 30°.

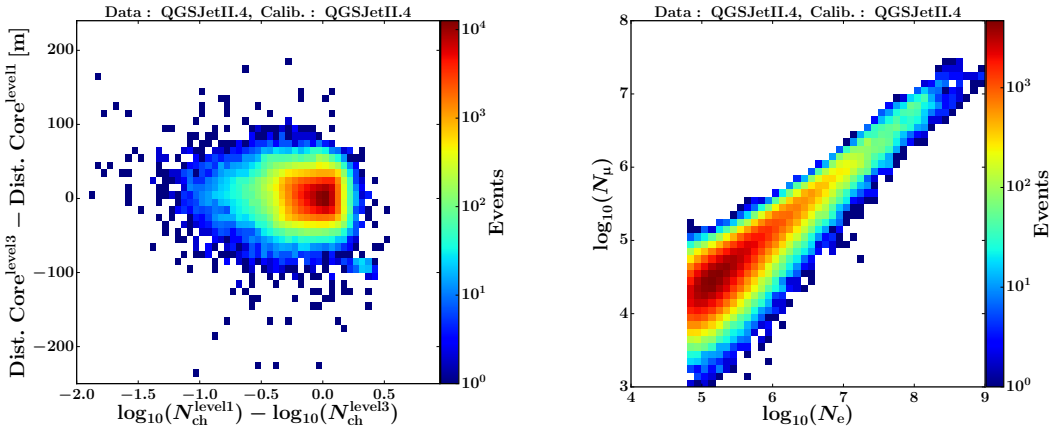

Figure 5.9.: Left: The difference in the distance of the core positions as reconstructed in level1 and level3 is plotted against the difference in the reconstructed number of electrons. Right: The N_μ vs N_e distribution for events located in KASCADE-Grande is shown. Both distributions have been obtained before the level1-level3 cuts have been applied using simulated data.

needed in case of KASCADE. The cut on the difference between $\log_{10}(N_{\mathrm{ch}}^{\mathrm{level1}})$ and $\log_{10}(N_{\mathrm{ch}}^{\mathrm{level3}})$ is the same that is used for events located in KASCADE-Grande.

5.2.2. N_e and N_μ

The reconstruction of N_e and N_μ follows the same procedure as described in section 4.1.2. However, this analysis is based on the combined information from both detectors instead of KASCADE only and in case of the reconstruction of N_e a slightly different parameterization for the NKG-like function (Eq. 5.1) is used, which is $a = 1.6$, $b = 3.5$, and $r_0 = 20\,\mathrm{m}$ at level 2 and $a = 1.6$, $b = 3.4$, and $r_0 = 30\,\mathrm{m}$ at level 3.

$$\rho = N_e \cdot c(s) \cdot (\frac{r}{r_0})^{s-2} \cdot (1 + \frac{r}{r_0})^{s-4.5} \tag{5.1}$$

with

$$c(s) = \frac{\Gamma(b-s)}{2\pi r_0^2 \Gamma(s-a+2)\Gamma(a+b-2s-2)} \tag{5.2}$$

In addition, the size dependent parameterization of the shape of the muon lateral density function has been updated and N_μ has been obtained without applying the truncation. The parameterization for the reconstruction of muons is $a = 1.5$, $b = 3.7$, and $r_0 = 420\,\mathrm{m}$.

The combination results in some advantages compared to the standalone analyses, which are discussed later in this chapter. But first, the quality of the reconstruction will be compared to the standalone analyses.

The reconstruction accuracies of the KASCADE standalone analyses are shown in figure 5.10 for proton and iron induced showers. The accuracy for N_e at 10^5 electrons is about 6 % reaching values below 3 % towards higher energies. In the range from $10^{4.0}$ to $10^{5.5}$ muons (truncation applied), the accuracy changes from about 17 %

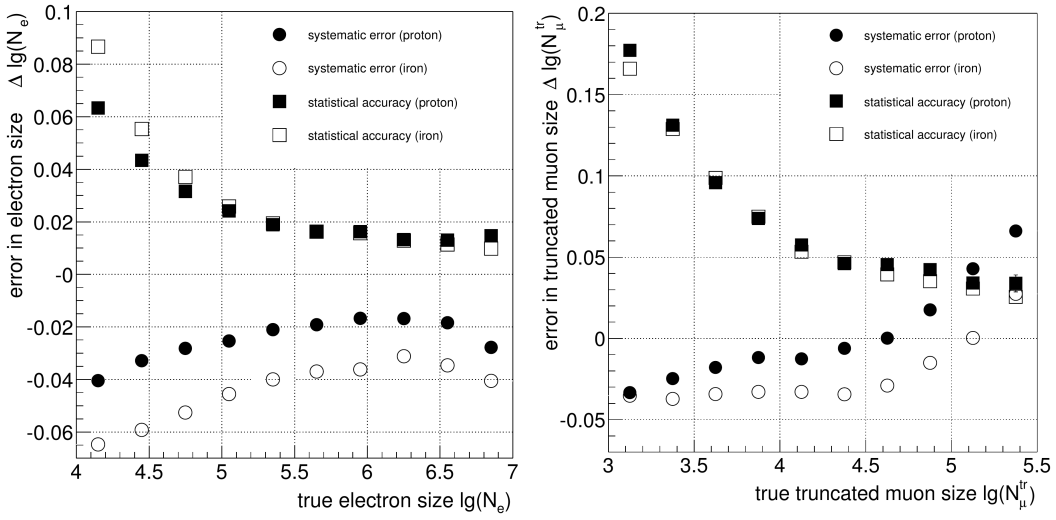

Figure 5.10.: The reconstruction accuracies of KASCADE for N_e (left panel) and N_μ^{tr} (right panel), see [57].

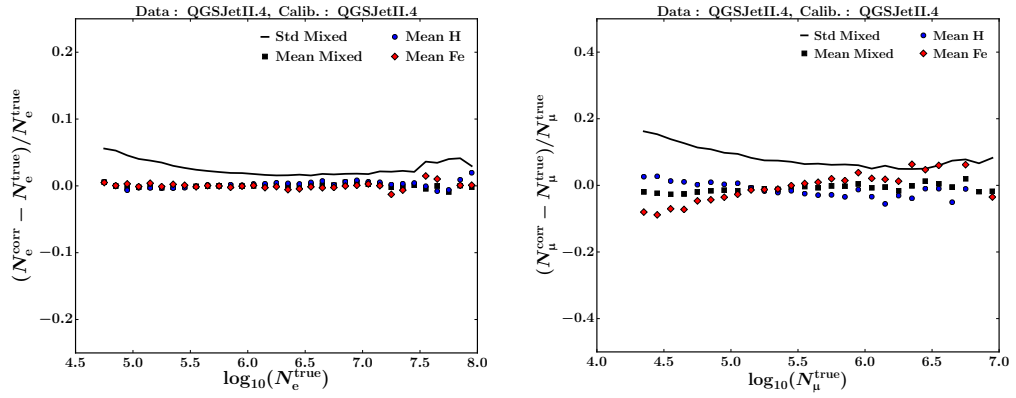

Figure 5.11.: The reconstruction accuracies of N_e (left panel) and N_μ (right panel) for events located inside KASCADE and reconstructed with the combined approach. The mean deviations, i.e. the systematic uncertainties are shown for H, Fe, and a mixed composition of 20 % each of H, He, C, Si, and Fe. In addition, the standard deviation is shown in case of the mixed composition.

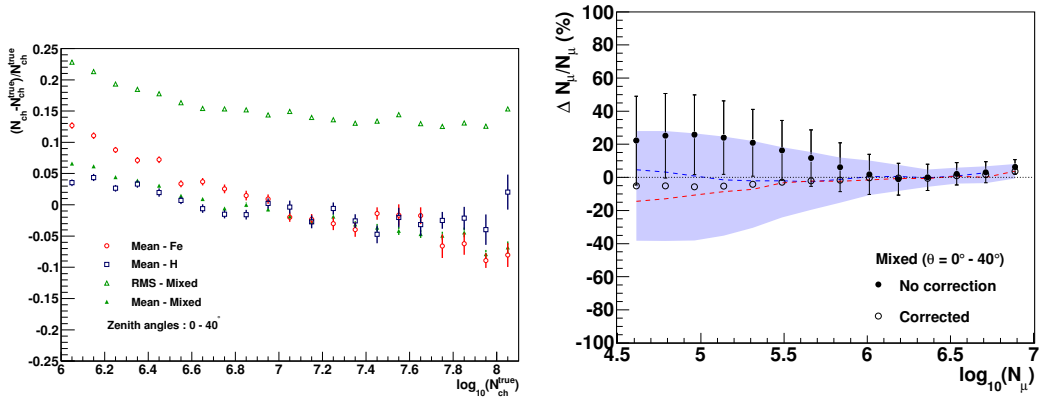

Figure 5.12.: The reconstruction accuracies of KASCADE-Grande for N_{ch} (left panel) and N_{μ} (right panel, see [68]). In both cases, "mixed" refers to a mixed composition of 20 % each of H, He, C, Si, and Fe. In case of N_{ch}, the mean deviations from the true number of charged particles are shown for H, Fe, and the mixed composition. In addition, the standard deviation (labeled as RMS) is shown for the mixed composition. For the number of muons, the mean deviations are shown for H (blue dashed line), Fe (red dashed line), and the mixed composition. For the latter, the mean deviations are shown before a correction for known systematic effects has been applied (solid circle, error bars show the standard deviation) and after the correction has been applied (open circles, the shaded band shows the standard deviation relative to the mean values).

to 7 %. In figure 5.11 the same information is shown for the combined analysis for events located inside the KASCADE array. The differences to the true N_e and N_{μ} are given in percent. The accuracy for N_e ranges from about 4.5 % at 10^5 electrons to below 2 % towards higher energies. Due to the truncation, the corresponding N_{μ} range is shifted by approximately 0.5 in log scale, therefore, the relevant range is from $10^{4.5}$ to $10^{6.0}$ muons. Within this range, the accuracy changes from about 14 % to 5 %.

The reconstruction accuracies of the KASCADE-Grande standalone analyses are shown in figure 5.12 for proton and iron induced showers, and a mixed composition of 20 % each of H, He, C, Si and Fe. The accuracy for N_{ch} improves from about 22 % to about 15 % within the energy range of KASCADE-Grande. The threshold of full efficiency corresponds to about $10^{5.3}$ muons. At this number of muons, N_{μ} is reconstructed with an accuracy of about 25 % improving towards higher energies to about 5 %. The same information is given in figure 5.13 for the combined analysis and for events located in KASCADE-Grande. The accuracy for N_e [3] improved significantly compared to the standalone analyses. At the threshold, N_e is reconstructed with an accuracy of about 16 %. At higher energies an accuracy of better than 9 % can be reached. The reconstruction of the number of muons is improved even more. At $10^{5.3}$ muons an accuracy of about 13 % is achieved and about 5 % at 10^7 muons.

[3] Note that N_e is used in the combined analysis instead of N_{ch}. The accuracies should be comparable, since the electromagnetic component is dominant in the relevant energy range.

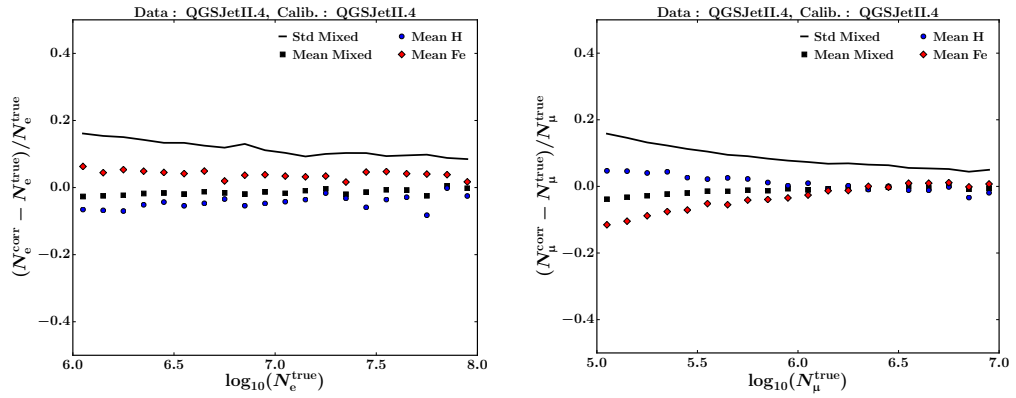

Figure 5.13.: The reconstruction accuracies of the combined analysis for events located inside KASCADE-Grande and for N_e (left panel) and N_μ (right panel). The mean deviations are shown for H, Fe, and a mixed composition of 20 % each of H, He, C, Si, and Fe. In addition, the standard deviation is shown in case of the mixed composition.

As shown on the r.h.s. of figure 5.12, corrections have been applied to the number of muons, which depend on the muon number itself, on the distance of the muon detectors to the shower core and the zenith angle. The fixed slope of the muon lateral density function introduced a systematic overestimation of the number of muons towards large distances of the shower core to the muon detectors. This is explained in more detail in [68].

A similar correction has been applied in the combined analysis additionally taking into account N_e. The procedure of the correction is explained in Appendix B. In order to check if this correction works also for an independent data set, the corrections derived using data simulated with QGSJetII4 have been applied to EposLHC based simulations. The result is shown and discussed in Appendix B. There it is shown that the corrections also work in the light of a different hadronic interaction model.

5.2.3. Efficiency

The probability to register an event depends on the number and the spread of particles that reach the ground (secondary particles have to hit enough detectors). This probability is called the trigger efficiency, e.g. the percentage of events that trigger the data taking. If an event has triggered the detector, its successful reconstruction is not guaranteed. The efficiency of interest is, therefore, the combined trigger and reconstruction efficiency. The procedure of deriving this efficiency and the corresponding cuts in the N_e-N_μ plane are explained in the following.

The efficiency is derived using simulated air showers. The number of generated events for a certain energy, fiducial area and range of arrival direction is known and is compared to the number of events that have been successfully triggered and reconstructed with their core inside the same fiducial area, their arrival direction within the same range and that fulfill all quality criteria.

Classically, the efficiency is defined as the fraction of the events hitting the detector that get fully reconstructed and are part of the final selection. Hence, it can only take

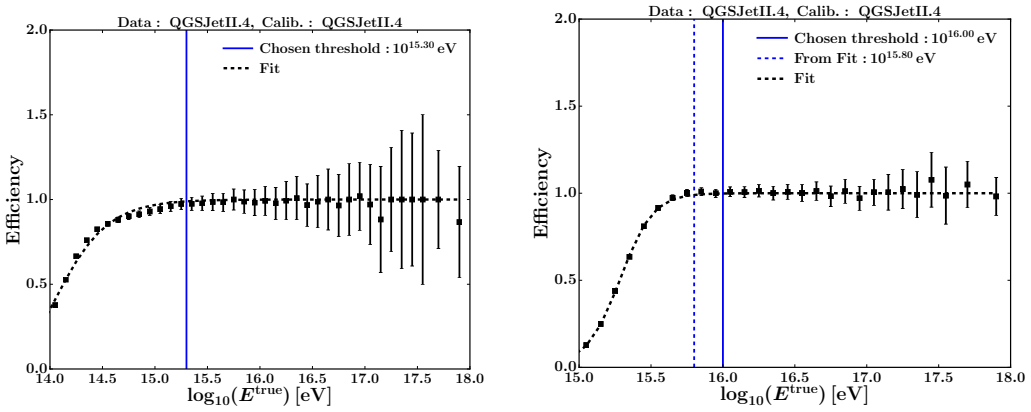

Figure 5.14.: The combined trigger and reconstruction efficiency as a function of the true energy. Left: For events located in KASCADE. Right: For events located in KASCADE-Grande.

values smaller or equal to unity. However, the classical definition of the efficiency is not representative for the "effective efficiency". While the classical definition takes into account the events that are not part of the final selection, because the core has been reconstructed to lie outside of the selected area, it ignores that events might be reconstructed to lie within the selected area, although the true core position lies outside of it. These events fulfill all selection criteria, they have been reconstructed with the same quality as the events that have hit the detector just within the selected area and their core positions are reconstructed with the same uncertainties. Hence, the loss of events due to them being pushed out of the selected area is compensated by equally well reconstructed events being pulled into the area. However, the probability to pull events into the area is larger than the probability to push events out of the area. This can be explained by defining a "border region" at the left and the right side of a line that follows the contour of the area. The two regions have the same width defined by the uncertainty on the position of the core, hence, the region on the outside of the area is slightly larger than the region on the inside of the selected area. This effect has to be taken into account, which is why the "effective efficiency" defined above differs from the classical one used in e.g. elementary particle physics.

Angles	Location	QGSJetII2	QGSJetII4	EposLHC
00.00-16.78°	KASCADE	$10^{15.2}$ eV	$10^{15.3}$ eV	$10^{15.3}$ eV
	Grande	$10^{15.8}$, $10^{16.0}$ eV	$10^{15.8}$, $10^{16.0}$ eV	$10^{15.8}$, $10^{16.0}$ eV
16.78-24.09°	KASCADE	$10^{15.3}$ eV	$10^{15.4}$ eV	$10^{15.4}$ eV
	Grande	$10^{15.8}$, $10^{16.0}$ eV	$10^{15.9}$, $10^{16.0}$ eV	$10^{15.9}$, $10^{16.0}$ eV
24.09-30.00°	KASCADE	$10^{15.4}$ eV	$10^{15.6}$ eV	$10^{15.6}$ eV
	Grande	$10^{16.0}$ eV	$10^{16.0}$ eV	$10^{16.0}$ eV

Table 5.1.: Energy thresholds for full efficiency are listed depending on the zenith angle range, whether the shower was reconstructed in KASCADE or KASCADE-Grande, and hadronic interaction model. If two values are given, then the first one corresponds to the fitted threshold, the second one corresponds to the threshold chosen.

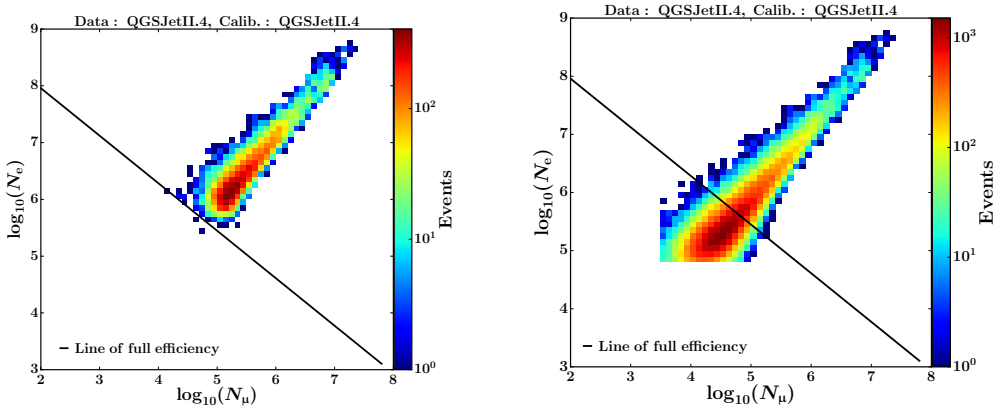

Figure 5.15.: The "full efficiency cut" (solid line) on top of the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution for simulated events located in KASCADE-Grande. This is shown on the l.h.s. for events with energies above the threshold of full efficiency and for all events on the r.h.s.

Ideally, this efficiency is equal to unity, values significantly smaller than one could indicate a possible underestimation of the reconstructed energy spectrum and values significantly above one could indicate a possibly overestimated number of events.

In order to take the shower attenuation into account [4], the analysis is split into three zenith angle ranges, namely 0°-16.78°, 16.78°-24.09°, and 24.09°-30.00°. The efficiency for the first zenith angle range (0°-16.78°) using QGSJetII4 as the hadronic interaction model is shown in figure 5.14.

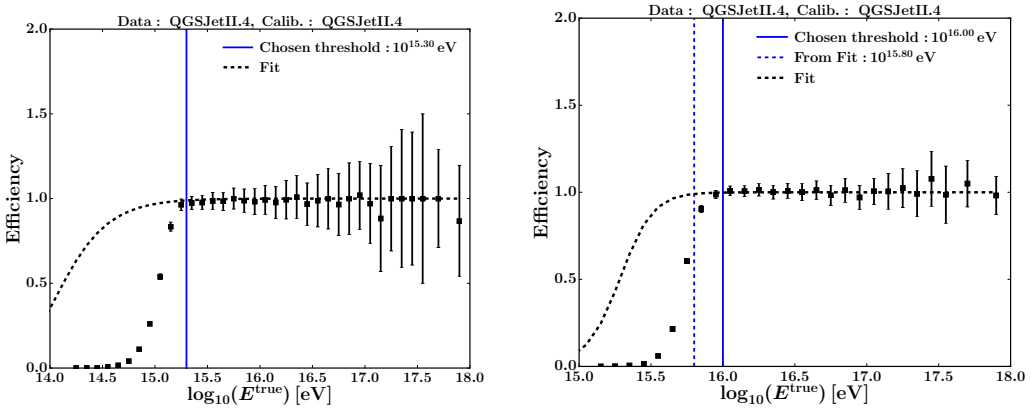

Figure 5.16.: The combined trigger and reconstruction efficiency as a function of true energy with the "full efficiency cut" applied. Left: For events located in KASCADE. Right: For events located in KASCADE-Grande. The dashed line labeled "Fit" corresponds to the line shown in Fig. 5.14.

The threshold of full efficiency increases slightly with increasing zenith angle, therefore, figure 5.14 shows the lowest possible threshold for which full or almost full efficiency is maintained. Values above 1.0 are due to events reconstructed within

[4] The shower attenuation increases towards larger zenith angles, because the shower has to pass more atmosphere.

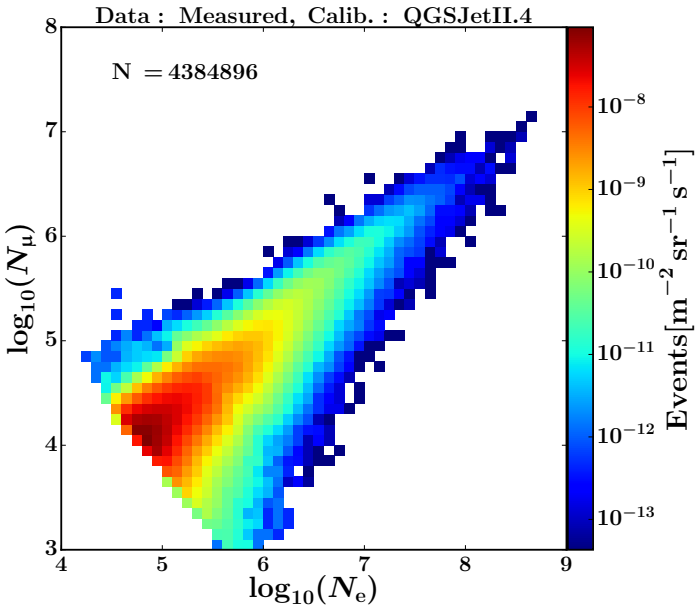

Figure 5.17.: The N_μ vs N_e distribution for events with a zenith angle within 0 ° to 30 °. All selection criteria have been applied. The distribution is based on all events, i.e. on those measured in KASCADE and KASCADE-Grande. The different "lines of full efficiency" for the two areas and the three zenith angle ranges have been taken into account.

the respective fiducial area and angular range, although they truly lie outside of it. Values below 1.0, but above the chosen threshold are mainly caused by events migrating to a higher zenith angle range. The thresholds obtained for the zenith angle ranges considered are listed in table 5.1 for events located in KASCADE and KASCADE-Grande, and for the three hadronic interaction models.

The energy threshold is used to define a "full efficiency cut" in the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ plane. Figure 5.15 shows this cut on top of the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution for showers induced by particles with energies above the threshold (l.h.s.) and on top of the same distribution for all simulated showers that have passed the quality criteria described in the last section (r.h.s.). This is only shown for events located in KASCADE-Grande. A similar cut has been derived for events located in KASCADE.

For iron primaries, a stricter selection is possible, the larger shower-to-shower fluctuations e.g. for protons, however, result in the line shown in figure 5.15 being selected. As can be seen in the plot on the r.h.s. of figure 5.15 as well as in figure 5.16, the cut cannot remove all showers induced by primaries with energies below the threshold, but they are sufficiently suppressed without the events above threshold being influenced at all.

5.2.4. The Event Sample

Like the standalone analyses the combined analysis is based on the number of electrons and muons reaching the observation level, as well as on their correlation.

The two-dimensional shower-size spectrum is shown in figure 5.17 for all events located inside the fiducial area with a zenith angle within 0 ° to 30 °. The different

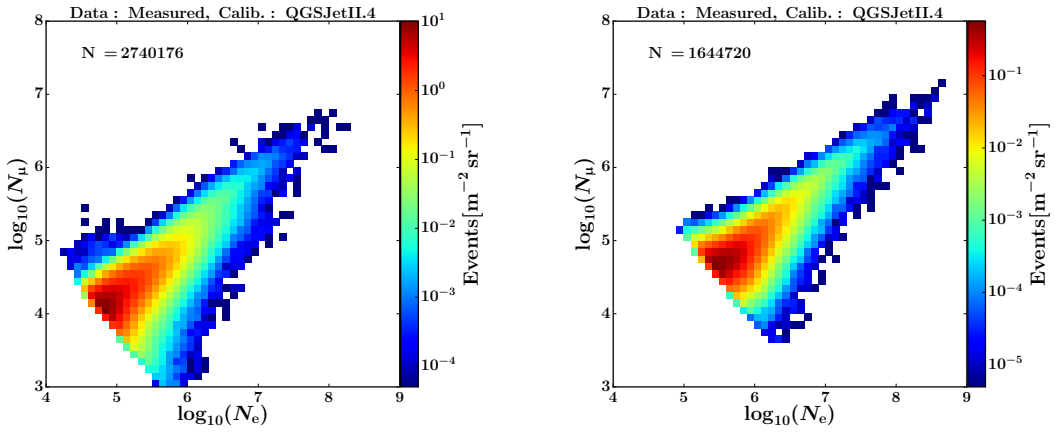

Figure 5.18.: The N_μ vs N_e distribution for events with a zenith angle within 0 ° to 30 °. This is shown on the l.h.s. for events located in KASCADE and on the r.h.s. for events located in KASCADE-Grande. In both cases, all selection criteria have been applied.

"lines of full efficiency" for the two areas and the three zenith angle ranges have been taken into account. In total, 4384896 events have been analyzed in this work. The transition from "KASCADE-only" to the region where both arrays contribute is very smooth. Even at the edges of the distribution, no hard transition is visible, despite the slightly different resolution for events located in the respective other area.

Nonetheless, in order to take the smaller differences of the reconstruction procedure for events located in KASCADE and KASCADE-Grande into account, the data set has been split based on the location of the shower core. In particular the measurement of the muons differ for those sets. In one they are measured close to the shower core (up to 200 m), in the other farther way (200 - 700 m). The two-dimensional shower-size spectra for events located in KASCADE and KASCADE-Grande are shown in figure 5.18 for events with a zenith angle within 0 ° to 30 °. On the l.h.s. it is shown for events located in KASCADE and on the r.h.s. for events with core positions within KASCADE-Grande.

In case of KASCADE, 2740176 events fulfill all quality criteria. Despite the significantly larger fiducial area, a smaller number of events, namely 1644720, are selected for KASCADE-Grande. The reason for a smaller number is the higher energy threshold combined with the steep energy spectrum.

For KASCADE the number of showers fulfilling the criteria decreases towards larger zenith angles. This is due to the increasing energy threshold, because more inclined showers have to pass more atmosphere and, therefore, the shower is more attenuated before it reaches the observation level.

In case of KASCADE-Grande, the picture is reversed, i.e. the number of events increases towards larger zenith angles. This is due to KASCADE-Grande being triggered mainly by the muonic component, especially in case of relatively small showers at the energy threshold. Hence, the larger zenith angle results in an elongated footprint which increases the chance to hit enough stations to trigger the data taking. However, the combined trigger and reconstruction threshold still increases

towards larger zenith angles, because the decreasing number of electrons reaching the detectors does not always allow for a successful reconstruction.

5.3. Analysis

This section describes the performance and results of the combined analysis. Results on energy spectra and mass composition will be presented and discussed for three different hadronic interaction models, namely QGSJetII2, QGSJetII4, and EposLHC. The last two models have already been fine tuned including the first LHC data. The first model is included to compare the results of the combined analysis with the previous standalone analyses, for which results are available.

The outline of the analysis is as follows.

In section 5.3.1 the procedure of deriving the energy of the primary particle is explained. The resulting all-particle energy spectra are presented and differences to the results of the standalone analyses are discussed. It is shown that the mass composition of cosmic rays is needed to be known to correct for mass dependent misreconstructions of the fluxes of the various elements.

Based on the parameter k (Eq. 4.5) the event sample has been divided in light and heavy mass groups. The corresponding procedure and the results are discussed in section 5.3.2. It is shown that with this approach the composition cannot be reconstructed with high enough detail in order to correct for the mass dependent misreconstructions. Although the separation in two mass groups works reliably, separating the medium mass primaries from the heavy mass group is not possible without taking the width of the distribution of k into account, which is ignored in the procedure used for the separation.

Hence, a different approach is needed (section 5.3.3) and has been developed that results in a reconstruction of the composition with a degree of detail never reached before in this energy range. However, due to differences between simulations and measurements that have been revealed by the present work, it is not possible to successfully apply this method to measured data.

In section 5.3.4, the results and findings are summarized.

5.3.1. The Energy Spectrum

Reconstruction of the primary energy

The primary energy is reconstructed on an event-by-event basis according to equation 4.4, which uses the k parameter (Eq. 4.5) to take the mass dependence into account. The parameterization of k as a function of $\log_{10}(N_e)$ is shown in figure 5.19 for events located in KASCADE (l.h.s.) and KASCADE-Grande (r.h.s.) using QGSJetII4 based simulations.

For a certain reconstructed number of electrons the fits of the mean values of $D = \log_{10}(N_e)$-$\log_{10}(N_\mu)$ for proton (D_H) and iron (D_{Fe}) induced showers define $k = 0$ and $k = 1$, respectively. Between these two values, k scales linearly with D (Also beyond $k = 0$ and $k = 1$, which are defined by the respective mean values of D and not by the minimum and maximum values.).

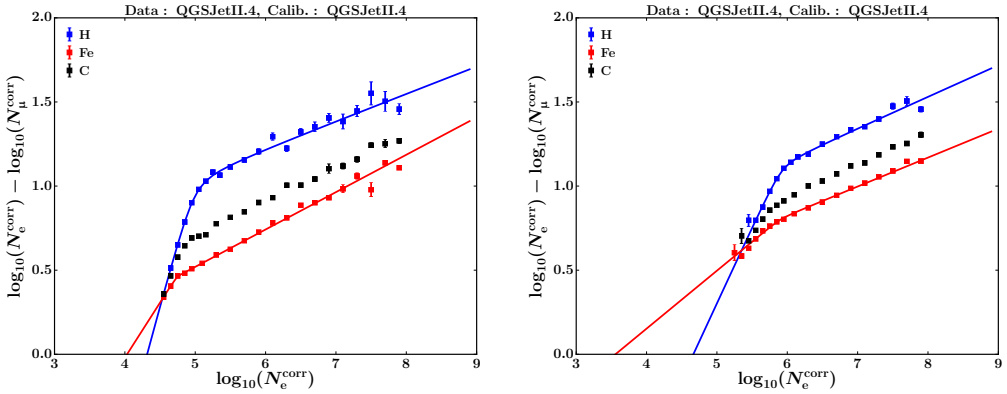

Figure 5.19.: $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ is shown as a function of $\log_{10}(N_e)$ for QGSJetII4 simulations for the first zenith angle range. The lines correspond to the results of a fit to the respective data points. Left: Core inside KASCADE. Right: Core inside KASCADE-Grande.

The change of slope in the mean values of D as a function of $\log_{10}(N_e)$ is caused by the efficiency cut applied in the N_e-N_μ plane. Because of this, small N_μ values are cut away below some N_e threshold. This threshold depends on the mass of the primary, as the efficiency is defined for a certain threshold energy and the mean N_e for a certain energy gets smaller with increasing mass. In addition, the intrinsic shower-to-shower fluctuations are much larger for light primaries compared to heavy primaries, which explains why proton induced showers are affected not only towards higher N_e, but also more severely.

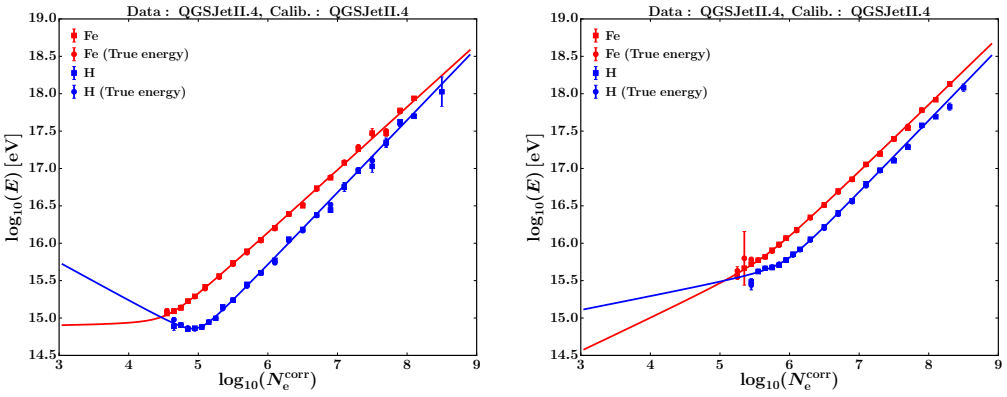

Figure 5.20.: The mean values of $\log_{10}(E^{true})$ are shown as a function of $\log_{10}(N_e)$ for QGSJetII4 simulations for the first zenith angle range. Left: Core inside KASCADE. Right: Core inside KASCADE-Grande. Circles show the mean of the true energy, i.e. the data points used for the fits (lines). Rectangles correspond to the reconstructed mean values.

Figure 5.20 shows the energy calibration function, which is again given by two lines connected by a hyperbola [71]. It is nicely visible how the chosen threshold in the N_e-N_μ plane cuts into the energy distribution for low $\log_{10}(N_e)$ values. However, this direct influence vanishes above the chosen energy threshold. Since the mean energies for a certain number of electrons is different for light and heavy elements,

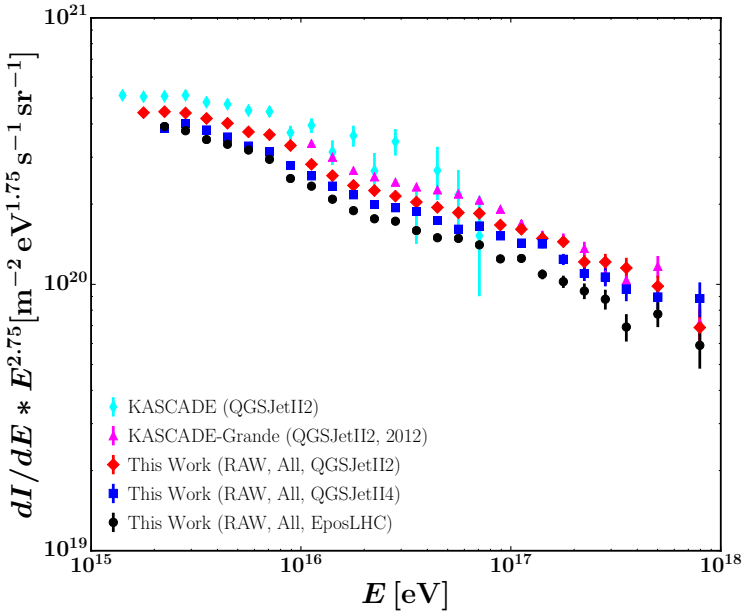

Figure 5.21.: Measured all-particle energy spectra reconstructed based on simulations using different hadronic interaction models. For comparison, all-particle spectra of the standalone analyses using QGSJetII2 have been added.

k is used to scale the energy linearly between the energy corresponding to a proton and iron primary, respectively.

The circles shown in figure 5.20 correspond to the mean simulated energy in the $\log_{10}(N_\mathrm{e})$ bin, i.e. the true energy for iron and proton simulations. The squares show the mean reconstructed energy for the same iron and proton induced events, including the mass dependence according to the k parameter. The true mean for iron and proton induced showers are well reconstructed, which is expected, because k was also defined using proton and iron induced showers. The means for intermediate primaries and the spread of the reconstructed energy for all primaries are a different matter, as will be discussed later.

The measured primary energy spectrum

The measured all-particle energy spectra resulting from the application of the above described method are compiled in figure 5.21.

An interesting observation is that the shape of the reconstructed spectra does not seem to depend strongly on the hadronic interaction model that was chosen for the calibration. The absolute flux, however, does depend on that choice. It is argued in [65] that the main reason for this is the difference in the predicted number of muons. For example, EposLHC predicts significantly more muons compared to QGSJetII2. Therefore, the measured data seems to have a lighter composition, because the events seem to be relatively "muon-poor". As can be seen in figure 5.20, a primary with a lower mass, such as a proton compared to an iron nucleus, gets assigned a lower energy for the same measured $\log_{10}(N_\mathrm{e})$. This results in the spectra being shifted towards lower energies.

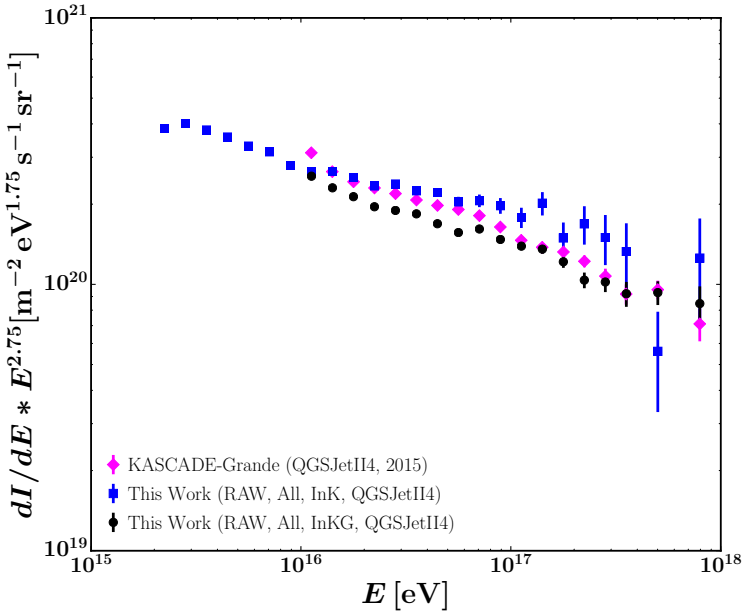

Figure 5.22.: Measured all-particle spectra for QGSJetII4 compared for the KASCADE-Grande standalone and the combined reconstruction. For the latter the spectra for events located in KASCADE and for those located in KASCADE-Grande are shown separately. In both cases the raw spectra are shown, i.e. the mass dependent misreconstructions have not been corrected for.

Taking a look at the results of the standalone analyses, the flux measured by KASCADE seems to connect well to the lower end of the KASCADE-Grande spectrum, however, it starts to diverge from thereon. Although the spectra of the combined analysis do not seem to be affected by the same systematic uncertainties, taking a look at figure 5.22 shows that they do. The larger flux reconstructed for events located in KASCADE is suppressed by the much larger number of events located in KASCADE-Grande, which seem to be following a different spectrum. This difference between the areas covered by the two arrays affects the results for all hadronic interaction models used in this analysis, hence, the comment on the lower absolute flux being due to a different shower development is still valid.

Ignoring the differences to the standalone analyses for a moment, it is worth noting that the spectra shown in figure 5.21, exhibit all the features that have been found already in the standalone analyses. One feature is the knee, known since 50 years [72], which has also been found by other experiments, such as the EAS-TOP installation [13, 14]. It is best visible in case of QGSJetII2, because of the additional energy bin available at low energies, but can also be found in the spectra of the other two models. A concavity at a few 10^{16} eV is visible for all models, the standalone analyses, and also in the separate spectra measured for events located in KASCADE and KASCADE-Grande (see Fig. 5.22). The same is true for a knee-like structure at about 10^{17} eV, which was found by KASCADE-Grande to be due to a structure in the spectrum of heavy primaries. Also this structure has been confirmed by other experiments, such as IceTop [73] and Tunka [74].

For a more detailed and accurate discussion of the results (especially when comparing them to astrophysical models), the differences to the standalone analyses have to be understood as well as the differences of the result obtained for events located in KASCADE to the one obtained for events within KASCADE-Grande. The latter will not be discussed in this section, as it will only become clear once an attempt to measure the composition has been made, which is discussed in the next two sections. Therefore, the differences to the standalone analyses are explained next, the final discussion of the results will be postponed to later (when the current limitations are clear), taking place in chapter 6.

Explaining the difference to the standalone analyses

Comparing the results of the standalone analyses with the results of the combined analysis, the reconstructed spectrum of the latter is systematically lower, but flatter [5], especially at lower energies. A large part of this difference between the results of the combined analysis and the standalone analyses is due to the fact that the latter includes corrections for the uncertainties and biases of the reconstruction of the energy, the former, however, has not been corrected for these effects and is thus labeled as "RAW".

The coefficients for the parameter k and the energy have been fine tuned for the standalone analysis of KASCADE-Grande to equally well reconstruct the simulated spectra for proton and iron primaries. After ensuring this, the spectrum has been unfolded to correct for bin-to-bin migrations among the energy bins, due to the non-zero bias on the reconstructed energy and due to the uncertainty of the reconstructed energy being larger than the bin-sizes. One main concern for the present analysis was that the spectrum should be reproducible in an exact manner, i.e. if the analysis is performed again, no user input should result in differently reconstructed spectra. While the required fine tuning of the parameters applied in the KASCADE-Grande analysis was not very extensive, doing it again will result in very similar - yet different - spectra. To avoid this, the present analysis does not use any user supplied input such as start parameters for the fits (which are extracted from the data itself) or even a manual fine-tuning of the energy calibration (which, in addition, does not work at lower energies as will be shown in a moment).

Without a correction, the deviation of the reconstructed flux from the true one depends on the mass of the primary, which can be seen in figure 5.23. Shown are the predictions of the spectra for five different nuclei according to the H4a [29, 30] model (lines) and the reconstructed simulations, which have been weighted in a way that the true simulated spectra match the model predictions, hence, the markers represent the H4a model predictions after the detector and reconstruction effects have been folded into these.

The flux of iron primaries is reasonably well reconstructed, while the fluxes of lower mass particles are generally too low. As shown in figure 5.20, the mean energies as a function of $\log_{10}(N_e)$ are reconstructed quite accurately, therefore, the steep

[5] A flat or hard spectrum is described by a larger (less negative) index of the underlying power law compared to a steep spectrum. If two spectra match each other at high energies a flat spectrum contributes less at low energies compared to the steep spectrum (as it is the case here). However, if they match each other at low energies, the flat spectrum becomes larger (higher flux) than the steep spectrum towards high energies.

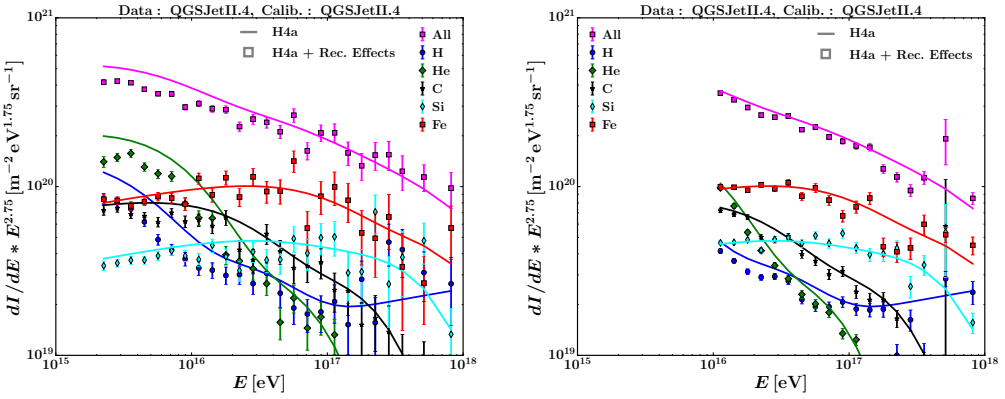

Figure 5.23.: Simulated spectra weighted to match the flux and composition predicted by the H4a [29, 30] model. Lines describe the true, simulated flux. Markers correspond to the reconstructed flux. This is shown on the l.h.s. for events located in KASCADE and on the r.h.s. for events located in KASCADE-Grande.

spectrum should lead to a systematic overestimation of the flux towards higher energies. In this case, the lower flux could only be explained for the first few energy bins, because these are affected by the N_e-N_μ cut, which significantly suppresses the flux below the chosen energy threshold (as was shown in figure 5.16) and, therefore, the shift of events to higher energies cannot be (over-)compensated. The different behavior for iron and proton primaries could then be explained by the much smaller shower-to-shower fluctuations for heavier primaries. In fact the N_e-N_μ distribution for iron is so narrow, that the cut does not notably influence the energy reconstruction above the threshold.

This scenario cannot explain the underestimated flux at higher energies, however. The most likely reason can be explained by the interplay of the fits shown in figures 5.19 and 5.20 and the N_e-N_μ cut together with the mass- and energy dependent degree of the shower-to-shower fluctuations.

The symptom which ultimately leads to the underestimation for light primaries is displayed in figure 5.24. Shown is the ratio of the $\log_{10}(E)$ vs $\log_{10}(N_e)$ distribution to the $\log_{10}(E^{\text{true}})$ vs $\log_{10}(N_e)$ distribution for events located inside KASCADE. For iron primaries (r.h.s.) the mean and the spread of the true energies is quite well reconstructed, which explains the sufficient reconstruction of the iron flux. For protons (l.h.s.), on the other hand, the reconstructed spread is much too small compared to the spread of the true energy. In fact, the reconstructed spread is very similar to the spread for iron, modified slightly by the different energy distributions in the $\log_{10}(N_e)$ bins. In order to correctly reconstruct the flux, a good description of the spread is needed and a slightly underestimated mean. The latter ensures that the overestimation of the flux due to the larger statistics in lower energy bins is countered.

In order to "fix" this manually, one would have to tune the calibration of the k parameter and/or the one of the energy. Figure 5.25 shows the influence of both the energy calibration (which defines the H and Fe "Baselines" shown on the r.h.s. of figure 5.25) and the calibration of D = $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ on k and on the energy

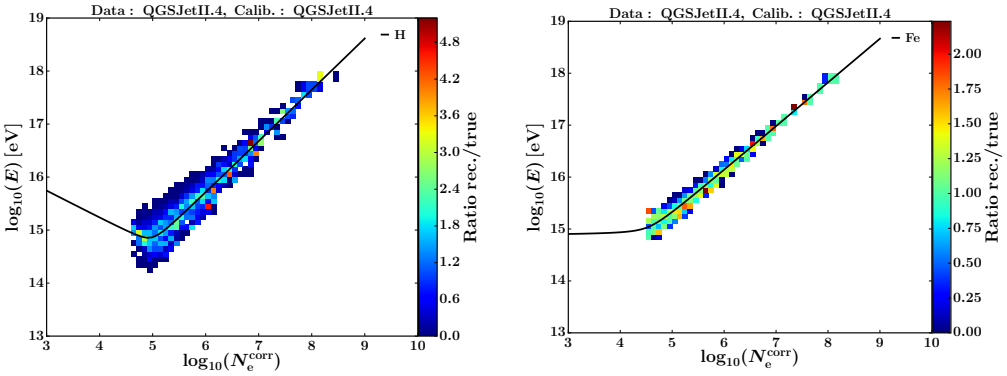

Figure 5.24.: Shown is the ratio of the $\log_{10}(E)$ vs $\log_{10}(N_e)$ distribution to the $\log_{10}(E^{true})$ vs $\log_{10}(N_e)$ distribution for events located inside KASCADE. The lines correspond to the energy fits. Left: Proton primaries. Right: Iron primaries.

for the $\log_{10}(N_e)$=5.05 bin. On the l.h.s. the development of k as a function of the measured value of D is shown. The values for D for iron and proton are taken from the corresponding fits evaluated at $\log_{10}(N_e)$=5.05. The values for proton are varied within -0.2 and +0.2 to show the influence of the reference value of D on the development of k. On the r.h.s. the change in k is translated into the change of the energy using the energy fits at the same number of electrons. In addition, the black lines show a shifted proton energy and the corresponding change of the energy with D. This is done to illustrate that the change of the energy as a function of D can be changed either by shifting one of the energy baselines or by changing the D reference values.

Which calibration should be changed depends on the other $\log_{10}(N_e)$-bins, because one would not change the values independently for each $\log_{10}(N_e)$-bin, one would rather change the calibrations which affect all $\log_{10}(N_e)$-bins. However, the changes made via one parameter can be partly compensated by the manipulation of the other parameter, which develops differently as a function of $\log_{10}(N_e)$. Already at this point, it is obvious that this procedure will not give optimal results for all $\log_{10}(N_e)$-bins and if the necessary adjustments are extensive, the fine tuning of the parameters can get quite complicated.

In addition, one cannot change the reconstruction of the energy for protons and keep it unaffected for iron. For example, keeping D_{Fe} and the energy baseline for iron fixed will only preserve the reconstructed energy for those events having a k value of 1 (which are the same events, as k=1 is defined by D = D_{Fe}). The spread will be changed and whether the mean will be changed depends on the distribution of D and its energy dependence.

For the problem at hand, this means that one will have to find a compromise between keeping the good results for iron and improving the results for protons. This could be done in the following way. The l.h.s. of figure 5.26 displays the fraction of events with a true energy above the threshold of all events above the N_e-N_μ cut. Included are only proton induced showers located inside KASCADE. Taking a look at the $\log_{10}(N_e)$=5.05 bin, it is clear that the events at large D values, i.e. small

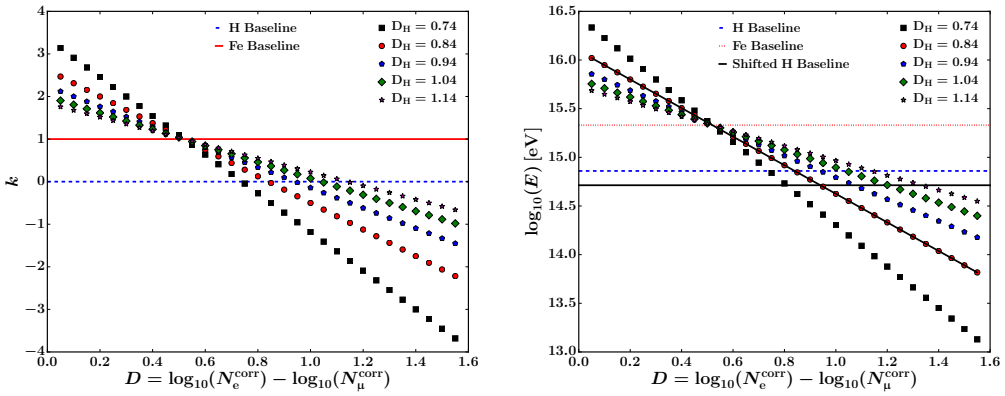

Figure 5.25.: Left: The dependency of the k parameter on D = $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ is shown for different reference values D_H, which defines at which D value k is defined to take the value 0. Right: The dependence of k on D is translated to the dependence of the energy on D for the same choices of D_H. In addition, it is demonstrated that a shift of the proton energy baseline can produce the same changes as a shifted D_H.

$\log_{10}(N_\mu)$ values, are not events with energies above the threshold. Those events are concentrated at low D values. The distribution of D for events within this $\log_{10}(N_e)$ range is shown on the r.h.s. The much larger number of events for protons is due to the lower energy events dominating almost the entire distribution. The task is to raise the assigned energy for those events in the D-region that is overlapping with the D-distribution of iron, because these are the events that have truly been induced by protons with an energy above the threshold of full efficiency. Keeping D_{Fe} fixed and shifting D_H to larger values will achieve that. However, it will also increase the energy of low-energetic events and it will lower the energy for events with k values above 1.

Therefore, what would be achieved by the above strategy is the underestimation of the proton flux being compensated to some extent, whereas the iron flux starts to be underestimated. Hence, what is not achieved is a good reconstruction of the energy, the mean energy for protons will be shifted towards higher energies, but the energy spreads will decrease for protons and for iron, as the steepness of the k parameter is decreased. At some point, the fluxes of proton and iron primaries will be affected by similar systematic uncertainties [6] (in terms of flux), which could be corrected for - independently of the mass.

This procedure has been applied in case of the standalone analyses of KASCADE-Grande [15, 17, 68]. However, it cannot be applied here [7], because the changes to the calibration of the energy and of D necessary to achieve a similar underestimation of the flux for iron and proton primaries are too extensive at low energies. Already

[6] They will be similar, but not be the same. While for iron primaries, the reduction in the spread of k results in an underestimation of the flux, the underestimation of the flux for proton primaries is not compensated by adjusting the spread, but by shifting the mean of k to a larger value.

[7] In principle, it could be applied for the events located in KASCADE-Grande, but would introduce systematic uncertainties that are different for events located in KASCADE and for events located in KASCADE-Grande.

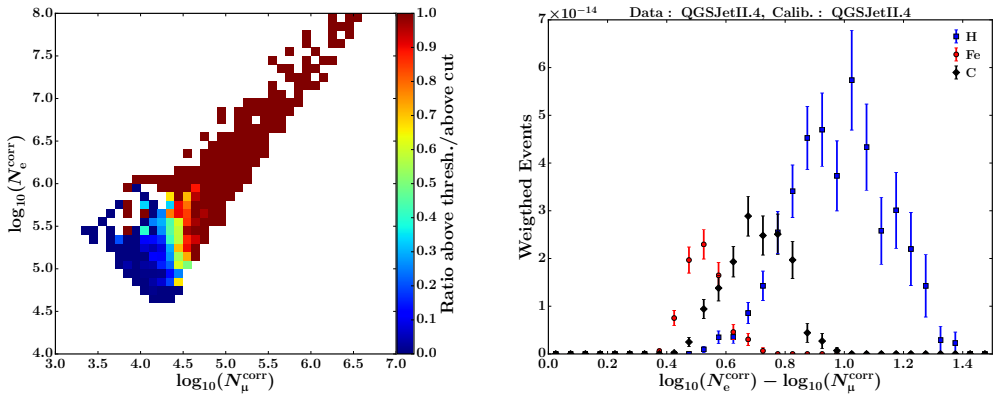

Figure 5.26.: Left: The $\log_{10}(N_e)$ vs $\log_{10}(N_\mu)$ distribution for simulated proton events with energies truly above the energy threshold is shown divided by the distribution corresponding to all events above the N_e-N_μ cut line. Right: The $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution for events with $\log_{10}(N_e)$ values within the bin of $5.0 \leq \log_{10}(N_e) \leq 5.1$. The events have been weighted so that the true spectra follow a power law with index –3.0 (they have been simulated with an index of –2.0 to save processing time.). Both: All events are located in the KASCADE array.

for helium, the spread of k is much smaller compared to the spread for protons. However, helium would be affected by changing the calibration similarly to protons, because the mean value of k for helium is close to the one for protons. Silicon primaries, although they are similar to iron regarding the spread of k, would be affected differently from iron primaries. While for iron primaries the mean of k and the mean energy for a certain number of electrons are fixed (the values for k and the energy for iron are the respective reference values), the mean value for k will be pushed to larger values for silicon, increasing the mean reconstructed energy (over)compensating the underestimation of the flux to some extent, which needs to be similar to the underestimation for iron as the goal is to remove the mass dependence. Hence, any changes to the energy calibration for protons will also affect intermediate mass primaries.

The reason why it was possible to apply this method in case of KASCADE-Grande is that for $\log_{10}(N_e) > 6.0$, i.e. roughly the threshold of full efficiency for KASCADE-Grande, the spread of N_e and N_μ and, hence, the spread of k and the energy are significantly smaller than at about 10^{15} eV. Therefore, the changes to the calibration of D and the calibration of the energy needed to achieve an approximately mass independent misreconstruction of the flux are much smaller in case of KASCADE-Grande.

To achieve better results, a different, non-linear definition of k would be needed, so that k can change faster close to zero and, therefore, one could map the much larger spread of the reconstructed energy for protons, however, this non-linear definition would have to take into account also the different spreads for intermediate mass primaries.

Instead of deriving a more complicated procedure to reconstruct the energy, the strategy for the correction of the mass dependent misreconstructions is to use a

measurement of the mass composition, which is needed anyway in order to restrict the various astrophysical models.

Summary

To summarize this section, the deviations to the results of the standalone analyses can be explained by this mass dependent misreconstruction of the energy (and by that of the flux), which has been corrected for in case of the standalone analyses. In addition, the mass dependent misreconstruction is different for events located in KASCADE and for those in KASCADE-Grande and is, therefore, in itself mass sensitive. Since this mass dependence cannot be "removed", the composition has to be known as accurately as possible to be able to retrieve the correct flux. An approach using the k parameter will be discussed in the next section, however, it is important to note again that a correction will remove the differences to the previous results, which already include corrections, but the difference between KASCADE and KASCADE-Grande (already visible for the standalone analyses) will be shown to be of a different nature.

5.3.2. Separation into Mass Groups

The goal is to separate the events into at least two mass groups resulting in two spectra, one containing light primaries, i.e. mainly H + He, the other containing the rest, which is - in terms of available simulations - C, Si, and, Fe and called heavy primaries. A separation into mass groups will serve two purposes. A composition estimate is needed to be able to distinguish between the various astrophysical models on the origin and propagation of cosmic rays. The second purpose is that an estimate of the composition is needed to correct for the mass dependent misreconstructions shown in the previous section.

With the KASCADE-Grande standalone analysis, a separation into three or more mass groups was not possible because of the reconstruction uncertainties of N_e and N_μ. With the increased accuracy of the combined analysis, this could have changed and will be investigated in the following.

The method

The procedure to assign each event to a mass group is rather simple. The mean k values for helium and carbon are displayed in figure 5.27 together with the mean values of k for silicon. In case of the separation into two mass groups, the value of k for an event is compared to the line fitted to the mean values of k for He and C. If the value of k for the event is larger or equal to the line, the event is added to the heavy mass group, otherwise it is added to the group of light primaries.

Even for simulated data, this separation is not expected to give the same results for events with cores inside KASCADE and for those located in KASCADE-Grande, because the difference of the widths of the distributions of k in each energy bin and the different mass-dependent misreconstructions are ignored in this method. While the mean values of k are - in principle - composition independent, the amount of contamination of the mass groups - due to the widths of k - is composition dependent. This describes the basic problem with this approach, reducing to the only possible statement that the events in the light mass group are predominantly of lower mass

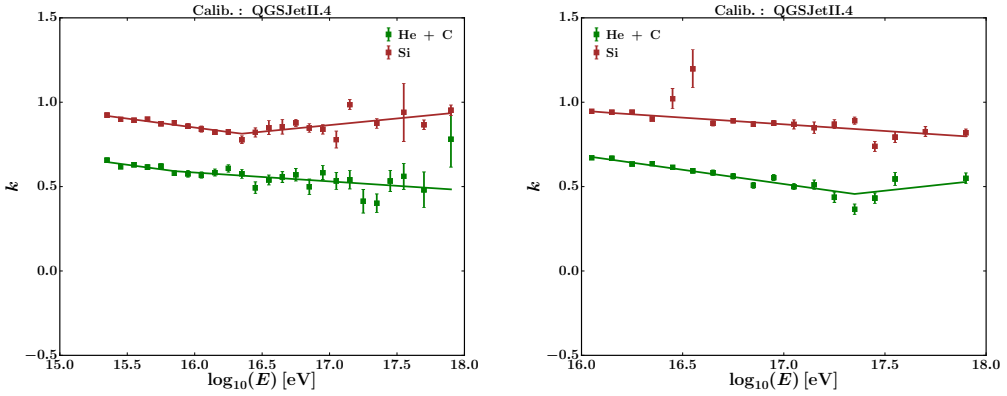

Figure 5.27.: The fitted mean k values as a function of reconstructed energy for QGSJetII4 simulations and for zenith angles between 0 and 20.70°. Left: Core inside KASCADE. Right: Core inside KASCADE-Grande

compared to the events contained in the heavy mass group. There is no way to answer e.g. the question "How heavy is heavy?" or "Is heavy equal to iron or maybe silicon?". In addition, a spectrum for the light mass group derived using QGSJetII2, which is defined as consisting mainly of H + He, would correspond to a pure proton component if nature is instead described by Epos1.99 [17]. The main reason is that Epos1.99 predicts more muons compared to QGSJetII2. Therefore, applying the calibrations obtained using QGSJetII2 to data generated with Epos1.99 results in helium primaries seeming to be medium mass primaries, hence mainly protons remain in the light mass group. Because of these difficulties, the knee identified in the data of KASCADE-Grande of the heavy mass group has always been defined as just that, a "heavy" knee, although because of the location of the structure in the energy spectrum, it is quite possibly "the iron knee". Viewed in this context, the discovery of a knee-like structure in the spectrum of heavy primaries has still been a very important result, despite the fact that it has not yet been proven to be caused by an iron component.

Testing the method with simulations

In order to test the method, the QGSJetII4 simulations have been weighted using the H4a model. In previous analyses, the tests were based on a constant, simulated composition of 20 percent of each primary using a single power law with an index of -3.0 [8]. However, as the widths of the k parameter depend on the energy and the mass of the primary particle, a more realistic composition assumption is used to test whether the method can successfully reconstruct the chosen mass groups taking the composition dependent contamination of the respective other mass group into account.

The energy spectra of the two mass groups are shown for simulated data in figure 5.28. For events located in KASCADE (l.h.s.) the large shower-to-shower fluctuations result in a significant shift of events induced by light primaries to the heavy

[8] The simulations follow a power law with index –2.0 to save processing time, while ensuring that the high energy bins are sufficiently populated too. Hence, the simulations have been weighted to represent a single power law with an index of –3.0, being closer to the measured spectrum.

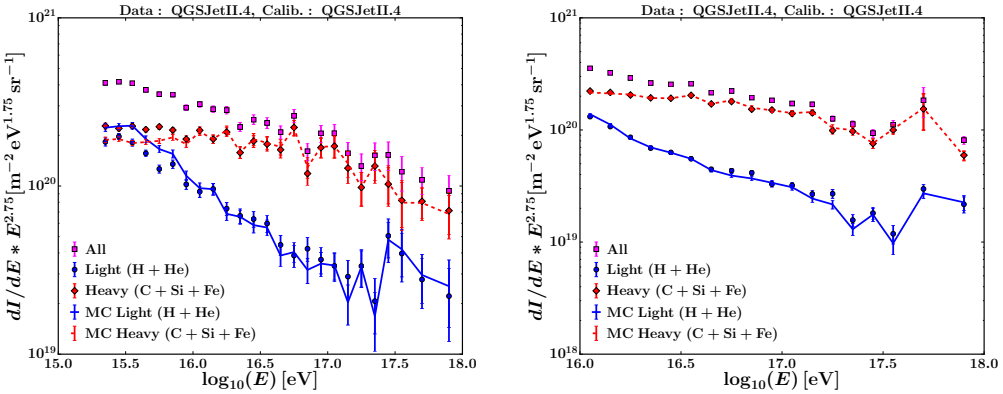

Figure 5.28.: The reconstructed test spectra of all particles and the light and heavy mass group using QGSJetII4 simulations. The simulations have been weighted according to the H4a model. Left: Events located inside KASCADE. Right: Events located in KASCADE-Grande.

mass group. Beyond 10^{16} eV, the fluxes of light and heavy primaries are reasonably well reconstructed, as is the case for KASCADE-Grande (r.h.s.). The reference spectra for the light and heavy components (labeled MC Light and MC Heavy) are obtained using the same reconstructed energy, i.e. they are affected by the same mass dependent misreconstructions of the energy, however, the true mass was used to assign each event to the corresponding mass group. Hence, the spectra labeled "Light" and "Heavy" add up to the same all-particle spectrum as the spectra labeled with a prefixed "MC" do. The only difference is that the separation of the events into the two mass groups is based on k in case of the spectra labeled without the prefix "MC", while the lines have been obtained by separating the same events based on their true mass, which is known, because simulated data was used.

The method shows compatible performances for events located in KASCADE and KASCADE-Grande, if one takes the different mass-dependent misreconstruction of the flux into account. Also, the features predicted by the H4a model, such as the ankle-like features in the spectrum of light primaries and the knee-like feature in the spectrum of heavy primaries and – at low energies – in the spectrum of light primaries have been successfully reconstructed.

The assumption of a single power law and a mixed composition of 20 percent of each primary, usually used for this test, is shown for simulated data in figure 5.29. Seeming deviations of the all-particle spectra and the lines [9] from a single power law are caused by the misreconstructions explained in the previous section. This composition assumption can be used to test one important requirement for this method, which is that the method must not introduce any features in a featureless single power law spectrum. This is not the case for events within KASCADE as is shown on the l.h.s. of figure 5.29. However, it is also revealed that this test alone does not sufficiently describe the problem of one group contaminating the other. While there is still a small shift of events from the light to the heavy mass group visible below 10^{16} eV, it is not so obvious that a dominant light component will result in a

[9] The energy spectra represented by lines again add up to the same all-particle spectrum as the spectra labeled without the prefix "MC".

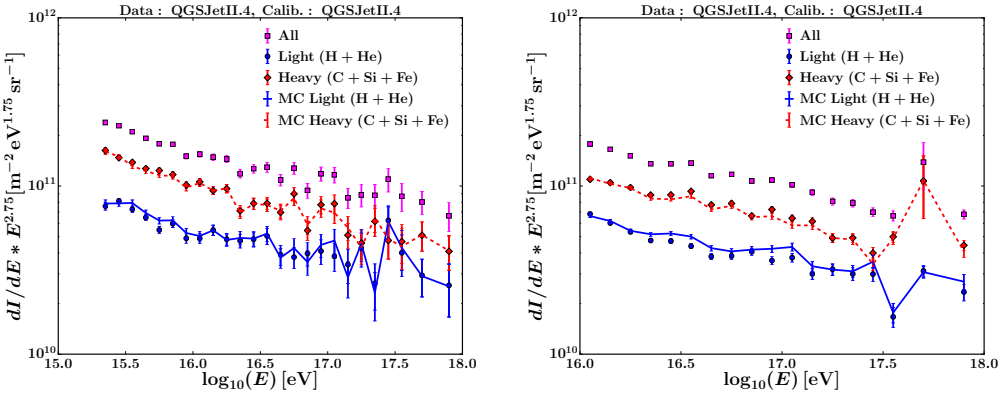

Figure 5.29.: The reconstructed test spectra of all particles and the light and heavy mass group using QGSJetII4 simulations. The simulations have been weighted to correspond to a single power law with index –3.0 and a mixed composition of 20 percent of each primary. Left: Events located inside KASCADE. Right: Events located in KASCADE-Grande.

significant migration to the heavy mass group.

On the other hand, taking a look at the r.h.s. of figures. 5.28 and 5.29, the picture is reversed to some degree. While the H4a model assumption does not reveal any problems, because the heavy component dominates the all-particle spectrum (but not as much as to significantly overestimate the light component), the single power law assumption reveals a shift of events from the light to the heavy mass group.

When trying to interpret deviations of the spectra of the light and heavy mass groups from single power laws regarding the validity or quality of the k parameter based approach, it cannot be stressed enough that at this point one must not evaluate the spectra labeled as “Light” and “Heavy” independently from the spectra labeled “MC Light” and “MC Heavy”. Judged only by the spectra labeled as “Light” and “Heavy” one might be tempted to see knee like structures in the spectrum of heavy primaries and ankle like structures in the spectrum of light primaries, however, these are mainly based on the mass dependent misreconstructions as can be seen looking at the spectra represented by the lines. Taking this into account, the remaining differences between the spectra of the light and heavy mass groups labeled with and without the prefix “MC” cannot be described as a statistically significant deviation from a single power law caused by the separation.

Nonetheless, the last point is very important to remember as long as the necessary corrections for the misreconstruction of the energy have not been applied, because, when interpreting measured data, there will be no reference to compare it against, one will have to judge the result based on what will be labeled as “Light” and “Heavy”.

Measured data and two mass groups

Unfortunately, applying the separation to measured data reveals (Fig. 5.30) that the results for events located in KASCADE and KASCADE-Grande are not consistent for real data. The muons measured for events inside KASCADE are those relatively near to the shower axis. For core positions inside KASCADE-Grande, the distance

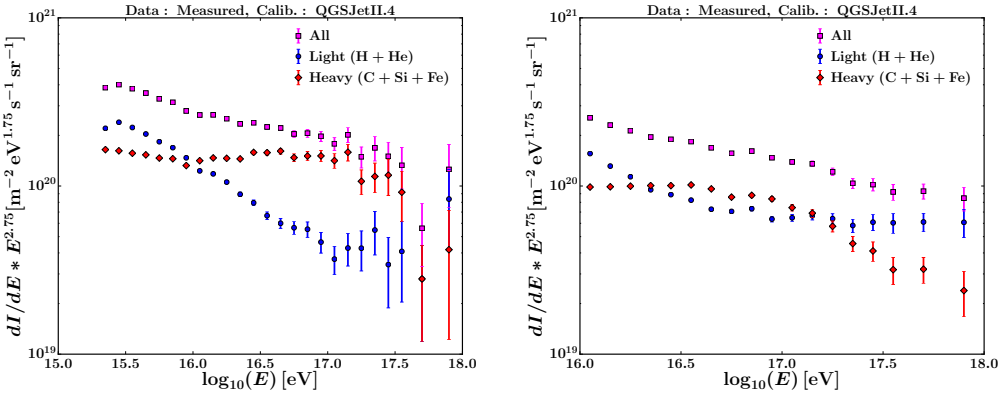

Figure 5.30.: The reconstructed flux of all particles and the light and heavy mass group for measured data. Left: Events located inside KASCADE. Right: Events located in KASCADE-Grande.

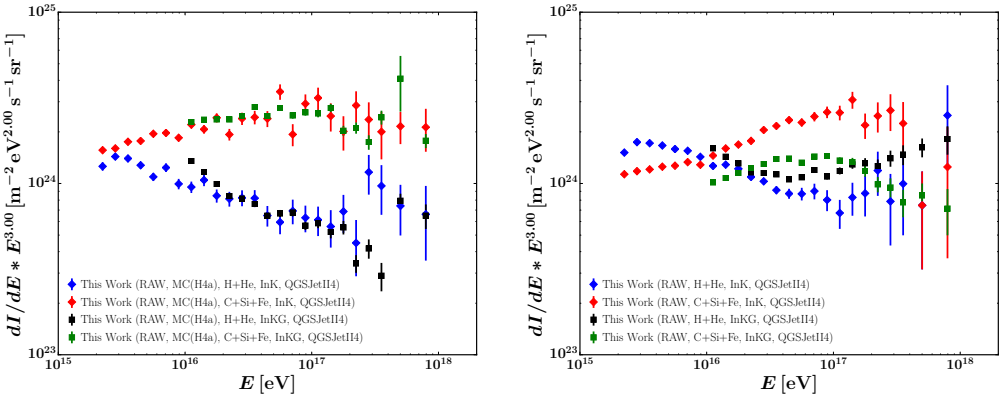

Figure 5.31.: The reconstructed flux of the light and heavy mass group for events located in KASCADE and KASCADE-Grande. Left: The events have been simulated using QGSJetII4 as the hadronic interaction model. The data has been weighted according to the H4a model. Right: The spectra have been obtained using measured data.

of the measured muons to the shower axis depends on the exact core position and the arrival direction. Therefore, the development of the muonic component, its spread and its energy distribution as a function of the distance to the shower core are very important. Obviously, the simulations do not describe nature sufficiently well. That this is the case is shown in figure 5.31. On the l.h.s. the reconstructed spectra for events located in KASCADE and KASCADE-Grande are shown for the light and heavy mass groups using simulated data. The simulated data that has been weighted according to the H4a model. There is no significant difference between the reconstructed spectra for the events located in KASCADE and KASCADE-Grande that could not be explained by the different mass-dependent misreconstructions for the events located in the respective other area. In any case, the observed differences are significantly smaller compared to what is visible on the r.h.s. of figure 5.31, which shows the measured spectra of the heavy and light mass groups for both selected areas. Possible explanations for the observed differences between simulations and

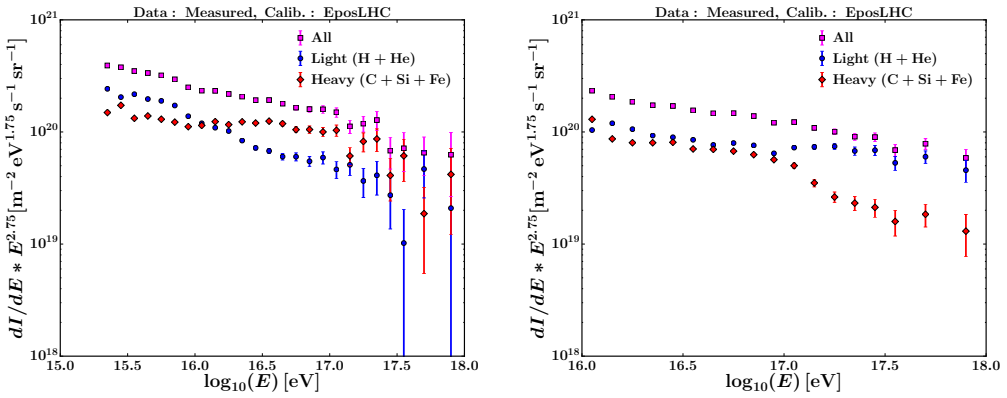

Figure 5.32.: The reconstructed flux of all particles and the light and heavy mass group for measured data using EPOS-LHC based calibrations. Left: Events located inside KASCADE. Right: Events located in KASCADE-Grande.

measured data are discussed in detail in chapter 6. The discrepancy between simulations and measurements regarding the muonic component and its lateral density function at observation level is briefly discussed in appendix C.

In addition to the different mass-dependent misreconstruction of the flux, the differently reconstructed composition, which is heavier for KASCADE, explains the larger flux found by KASCADE (see fig. 5.21) compared to KASCADE-Grande, which was already visible for the standalone analyses.

Especially in the light of the standalone results, it is important to note that both results, while showing a different absolute abundance of light and heavy primaries and being influenced by different mass-dependent systematic uncertainties, show similar features. Both heavy components exhibit a knee-like feature. Both light components seem to be influenced by the onset of another component. A definite interpretation of the results would require a working correction for the mass-dependent systematic uncertainties, however. Of course, the latter would leave the differences between simulation and measurement untouched. This is not a limitation of the method itself, therefore, the differences that are based on an inaccurate description of the measured data by the hadronic interaction models currently cannot be corrected for.

Using EposLHC based calibrations on measured data

Comparing the results based on QGSJetII4 calibrations with the results obtained using EposLHC to interpret the measurements (the latter being shown in figure 5.32), it can be seen that the composition does not differ too much at lower energies. This can already be guessed taking a look at the all-particle spectra shown in figure 5.21, where these deviate mainly at higher energies. At higher energies, however, the composition is reconstructed to be lighter in the case of EposLHC, which results in the lower flux. Again, events located in KASCADE result in a heavier composition compared to events located in KASCADE-Grande.

The differences between the two models are discussed in more detail in chapter 6.

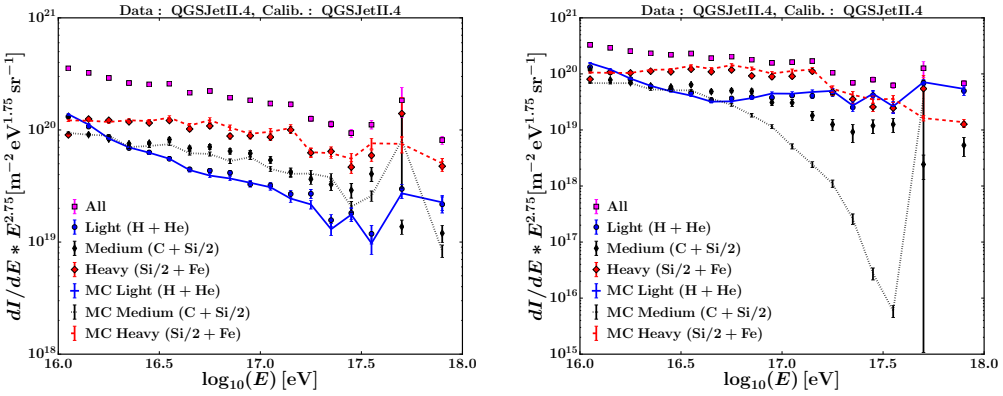

Figure 5.33.: The reconstructed flux of all particles and the light, medium, and heavy mass group for simulated data. This is shown for events located in KASCADE-Grande. Left: Events have been weighted according to the H4a model. Right: Events have been weighted according to the GST3 model.

Testing the possibility of separating three mass groups

Getting back to the problem of the mass-dependent misreconstructions, even if the simulations could describe the measured data, a more accurate description of the composition would be helpful in order to achieve a more precise correction for these misreconstructions. For example, it is still unknown if the heavy component consists mainly of carbon, silicon, or iron primaries. Without taking the widths of the distributions of k into account, it is not possible to separate protons from helium. Nonetheless, with the significantly improved reconstruction accuracy of N_e and N_μ, splitting the heavy mass group might be possible.

In order to split the group with a focus on a cleaner heavy mass data sample, a second separation line was introduced using the mean k values for silicon simulations (Fig 5.27).

The results for events located in KASCADE-Grande using QGSJetII4 based simulations are shown in figure 5.33. When using the H4a model to weight the simulated data (l.h.s.), the results show that the proton flux is well reconstructed, however, there seems to be a systematic shift of events from the heavy to the medium mass group. This can be explained by the mean k values for silicon being closer to the mean of k for iron than for carbon. Therefore, there is a shift from heavy to medium instead of a shift from medium to heavy, which one would expect from the wider k distribution for carbon in comparison to the one for iron.

The r.h.s. of figure 5.33 displays the results using the same data, but the GST3 model as a weight. The important difference to the H4a model is that this model does not predict any silicon. The shortcomings of the method are clearly visible, as the shift of events from the dominating heavy component to the medium component results in a significant overestimation of the flux of the latter, rendering the result useless for the task at hand, since there is no way to assert the correctness of the mass groups for measured data.

It is important to note that the separation into two mass groups is working. This

can be inferred from figure 5.33, because the spectrum of light primaries is well separated from the medium and heavy mass group. Recombining the medium and heavy mass groups will not affect the light mass group, of course, because they would simply be added. In addition, above 10^{17} eV, the helium component vanishes or has already vanished for the H4a and GST3 model respectively, indicating no change in the separability of the light component from the rest. The reason that the separation of the light primaries works, while the separation of the medium component from the heavy one fails, lies in the different spread of k for the relevant primaries and in the position of its mean relative to the corresponding separation line.

Summary

A separation of the events into two mass groups based on the k parameter is feasible. In order to further separate the contributions of the various elements, the widths of the k distributions have to be taken into account. Using only two mass groups does not provide enough information for a well working correction of the energy for its mass dependency. Because of this, a correction based on the k parameter is discarded in favor of a method that takes the width of the $\log_{10}(N_e)$ - $\log_{10}(N_\mu)$ distribution into account. This is subject of the next section. Nonetheless, due to the differences between measured data and simulations, the – in principle – more accurate method fails due to a tighter dependence on the simulation model used. Therefore, the two mass groups derived in this section will be used in chapter 6 for a discussion of both, the reliability of the hadronic interaction models and the validity of astrophysical models (within the current limitations shown in this and in the previous section).

5.3.3. Model based Composition Reconstruction

All previous attempts to separate the events measured with KASCADE-Grande into five mass groups (H, He, CNO, Si, Fe) either failed or had to be limited to energies way below 10^{18} eV. The reason was a limitation due to the decreasing number of events towards higher energies and due to the uncertainties in the reconstruction of N_e and N_μ giving rise to large uncertainties in the individual spectra.

Now, since both limitations have been reduced by the simultaneous use of the KASCADE and KASCADE-Grande detectors, two new approaches have been devised. The first approach takes into account the energy dependence of the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution and is discussed in this section. The second approach, which uses the total $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution only, turned out to give less reliable results and is discussed exclusively in appendix D. However, the first part of the following discussion also applies to the second method.

In order to distinguish the two methods, the first one is labeled “energy dependent $\log_{10}(N_e)$-$\log_{10}(N_\mu)$” approach, while the second method is labeled “total $\log_{10}(N_e)$-$\log_{10}(N_\mu)$” approach.

The methods are derived using QGSJetII4 as the hadronic interaction model, however, using e.g. EposLHC instead does not result in different conclusions regarding the performance of the methods.

Important note:
The GST3 and H4a models are only used to define and generate test data sets

based on simulations, which are then used in place of the measured data to test the procedures. At no point either of these models is used to be compared to measured data.

Common base of the two new approaches

KASCADE and KASCADE-Grande, being located next to each other, have been exposed to the same flux of cosmic rays. The goal of the methods described in this section is to derive a model of the flux of cosmic rays, which has to be valid for both detectors. The measured all-particle energy spectra for events in KASCADE and KASCADE-Grande and the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions for the same events are used as target distributions. Hence, the concept is to reconstruct the measured distributions with simulated data by deriving a theoretical model of the true flux and composition of cosmic rays, i.e. the simulated data, weighted according to the theoretical model, describes simultaneously the measured all-particle energy spectra and $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions for events located in KASCADE and KASCADE-Grande.

In order to assert the quality of a model, the detector response and the mass-dependent misreconstructions have to be taken into account. This is done based on simulations. At each step, the simulated data is weighted so that its true spectra for the various masses correspond to the model predictions. Applying the reconstruction procedures to these simulations will result in a simulated energy spectrum and $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution for KASCADE and for KASCADE-Grande, which have to be compatible with the measured counterparts in case that the model matches the spectrum and composition realized in nature.

The comparison of the simulated to the measured (or test-data) distributions is based on the implementation of a χ^2 test for comparing - possibly weighted - histograms, as it is included in the ROOT framework [10].

The methods are developed using simulated data also as the target (which will be substituted by the measurements in the end), as only then the true composition is known, which is needed to study the performance of the procedures. Due to the available number of simulated events being limited, it is not feasible to use independent data sets for the creation of an artificial "measured" spectrum and composition and for the reconstruction of it. In order to mitigate this limitation, 10 percent of the total data set have been randomly left out of the sample used to estimate the flux and composition. 10 percent, however, is a compromise to keep a high enough number of events at the highest energies.

Both methods are based on a few common assumptions/considerations:

1) The all-particle spectrum is composed of several populations of cosmic rays. Each of the constituents of a population is defined by one power-law with an exponential cutoff. The cutoff energy is defined for the proton component of a population, the cutoff of additional components is given by their charge multiplied by the proton cutoff (see section 2.3.2) [11].

[10] Chi2Test method of the TH1 class of ROOT [75]

[11] Of course, this assumption might not be correct for an extragalactic population, because the spectrum might be modified at the highest energies by the GZK-effect or photo-disintegration of

2) The summed contribution of all populations to the flux of a certain element is restricted by the fluxes measured by CREAM [33, 76] (see section 2.3.2). In addition, since the cutoff energies are defined by the proton component, the possibility of a population with a cutoff below 10^{15} eV has to be taken into account, as it might give additional contributions of heavier primaries in the KASCADE energy range. These would not be covered if only proton cutoffs at energies within the combined KASCADE and KASCADE-Grande energy range are considered. In order to account for this, a first population is introduced based on the CREAM data. The cutoff of this population is varied from 10^{14} eV to 4×10^{15} eV and enters as a fixed parameter into the fitting procedure, which is started independently for each "CREAM-cutoff". The start values are defined by a fit to the CREAM fluxes, using the fixed cutoff energy. However, while fitting the additional populations, the indices and normalizations of the underlying power-laws are non-fixed parameters, keeping only the "CREAM-cutoff" fixed. This is done to ensure that the contributions of the additional populations do not result in a systematic overestimation of the flux at energies covered by CREAM. Comparing the solutions based on different "CREAM-cutoffs", the one that best describes the four measured distributions is chosen as the final one. This way, the combined analysis may even be able to give hints at another population with a proton cutoff below its covered energy range.

3) The CREAM results for seven masses are considered, namely H, He, C, O, Mg, Si, and Fe. However, the simulated data set covers only H, He, C, Si, and Fe, therefore, oxygen is considered to be described by carbon-simulations and magnesium is assumed to be described by silicon-simulations. Since there is no way to disentangle these combinations (even by means of corresponding simulations for oxygen and magnesium), C + O, and Mg + Si enter the fitting procedure with a common power-law index, assuming that the acceleration and propagation mechanisms do not produce too different spectra for these combinations, apart from individual absolute fluxes.

Energy dependent $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ approach: Implementation and simulation based test

The first method is based on fitting the fractions of the five primaries by comparing the sum of their D = $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions with the measured [12] one, splitting the data into various energy ranges.

The procedure to obtain these fractions is the following. First, a single power law is fitted to the all-particle energy spectrum of the measured or – as in this case – of the test data. The simulated data is weighted according to this fit, because the simulated data has been generated assuming a power law with index –2, hence, the high energetic primaries in each energy bin would have a weight too high compared to the measured (or test) data. This power law is used as the same reference for all five primaries, which enter into the next step with the same absolute abundance. The distribution of D is then obtained for each primary and energy range. These distributions may have a different number of events for the five primaries and the

nuclei. However, since KASCADE and KASCADE-Grande did not cover this energy range, this is of no concern here. In any case, the fitted contributions of the individual primaries should only be interpreted within the covered energy range.

[12] Again, "measured" may correspond to simulation based test data

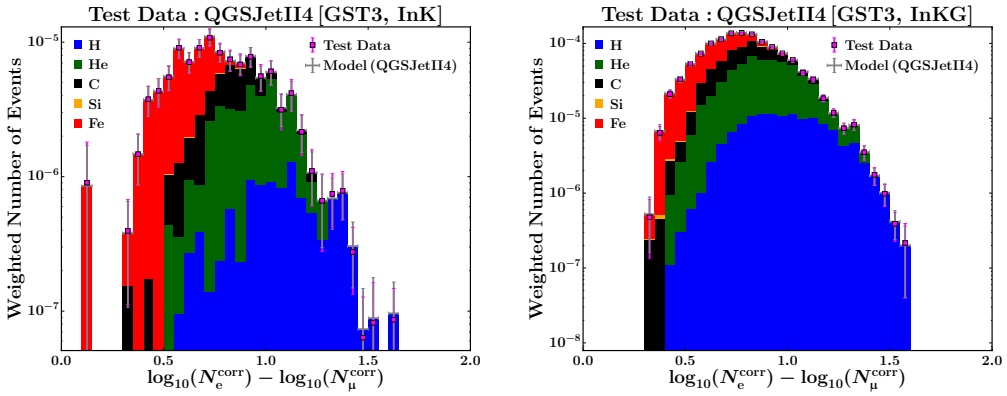

Figure 5.34.: The $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions for reconstructed energies within 10^{16} and $10^{16.1}$ eV. "Test Data" shows the distribution for the data, which is based on QGSJetII4 simulations weighted according to the GST3 model. "Model" represents the sum of the distributions of the five primaries. The relative abundance of each primary has been fitted comparing the distribution labeled "Model" to the distribution labeled "Test Data". Left: Core inside KASCADE. Right: Core inside KASCADE-Grande

same energy bin. The reason is that the same absolute abundance is defined by the true, i.e. simulated energy spectra, while the reconstructed energy is used to split the events into ranges of energy. Hence, the mass dependent misreconstructions of the reconstructed energy might result in a number of events per energy bin different for the five primaries. However, this is not a problem, because the same misreconstructions apply for measured data (assuming that the hadronic interaction model describes nature well enough). In order to adjust the relative contribution of the five primaries to the corresponding energy bin, their D distributions are scaled so that the sum of the five distributions matches the measured distribution of D. The scaling is applied to the effective number of events of a certain primary in the energy bin keeping the shape of the distribution of D untouched.

The result of this procedure is shown for reconstructed energies between 10^{16} eV and $10^{16.1}$ eV in figure 5.34 on the l.h.s. for events located in KASCADE and on the r.h.s. for events located in KASCADE-Grande. The markers labeled "Test Data" correspond to the QGSJetII4 based simulations weighted according to the GST3 model to define a composition used as a substitute for the data really measured. Labeled as "Model" is the sum of the contributions of the five primaries that have been weighted according to the above described procedure. The distribution of D for the test data is well reconstructed, however, in case of KASCADE-Grande, a small contribution of Si remains, although the GST3 model does not predict any contribution of Si.

In figure 5.35 the results of the fitted relative abundances of the five elements for several energy bins are compiled. The fractions are given relative to the contribution of protons. Circles show the resulting relative fractions for events located inside KASCADE, squares correspond to the results for events with core positions inside KASCADE-Grande. The solid lines are the true relative fractions according to the

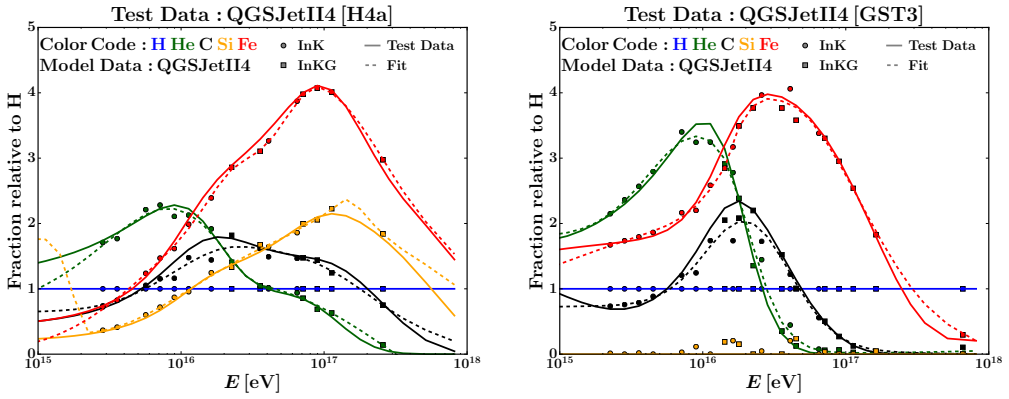

Figure 5.35.: The fitted fractions as a function of energy are shown relative to the proton contributions for QGSJetII4 simulations. Circles correspond to events located in KASCADE, squares show the results using events with core positions inside KASCADE-Grande. Solid lines describe the true fractions as they are predicted by the model used to define the test data. Dashed lines correspond to fits to the reconstructed fractions. Left: The H4a model was used to define the test composition. Right: The GST3 model was used to define the test composition.

H4a (l.h.s.) and GST3 (r.h.s.) models, representing the test data, i.e. the substitute for actually measured data. The dashed lines are sums of fitted gaussian-like distributions. These may have a different shape parameter to the left and the right side of the mean. In addition, negative contributions, i.e. subtractions, are allowed. This procedure could be improved by smoothing sharp changes as it can be observed at about 10^{17} eV in the estimated fractions of silicon in case of the H4a model based test data. In case of the GST3 model, a contribution of silicon primaries has been successfully removed from the initial composition of 20 % of each of the five primaries. The decision to ignore the remaining contributions of Si to single energy bins is based on the assumption that a spiky contribution of a primary to single or very limited ranges of energy bins is not likely to have an astrophysical origin.

Once these relative fractions have been obtained, the strategy is to iteratively translate these into particle fluxes based on the aforementioned populations. As a first step, a proton flux is fitted using the four measured distributions. The fluxes of the remaining particles are obtained by using the relative fractions. The result of the procedure is shown in figure 5.36. Labeled as “Test Data” are the reconstructed spectra based on the composition according to the GST3 model, which is shown as solid lines labeled “Test Truth”. Labeled as “Model” are the reconstructed spectra based on the model being fit. Dashed lines labeled “Model Estim.” correspond to the estimate of the model based on the fitted populations. Dash dotted lines labeled “Model Estim. (Frac.)” correspond to the model predictions that are based on the estimated proton flux and the fractions of the primary relative to the energy spectrum of protons shown in figure 5.35. On the l.h.s. of figure 5.36 only the proton flux is described by the population based fit. The description of the true proton flux, i.e. the flux of protons predicted by the GST3 model is fairly accurate within the covered energy range, however, at higher energies, the fitting procedure is dominated by the comparison of the energy distributions, as the main contributions to the to-

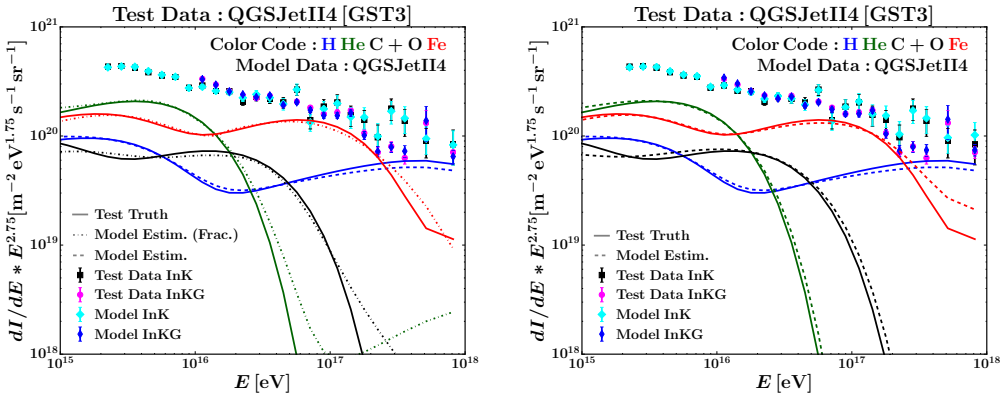

Figure 5.36.: Markers labeled with "Test Data" correspond to the reconstructed all-particle flux obtained by weighting the QGSJetII4 based simulations according to the predictions of the GST3 model, which are shown as solid lines (Test Truth). Markers labeled with "Model" show the reconstructed flux based on the current prediction of the fitted model, which is based on the dashed and dash dotted lines. The former are based on the fitted populations, the latter are based on the estimated flux of protons and the fractions of the corresponding primary relative to the contribution of protons. This is shown on the l.h.s. for a fitted proton flux, upon which the spectra of the other primaries are based. On the r.h.s. all spectra have been substituted by population based fits. The "CREAM-cutoff" was fixed at $10^{14.05}$ eV.

tal $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions are predominantly of lower energy. Therefore, the overestimation of the iron fraction (Fig. 5.34, r.h.s.) enforces a lower contribution of proton primaries. An attempt to remedy this problem could be to split the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions into two parts, one consisting of events below e.g. 10^{17} eV, the other containing the rest. However, due to the low number of events above 10^{17} eV in case of KASCADE, the resulting distributions would result in too many bins having a too low content for the χ^2 method used. Possible ways to avoid that problem could be to use only KASCADE-Grande at the highest energies or to try a differently defined comparison of the histograms that works better with low content bins. In principle, one might also start to merge $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ bins, however, at high energies only a few bins would remain, effectively resulting in the loss of most of the information regarding the primary that contributes to these bins.

After the estimate of the proton flux has been obtained, the estimates of the remaining four primaries, which are - at this stage - based on the proton flux and the relative fractions, are iteratively substituted by fitting their corresponding populations. First, the helium estimate will be fitted by keeping the proton flux, the current estimates for C, Si, and Fe, and the cutoffs of the proton populations fixed. Therefore, the new estimate for the flux of helium primaries is based on a fit of the normalizations and slopes of the populations with cutoffs fixed and given by the corresponding proton cutoff multiplied by two, i.e. the charge of helium primaries. This procedure is performed next for C, then for Si, and finally for Fe. The final result is shown on the r.h.s. of figure 5.36. The regularization introduced by fitting the population based fluxes (given by dashed lines), results in a better description

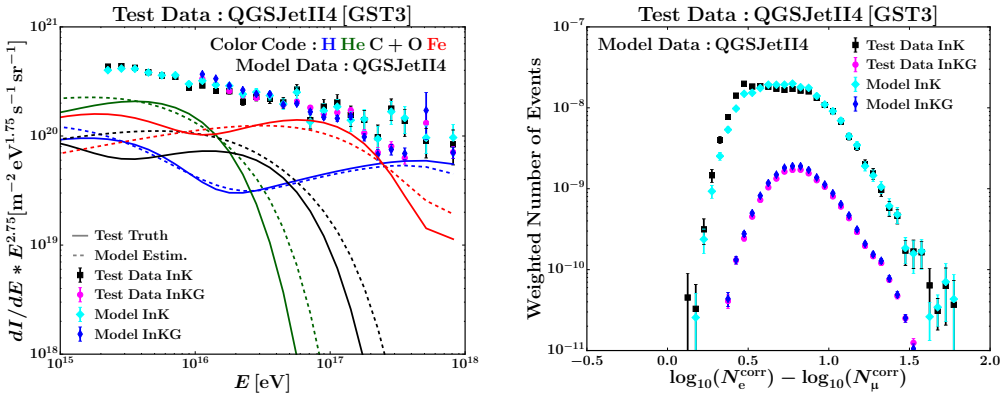

Figure 5.37.: The same information as shown on the r.h.s. of figure 5.36 is shown on the l.h.s. for a "CREAM-cutoff" fixed at $10^{15.40}$ eV. The corresponding total $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions are shown on the r.h.s. separately for events located in KASCADE and KASCADE-Grande. The test composition is set according to the GST3 model.

of the helium and carbon fluxes, however, the overestimation of the iron flux at the highest covered energies is enhanced. On the other hand, the additionally required iron population has been correctly introduced. The solid and dashed lines show the contributions of the various primaries to the all-particle spectra as predicted by the GST3 model and the model being fit, respectively. The all-particle spectra (markers) correspond to the total flux of all primaries as predicted by the respective model after the detector and reconstruction effects have been folded in by using QGSJetII4 based generated data.

The "CREAM-cutoff" found to fit the test data best is at $10^{14.05}$ eV. However, the cutoff defined for the first population of the GST3 model is at about $10^{14.08}$ eV. A possible improvement of the procedure could be to keep the current 0.05 step in $\log_{10}(E)$ to save processing time, but to add a second run using a smaller step size and a narrow range around - in this case - $10^{14.05}$ eV.

While the study of possible systematic uncertainties (also considering the fact that the data is a subset of the test data) has yet to be performed, it is for the first time that a study of the composition, based on KASCADE-Grande, could - in principle - be performed in this detail. This has only become feasible because of the significant improvement of the reconstruction of N_e and N_μ due to the combined analysis.

It has been mentioned that the CREAM based first population could reveal the necessity of an additional population with a proton cutoff below the knee. In case of the GST3 model, which predicts such a component, this can be checked by taking a look at the results based on a higher "CREAM-cutoff". In figure 5.37 the results with a cutoff of the first population at $10^{15.40}$ eV are shown. The flux of protons is reasonably well reconstructed, which is visible also in the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions shown on the r.h.s. of the same figure, as protons dominate mainly the right tail of the distributions. Keeping in mind that the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions are dominated by the low-energetic events, the deviations in case of KASCADE can be explained by the overestimation of the carbon flux and the underestimation of the iron flux. The first one results in the overestimation of the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$

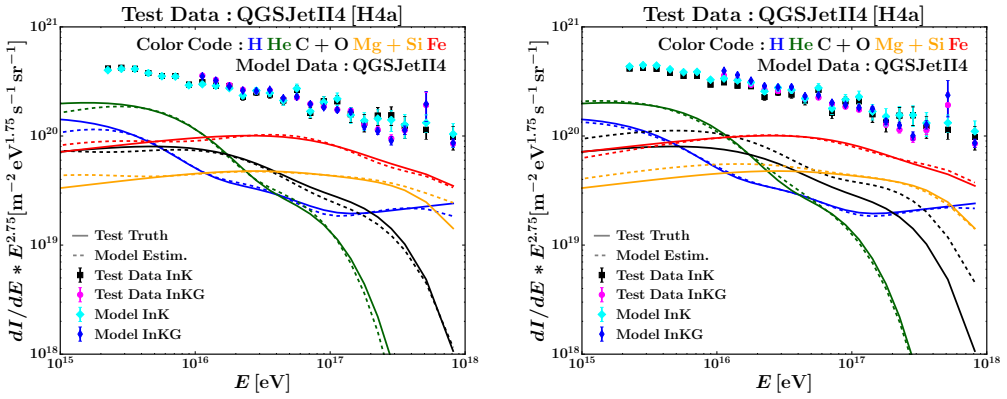

Figure 5.38.: The same information as shown on the r.h.s. of figure 5.36 is shown on the l.h.s. for a "CREAM-cutoff" fixed at $10^{14.10}$ eV and at the r.h.s. for a "CREAM-cutoff" fixed at $10^{15.40}$ eV. Instead of the GST3 model, the H4a model was used as the test composition.

distribution between 0.5 and 1.0, the latter explains the deviation below 0.5. For events located in KASCADE-Grande, the flux of proton, helium, and iron is not too different from the truth in the relevant energy range. Therefore, the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution is reasonably well described below 0.5 and beyond 1.0. The overestimation of the carbon flux, however, results in the visible overestimation in between. In case of the GST3 model, the necessity of a population below the covered energy range was successfully recognized.

As will be discussed next, the decision to use the upper end of the CREAM spectrum as a reference for the first population has a strong impact on the result, should the assumption be wrong. This can be tested by using e.g. the H4a model instead of the GST3 model, which seems to be tuned to the entire CREAM data instead of the upper end, i.e. above the kinks visible in their data, only.

The results for a "CREAM-cutoff" at $10^{14.1}$ eV are shown on the l.h.s. of figure 5.38. On the r.h.s. the results for a cutoff of the first population at $10^{15.4}$ eV is displayed. Although the cutoff of the first population is at $10^{15.6}$ eV according to the H4a model, fixing the cutoff at $10^{14.1}$ eV results in a better description of the composition and the all-particle flux. The reason why an early cutoff in the first population can still result in a reasonable description at higher energies is that the cutoff is early enough, so that the additional populations can still dominate the flux within the covered energy range without being influenced too much by the low energy population, even in case of iron. Hence, the influence of the first population is effectively suppressed, which is a problem regarding the interpretation of this early "CREAM-cutoff". Since the individual components of a population are connected to its proton component by charge dependent cutoffs, a contribution of medium or heavy mass primaries at low energies can only be modeled, if the contribution of protons that corresponds to the population is allowed to have a cutoff below the energy range covered by KASCADE. In order to take this possibility into account, the "CREAM-cutoff" was introduced to the model. A shift of the "CREAM-cutoff" to energies below 10^{15} eV implies that there is an additional population with a proton cutoff in the energy range between what is covered by CREAM and KASCADE. However, in case of the H4a model, the

"CREAM-cutoff" has been shifted to low energies, because the assumptions regarding the CREAM data (see section 2.3.2) are incompatible to the assumptions made here for the fit of the first population. Hence, the interpretation of the early cutoff of the first population as a hint at an additional low-energy population could be wrong, if the assumptions regarding the CREAM data should prove to be wrong.

In order to understand why the "CREAM-cutoff" is chosen as it is, instead of one closer to the true one, the r.h.s. of Fig 5.38 can be used. The flux for proton and helium are very well reconstructed. The reason is that the kink in the CREAM data is below the CREAM energy range for protons and helium, therefore, the H4a model is tuned to the same low energy data. In addition, during the fitting of the proton and helium fluxes, the "active" estimates for the flux of the heavier primaries are still based on the well estimated proton flux and the fitted fractions shown on the l.h.s. of figure 5.35. The remaining question is, therefore, why the iron flux is so well described (within the covered energy range), while the fluxes for carbon and silicon are overestimated. The reason is connected to the flux of carbon and silicon actually being the sum of C+O and Mg+Si, respectively. Therefore, the fit of two (e.g. C+O or Mg+Si) fluxes are condensed into the fit of only one flux for the comparison of the histograms, however, the comparison to the CREAM results enter individually. Hence, in the current implementation, the influence of the CREAM data is too large, if the assumption regarding it is invalid and the cutoff is fixed at an energy too high for the first population to be sub-dominant at higher energies.

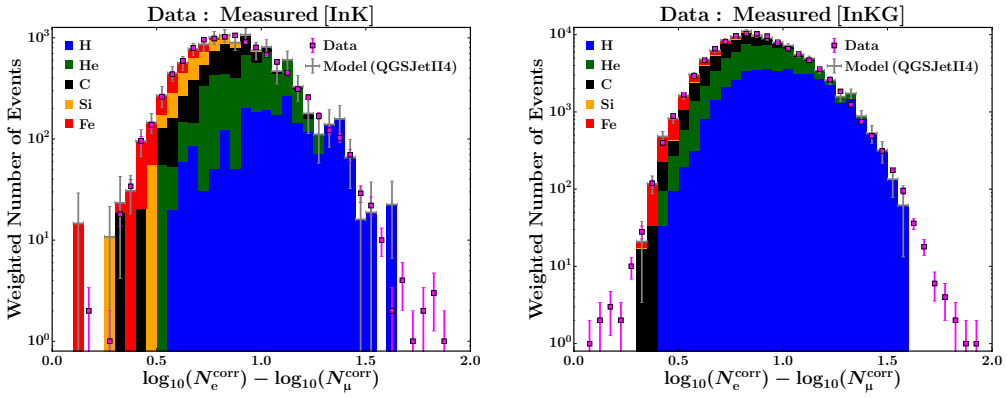

Figure 5.39.: The $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions as measured and as predicted by the model being fit are shown for reconstructed energies within 10^{16} and $10^{16.1}$ eV. "Data" shows the target distribution, which is obtained from measured data (each measured event has the weight 1). "Model" represents the sum of the contributions of the five primaries based on QGSJetII4 simulations. Left: Core inside KASCADE. Right: Core inside KASCADE-Grande.

Applied to measured data

Taking the discussion above into account, the procedure is applied to the measured data. However, as shown in figure 5.39, the discrepancy in the reconstructed number of muons for events located in KASCADE and KASCADE-Grande results in an energy-binned fraction estimation which fails due to a too different composition

result. As shown in the last chapter, the estimate for events inside KASCADE is heavier than it is for events located in KASCADE-Grande. This is also the case here. While for showers hitting KASCADE-Grande no contribution of silicon was found, a significant contribution of silicon is estimated for the showers with core positions within KASCADE, while the relative contribution of protons is smaller.

Interesting is that the tails of the measured distributions are not populated by the simulations. This can have several reasons. Maybe the number of simulated events is not high enough for the low statistics bins to be populated, or the shower-to-shower fluctuations are underestimated in the simulations, or the reconstructed composition is yet too heavy, therefore, the measured $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution corresponds to a lower energy bin with larger fluctuations (which could also be achieved by an underestimation of N_e in the simulations). In any case, since both measured distributions are reasonably well described (ignoring the tails), it seems possible to get a composition estimate for KASCADE and KASCADE-Grande individually, but not a consistent one for the combined array.

The approach described above has some drawbacks. For example, it depends on the quality of the estimates of the fractions of the contributions of the various primaries. Already a small overestimation of the iron flux at higher energies resulted in the underestimation of the proton flux, which could not recover, as it is fitted only once and the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution is not influenced enough by the high energy events. The positions of the cutoffs, which are fixed by the fit for protons, are not adjusted during the fits for the other primaries. The whole procedure is, therefore, based on the assumption that the fit of the fractions is already a good first estimate of the composition and that the main concern is to translate these fractions into fluxes while regularizing non-physical fluctuations.

Another, simpler approach does not rely on this assumption. Instead, the entire populations are fitted at once. However, it turned out that this alternative approach is too flexible in order to give similarly reliable results. Nonetheless, the results are useful to discuss the influence of a muon lateral density function, which describes simulated data, but which is too flat compared to measured data. Hence, a detailed discussion of the second approach is given in appendix D where it is used for a discussion on the influence of the lateral density function.

Summary of the model based composition reconstruction

Two methods for the determination of a consistent composition estimate in the energy range of 10^{15} eV to 10^{18} eV have been derived using a model based approach. The first method gives stable results, as long as the measurements are actually described by the simulations. Since this is not the case, a second approach was tested, which does not depend as strictly on the simulations. However, the second method does not provide stable results, as it turns out to be too flexible (see appendix D for details). Therefore, the combined analysis is capable of producing detailed composition reconstructions once the simulations have improved enough to describe nature sufficiently well, so that the first method can be used.

Since the latter is not the case, it is not so important at this point whether the assumptions regarding the CREAM data and the population based approach turn out to be valid or not. What is important right now, especially regarding KCDC

(see Part 2), is that KASCADE and KASCADE-Grande, when combined, provide the basis to reconstruct the composition with a precision not reached by any other experiment in the covered energy range up to date.

Regarding the reconstruction of the energy spectrum and mass composition this precision cannot be fully utilized at the moment due to unknown systematic differences between the interaction models and nature, however, it is this precision that makes studying the differences among the models possible [13].

5.3.4. Summary

The reconstruction of the energy spectrum revealed the need to correct for composition dependent misreconstructions, which, in addition, depend slightly on whether the event was located in KASCADE or KASCADE-Grande. A detailed reconstruction of the composition could have been used to correct for the misreconstruction, however, the composition cannot be determined at present due to the unknown systematic differences of the interaction models and nature.

Using the standard KASCADE-Grande method to distinguish between light and heavy primaries – based on the correlation of N_e and N_μ via the parameter k – was found to yield different results for events located in KASCADE and for events located in KASCADE-Grande. This is partly based on the different systematic misreconstructions of the fluxes of the various primaries and the fact that only the mean values of the parameter k are used. Therefore, the different widths of the k-distributions for different primaries as well as for KASCADE and KASCADE-Grande due to their different reconstruction accuracies of N_e and N_μ are ignored. The different widths give rise to different degrees of contamination of the respectively other mass group.

This part can be studied using simulations, however, the dominating contribution to the difference between KASCADE and KASCADE-Grande is not based on inaccuracies or misreconstructions due to the method applied and is only revealed when using measured data. It is based on the fact that simulations are incompatible with measured data. This results in the measured composition being heavier for events located in KASCADE compared to events located in KASCADE-Grande. Since heavier primaries are assigned a larger energy, this behavior explains the systematically higher flux reconstructed with KASCADE compared to KASCADE-Grande that was already visible in the standalone analyses, which used two different reconstruction procedures.

It was argued (mainly in appendix D) that a too flat muon lateral density function, which has a fixed slope based on simulations, could lead to a systematically higher reconstructed number of muons for events within KASCADE compared to events located in KASCADE-Grande. Using the constant-intensity-cut method, it is revealed in appendix C that indeed a different number of muons is reconstructed depending on the energy of the primary and the distance of the muon detectors to the shower. This behavior is not visible for simulated data, which means that the chosen parameterization of the lateral density function works for simulated data.

[13] Although this precision does not help determining the unknown absolute energy and mass scale.

Even in absence of the systematic differences between the simulations and the measured data, a more detailed composition measurement is needed for an accurate estimation of the energy spectrum and mass composition than can be provided by using the k-parameter based approach. Therefore, although a consistent reconstruction of the composition is not possible at present, the analysis' method has been further developed in order to be ready to be applied once the simulations have evolved enough to give predictions in better agreement with measured data.

The main approach was to use the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions for several energy ranges in order to find the fractions of H, He, C, Si, and Fe primaries needed to reproduce the measured distributions for KASCADE and KASCADE-Grande by simulated data. As a mean to regularize the result, a population based model estimation was chosen together with an initial composition based at low energies on CREAM data. By folding the detector response and reconstruction effects into the model of the true flux, a comparison of weighted simulations to the measured data was performed, simultaneously using the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions as reconstructed for events located in KASCADE and KASCADE-Grande together with their reconstructed all-particle fluxes.

It was found that this method works well. This has been tested by generating artificial "measured" data by using the H4a and GST3 composition and absolute flux assumptions.

However, while the method gives satisfactory results for simulations, the measured fractions of the different primaries are too different for KASCADE and KASCADE-Grande in order to provide a reasonable result for measured data. Therefore, a second method was devised which was designed to be less strongly coupled to the simulations. Lacking the strong guidance based on the energy dependent fractions of the first method, this approach proves to be too flexible. For simulations, reasonable results have been obtained only in case of the GST3 model.

Concluding, the combined analysis has led to a quality of the reconstruction that the flux and composition can be determined with a much higher accuracy than what has been provided up to now in the covered energy range. However, since the k-parameter based approach does not provide enough detail on the composition, a different method had to be devised, which has a tighter coupling to the simulations. The latter is, however, the main reason why a consistent composition estimate in the energy range of 10^{15} eV to 10^{18} eV is currently not possible for all the measured events located in KASCADE and KASCADE-Grande simultaneously.

Nonetheless, it has been demonstrated what will be possible in the future regarding the estimation of the energy spectrum and mass composition and what is possible even now regarding the tests of the validity of hadronic interaction models. The latter will take place in the next chapter.

6. Discussion

The goal of the combined analysis of KASCADE and KASCADE-Grande was to improve the reconstruction of the shower parameters in order to get a consistent and more detailed estimate of the cosmic ray composition from 10^{15} eV to 10^{18} eV. This goal has been reached to a large extent. The combined reconstruction has led to a significantly improved estimate of the main shower variables, namely N_e and N_μ. Ignoring the limitations imbued by the hadronic interaction models, the combined analysis was shown to be capable of reconstructing the composition and energy spectrum of cosmic rays much more accurately than the standalone analyses. The latter is especially true for KASCADE-Grande, KASCADE benefits mainly from an increased energy range.

Before the actual discussion of the results starts a brief outline is given, which might help in following the more detailed discussion afterwards.
The first part of the discussion focuses on the differences between the hadronic interaction models EposLHC and QGSJetII4. It is shown that the predictions of the two models are compatible for light primaries, but deviate from one another regarding heavy primaries.

The observed differences between the events located in KASCADE and the events located in KASCADE-Grande are caused by deficiencies in the description of hadronic interactions by the two models. It is discussed how far the differences can be explained using the available simulations and how much has to be attributed to differences between simulations and measurement. It is shown, that possible deviations visible within the simulations cannot explain the differences between the events located in the areas covered by the two detectors.

There are two ways how the discrepancy between KASCADE and KASCADE-Grande can be related to differences in the prediction of the simulations and measurements. One way is related to the difference of the predicted mass scale to the one realized in nature depending on which part of the muonic component is measured, i.e. the one close to the core or farther away from it. This would shift events from one mass group to the other mass group depending on which detector array is considered as the reference.

The other way how this discrepancy may be explained is a – possibly mass dependent – difference of the predicted energy scale to the one realized in nature, again depending on which part of the muonic component is measured. This could result in a general shift of the reconstructed energy depending on where the core of the event is located relative to the muon detectors of KASCADE.

Both possibilities are discussed. At the end, taking the discussion of the above into account, a short discussion of the findings of other experiments is given and the results of the combined analysis are compared to the predictions of astrophysical models.

6.1. Hadronic Interaction Models

Although the results regarding the energy spectrum and mass composition of cosmic rays differ for events located in KASCADE (InK) and for events located in KASCADE-Grande (InKG), one very important observation can be made, where both detectors show a consistent picture. The following discussion of this observation may help to shed some light on the origin of the differences between QGSJetII4 and EposLHC.

Figure 6.1 shows the energy spectra for all particles and the light and heavy mass group reconstructed for QGSJetII4 (markers) and EposLHC (dashed lines) based simulated data. Prior to the reconstruction, the simulations have been weighted, so that their true spectra match the predictions of the H4a model. On the l.h.s. this is shown for events located in KASCADE and on the r.h.s. for events located in KASCADE-Grande. The plots show how the true spectrum is modified by mass-dependent misreconstructions and contaminations of the respective other mass group. Within the statistical uncertainties, which are shown for better readability only in case of QGSJetII4, both models exhibit comparable effects. Especially at high energies, the tenfold oversampling [1] of the simulated showers may yield larger differences, if a strongly fluctuating event is present several times in one data set. Nonetheless, the events located in KASCADE and KASCADE-Grande, respectively, give similar results reproducing the input.

Comparing the same information obtained for measured data (Fig. 6.2) reveals that EposLHC and QGSJetII4 yield essentially the same spectra for light elements. Although they are not exactly the same, of course, this result presents good evidence that the difference between the all-particle spectra is almost exclusively caused by the heavy mass group. While these differences can be attributed to differences in the modeling of hadronic interactions, both models give similar deviations when comparing the results for events located in KASCADE to the results obtained for events located in KASCADE-Grande.

The agreement on the energy spectra of light primaries is quite interesting, because the models are tuned to the same data, which is mainly based on proton-proton collisions. A possible explanation is, therefore, a different extrapolation of proton-proton to proton-nucleus and nucleus-nucleus interactions. It was already mentioned

[1] Oversampling is a method used to increase the effectively available number of simulated events, by reusing the same simulated air-showers several times while distributing them randomly at different positions of the detector area.

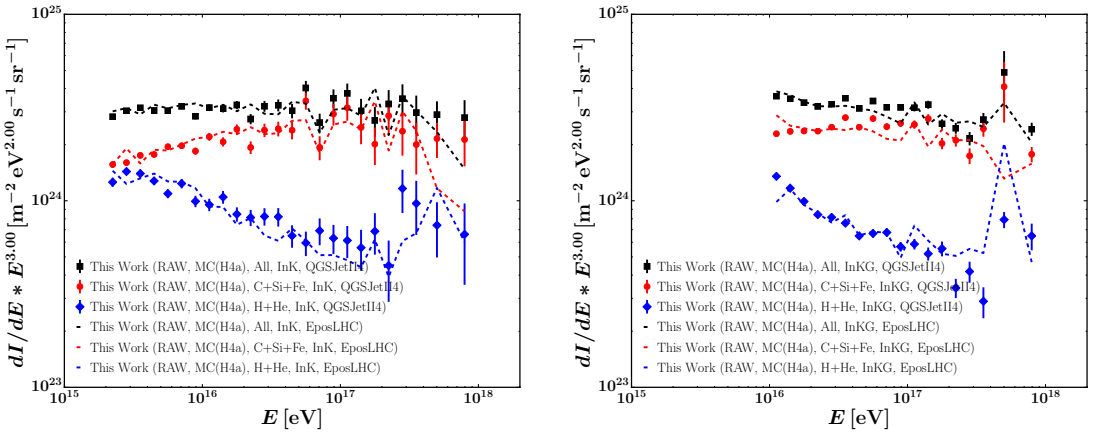

Figure 6.1.: The energy spectrum for all particles and the light and heavy mass groups are shown for simulated events. Prior to the reconstruction, the simulations have been weighted, so that their true spectra match the predictions of the H4a model. Dashed lines correspond to simulations based on EposLHC. Markers show the spectra obtained for QGSJetII4 based simulations. This is shown on the l.h.s. for events located in KASCADE and on the r.h.s. for events located in KASCADE-Grande.

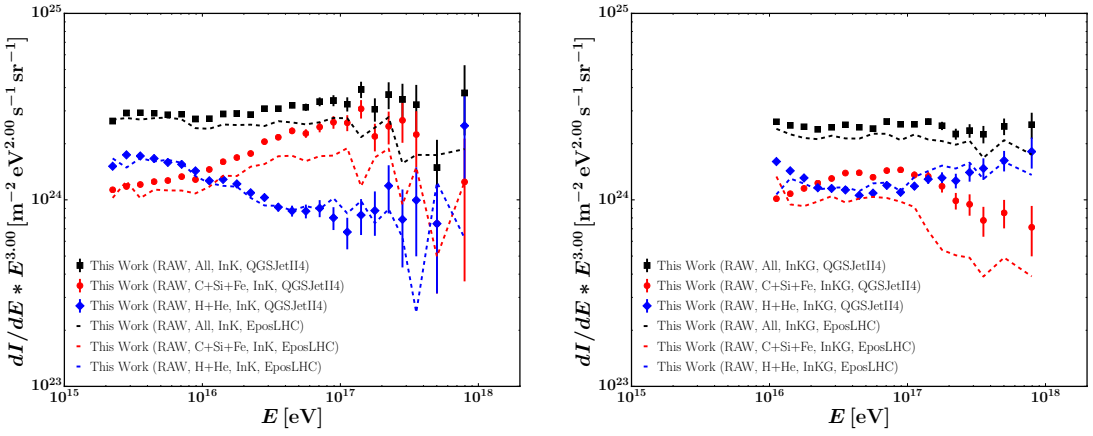

Figure 6.2.: The energy spectrum for all particles and the light and heavy mass groups are shown for measured events. Dashed lines correspond to calibrations based on EposLHC. Markers show the spectra obtained for QGSJetII4 based calibrations. This is shown on the l.h.s. for events located in KASCADE and on the r.h.s. for events located in KASCADE-Grande.

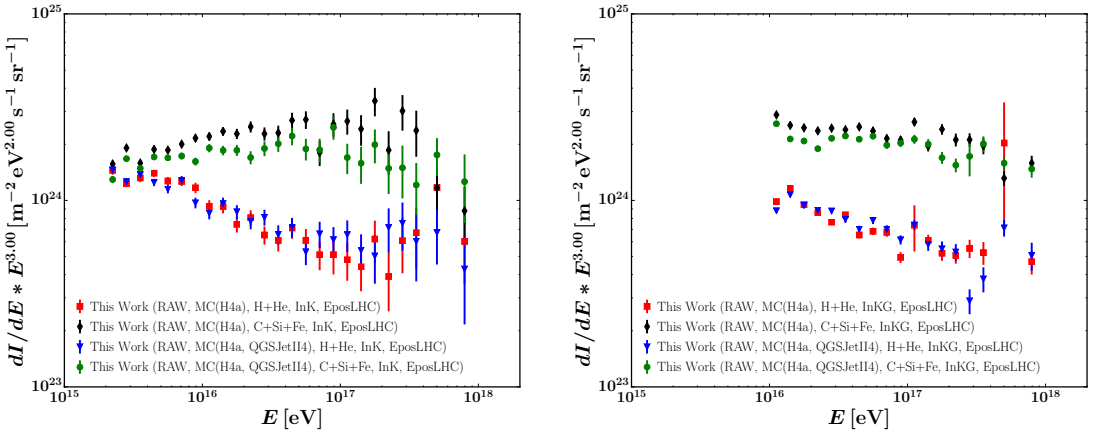

Figure 6.3.: The energy spectra for the light and heavy mass groups are shown for EposLHC based calibrations used to interpret both QGSJetII4 based simulations (light: blue triangles, heavy: green circles) and EposLHC based simulation (light: red squares, heavy: black diamonds), respectively. This is shown on the l.h.s. for events located in KASCADE and on the r.h.s. for events located in KASCADE-Grande.

in section 3.2 that this extrapolation is not entirely accurate. While the longitudinal momentum fraction distributions for neutral pions and rho-mesons in the interactions $\pi^+ + p \rightarrow \pi^0$ and $\pi^+ + p \rightarrow \rho^0$ seem to be sufficiently well estimated [47], NA61/SHINE measurements [46] of the $\pi^- + C$ are not so well described.

This picture is even more striking if one considers that both models have been tuned to the 7 TeV LHC data. This is equivalent to about 2.5×10^{16} eV protons hitting protons at rest. Up to this energy, there are essentially no deviations visible in the spectra of light primaries, neither for simulations nor for measured data [2]. In addition, the deviations among the models regarding the heavy primaries seem to be increasing with energy. However, the true reason could actually be unrelated to the energy itself, but to the mean mass of the heavy mass group. According to the current knowledge (and to the H4a and GST3 models), the heavy mass group should become heavier with energy, because the populations of medium mass particles reach their cutoff energies one after another. The difference in the extrapolation to nucleus-nucleus interactions would then increase as the number of participants rises.

The following discussion is dedicated to determine whether the above can be confirmed using simulations. This will mainly involve the comparison of the simulated spectra as reconstructed with calibrations based on the same hadronic interaction model to the ones reconstructed using the calibrations derived from the respective other model.

As a first step, the results of the combined analysis for events located in KASCADE and KASCADE-Grande are compared separately. In the following, when simulations are used as test-data, they have been weighted prior to the reconstruction, to make their true spectra match the H4a composition assumption.

[2] Except in the first energy bin for KASCADE-Grande. This is covered by the simulations, however, and is not due to differences in the models.

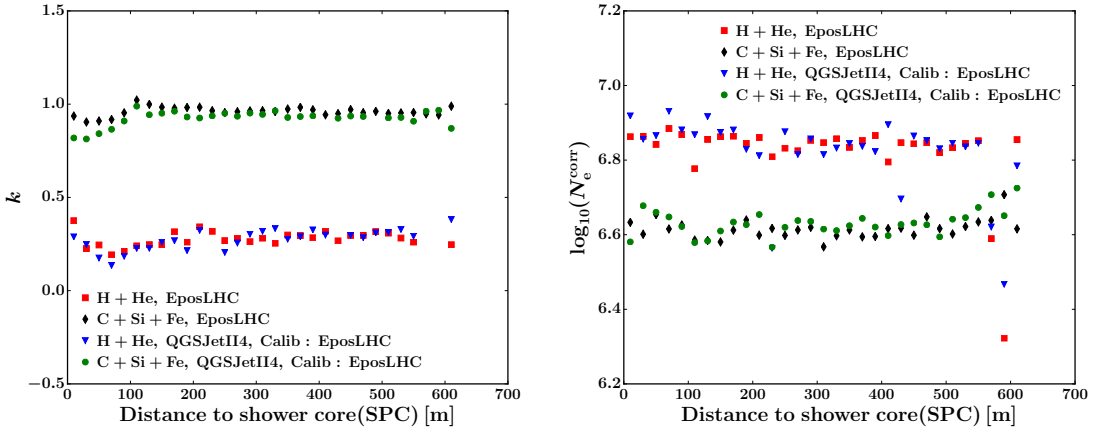

Figure 6.4.: The mean values of k (l.h.s.) and $\log_{10}(N_e)$ (r.h.s.) are shown as a function of the distance of the center of KASCADE (i.e. to the muon detectors) to the shower core in shower coordinates. This is displayed for events with a reconstructed energy above $10^{16.5}$ eV using EposLHC based calibrations and for data simulated using both EposLHC and QGSJetII4.

Figure 6.3 shows the spectra of light and heavy primaries as they are reconstructed using EposLHC based calibrations. Using the same simulations as test-data (light: red squares, heavy: black diamonds) yields essentially the same result for light primaries as it is obtained for QGSJetII4 based test-data (light: blue triangles, heavy: green circles). This is especially true up to a few 10^{16} eV. For KASCADE-Grande this seems to be the case up to the highest energies covered. Depending on the cause of it, reconstructing a lower flux of heavy primaries when using QGSJetII4 instead of EposLHC based simulations (where the true flux is the same) can be achieved to some extent without notably influencing the light component. This could well be the case at low energies for KASCADE and in general for KASCADE-Grande. For the latter, the differences in the heavy component seem to be smaller than for KASCADE. However, due to the limited number of events, this cannot be claimed with certainty. On the other hand, the larger flux of light primaries for KASCADE using QGSJetII4 would fit to a shift of events from the heavy mass group to the light one, which would be accompanied by a lower reconstructed energy for the events being shifted.

In fact, the main contributions to the discrepancy can be estimated from what is displayed on the l.h.s. of figure 6.4. Shown is the mean value of k for the light and heavy mass groups as a function of the distance of the center of KASCADE to the shower core. Up to 90 m, the shower is considered to have hit KASCADE, for larger distances the calibrations obtained for KASCADE-Grande are used. The result is shown for simulated data based on EposLHC and QGSjetII4. In both cases, the calibration obtained using EposLHC based data have been used. Only events with a reconstructed energy exceeding $10^{16.5}$ eV are considered.

At larger distances, there is no visible difference between the two models in case of light primaries. At lower distances, i.e. for showers that have hit KASCADE, there might be a tendency for the QGSJetII4 based simulations to yield smaller values of

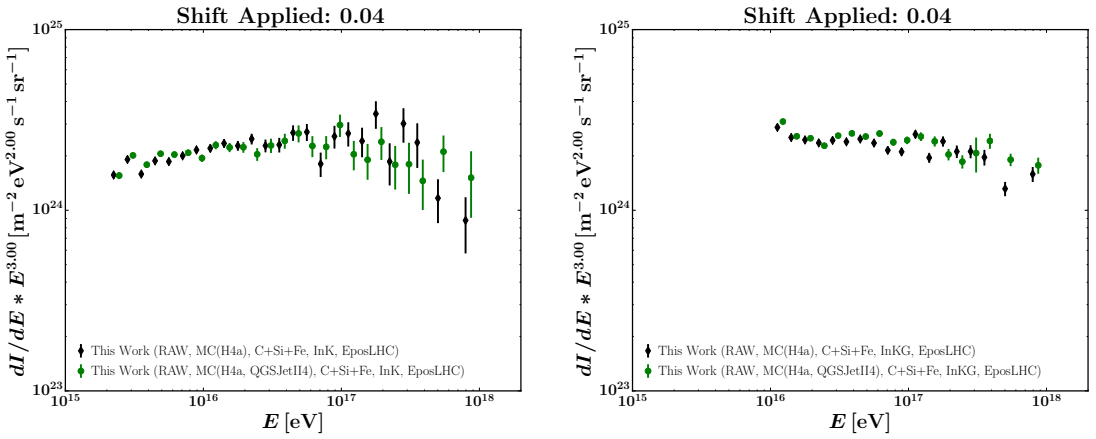

Figure 6.5.: The energy spectra of heavy primaries are shown as reconstructed for events located in KASCADE (l.h.s.) and KASCADE-Grande (r.h.s.) using the EposLHC based calibrations. This is shown for data simulated with EposLHC (black diamonds) and QGSJetII4 (green circles). In case of the latter, a general shift in $\log_{10}(E)$ of plus 0.04 has been applied.

k, when interpreted with EposLHC. This is already an indication that there should be no significant contribution of heavy primaries at $10^{16.5}$ eV, because they would mainly populate the distribution of k at larger values. This is no strong restriction towards higher energies, however, as the distribution of k will be dominated by the events close to $10^{16.5}$ eV. In order to lower the value of k, one can increase the prediction of N_e or decrease the prediction of N_μ in the model. However, in this case an increase of N_e is preferable, as it changes the energy only slowly. This is because the lower value of k decreases the energy, while the higher value of N_e increases the energy. The impact of N_μ on k is a bit larger than that of N_e, however, it results in a faster decrease of the reconstructed energy. This is so, because when N_e is fixed, the same values of the energy calibration are used. For example, for k equal to zero (i.e. for a proton) a value of $\log_{10}(N_e)$ of 7.0 and $\log_{10}(N_\mu)$ of 5.6 corresponds to an energy of about $10^{16.68}$ eV. In order to shift this event to the heavy mass group, one would have to decrease the prediction of $\log_{10}(N_e)$ by 0.3 increasing k by 0.57. The resulting energy, when keeping the prediction of $\log_{10}(N_\mu)$ fixed, would be $10^{16.58}$ eV. Increasing $\log_{10}(N_\mu)$ by about 0.2 would result in a very similar value for k, however, the energy would be increased to about $10^{16.86}$ eV.

Therefore, in order to get a shift in k of about –0.05 a corresponding increase of the predicted value of $\log_{10}(N_e)$ of about 0.03 is needed. This is very close to what is needed to explain the shift visible on the r.h.s. of figure 6.4, which displays the mean $\log_{10}(N_e)$ for the same events and reconstruction that is shown on the l.h.s. of the same figure.

Although the difference between the spectra of heavy primaries seems to be more substantial, a global change of the energy scale for heavy primaries of only 0.04 (in $\log_{10}$) is enough to shift the two spectra for KASCADE and KASCADE-Grande on top of each other (Fig. 6.5). However, a smaller shift would give an even better match for KASCADE-Grande. For KASCADE this is the case at lower energies, while for higher energies a slightly larger shift could fit better, however, this cannot be

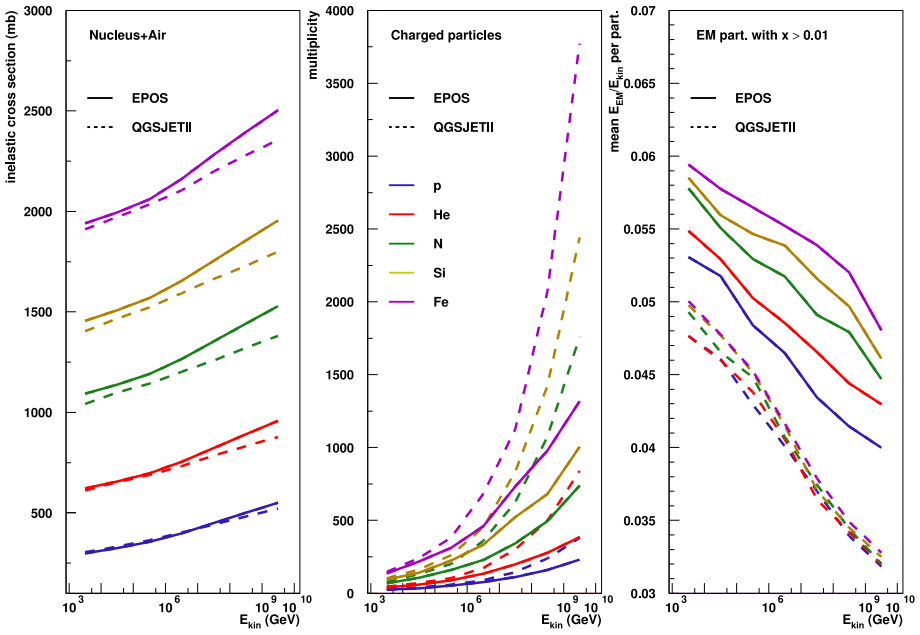

Figure 6.6.: Three properties of Nucleus-Air interactions as predicted by EposLHC (solid) and QGSJetII4 (dashed) are displayed as a function of the primary energy. Left: Inelastic cross sections. Middle: Multiplicity of charged particles. Right: Mean energy of electromagnetic particles relative to the primary energy and divided by the total number of electromagnetic particles. Only particles with $x_F > 0.01$ are considered [77].

judged with the available statistics. A shift by only 0.04 can be easily accomplished without any influence on the mass separation and, therefore, on the spectra of light primaries. Preferably, this would be accomplished by an increase of the predictions of both, the number of muons and the number of electrons. This way, k can remain constant, while the energy is increased. However, according to Fig 6.4, a small increase of k might actually be wanted, which could be achieved by increasing mainly the predicted values of $\log_{10}(N_\mu)$.

This shift in the energy or mass scale is not caused by deficiencies in the reconstruction procedures, but by differences in the description of the development of air-showers between EposLHC and QGSJetII4. Since the initial configuration of the primary mass and energy is known in case of simulations, the differences between the models can be studied by direct comparisons of the particle distributions.

In general, the predictions of both hadronic interaction models give extremely consistent results, especially in case of light primaries. The large impact of the shower-to-shower fluctuations on the reconstructed flux of light primaries (the light mass group in truth dominates at low energies) make this match even more striking, especially when considering that the results obtained using the respective other model are consistent for a large energy range. Nonetheless, at some point the degree of agreement will have to be checked at a more basic level, such as a direct analysis of the particle density distributions. In principle, the combined reconstruction would be very well suited for the task, however, that is beyond the scope of this work.

At this point, a comparison of the basic properties of the two models may help to see, if these fit into the general picture described above. For one, the predicted shower-

to-shower fluctuations in the number of electrons should not be very different for the two models. An important source of the fluctuations is given by the fluctuation of the shower maximum. Based on the discussion in [44], the fluctuations are mainly related to the variance in the depth of the first interaction and the variance in the multiplicity n_{mult} (Eq. 6.1, taken from [44]).

$$\mathrm{V}(X_{\mathrm{max}}) \propto 1/\sigma_{\mathrm{I}}^2 + \lambda_{\mathrm{e}}^2 \mathrm{V}(\log(n_{\mathrm{mult}})) \tag{6.1}$$

Taking a look at the l.h.s. plot of figure 6.6, one can see that the inelastic cross sections σ_{I} for proton-Air interactions are very similar at about 10^{15} eV. In addition, as reported in [47], the fluctuations of the multiplicity are now very similar for the two models. Therefore, while the impact of the remaining differences has still to be quantified by a detailed analysis, there is no immediate contradiction regarding this specific point.

The depth of the first interaction is strongly governed by the inelastic cross section. The cross section is significantly larger for heavy primaries compared to light primaries, which introduces the large difference in the measured number of electrons for the same primary energy, but different masses. It is interesting to see (Fig. 6.6) that the difference among the models regarding the inelastic cross section is small for protons and increases with the mass of the primary particle. This discrepancy increases even to larger energies. Judged only by the cross section, one would expect the shower maximum to be at higher altitudes in case of EposLHC compared to QGSJetII4. The difference, as said, increases with mass and energy of the primary [3]. This alone would, therefore, result in a lower number of electrons at observation level for data simulated using EposLHC. However, depending on how much energy is going to possible fragments of the nucleus (which is larger in case of QGSJetII4 than for EposLHC [77]), the higher multiplicity shown in the middle graph of figure 6.6 in case of QGSJetII4 could compensate the impact on the shower maximum of the different cross sections, however, its influence on the depth of X_{max} is smaller (its influence is logarithmic, see for example [44]). The main message is again a significant dependence of the model predictions on the mass of the primary.

The third plot shown in figure 6.6 displays the fraction of the primary energy transferred onto single electromagnetic particles (in this case mostly photons from π^0 decay) averaged over all electromagnetic particles with $\mathrm{x_F} > 0.01$ [4]. According to the Heitler model, the depth of the shower maximum is mainly governed by the electromagnetic particles arising from the first interaction [5]. These particles have a larger fraction of the primary energy in case of EposLHC. While this is true also for protons, the difference between the two models is again larger for heavy primaries.

[3] While the strongest impact is probably that of the first interaction, there are differences among the models regarding the fragmentation of nuclei, both in terms of what these fragments are and how the energy is distributed.

[4] $\mathrm{x_F}$ is called "Feynman-x". It is defined as $2p_{||}/\sqrt{s}$ with $\sqrt{s}$ being the energy of the projectile (assuming the target, i.e. a particle of the atmosphere is at rest) expressed in the center of mass reference system (CMS) and $p_{||}$ being the longitudinal momentum (CMS) of the corresponding produced particle.

[5] However, as demontrated in [44], a higher elasticity will result in a deeper penetrating shower, because the energy is transported deeper into the atmosphere. Usually, a higher elasticity was connected with a lower X_{max}, because less energy is transferred to the first generation of electromagnetic particles.

Being the lightest known meson, the pion is the most abundant of the species driving the hadronic cascade. The properties of pion-Air interactions (and proton-Air for that matter), such as cross section, multiplicity, and elasticity are quite similar for both models [77]. In the end, the shower maximum is a bit deeper in the atmosphere, and the number of muons is a bit larger for data simulated with EposLHC when compared to data simulated using QGSJetII4 [47], however, the energy distribution is different, QGSJetII4 predicting a larger mean muon-energy.

To summarize this part of the discussion, a dedicated analysis is needed to quantify how well the two models match one another. However, both, the findings of this work and the basic interaction properties suggest a better agreement of the two models for light primaries than for heavy primaries. The disagreements for heavy primaries seem to scale with mass and energy.

For the remaining part, it is assumed that the predictions of the EposLHC and QGSJetII4 models are compatible for light primaries.

6.2. KASCADE vs KASCADE-Grande

Considering the observed difference in the composition between KASCADE and KASCADE-Grande, it is worth testing whether it is expected from simulations, i.e. covered by known detector and reconstruction effects. Taking a look at figure 6.7, one can see that the spectra obtained for KASCADE and KASCADE-Grande are in general agreement to one another. In case of using the EposLHC based calibrations to reconstruct the spectra for the same model (l.h.s.) reveals, that there are tendencies for a slightly higher, but steeper reconstructed flux of light primaries for events within KASCADE-Grande compared to the events located in KASCADE. However, the differences are too small in order to explain the differences visible in figure 6.8 (l.h.s.), where the EposLHC based calibrations have been applied to measured data. In fact, for measured data, the reconstructed energy spectrum for the light component is less steep in case of KASCADE-Grande compared to KASCADE. The same can be said when using QGSJetII4 based calibrations on simulations based on the same model (Fig. 6.7, r.h.s.) compared to applying it on measured data (Fig. 6.8, r.h.s.).

Since this difference between KASCADE and KASCADE-Grande is not seen in simulations for either hadronic interaction model, it has to be caused by a general difference between simulated and measured data.

While only a dedicated analysis can give certainty here, it might be reasonable to assume that KASCADE-Grande is affected less by these general differences than KASCADE assuming that a differently predicted mass scale is the cause of the discrepancy between KASCADE and KASCADE-Grande. The reasoning is the following. An overestimation of the number of muons at KASCADE, increasing with energy, as it is discussed in appendix C, would result in an increasing number of events being shifted from the light to the heavy mass group. This would make ankle-like features in the light spectrum and knee-like features in the heavy spectrum less pronounced. This might be visible in figure 6.8. However, the difference between the spectra for events measured in KASCADE and KASCADE-Grande could also be interpreted in the opposite way, i.e. the increasing underestimation of the number of muons at KASCADE-Grande would result in an increasing number

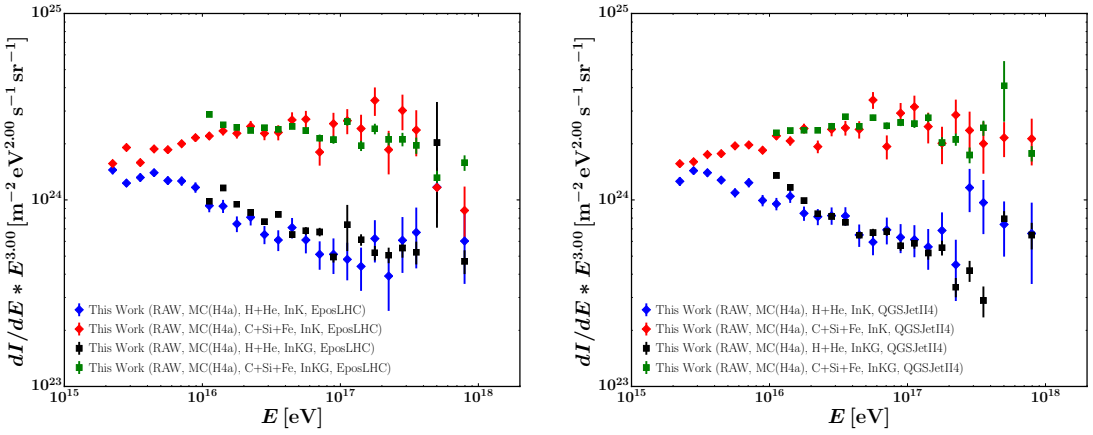

Figure 6.7.: The simulated energy spectra for events located in KASCADE and KASCADE-Grande are shown separately for the light and heavy mass groups. The corresponding EposLHC (l.h.s.) and QGSJetII4 (r.h.s.) based simulations have been weighted according to the H4a model.

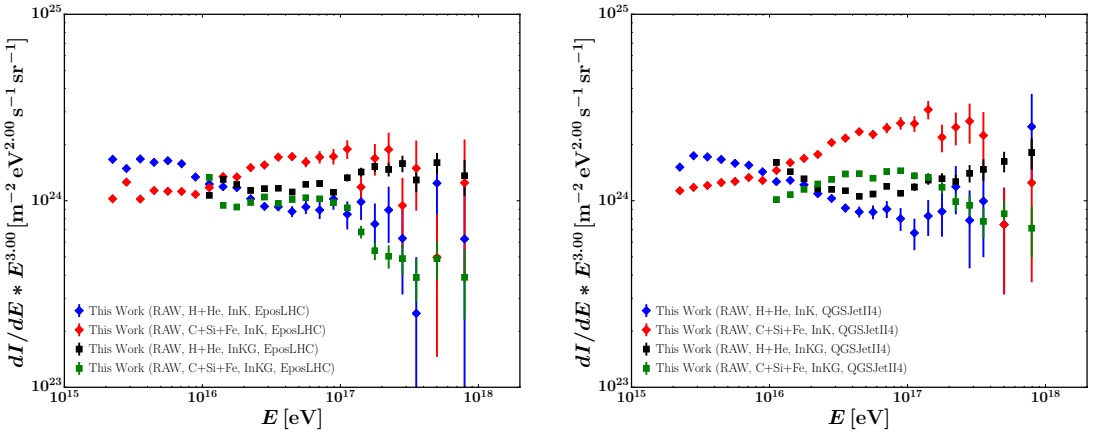

Figure 6.8.: The measured energy spectra for events located in KASCADE and KASCADE-Grande are shown separately for the light and heavy mass groups. The corresponding EposLHC (l.h.s.) and QGSJetII4 (r.h.s.) based simulations have been used for the calibrations.

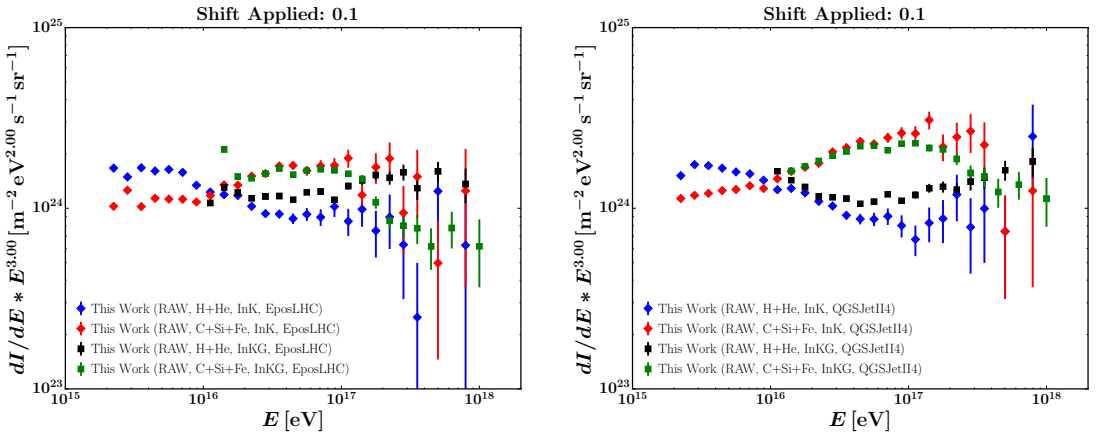

Figure 6.9.: The measured energy spectra for events located in KASCADE and KASCADE-Grande are shown separately for the light and heavy mass groups. The corresponding EposLHC (l.h.s.) and QGSJetII4 (r.h.s.) based simulations have been used for the calibrations. In addition, the energy of heavy primaries located in KASCADE-Grande has been shifted by 0.1 in $\log_{10}(E)$.

of events being shifted from the heavy to the light mass group. This would result in knee-like features in the spectrum of heavy primaries and ankle-like features in the spectrum of light primaries. For the interpretation of the differences both ways seem valid, however, as shown in figure 6.2, the measured spectra of light primaries are very similar for both models and both areas. Shifting a significant amount of heavy primaries to the light mass group should result in less agreeing spectra based on the two models. Therefore, the agreement of the two models on the light component is much easier to explain if one assumes that the KASCADE-Grande response to the showers is closer to the simulations. The spectrum of the light mass group would really consist mostly of light primaries, hence, it is described by both models in the same way. The corresponding spectra for KASCADE could even be in better agreement, because, by shifting events to the heavy mass group, the light mass group would correspond to a purer light composition.

If it turns out that this is the case, it would mean that the lateral density function as predicted by the models is too flat because the number of muons closer to the shower core are underestimated. Due to the fixed slope of the lateral density function during the reconstruction of the number of muons, this would always result in a compensation of the underestimated number of muons for KASCADE-Grande. On the other hand this would enhance the problem for KASCADE, because the number of muons would be overestimated for measured data, but underestimated for simulated data.

While a different mass-scale is a possible explanation for the observed difference between KASCADE and KASCADE-Grande, there is still one problem left. According to what is shown in appendix C, the discrepancy between KASCADE and KASCADE-Grande should be small at low energies. However, taking a look at figure 6.8 this can only be attributed to the light primaries, which show similar differences among the two detectors as it is predicted by the simulations. This is not

the case for heavy primaries. This observation leads to an explanation which might be more justified. To restore a match among the models for the heavy mass group a shift of the energy scale of 0.04 in $\log_{10}(E)$ was proposed. There is no reason that a global difference between measurement and simulation cannot have the same origin, hence, requiring another shift of the energy scale. In fact, in the discussion on the differences among the models, there have already been hints at a slightly different behavior of KASCADE and KASCADE-Grande, which was "solved" by the shift of 0.04. Even more so, there were hints at a slightly increasing difference also in the light component at energies beyond what is covered by the 7 TeV-LHC data (Fig. 6.3).

Introducing a shift in $\log_{10}(E)$ for heavy primaries located in KASCADE-Grande (Fig. 6.9) results for both, EposLHC (l.h.s.) and QGSJetII4 (r.h.s.) in a better agreement between the results obtained for events measured within KASCADE and those in KASCADE-Grande at lower energies. Towards higher energies, a larger shift would be needed. This could be attributed also to an increasingly heavier composition. A similar difference as observed between EposLHC and QGSJetII4 could lead to this behavior between both models and measured data. In addition, there is no reason to assume that both models, agreeing on the result of the light component, could not have a similar, common difference for the light component in the other direction. There is a clear energy dependence visible, which has previously been attributed to a shift of events from the heavy mass group (or the other way around, if one uses KASCADE-Grande as reference). One has to consider that a shift in $\log_{10}(E)$ of only 0.1 had this seemingly large effect (the spectrum has been scaled by E^3). Hence, only a small shift in $\log_{10}(N_\mu)$ is needed to account for this difference in the energy of the heavy particles. Since the difference is much smaller for the light component, an even smaller, energy-dependent shift would be needed. Of course, a shift of N_e and N_μ would change the evaluation of the calibrations, hence, the resulting spectra would look a bit different from what is achieved by a direct shift of the energy. In any case, a shift in the energy scale which depends on the mass of the primary particle could easily explain why the two results obtained for events located in KASCADE and KASCADE-Grande can be different without influencing the agreement of the two hadronic interaction models on the spectrum of the light component. Of course, combinations of different energy and mass scales are also possible.

Which air-shower array is closer to the truth cannot be decided without a dedicated analysis. Both cases are possible and solutions can be found for both versions that avoid a large contamination of the light mass groups. The latter is a requirement to be able to explain why both models can yield the same predictions for light particles although they differ on the predicted shower development for heavy primaries. However, the validity of these solutions depend on the composition and on how well the two models really match each other for the light mass group and how different they are already for medium mass primaries. This is beyond the scope of this work, however.

6.3. Other Experiments

The three main experiments covering at least the energy range from 10^{16} eV to 10^{18} eV are IceTop, Tunka-133, and KASCADE-Grande. Their main results on the

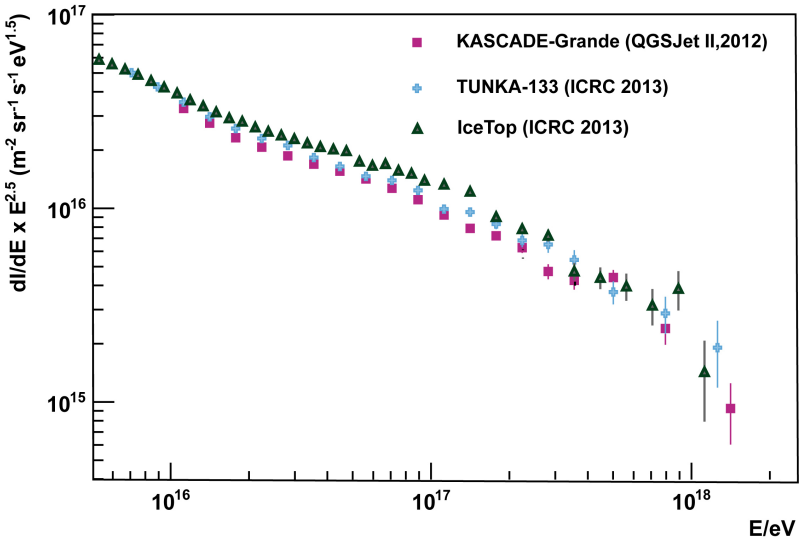

Figure 6.10.: The energy spectrum for all particles as reconstructed by IceTop, Tunka-133, and KASCADE-Grande. Picture taken from [78].

energy spectrum of cosmic rays are compiled in figure 6.10. IceTop consists of 81 ice-Cherenkov stations which are built on top of the IceCube in-ice detector at the South Pole. Tunka-133 is made of 133 photomultipliers that observe the Cherenkov light emitted during the shower development in the atmosphere. Tunka is located near to Lake Baikal in southern Siberia. While the results of Tunka-133 and KASCADE-Grande have been obtained taking the mass of the primary into account based on the measured data, the result of IceTop is based on a assumed composition that corresponds to the H4a model. The two main features in the shown energy range, namely the concavity at a few times 10^{16} eV and the knee-like structure at 10^{17} eV are visible in all three spectra. It is worth noting that a result for KASCADE-Grande obtained using SIBYLL2.1 [79] as the hadronic interaction model is considerably closer to the result shown for IceTop, which also uses SIBYLL2.1.

Tunka-133 has published a measurement of the cosmic-ray composition which is based on the separation of their data set in two mass groups. Their result is similar to the one obtained by the standalone analysis of KASCADE-Grande (see section 4.2.3).

The published results of IceTop are based on composition assumptions. Recently, the collaboration has shown on conferences that they are working on a measurement of the composition based on the measurement of muons in the in-ice detector of IceCube. However, their analysis is still ongoing and can currently not be used to be compared to the results of e.g. KASCADE-Grande.

In both cases, the currently achieved accuracy is below the one reached in the work presented here, however, especially IceTop benefits from being located at a higher altitude. Getting closer to the shower maximum reduces the impact of the shower-to-shower fluctuations significantly, which are predicted differently by the hadronic interaction models. On the other hand, the reduced impact of the hadronic interaction model reduces the capability to spot remaining inconsistencies among the models and measured data. This is where KASCADE and KASCADE-Grande, when combined, excel at. The capability to sample the muonic component, which is

especially sensitive to the evolution of the hadronic component in the atmosphere, at different distances to the shower core and at different threshold energies is unique in this energy range. While the stronger dependence on the hadronic interaction model is a hindrance for the successful reconstruction of the energy spectrum and mass composition, it is this dependence and the simultaneous use of KASCADE and KASCADE-Grande that revealed the remaining discrepancies among the models and between models and nature. KASCADE and KASCADE-Grande, when used separately, could not have achieved this. The final statement would have been that both models give consistent results for light primaries. Only by combining both detectors and by comparing the two results, a general difference between simulations and measurements, also in case of light primaries, was revealed. In addition, the improved reconstruction accuracies have improved the sensitivity to smaller differences between simlation and measurement.

Finally, it should be remarked that the three experiments employ three experimentally different observables of the showers. This fact can be used to test hadronic interaction models, when the analyses are combined. Indeed the three experiments have approved to work together on this.

6.4. Comparison of the Results with Astrophysical Models

As the final part of this discussion, the predictions of the H4a and GST3 models will be confronted with the measurements. Based on the above discussion on the differences between the hadronic interaction models and between KASCADE and KASCADE-Grande, the predictions will be compared to the results of KASCADE using the QGSJetII4 hadronic interaction model. The choice of the latter is based on the simple fact that more simulated events are available compared to EposLHC. The decision to use KASCADE instead of KASCADE-Grande is based on the larger energy range covered by KASCADE.

The measured spectra for all particles and the light and heavy mass groups are shown in figure 6.11 for both astrophysical models. In the top panel, the predictions of the H4a model are shown, where the detector and reconstruction effects have been introduced using the data simulated with QGSJetII4.

At first glance, it might be interesting to note that the predicted and the measured spectra seem to follow the same distribution for light primaries. The only difference seems to be the absolute flux of this component, which is not surprising as the results of the standalone analysis of KASCADE and KASCADE-Grande have been available at the time the H4a model was published. Therefore, it is more interesting to see that the results for energies higher than the standalone analysis (10^{17} eV) are also in good agreement (apart from a normalization).

While the relative abundance of the elements in the light mass group may very well be the one assumed in the H4a model, the heavy mass group seems to be different from the model predictions, at least at the lower energies. This could be due to the assumption made regarding the CREAM data. A less steep heavy spectrum would be well within the uncertainties of the corresponding CREAM measurements.

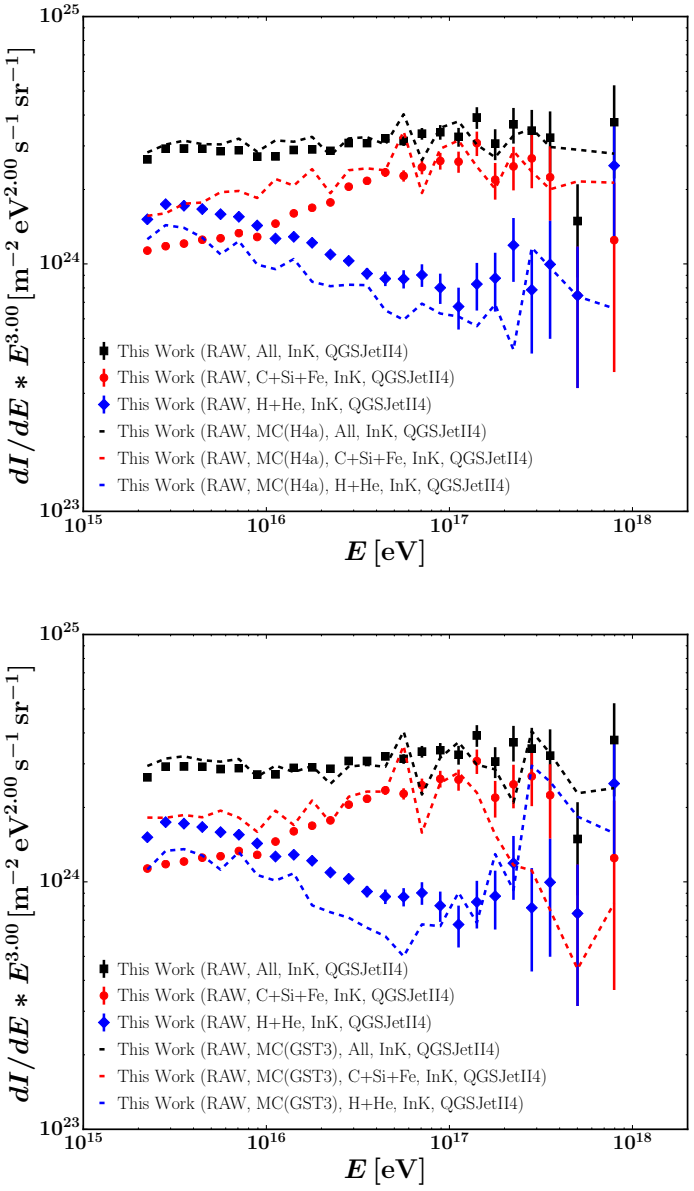

Figure 6.11.: The energy spectrum for all particles and the light and heavy mass groups are shown for events measured in KASCADE (markers). Dashed lines correspond to the spectra based on QGSJetII4 simulations that have been weighted to match the predictions of the H4a (top panel) and GST3 (bottom panel) models.

A simple change of the slope and normalization of the heavy components could already suffice in order to describe the data.

Comparing the predictions of the GST3 model (bottom panel of Fig. 6.11) with the measured results, one can see that the additional iron component at about 10^{15} eV does not agree well with the measurements. However, taking a look at the shifted results of KASCADE-Grande (r.h.s. of figure 6.9), the model predictions would actually fit reasonably well. Even the predicted extragalactic iron component could be present in the data. More detailed tests will show if it is reasonable, however, there is no reason to believe, that only KASCADE-Grande (i.e. muon measurements at a larger distance to the core) requires a change of the predictions of the hadronic interaction models.

In any case, it does not seem that the models fit completely to the KASCADE and KASCADE-Grande results. However, it cannot be said at the moment, how much would have to be changed by the astrophysical model [6] and how much better the description will get once the hadronic interaction models have improved.

Ankle, Dip or Mixed

Distinguishing between the Ankle, Dip, and Mixed composition models (see section 2.3.1) is currently not possible. If the results for the events located in KASCADE-Grande turn out to be more reliable, the Dip model might still be a valid option. However, if the flattening in the spectrum of heavy primaries is not caused by a shift of events from the light mass group, a mixed composition model could be preferable. In either case, the transition would occur below the ankle, hence the Ankle model could be discarded.

For KASCADE-Grande the interpretation does not change depending on which of the two hadronic interaction models is employed. However, for events located in KASCADE the energy at which the spectra of the light and the heavy mass groups cross each other certainly depends on the model used. The result based on QGSJetII4 seems to allow for a higher transition energy compared to EposLHC and could be compatible with the Ankle model and incompatible with the Dip model. However, in either case, due to the limited statistics at high energies, it is not clear how the spectra would have to be extrapolated towards higher energies.

In the end, one should also not forget that the results shown here are raw spectra, i.e. they are still influenced by mass dependent misreconstructions. This is one part of the reason why a detailed comparison to the results of other experiments is omitted here. The other part is the unknown systematic uncertainty introduced by the hadronic interaction models. It is not possible to decide whether the true energy spectrum and mass composition is between the results of KASCADE and KASCADE-Grande or if it is beyond either. In addition, the absolute energy and mass scale does not necessarily correspond to the one predicted by either of the two hadronic interaction models employed. Again the truth could lie between the two predictions or beyond either. Hence, it is not possible to meaningfully estimate the systematic uncertainties, because differences between the actual number of muons and electrons produced in the air-shower induced by a primary of a certain energy

[6] And what impact this would have on the number and distribution of sources and the propagation and acceleration mechanisms.

and mass to the numbers predicted by the models could make these events look like a particle of a different mass and/or energy without revealing the true extent of the discrepancies. A comparison of the results of different experiments is only meaningful, if the corresponding systematic uncertainties are known accurately. Unreliable estimates of the systematic uncertainties are likely to lead to wrong conclusions regarding the compatibility to other results or astrophysical models. Hence, the quantification of systematic uncertainties is also omitted, as these would have to be based on assumptions that cannot be verified.

For the standalone analyses that are based on the parameter k, a procedure of estimating the systematic uncertainties has been developed [68]. Taking this as a basis and taking into account the better reconstruction accuracy for N_e and N_μ an estimate for the systematic uncertainties can be obtained if one is willing to ignore uncertainties related to the hadronic interaction models. Hence, if the respective hadronic interaction model is assumed to be correct, it is reasonable to assume that a systematic uncertainty of the flux of less than 15 % can be achieved once the mass-dependent misreconstructions have been corrected for.

In any case, although it is not possible to infer whether there are global differences between the predictions of the EposLHC and QGSJetII4 models and nature (also for the respective light components, which could both be inaccurate in the same way), it is possible to test how well the two models agree in case of light primaries and why (and how much) the two models differ regarding their predictions for the heavier primaries.

Appendices

A. Coefficients

The parameters obtained for the calibration of the energy and k parameter are listed below. The function is given by two lines connected via a hyperbola [71] as defined in Eq. A.1

$$
\begin{aligned}
f(x) &= a + \beta_1 \cdot x_d + \beta_2 \cdot \sqrt{x_d^2 + 0.25 \cdot r^2} \\
x_d &= x - x_t \mid x_t : \text{ coordinate of transition} \\
\beta_1 &= (s_1 + s_2)/2 \mid s_{1,2} : \text{slopes before/after transition} \\
\beta_2 &= (s_2 - s_1)/2 \mid r : \text{radius of hyperbola}
\end{aligned}
\tag{A.1}
$$

Proton (H)						
Angles	Location	a	r	s_1	s_2	x_t
00.00-16.78°	KASCADE	1.0626	0.3828	1.5721	0.1619	4.9799
	Grande	1.1426	0.2941	0.9057	0.1876	5.9230
16.78-24.09°	KASCADE	1.0179	0.2221	1.5120	0.1628	5.0550
	Grande	1.1547	0.6641	0.9813	0.1207	5.8344
24.09-30.00°	KASCADE	0.9646	0.0052	1.3737	0.1659	5.2249
	Grande	1.0454	0.5472	0.8369	0.1381	5.7650
Iron (Fe)						
Angles	Location	a	r	s_1	s_2	x_t
00.00-16.78°	KASCADE	0.4607	0.0000	0.6548	0.2221	4.7343
	Grande	0.8095	0.1844	0.3447	0.1719	5.9055
16.78-24.09°	KASCADE	0.4023	0.1171	0.8727	0.2194	4.7271
	Grande	0.6512	0.1467	0.5972	0.1910	5.5041
24.09-30.00°	KASCADE	0.3316	0.0000	272808.8178	0.2154	4.7500
	Grande	0.5479	0.0004	0.5382	0.1897	5.3767

Table A.1.: Parameter values for the calibration of the k parameter based on QGSJetII4. The extremely large slope before the bending (s1) in case of iron does not affect the reconstruction. It is caused by the very sharp transition (r=0) that occurs directly at the first data point, hence, the slope before the bending is not restricted by any data point. In this sense the effective calibration follows a single line and a change of the slope below the first data point would be a cosmetic measure that does not affect the reconstruction. However, no manual adjustments are made in the present work, even if they would be of cosmetic nature, only.

Proton (H)						
Angles	Location	a	r	s_1	s_2	x_t
00.00-16.78°	KASCADE	14.7189	0.3828	-0.5135	0.9694	4.9848
	Grande	15.6334	0.6730	0.1822	0.9654	5.9297
16.78-24.09°	KASCADE	14.8075	0.2802	-0.9783	0.9663	4.9995
	Grande	15.6683	0.2980	0.1394	0.9274	5.8202
24.09-30.00°	KASCADE	15.1226	0.0000	-0.6374	0.9775	5.2231
	Grande	15.5867	0.6985	0.1232	0.9912	5.7139
Iron (Fe)						
Angles	Location	a	r	s_1	s_2	x_t
00.00-16.78°	KASCADE	14.9025	0.6714	0.0088	0.8444	4.5500
	Grande	15.6987	1.0717	0.4377	0.9048	5.6327
16.78-24.09°	KASCADE	15.0173	0.3429	-1.4528	0.8358	4.5500
	Grande	15.6727	0.4461	0.2376	0.8509	5.3915
24.09-30.00°	KASCADE	15.3201	0.0623	-7.0642	0.8248	4.7500
	Grande	15.3638	0.7425	-0.3729	0.8659	4.9532

Table A.2.: Parameter values for the calibration of the energy based on QGSJetII4.

Proton (H)						
Angles	Location	a	r	s_1	s_2	x_t
00.00-16.78°	KASCADE	1.0623	0.4374	1.6724	0.1469	4.9944
	Grande	1.1598	0.4660	0.9332	0.1716	5.9442
16.78-24.09°	KASCADE	0.9758	0.2965	1.4824	0.1521	5.0399
	Grande	1.0675	0.3444	0.8364	0.1702	5.8601
24.09-30.00°	KASCADE	0.9326	0.2401	1.4282	0.1449	5.1674
	Grande	0.9885	0.6506	1.0370	0.1600	5.6578
Iron (Fe)						
Angles	Location	a	r	s_1	s_2	x_t
00.00-16.78°	KASCADE	0.4752	0.0851	0.9155	0.2103	4.7603
	Grande	0.7553	0.2115	0.6608	0.1647	5.6557
16.78-24.09°	KASCADE	0.3964	0.0384	1.3605	0.2101	4.6990
	Grande	0.6450	0.0856	0.5089	0.1816	5.5304
24.09-30.00°	KASCADE	0.3698	0.0836	0.7017	0.2050	4.9014
	Grande	0.5519	0.1159	0.5103	0.1849	5.4382

Table A.3.: Parameter values for the calibration of the k parameter based on EposLHC.

Proton (H)						
Angles	Location	a	r	s_1	s_2	x_t
00.00-16.78°	KASCADE	14.7514	0.3138	-0.7825	0.9243	4.9838
	Grande	15.6153	0.4842	0.0847	0.9056	5.8634
16.78-24.09°	KASCADE	14.8857	0.2409	-0.6244	0.9513	5.0367
	Grande	15.1632	0.9183	-0.9228	0.9849	5.4500
24.09-30.00°	KASCADE	15.0809	0.1730	-0.8196	0.9590	5.1336
	Grande	15.2147	0.9199	-0.8777	0.9619	5.3561
Iron (Fe)						
Angles	Location	a	r	s_1	s_2	x_t
00.00-16.78°	KASCADE	14.9834	0.2803	-1.3012	0.8227	4.6147
	Grande	15.5116	1.0311	0.0657	0.9241	5.4792
16.78-24.09°	KASCADE	15.0722	0.2127	-2.0666	0.8184	4.5974
	Grande	15.7545	0.3435	0.4744	0.8342	5.4727
24.09-30.00°	KASCADE	15.3025	0.2644	-1.0292	0.8142	4.7500
	Grande	15.8074	0.3394	0.5245	0.8312	5.4016

Table A.4.: Parameter values for the calibration of the energy based on EposLHC.

Proton (H)						
Angles	Location	a	r	s_1	s_2	x_t
00.00-16.78°	KASCADE	1.0740	0.4236	1.7736	0.1769	4.8289
	Grande	0.9438	0.0000	14.0754	0.2932	5.3500
16.78-24.09°	KASCADE	1.0119	0.3735	1.7645	0.1813	4.8373
	Grande	1.1058	0.1447	0.8830	0.2505	5.8605
24.09-30.00°	KASCADE	0.9752	0.3056	1.5033	0.1820	4.9110
	Grande	1.0249	0.4169	0.8878	0.2260	5.6530
Iron (Fe)						
Angles	Location	a	r	s_1	s_2	x_t
00.00-16.78°	KASCADE	0.4609	0.3267	1.4672	0.2295	4.4567
	Grande	0.8510	0.0704	0.3687	0.1944	5.8457
16.78-24.09°	KASCADE	0.4270	0.0000	0.6117	0.2410	4.6500
	Grande	0.7386	0.0003	0.5839	0.1956	5.6112
24.09-30.00°	KASCADE	0.3558	0.3906	1.1132	0.2188	4.5125
	Grande	0.6092	0.1505	0.5774	0.1974	5.3662

Table A.5.: Parameter values for the calibration of the k parameter based on QGSJetII2.

Proton (H)						
Angles	Location	a	r	s_1	s_2	x_t
00.00-16.78°	KASCADE	14.6026	0.3390	-0.7727	0.9536	4.8475
	Grande	15.2998	0.0000	-6.5429	0.8417	5.3806
16.78-24.09°	KASCADE	14.7037	0.2787	-0.7002	0.9626	4.8657
	Grande	15.7324	0.1303	0.2734	0.8684	5.8585
24.09-30.00°	KASCADE	14.3969	0.6904	-2.3593	1.0220	4.6643
	Grande	15.3546	0.8169	-0.3000	0.9410	5.4871
Iron (Fe)						
Angles	Location	a	r	s_1	s_2	x_t
00.00-16.78°	KASCADE	14.8776	1.5924	0.4101	0.8680	4.6520
	Grande	15.4744	0.5663	-0.5709	0.8739	5.3705
16.78-24.09°	KASCADE	14.9233	0.7465	0.2119	0.8175	4.4690
	Grande	15.9807	0.0006	0.6791	0.8359	5.7829
24.09-30.00°	KASCADE	15.0953	0.5291	0.0309	0.8337	4.5550
	Grande	15.6256	0.5536	0.3347	0.8539	5.2477

Table A.6.: Parameter values for the calibration of the energy based on QGSJetII2.

B. N_e and N_μ Corrections

The quality of the reconstruction of N_e and N_μ depends on the numbers itself, the distance of the shower core in shower coordinates relative to the muon detectors, and the zenith angle. The latter two are more important for events located in KASCADE-Grande, as some simplifications regarding the slope of the muon lateral density function are in place, which result in the shape of the lateral density function being a fixed parameter. Therefore, the lateral density function cannot be expected to describe proton and iron like showers equally well and a mass dependence is to be expected, which can be accounted for using N_e and N_μ. The dependence on the distance can be explained by the limited range of sampling points provided by the KASCADE muon detectors, hence, if the lateral density function is too steep, the total number of muons will be overestimated for large distances and an underestimation of the same number will be obtained for events close to the muon detectors. In the following, the dependency of how accurately N_e and N_μ are reconstructed on these four quantities is presented and the results of the correction procedure are shown.

The dependence of the reconstruction of N_μ on the mass of the primary is visible in the two-dimensional shower-size-spectrum shown in figure B.1, where the color bar shows the mean deviation of the reconstructed number of muons from its true value. This is shown on the l.h.s. for events located in KASCADE, and on the r.h.s. for events with core positions within KASCADE-Grande.

It is important to note that the minimum/maximum values shown on the color bar have been chosen explicitly to see also the evolution of the deviations, where these are small. Therefore, the dark blue and red regions may have values beyond ± 0.1.

For the standalone analysis of KASCADE-Grande a correction of the number of muons was based on the same quantities, except for the number of electrons. The correction was based on a simple three-dimensional matrix filled with the mean differences of the reconstructed number of muons to the true simulated one. In addition, a smoothing of these values had been applied in order to avoid significantly different correction values for neighboring cells.

While for the standalone analysis N_{ch} was used, which does not depend on the

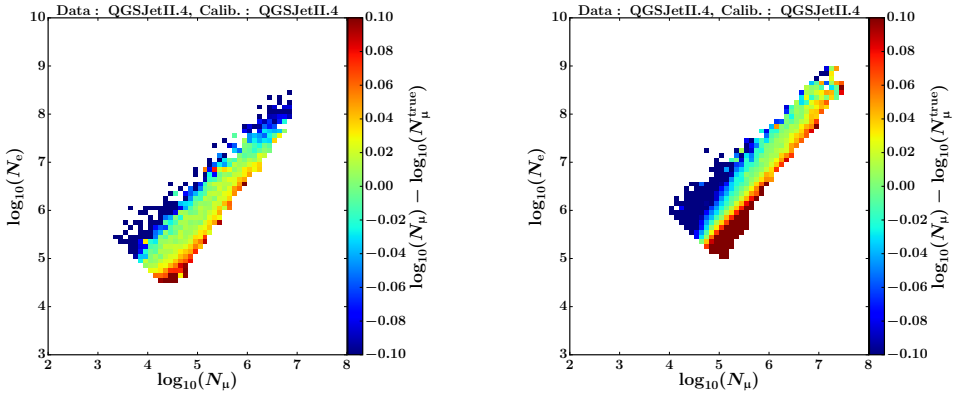

Figure B.1.: The deviation of the reconstructed number of muons from their true value is shown in the two-dimensional N_e-N_μ plane. Left: Events were located in KASCADE. Right: Events were located in KASCADE-Grande

reconstruction of the muon lateral density function, the combined analysis employs the number of electrons, which is reconstructed taking the muon lateral density function into account. Hence, a systematic misreconstruction of the number of muons results also in a corresponding misreconstruction of N_e. This effect is taken into account, in addition to a certain dependence on the mass of the primary, by using both N_e and N_μ. However, introducing a fourth dimension to the matrix based approach is not feasible, because the amount of simulated events necessary to sufficiently populate the cells of a four-dimensional matrix is not available.

Instead, the correction applied in the present analysis is based on Gradient Boosted Regression Trees (GBRT) as implemented in the scikit-learn Python package [80]. The various hyper-parameters, such as e.g. the number of estimators in the ensemble, maximum depth of an individual tree, or the minimum observations per leaf, have been inferred by using cross-validation. The procedure is as follows.

First the simulated data available is randomly split into five subsets. It is then checked if each of these five sets contain enough events at high energies to ascertain the test set is representative also for the high-energy events.

One after the other each set is kept as a test sample, while the remaining four sets are used together to train the ensemble of trees for various combinations of the hyper-parameters. The performance of each set of parameters is judged by the Mean Squared Error (MSE) comparing the simulated number of muons and electrons to the respective estimates after the correction has been applied. For each combination of hyper-parameters, the number of estimators is chosen to be the number where the MSE of the test sample is minimized. While for the training sample, the MSE keeps improving even at a high number of estimators, the MSE for the test sample starts to increase again at some point. This behavior is called overfitting, i.e. the ensemble starts to describe the fluctuations present in the training sample while losing the capability to give reasonable predictions for independent data, i.e. for the test sample.

Averaging the performance of each set of hyper-parameters for the five different test samples, the combination of parameters showing the best overall performance is

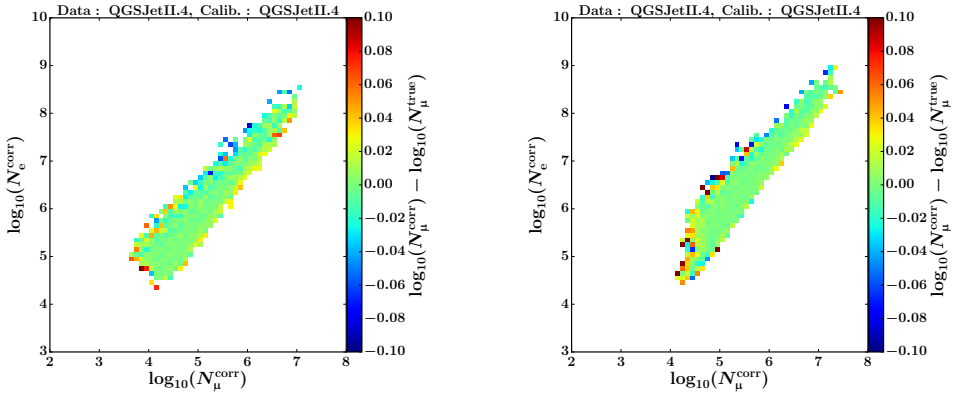

Figure B.2.: The deviation of the corrected number of muons from their true value is shown in the two-dimensional N_e-N_μ plane. Left: Events were located in KASCADE. Right: Events were located in KASCADE-Grande.

chosen to train the final ensemble using all available simulated events. At this point it is assumed that the final ensemble is not suffering from overfitting and that it benefits from the additional input. In order to test this assumption, a completely independent dataset is used to compare the performance of the ensemble of estimators when applied on the original data to the performance of the ensemble when applied on the data generated by a different hadronic interaction model. In the following, the former dataset is based on QGSJetII4, while the latter was generated using EposLHC, i.e. the ensemble has been trained using simulated data based on QGSJetII4 as the hadronic interaction model. Nonetheless, the general picture shown in the following is the same when using EposLHC to train the corrections.

After the correction has been applied, there is no significant mean deviation remaining. This is demonstrated in figure B.2. The prominent change of the shape of the two-dimensional shower-size spectrum in case of the events being located in KASCADE-Grande is due to the strongly overestimated number of electrons for events having energies below the energy threshold. That these are events below the threshold can be seen in figure B.3. On the l.h.s. the events with energies above the energy threshold are shown. The line corresponds to the selection applied in the N_e-N_μ plane, which is explained in section 5.2.3. On the r.h.s. of figure B.3 only events with combinations of $\log_{10}(N_e)$ and $\log_{10}(N_\mu)$ above this line are considered. Shown is the mean deviation of N_e from its true value. It is visible that the events that have passed the selection criteria but are not included in the selection shown on the l.h.s. of figure B.3 are the ones that show large deviations of N_e from its true value. These are the same events that populate – after having applied the correction – the region below the "Line of full efficiency" on the r.h.s. of figure B.2.

Until now, only the mean deviations have been shown for two of the four dependencies. In the following the mean deviation, i.e. the bias in the reconstruction is shown together with the standard deviation (the statistical uncertainty of a single measurement; in the following named std) of the difference between the reconstructed and the true (simulated) value of N_e and N_μ respectively. The values of the mean and std of these differences before and after the corrections have been applied are compared together.

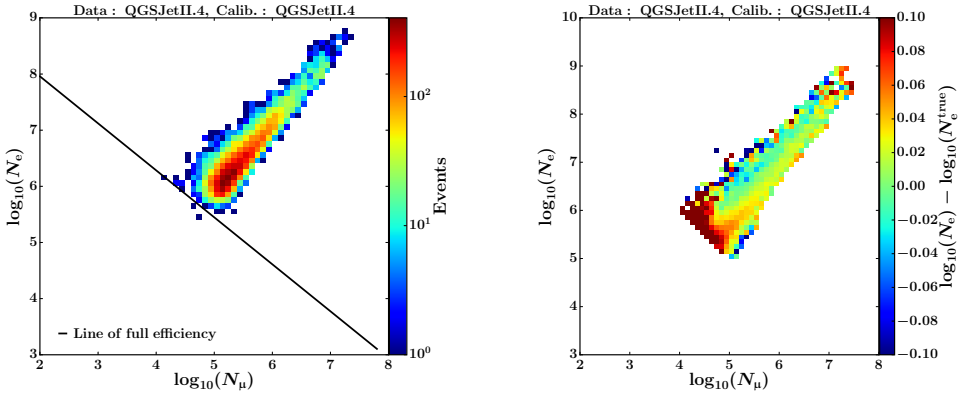

Figure B.3.: Left: The two-dimensional shower-size spectrum for events with energies above the threshold is shown. Events have been located in KASCADE-Grande. Right: The deviation of the reconstructed number of electrons from their true value is shown in the two-dimensional N_e-N_μ plane for events located in KASCADE-Grande.

This comparison is shown only for showers located in KASCADE-Grande, as the needed corrections are more extensive for KASCADE-Grande than they are for KASCADE, which can already be seen in figure B.1.

Figure B.4 shows the mean and std of the deviations of N_e (labeled "Not Corr.") and N_e^{corr} (labeled "Corrected") from N_e^{true} as a function of N_e^{true}. In both cases the same mixed composition of 20 percent each of H, He, C, Si, and Fe was assumed. Apart from a slightly reduced imprecision towards larger number of electrons, the correction does not significantly change the mean and std as a function of N_e^{true}. Applying the correction to data simulated with the EposLHC model (r.h.s.) yields essentially the same results as applying them to the data simulated with QGSJetII4 (l.h.s.), the model that was used to obtain the correction.

However, a clear improvement compared to the non-corrected values becomes visible when comparing the mean and std of the deviations of N_e and N_e^{corr} from N_e^{true} as a function of the distance of the shower core to the center of KASCADE in coordinates of the shower plane (Shower Plane Coordinates – SPC) and the zenith angle. This is shown in figures B.5 and B.6, respectively for the same mixed composition. A similar performance is obtained for both hadronic interaction models, especially in terms of the standard deviation. The means are slightly different, however, this is expected. While the composition assumption is the same, the different description of the shower development results in already different non-corrected values.

Figure B.7 shows the mean and std of the deviations of N_μ (labeled "Not Corr.") and N_μ^{corr} (labeled "Corrected") from N_μ^{true} as a function of N_μ^{true}. The mean values are changed mainly at higher numbers of muons while the standard deviation is improved especially at lower number of muons. Again, applying the correction to data simulated with the EposLHC model (r.h.s.) yields essentially the same results as applying them to the data simulated with QGSJetII4 (l.h.s.).

When viewed as a function of the distance in SPC (Fig. B.8), the increasing bias to a larger reconstructed number of muons is removed. This is the case for both models

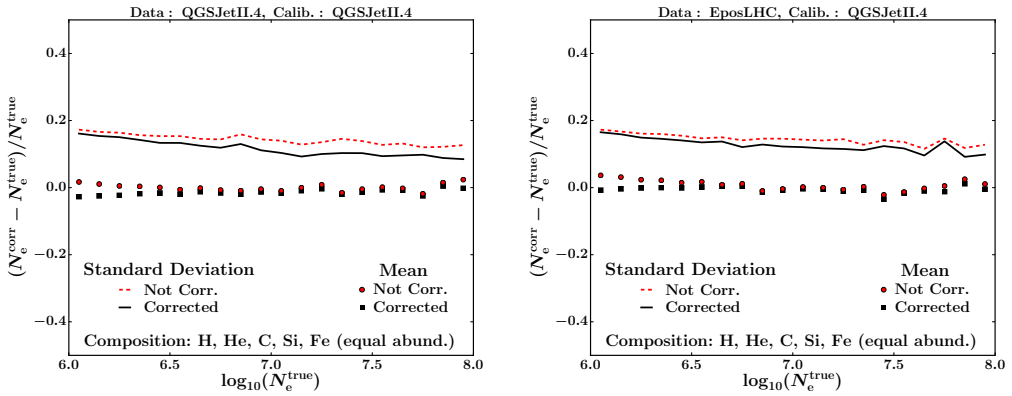

Figure B.4.: The mean and std of the deviation of the number of electrons from its true value is shown as a function of the true number of electrons. The original, non-corrected values are shown in comparison to the corrected ones. Left: The data has been simulated with QGSJetII4. Right: The data has been simulated with EposLHC. Both: Corrections have been obtained using QGSJetII4. A composition of 20 percent each of H, He, C, Si, and Fe has been assumed.

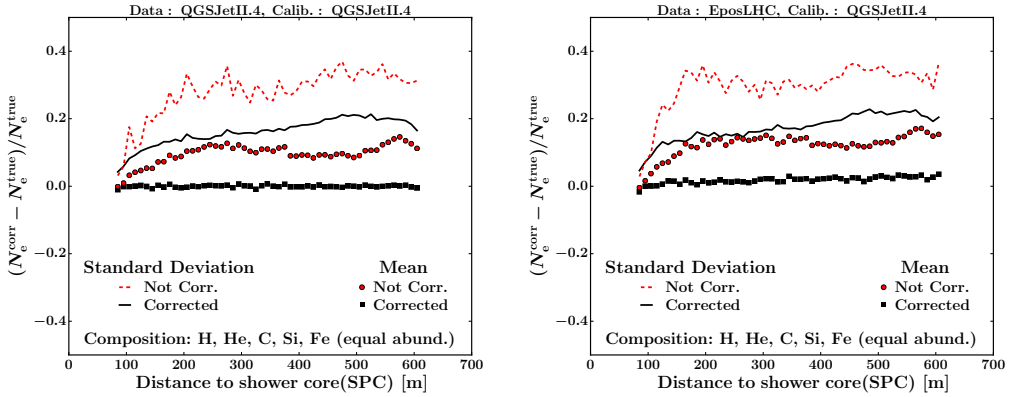

Figure B.5.: The mean and std of the deviation of the number of electrons from its true value is shown as a function of the distance of the shower core to the KASCADE array in shower coordinates. The original, non-corrected values are shown in comparison to the corrected ones. Left: The data has been simulated with QGSJetII4. Right: The data has been simulated with EposLHC. Both: Corrections have been obtained using QGSJetII4. A composition of 20 percent each of H, He, C, Si, and Fe has been assumed.

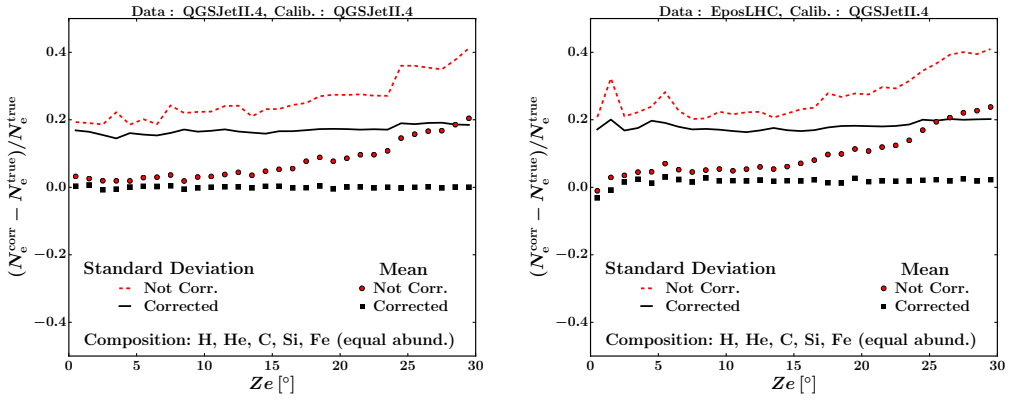

Figure B.6.: The mean and std of the deviation of the number of electrons from its true value is shown as a function of the zenith angle. The original, non-corrected values are shown in comparison to the corrected ones. Left: The data has been simulated with QGSJetII4. Right: The data has been simulated with EposLHC. Both: Corrections have been obtained using QGSJetII4. A composition of 20 percent each of H, He, C, Si, and Fe has been assumed.

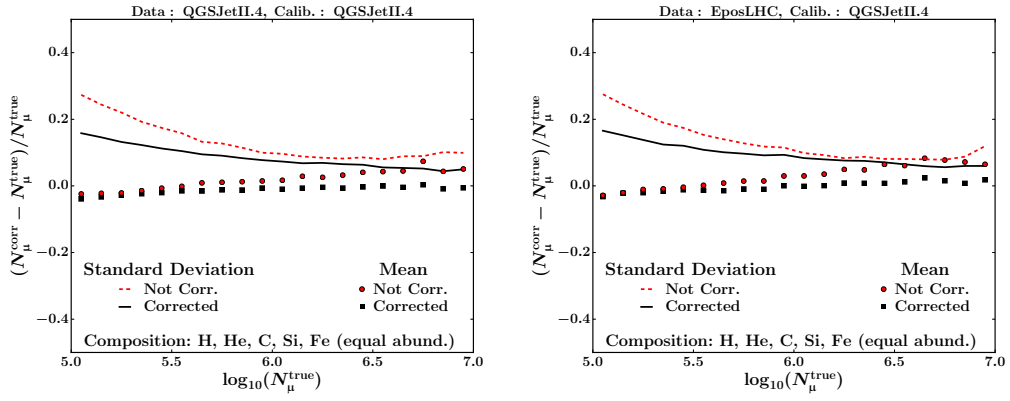

Figure B.7.: The mean and std of the deviation of the number of muons from its true value is shown as a function of the true number of muons. The original, non-corrected values are shown in comparison to the corrected ones. Left: The data has been simulated with QGSJetII4. Right: The data has been simulated with EposLHC. Both: Corrections have been obtained using QGSJetII4. A composition of 20 percent each of H, He, C, Si, and Fe has been assumed.

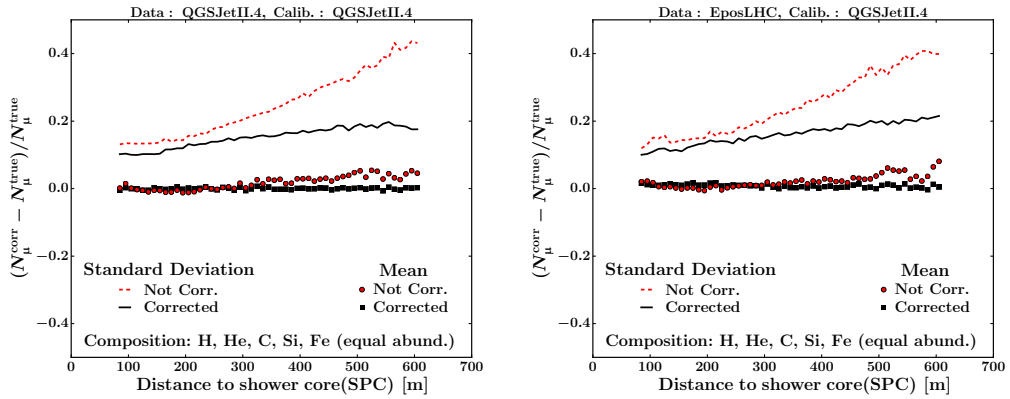

Figure B.8.: The mean and std of the deviation of the number of muons from its true value is shown as a function of the distance of the shower core to the KASCADE array in shower coordinates. The original, non-corrected values are shown in comparison to the corrected ones. Left: The data has been simulated with QGSJetII4. Right: The data has been simulated with EposLHC. Both: Corrections have been obtained using QGSJetII4. A composition of 20 percent each of H, He, C, Si, and Fe has been assumed.

(QGSJetII4: l.h.s., EposLHC: r.h.s.). Also the precision has improved significantly towards larger distances.

The deviations of N_μ and N_μ^{corr} from N_μ^{true} are shown as a function of the zenith angle in figure B.9. There is no dependence of the deviations on the zenith angle visible. The bias and precision have both been improved by the correction, however, independently of the zenith angle. Also as a function of the zenith angle, the performance of the correction are very similar for the two hadronic interaction models.

Concluding, the corrections of N_e and N_μ show satisfying performances. The slightly different values obtained for EposLHC can be attributed to the differences among the models, since they are already visible in the non-corrected values. Hence, there are no significant indications that the estimators suffer from overfitting the data simulated with QGSJetII4.

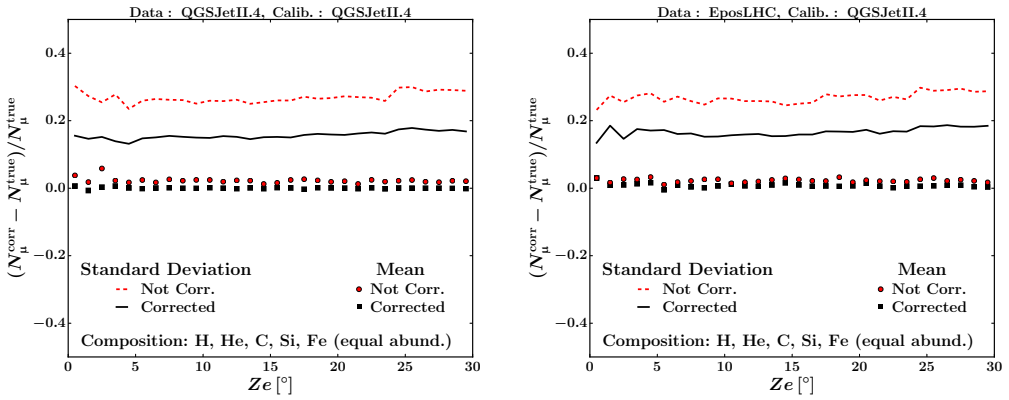

Figure B.9.: The mean and std of the deviation of the number of muons from its true value is shown as a function of the zenith angle. The original, non-corrected values are shown in comparison to the corrected ones. Left: The data has been simulated with QGSJetII4. Right: The data has been simulated with EposLHC. Both: Corrections have been obtained using QGSJetII4. A composition of 20 percent each of H, He, C, Si, and Fe has been assumed.

C. Muonic Component - Measurements vs Simulation

The differences visible in the reconstructed composition for events located inside KASCADE and the one found for events with core positions within KASCADE-Grande can only be partly addressed to the mass dependent misreconstructions that are slightly different for events located in the two detector arrays. An additional contribution is assumed to be due to the lateral distribution of muons being steeper in the measured data compared to the simulations. Figure C.1 illustrates how an lateral density function that fits to the simulated data but is too flat compared to measured data influences the reconstructed number of muons.

The black line corresponds to an arbitrary lateral density function assumed to be the true one, i.e. the one that is realized in nature. In addition, a flatter lateral density function (one that is predicted by simulations) is shown, which has been shifted up/down in order to match the lower/upper part of the distance range. In contrast to the reconstruction of N_e, the slope of the lateral density function is a fixed parameter in the reconstruction of the number of muons. Since only KASCADE was equipped with shielded detectors, only a small part of the shown distance range is covered by sampling points for the fit. This results in the shown overestimation of the number of muons in the case that the shower was near the center of KASCADE. In the event that the core is closer to the far edge of KASCADE-Grande, the number of muons is underestimated. This behavior can nicely explain the observation that for events located in KASCADE a heavier composition is reconstructed.

A simple way to check if this scenario is plausible [1] is based on the assumption that the same flux in the muon spectra, measured for different radii bins around the center of KASCADE, corresponds to the same primary energy and composition. If the lateral density function is well described by the simulations, the showers should yield the same mean number of muons, assuming that the number of events is high enough so that shower-to-shower fluctuations can be ignored.

[1] The idea was suggested in discussion with Juan Carlos Arteaga and Hugo Medina Guzman from Morelia, Mexico

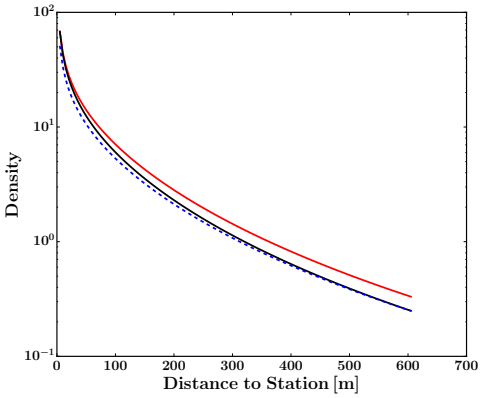

Figure C.1.: Comparison of different lateral density functions. The first one (black line) is compared to a flatter distribution, which is shifted up/down to match the lower end (red, solid) and the upper end (blue, dashed), respectively.

The muon spectra obtained for several distance ranges are shown in figures C.2 and C.3 based on the calibrations using QGSJetII4 and EposLHC, respectively. On the l.h.s. the spectra of the corresponding simulations are shown, on the r.h.s. the spectra of measured data based on the same model are displayed [2]. In addition lines of constant intensity are added. These lines are used to find the corresponding number of muons for each distance range.

Taking into account the decreasing number of events towards larger N_μ values, both sets of simulations show a relatively flat evolution of the mean number of muons with the distance. At low energies, this can also be said about the measured data, however, at larger values of $\log_{10}(N_\mu)$, a trend to a decreasing number of muons as a function of the distance gets visible. This indicates, in addition to the heavier composition reconstructed with KASCADE that the lateral density function for muons is fixed to be too flat, when compared to measured data, while the parameterization works for simulated data.

As this is only meant to be a quick cross check some interesting aspects can be derived. A study of the influence of the composition would help with the decision, whether the better description at lower energies is due to the different composition (the slope being fixed could result in a better description of the lateral density function for light/heavy primaries compared to the respective other particles) or if the shape being parameterized with N_e results in a better overall compatibility with the measured lateral density function.

As can be seen from figure C.1, covering a certain distance range for the fit of the lateral density function will minimize the overall over-/under-estimation of the number of muons. Within a KASCADE-Grande analysis [3] it was found that figure C.4 and figure C.5 can be used to correct to some extent for the effect of the unknown lateral density function. It was noticed that, almost independent of the degree by which the assumed lateral density function is too flat or too steep, there exists a

[2] Based on a model in the sense that the muon correction based on the respective model was used.

[3] publication in preparation

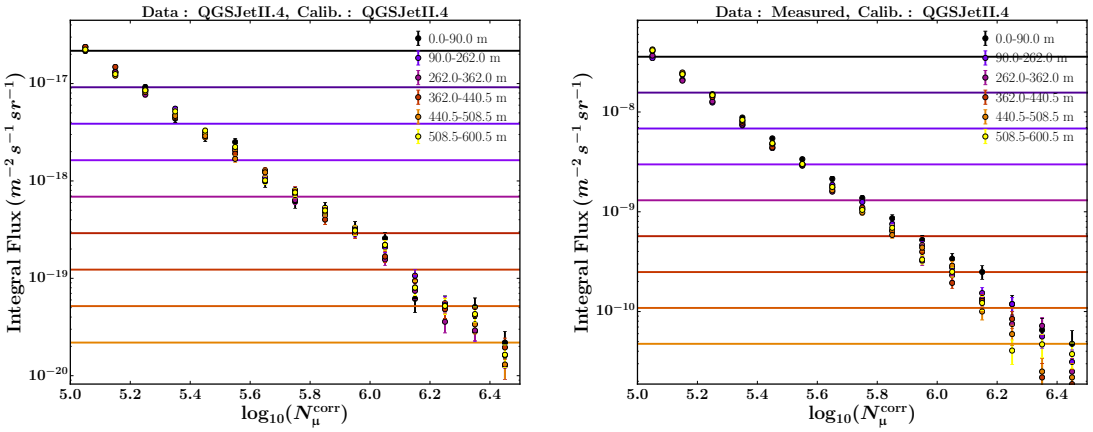

Figure C.2.: The integrated flux of muons measured within several distance ranges are shown together with lines of constant intensity. Left: Simulated data based on QGSJetII4. Right: Measured data, calibrations based on QGSJetII4.

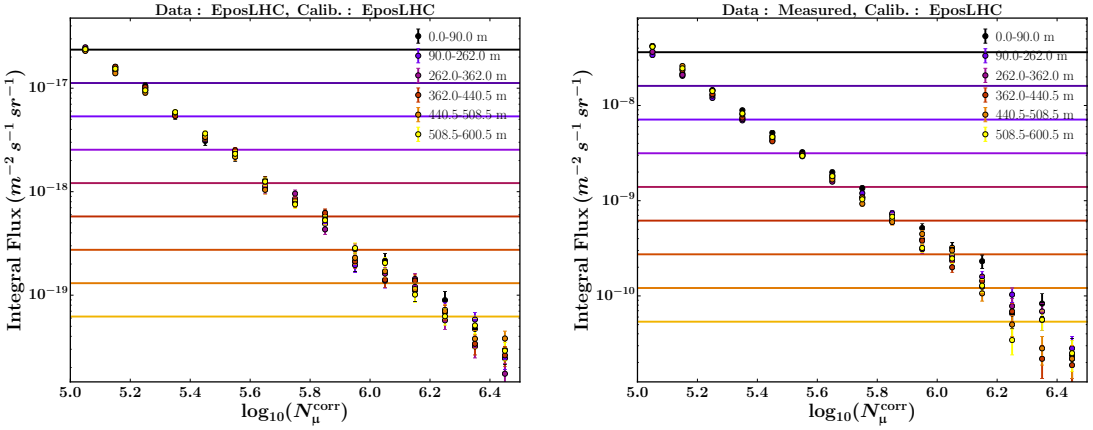

Figure C.3.: The integrated flux of muons measured within several distance ranges are shown together with lines of constant intensity. Left: Simulated data based on EposLHC. Right: Measured data, calibrations based on EposLHC.

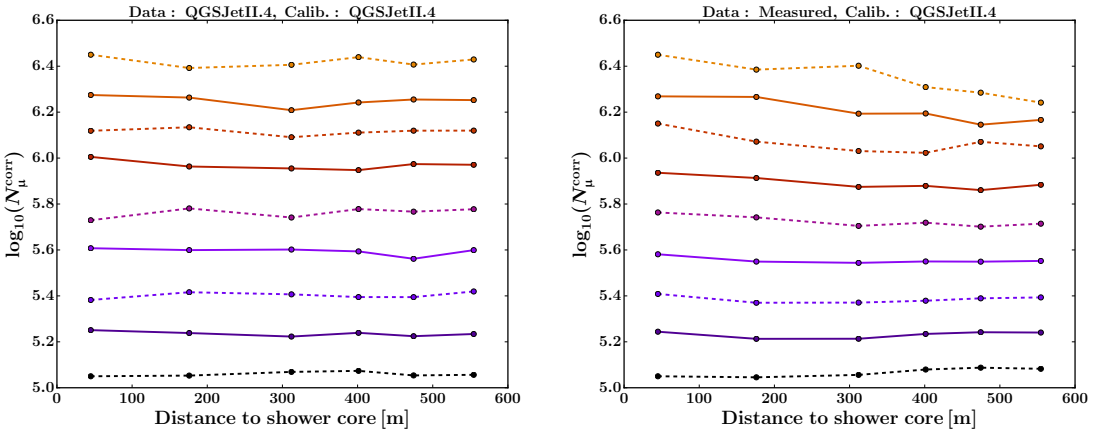

Figure C.4.: The integrated flux of muons measured within several distance ranges are shown together with lines of constant intensity. Left: Simulated data based on QGSJetII4. Right: Measured data, calibrations based on QGSJetII4.

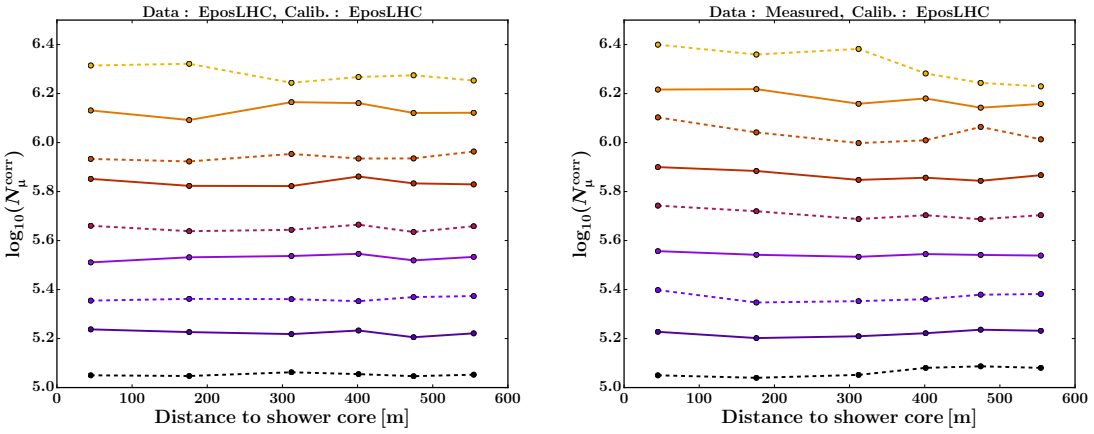

Figure C.5.: The integrated flux of muons measured within several distance ranges are shown together with lines of constant intensity. Left: Simulated data based on EposLHC. Right: Measured data, calibrations based on EposLHC.

common line in the $\log_{10}(N_\mu)$ vs Distance plane that minimizes the bias of the reconstructed number of muons, when $\log_{10}(N_\mu)$ at the intersection of that line with the corresponding line of figure C.4 is used instead of the reconstructed value. The procedure to find the correct intersection would be to identify the line of figure C.4 that corresponds to the reconstructed number of muons and distance and follow this line until the common line in the $\log_{10}(N_\mu)$ vs distance plane is reached.

Of course, such a correction would be based on a more sophisticated approach for getting the distributions shown in figure C.4, that at least takes the uncertainties of the muon flux into account.

D. Alternative Model-based Approach for Obtaining the Composition

Total $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ approach: Implementation and simulation based test

In this approach, the entire populations are fitted at once, instead of iteratively as described in section 5.3.3. Note that the common basis of the two approaches are described already in section 5.3.3.

The main problem with this alternative approach is that only the total $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions enter into the fit, which will effectively result in the fit being dominated by low-energy events, while for the rest only the restriction according to the energy spectrum is of importance, which, on its own, is a much weaker restriction. In addition, it was found during early tests that this method tends to introduce too many populations, which is partly due to the lack of influence of the high-energy events to the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution. This effect is taken care of by repeating the fitting procedure for several assumptions on the maximum number of populations for the various primaries. The assumption giving the result that best describes the measured data is chosen for the final result. These assumptions are:

3-3-3-3-3 populations for H-He-C-Si-Fe: Two galactic and one extra-galactic component with proton cutoffs above 10^{15} eV for each primary. Possibly an additional CREAM component, depending on whether the "CREAM-cutoff" is below 10^{15} eV.

2-2-2-2-2 populations for H-He-C-Si-Fe: One galactic and one extra-galactic component with proton cutoffs above 10^{15} eV for each primary. This is the case where the additional galactic component B proposed by Hillas is not needed. Possibly an additional CREAM component, depending on whether the "CREAM-cutoff" is below 10^{15} eV.

3-2-2-2-2(3) populations for H-He-C-Si-Fe: Two galactic components with proton cutoffs above 10^{15} eV for each primary and an additional extra-galactic proton (and

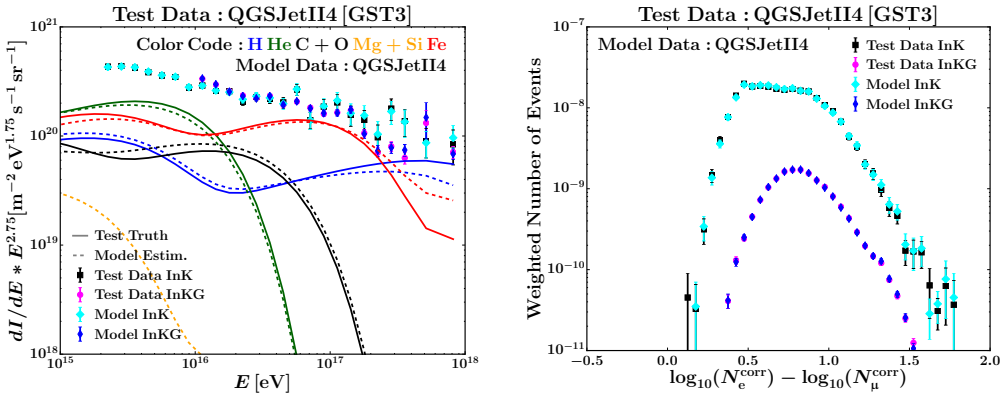

Figure D.1.: The same information as shown on the r.h.s. of figure 5.36 is shown on the l.h.s. for a "CREAM-cutoff" fixed at $10^{14.10}$ eV using the simultaneous fit method. The corresponding total $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions are shown on the r.h.s. separately for events located in KASCADE and KASCADE-Grande. The test composition corresponds to the GST3 model predictions. The maximum number of populations is 2-1-1-1-2.

iron) component. Possibly an additional CREAM component, depending on whether the "CREAM-cutoff" is below 10^{15} eV.

2-1-1-1-1(2) populations for H-He-C-Si-Fe: One galactic component with proton cutoff above 10^{15} eV for each primary and an additional extra-galactic proton (and iron) component. Possibly an additional CREAM component, depending on whether the "CREAM-cutoff" is below 10^{15} eV.

Using this approach for obtaining a population based model, some key points have to be rechecked, namely:

Q1) Assuming that the first population can be estimated using CREAM data, can the method be used to successfully reconstruct the true spectra?

Q2) Is the method able to correctly infer the necessity of a population below the covered energy range?

Q3) Assuming the previous requirements are met, what happens if the measured data is better described by simulations using a different hadronic interaction model?

Q4) Does the method successfully recover from the first population being obtained using CREAM data, should the assumptions regarding the latter be invalid?

In order to address questions Q1) to Q3), the GST3 model will be used to generate the test data. The model itself will be fit using QGSJetII4 based simulations. In case of point Q3), the test data will be generated using EposLHC instead of QGSJetII4. Question Q4) will be addressed using the H4a model instead of the GST3 model, using QGSJetII4 based simulations for both the test data and the model.

While the first method gives more accurate results, also the second method is able to provide an estimate that is close to the true composition (l.h.s. of Fig. D.1). Except for a small contribution of Mg+Si, which is forced by the CREAM data, no additional

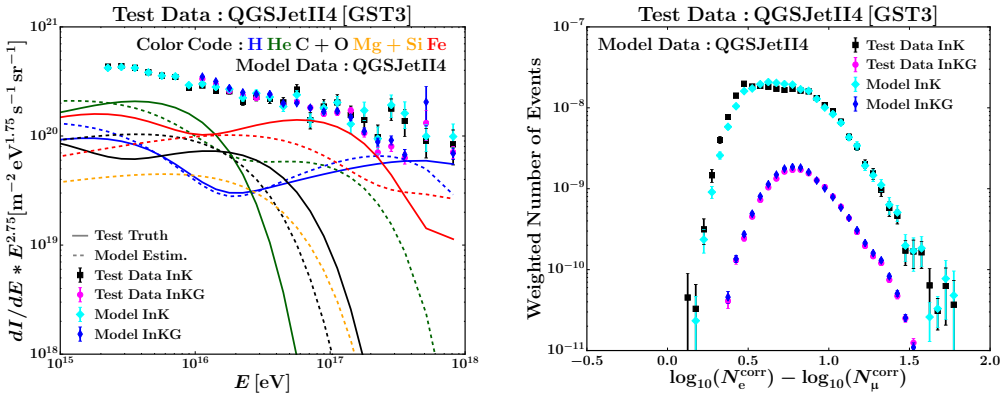

Figure D.2.: The same information as shown on the r.h.s. of figure 5.36 is shown on the l.h.s. for a "CREAM-cutoff" fixed at $10^{15.40}$ eV using the simultaneous fit method. The corresponding total $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions are shown on the r.h.s. separately for events located in KASCADE and KASCADE-Grande. The test composition corresponds to the GST3 model predictions. The maximum number of populations is 3-2-2-2-3.

contribution is added. In general, the method also suffers from the lack of influence of the higher energetic events to the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions. The overestimated proton flux and - in principle - the underestimated iron flux at low KASCADE energies is reflected in the corresponding $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution shown on the r.h.s. of figure D.1. The two dominating components at about 10^{16} eV are well described, resulting in a good description of the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution for events located inside KASCADE-Grande. Due to the larger statistics (because of the significantly larger area, despite the lower flux), the fit seems to be dominated by the KASCADE-Grande measurements, because, at that energy, the deviations in case of KASCADE dominate mainly in the low-statistics region of the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution. Interestingly, also the result using the simultaneous fit shows an overestimation of the iron flux at the largest covered energies. This is indirectly caused by the fit being dominated by the estimate at about 10^{16} eV. In order to get a good estimate of the iron flux at that energy, the too steep energy spectrum of the first population of iron has to be combined with a harder contribution to the second population and fine-tuned with the third population. Even though the cutoffs are roughly at the correct energy, the too hard spectrum cannot match the steeper fall of the iron spectrum. The proton spectrum, which is sub-dominant at the relevant energy of about 10^{16} eV, has a too steep third population with a too early cutoff, which is enforced to compensate the overestimation of the iron flux. Correspondingly, to provide enough events at high energies, this component, being too steep, has a too large contribution at lower energies. Therefore, this component adds to the overestimation of the proton flux at lower energies, where it is sub-dominant and, hence, not restricted enough by the KASCADE measurements.

What happens if the "CREAM-cutoff" is shifted to too high energies is shown in figure D.2. The iron flux, which is the second largest contribution to the all-particle flux up to about $10^{16.1}$ eV, is strongly underestimated. This is immediately visible in the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution for events located in KASCADE. The reason

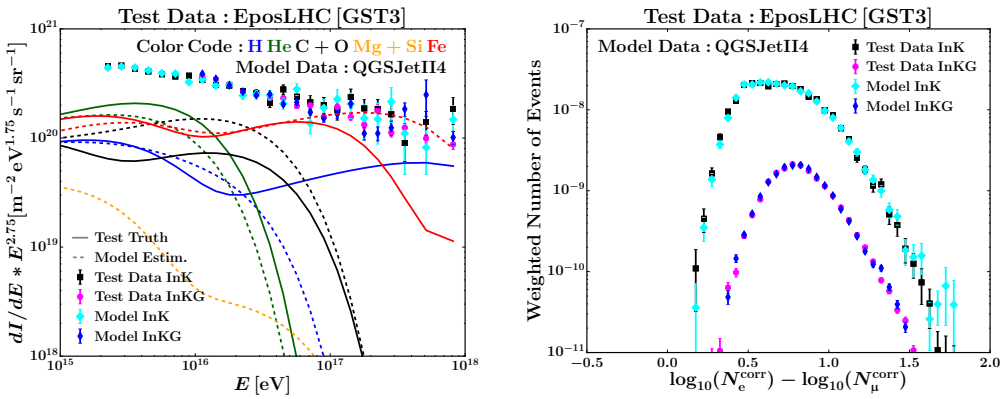

Figure D.3.: The same information as shown on the r.h.s. of figure 5.36 is shown on the l.h.s. for a "CREAM-cutoff" fixed at $10^{14.20}$ eV using the simultaneous fit method. The corresponding total $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions are shown on the r.h.s. separately for events located in KASCADE and KASCADE-Grande. The test composition corresponds to the GST3 model predictions. The maximum number of populations is 2-1-1-1-2. The target is obtained using EposLHC simulations, the model is fitted using QGSJetII4 simulations.

is again the dominating influence of the events located in KASCADE-Grande combined with the iron cutoff of the first population being beyond 10^{16} eV. Therefore, in order to stay compatible with CREAM at lower energies and the iron contribution at about 10^{16} eV, a flat spectrum is needed, which cannot match the iron flux at low KASCADE energies. The stronger influence of the CREAM data in case of C+O forces a too strong contribution of C+O, which is the main reason for the overestimation in the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions at medium $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ values. Although a "fake" Mg+Si contribution is enforced by the CREAM data, it does not have a strong influence on the result, because its fraction of the all-particle flux is quite small. In any case, a solution with an assumed cutoff of the first population at a too high energy has been successfully discarded.

Q3) is addressed in figure D.3. In general, the spectra and the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions are reasonably well matched. The apparent shift of the QGSJetII4 based $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions to larger $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ values relative to the EposLHC based ones can be attributed to the different muon content predicted by Epos together with the energy being interpreted by QGSJetII4 based calibrations. The estimated composition is too heavy compared to the true one, which is to be expected taking into account the results presented in section 5.3.2, i.e. the lighter composition obtained using EposLHC based calibrations to interpret the measured events.

When using the H4a based target (Fig. D.4), the problem of the too weak influence of the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions and them being dominated by the corresponding low-energetic events gets obvious. As shown for the first method in figure 5.38, a low enough "CREAM-cutoff", which is, in this case, at $10^{14.10}$ eV, should make a recovery possible, should the true cutoff be at higher energies. However, the first method strongly restricts the possible solutions by including the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$

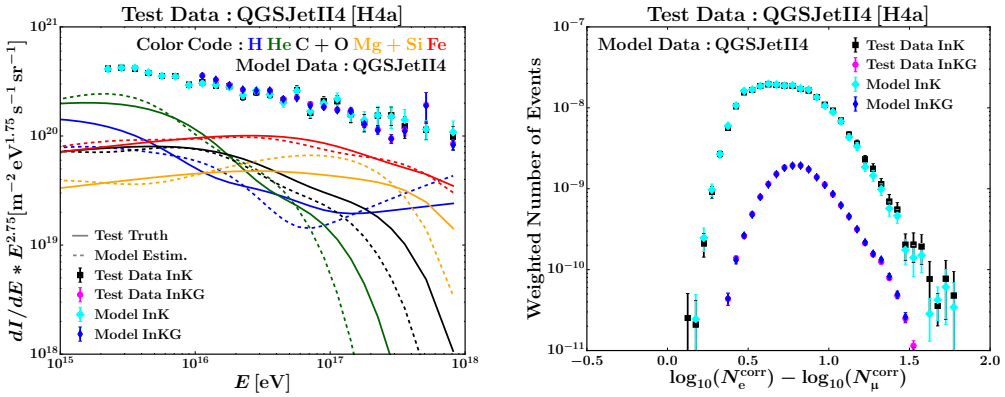

Figure D.4.: The same information as shown on the r.h.s. of figure 5.36 is shown on the l.h.s. for a "CREAM-cutoff" fixed at $10^{14.10}$ eV using the simultaneous fit method. The corresponding total $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions are shown on the r.h.s. separately for events located in KASCADE and KASCADE-Grande. The test composition corresponds to the H4a model predictions. The maximum number of populations is 3-2-2-2-3.

distributions for various energy bins. Good results are obtained near 10^{16} eV, which was already found to be the region that dominates the fitting procedure. Correspondingly, the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution for events located in KASCADE-Grande is quite well reconstructed, with a slightly to large contribution at high $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ values, which originate from the proton excess relative to the true flux. It is obvious that the populations are tuned to the energy spectra at lower and higher energies together with a reasonable description of the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution for events located in KASCADE. However, the significant lack of protons at energies below 10^{16} eV clearly results in a mismatch of the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution at larger $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ values.

Application to measured data

When applying this method to measured data, one has to consider that this method is dominated by the comparison of the energy spectra. The $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions are more sensitive to the composition, however, they are dominated by the primaries that have energies near to the energy threshold of the respective detector. Should the method fail to find a valid, i.e. plausible estimate of the energy spectrum and the mass composition, it will be revealed e.g. by a mismatch in one or both of the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions if the estimate of the composition is wrong at the corresponding energy threshold or by a mismatch between the estimated and measured energy spectra.

At energies between the energy thresholds of the two detectors, there is no room for non-physical combinations of populations that succeed to describe well both measured $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions and the measured energy spectra for KASCADE and KASCADE-Grande [1]. Otherwise an extremely hard energy spec-

[1] The difference between KASCADE and KASCADE-Grande for measured data is ignored. The method is derived, of course, on the assumption that the simulated data is compatible to the measured data.

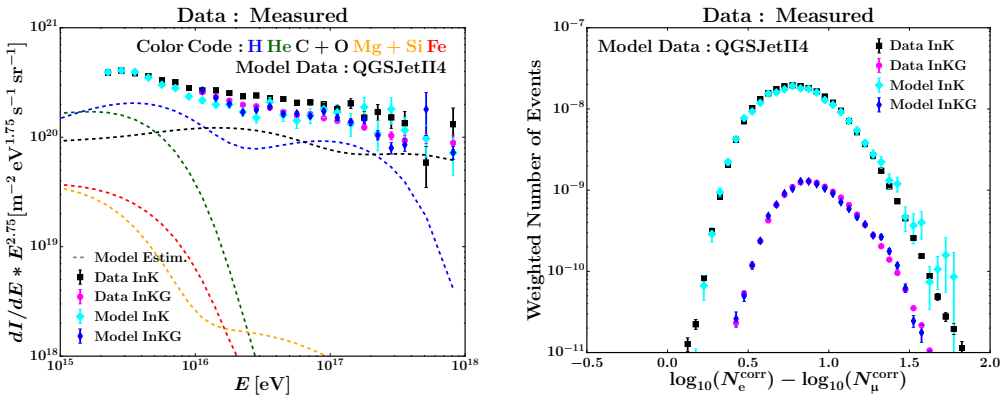

Figure D.5.: The same information as shown on the r.h.s. of figure 5.36 is shown on the l.h.s. for a "CREAM-cutoff" fixed at $10^{14.20}$ eV using the simultaneous fit method. The corresponding total $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions are shown on the r.h.s. separately for events located in KASCADE and KASCADE-Grande. The target corresponds to the measured data. The model is based on QGSJetII4 simulations. The maximum number of populations is 3-3-3-3-3.

trum with a cutoff between the two energy thresholds [2] would have to be compensated by a corresponding drop of the flux of the other primaries in order to keep the energy spectra from being different from the estimate. In order to match the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution for KASCADE-Grande an additional population would be needed to change the composition again. Such a behavior cannot be described by the model, because the number of populations is limited and the cutoffs of the various elements that contribute to a certain population are connected by their charge. However, this scenario is highly unlikely to be realized in nature, hence, if both measured $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions and the measured energy spectra for KASCADE and KASCADE-Grande are described well, it is safe to assume that the model is valid up to energies of a few 10^{16} eV. Beyond an energy of 10^{16} eV, the fitting procedure of the method relies only on the comparison of the energy spectra. Hence, the result has to be judged by how plausible it is, e.g. a change of the contribution of one element compensated by corresponding changes of the contributions of the other elements might hint at a deviation from the true composition. Another example would be the prediction of unnaturally hard spectra that cannot be explained even by exotic assumptions for the corresponding sources of the component.

Based on the assumption that the population based approach will either work or reveal inconsistencies as described above, the method is applied to measured data.

Figures D.5 and D.6 show the results for the measured data using simulations based on QGSJetII4 and on EposLHC, respectively.

As shown before, the KASCADE and KASCADE-Grande measurements are incompatible with simulations when compared together. This is again visible here. Using

[2] The spectrum has to be extremely hard, because it cannot contribute much to the energy spectrum at lower energies. The latter would be revealed by a mismatch of the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution for KASCADE.

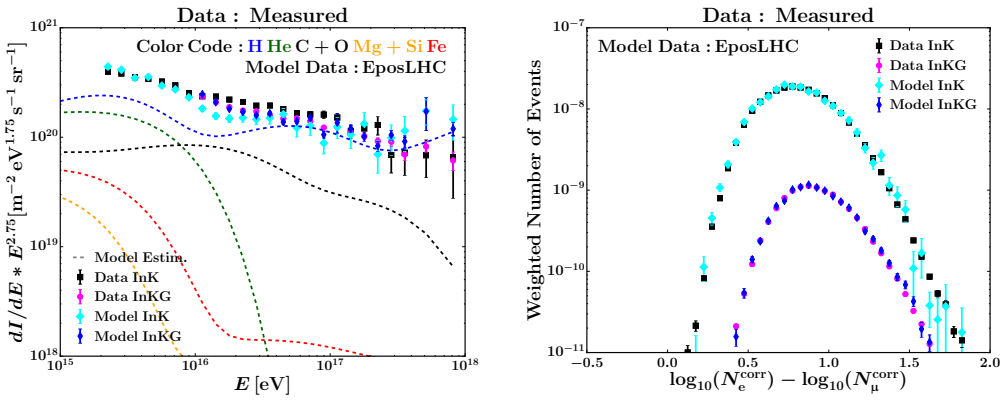

Figure D.6.: The same information as shown on the r.h.s. of figure 5.36 is shown on the l.h.s. for a "CREAM-cutoff" fixed at $10^{14.00}$ eV using the simultaneous fit method. The corresponding total $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distributions are shown on the r.h.s. separately for events located in KASCADE and KASCADE-Grande. The target corresponds to the measured data. The maximum number of populations is 3-2-2-2-2. The model is based on EposLHC simulations.

an EposLHC based target and QGSJetII4 for the model gives a reasonable result for KASCADE and KASCADE-Grande. In appendix B it is demonstrated that the muon correction based on QGSJetII4 works also for the EposLHC based simulations. Together with the information from appendix C this strongly suggests that the muon lateral density functions are more or less compatible among the models, but incompatible with the measured data. Assuming this is correct, the question whether there is a tendency for the number of muons near the shower core to be underestimated or the number of muons far away to be overestimated remains, as both (or a combination of both) result in a lateral density function too flat compared to the one realized in nature.

Because of the muons being measured only at KASCADE and the slope of the lateral density function being fixed, a too flat lateral density function will always result in a larger reconstructed number of muons for events inside KASCADE compared to events located in KASCADE-Grande. In fact, one can make an even stronger statement: N_μ will always be too large for events in KASCADE and too small for events located in KASCADE-Grande, as the measured data is the reference. The real question is, therefore, whether this effect is made worse (better) - by the interpretation with simulations - for events in KASCADE or for those in KASCADE-Grande.

The reason for this question is the following: Within the simulation, the lateral density function works for events within KASCADE and KASCADE-Grande alike [3]. Therefore, having measured the density close to the shower axis will give - within reconstruction uncertainties - the same N_μ as would be obtained if the shower would have hit KASCADE-Grande. For the following argument, it is assumed that the lateral density function is too flat and corresponds to the black, continuous line

[3] at least after the muon correction has been applied, which corrects for some assumptions made in the reconstruction that are only approximately correct.

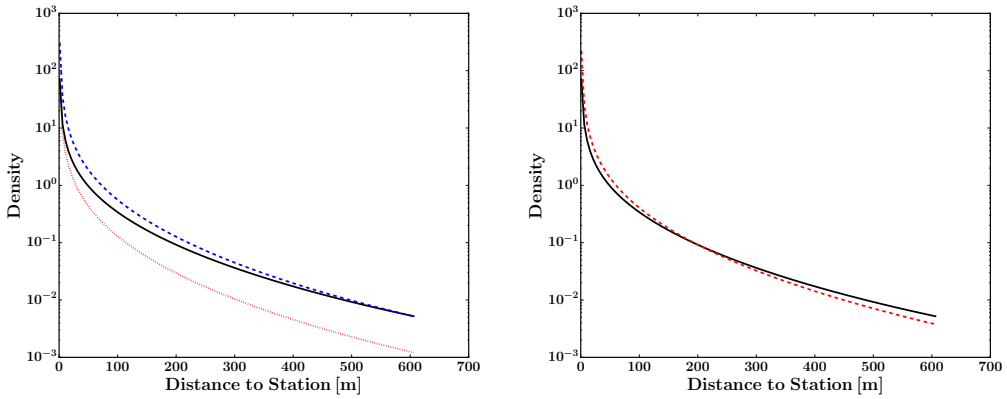

Figure D.7.: A fictive, flat lateral density function (black line) compared to a steeper lateral density function normalized to the black one at low distances (red line, l.h.s.), at medium distances (red line, r.h.s.), and at large distances (blue line, l.h.s.).

shown in figure D.7, which is a fictive lateral density function meant only for illustration. Assuming, in addition, that the lateral density function is too flat, because the density at low distances to the axis is underestimated and would - in truth, i.e. in nature - have values at the blue, dashed line, then, a shower hitting KASCADE will be reconstructed having a too small number of muons, i.e. the simulation should predict more muons. If the shower has hit KASCADE-Grande, the lateral density function compatible with the simulations is still too flat and the number of muons is also underestimated [4]. In this case, the overestimation of the measured N_μ for events within KASCADE will make the measured composition look even heavier as the simulation predicts too few muons. The underestimation for measured events within KASCADE-Grande will compensate to some extend the wrong predictions of the simulation.

On the other hand, assuming that the lateral density function is too flat, because the density is overestimated at distances farther away from the axis, the lateral density function realized in nature may correspond to the red, dash-dotted line. In this case, a shower hitting KASCADE-Grande will naturally result in an overestimated number of muons. Likewise, as the lateral density function compatible to the simulations is still too flat, a shower hitting KASCADE will also be reconstructed with a too large N_μ, i.e. the simulations should predict fewer muons. Therefore, the picture is reversed. While for events within KASCADE, the overestimated, measured N_μ will actually reduce the degree by which the estimated composition is too heavy, the composition obtained for events within KASCADE-Grande will seem to be even lighter.

Taking a look again at figures D.5 and D.6, the estimated composition, being dominated by the predictions of KASCADE-Grande seems to be very light (even ignoring

[4]The difference to the measured events is that for simulations, always the black, simulation based lateral density function is reconstructed, independent of the location of the core of the shower (the red/blue lines remain the true distributions realized in nature), while for measured data the true lateral density function is steeper and, therefore, the normalization of the fitted lateral density function shifts down towards larger distances (again, the true lateral density function is still the true lateral density function)

the exact predicted fluxes, which might be unstable at high energies, as was shown in figure D.4). In fact, if contributions of Si and Fe would not be forced by the CREAM data, dropping these would not notably influence the result. The energy spectrum for events within KASCADE would need a significantly heavier composition to be reproduced successfully, because these would be reconstructed with a higher energy. Assuming that there should be a contribution of e.g. iron, one might argue that the predictions of KASCADE could be more accurate compared to the one based on KASCADE-Grande. In that case, the underestimation of the measured N_μ at KASCADE-Grande should be accompanied by a predicted N_μ that is too large, hence, resulting in a very light composition. Therefore, the muon density at larger distances would be overestimated.

However, one should not forget a few important points:

1) The degree by which the lateral density function is too flat is likely to vary with energy, even for a composition constant with energy. A varying composition could result in the fixed lateral density function being more (less) compatible with the measurements introducing another energy dependence.

2) As depicted on the r.h.s. of figure D.7, a simultaneous under- and over-estimation at KASCADE and KASCADE-Grande, respectively, is also possible. Again, independently of the distance to the muon detectors, always the black lateral density function would be reconstructed, while the red, dashed one would always correspond to the true lateral density function, i.e. the one that should be predicted by the simulations. Therefore, a certain balance of under- and over-estimation could - in principle - result in a too flat lateral density function, but a correct prediction of N_μ. Hence, the discrepancy between KASCADE and KASCADE-Grande could be attributed to the measured data having a steeper lateral density function only. However, it is not expected that this balance in the simulations would be achieved for all energies and this large difference would require the true lateral density function to be even steeper.

3) There are other factors which influence N_μ and also N_e other than the mere lateral distribution, such as the shower-to-shower fluctuations.

The most important point is the following:

4) This method gives a distorted picture. Taking a look at figure 5.39, using only a limited energy range and without restrictions on the composition enforced by the populations and the respective other detector, a largely abundant helium component is found to contribute also to the $\log_{10}(N_e)$-$\log_{10}(N_\mu)$ distribution of KASCADE-Grande. Accordingly, the proton and carbon fractions in figure D.5 should be lower and a contribution of iron should be introduced. While there is a significant mismatch between KASCADE and KASCADE-Grande, and while the discussion on the lateral density function remains largely valid, the differences between KASCADE and KASCADE-Grande are not as dramatic as implied by the results obtained using the method described in this appendix.

References

[1] V.F. Hess. "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten". In: *Phys. Z.* 13 (1912), p. 1084.

[2] W. Bothe and W. Kolhörster. "Das Wesen der Höhenstrahlung". In: *Z. Phys.* 56 (1929), pp. 751–777. DOI: 10.1007/BF01340137.

[3] B. Rossi. "Über die Eigenschaften der durchdringenden Korpuskularstrahlung im Meeresniveau". In: *Z. Phys.* 82 (1933), p. 151. DOI: 10.1007/BF01341486.

[4] P. Auger et al. "Extensive Cosmic-Ray Showers". In: *Rev. Mod. Phys.* 11 (3-4 July 1939), pp. 288–291. DOI: 10.1103/RevModPhys.11.288.

[5] D. J. Bird et al. "Evidence for correlated changes in the spectrum and composition of cosmic rays at extremely high energies". In: *Phys. Rev. Lett.* 71 (21 Nov. 1993), pp. 3401–3404. DOI: 10.1103/PhysRevLett.71.3401.

[6] P. Blasi. "Origin of Galactic Cosmic Rays". In: *Nucl. Phys. B, Proc. Suppl.* 239–240 (2013), p. 140. ISSN: 0920-5632. DOI: 10.1016/j.nuclphysbps.2013.05.023.

[7] R. Aloisio, V. Berezinsky, and A. Gazizov. "Transition from galactic to extragalactic cosmic rays". In: *Astropart. Phys.* 39–40 (2012), pp. 129–143. DOI: 10.1016/j.astropartphys.2012.09.007.

[8] M. Unger. "Cosmic Rays above the Knee". arXiv:0812.2763 [astro-ph]. 2008. URL: http://arxiv.org/abs/0812.2763.

[9] A. Aab et al. "Searches for Anisotropies in the Arrival Directions of the Highest Energy Cosmic Rays Detected by the Pierre Auger Observatory". In: *ApJ* 804.1 (2015), p. 15. DOI: 10.1088/0004-637X/804/1/15.

[10] J. Blümer, R. Engel, and J. R. Hörandel. "Cosmic rays from the knee to the highest energies". In: *Prog. Part. Nucl. Phys.* 63.2 (2009), pp. 293–338. DOI: 10.1016/j.ppnp.2009.05.002.

[11] P. Blasi. "Recent developments in cosmic ray physics". In: *Nucl. Phys. B, Proc. Suppl.* 256–257 (2014), pp. 36–47. ISSN: 0920-5632. DOI: 10.1016/j.nuclphysbps.2014.10.004.

[12] T. Antoni et al. “KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems”. In: *Astropart. Phys.* 24 (2005), pp. 1–25. DOI: 10.1016/j.astropartphys.2005.04.001.

[13] M. Aglietta et al. “The cosmic ray primary composition in the knee region through the EAS electromagnetic and muon measurements at EAS-TOP”. In: *Astropart. Phys.* 21.3 (2004), pp. 583–596. DOI: 10.1016/j.astropartphys.2004.01.005.

[14] M. Aglietta et al. “The cosmic ray primary composition between 10^{15} and 10^{16} eV from Extensive Air Showers electromagnetic and TeV muon data”. In: *Astropart. Phys.* 20.6 (2004), pp. 641–652. DOI: 10.1016/j.astropartphys.2003.10.004.

[15] W.D. Apel et al. “Kneelike Structure in the Spectrum of the Heavy Component of Cosmic Rays Observed with KASCADE-Grande”. In: *Phys. Rev. Lett.* 107.171104 (2011). DOI: 10.1103/PhysRevLett.107.171104.

[16] W. F. Hanlon. Last visited on 05.06.2016. URL: http://www.physics.utah.edu/~whanlon/.

[17] W.D. Apel et al. “Ankle-like feature in the energy spectrum of light elements of cosmic rays observed with KASCADE-Grande”. In: *Phys. Rev. D* 87.081101(R) (2013). DOI: 10.1103/PhysRevD.87.081101.

[18] K. Greisen. “End to the Cosmic-Ray Spectrum?” In: *Phys. Rev. Lett.* 16 (17 Apr. 1966), pp. 748–750. DOI: 10.1103/PhysRevLett.16.748.

[19] G. T. Zatsepin and V. A. Kuz’min. “Upper Limit of the Spectrum of Cosmic Rays”. In: *Sov. J. Exp. Theor. Phys. Lett.* 4 (1966), p. 78. URL: http://adsabs.harvard.edu/abs/1966JETPL...4...78Z.

[20] R.U. Abbasi et al. “Study of Ultra-High Energy Cosmic Ray composition using Telescope Arrays Middle Drum detector and surface array in hybrid mode”. In: *Astropart. Phys.* 64 (2015), pp. 49–62. DOI: http://dx.doi.org/10.1016/j.astropartphys.2014.11.004.

[21] A. Aab et al. “Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications”. In: *Phys. Rev. D* 90 (12 Dec. 2014), p. 122006. DOI: 10.1103/PhysRevD.90.122006.

[22] H. V. Klapdor-Kleingrothaus and K. Zuber. *Teilchenastrophysik*. ISBN: 3-519-03094-2. B.G. Teubner Stuttgart, 1997. ISBN: 3519030942.

[23] M. G. Aartsen et al. “Search for Prompt Neutrino Emission from Gamma-Ray Bursts with IceCube”. In: *ApJL* 805.1 (2015), p. L5. DOI: 10.1088/2041-8205/805/1/L5.

[24] M. S. Longair. *High Energy Astrophysics*. 3rd ed. ISBN: 978-0-521-75618-1. Cambridge University Press, Mar. 2011. ISBN: 9780521756181.

[25] E. Fermi. “On the Origin of the Cosmic Radiation”. In: *Phys. Rev.* 75 (8 Apr. 1949), pp. 1169–1174. DOI: 10.1103/PhysRev.75.1169.

[26] R. Alves Batista et al. “CRPropa 3 - a Public Astrophysical Simulation Framework for Propagating Extraterrestrial
Ultra-High Energy Particles”. arXiv:1603.07142 [astro-ph.IM]. 2016. URL: http://http://arxiv.org/abs/1603.07142.

[27] E. Khan et al. "Photodisintegration of ultra-high-energy cosmic rays revisited". In: *Astropart. Phys.* 23.2 (2005), pp. 191–201. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2004.12.007.

[28] V. N. Zirakashvili et al. "Propagation of ultra-high-energy cosmic rays in Galactic magnetic fields". In: *Astron. Lett.* 24 (1998), pp. 139–143.

[29] T.K. Gaisser. "Spectrum of cosmic-ray nucleons, kaon production, and atmospheric muon charge ratio". In: *Astropart. Phys.* 35 (2012), pp. 801–806. DOI: 10.1016/j.astropartphys.2012.02.010.

[30] S. Tilav T.K. Gaisser T. Stanev. "Cosmic Ray Energy Spectrum from Measurements of Air Showers". arxiv:1303.3565v1 [astro-ph.HE]. 2013. URL: http://arxiv.org/abs/1303.3565v1.

[31] A.M. Hillas. "Cosmic Rays: Recent Progress and some Current Questions". arXiv:astro-ph/0607109. 2006. URL: http://arxiv.org/abs/astro-ph/0607109v2.

[32] B. Peters. "Primary Cosmic Radiation and Extensive Air Showers". In: *Il Nuovo Cim.* 22.4 (1961), p. 800. DOI: 10.1007/BF02783106.

[33] H.S. Ahn et al. "Energy Spectra of Cosmic-ray Nuclei at High Energies". In: *ApJ* 707.1 (2009), p. 593. DOI: 10.1088/0004-637X/707/1/593.

[34] J. Matthews. "A Heitler model of extensive air showers". In: *Astropart. Phys.* 22 (2005), pp. 387–397. DOI: 10.1016/j.astropartphys.2004.09.003.

[35] P. K. F. Grieder. *Extensive Air Showers*. Vol. 1. ISBN: 978-3-540-76940-8. Springer, 2010.

[36] P. K. F. Grieder. *Extensive Air Showers*. Vol. 2. ISBN: 978-3-540-76940-8. Springer, 2010.

[37] D. Heck et al. "CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers". In: *Report FZKA* 6019 (1998). Last visited on 06.06.2016. URL: http://www-ik.fzk.de/corsika/.

[38] A. Fasso' et al. "The FLUKA code: present applications and future developments". arXiv:physics/0306162 [physics.comp-ph]. 2003. URL: http://arxiv.org/abs/physics/0306162.

[39] S. Ostapchenko. "QGSJET-II: towards reliable description of very high energy hadronic interactions". In: *Nucl. Phys. B, Proc. Suppl.* 151.1 (2006), pp. 143–146. DOI: 10.1016/j.nuclphysbps.2005.07.026.

[40] S. Ostapchenko. "Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model". In: *Phys. Rev. D* 83 (1 Jan. 2011), p. 014018. DOI: 10.1103/PhysRevD.83.014018.

[41] S. Ostapchenko. "QGSJET-II: physics, recent improvements, and results for air showers". In: *EPJ web conf.* 52 (2013), p. 02001. DOI: 10.1051/epjconf/20125202001.

[42] T. Pierog et al. "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider". In: *Phys. Rev. C* 92 (3 Sept. 2015), p. 034906. DOI: 10.1103/PhysRevC.92.034906.

[43] R. Brun and F. Carminati. "GEANT-Detector Description and Simulation Tool". In: *CERN Program Library Long Writeup* (1993).

[44] R. Ulrich, R. Engel, and M. Unger. "Hadronic multiparticle production at ultrahigh energies and extensive air showers". In: *Phys. Rev. D* 83 (5 Mar. 2011), p. 054026. DOI: 10.1103/PhysRevD.83.054026.

[45] T. Pierog and K. Werner. "Muon Production in Extended Air Shower Simulations". In: *Phys. Rev. Lett.* 101 (17 Oct. 2008), p. 171101. DOI: 10.1103/PhysRevLett.101.171101.

[46] A. E. Herve et al. "Results from Pion-Carbon Interactions Measured by NA61/SHINE for Improved Understanding of Extensive Air Showers". In: *PoS* ICRC2015.330 (2015).

[47] T. Pierog et al. "Future of Monte Carlo simulations of atmospheric showers". In: *EPJ web conf.* 89 (2015), p. 01003. DOI: 10.1051/epjconf/20158901003.

[48] W.D. Apel et al. "Muon production height studies with the air shower experiment KASCADE-Grande". In: *Astropart. Phys.* 34.6 (2011), pp. 476–485. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2010.10.016.

[49] W D Apel et al. "Test of interaction models up to 40 PeV by studying hadronic cores of EAS". In: *J. Phys. G: Nucl. Part. Phys.* 34.12 (2007), p. 2581. DOI: 10.1088/0954-3899/34/12/005.

[50] W. D. Apel et al. "A test of the hadronic interaction model EPOS with air shower data". In: *J. Phys. G: Nucl. Part. Phys.* 36.3 (2009), p. 035201. DOI: 10.1088/0954-3899/36/3/035201.

[51] T. Antoni et al. "Test of high-energy interaction models using the hadronic core of EAS". In: *J. Phys. G: Nucl. Part. Phys.* 25 (1999), pp. 2161–2175. URL: http://stacks.iop.org/0954-3899/25/i=10/a=313.

[52] T. Antoni et al. "Geometric structures in hadronic cores of extensive air showers observed by KASCADE". In: *Phys. Rev. D* 71 (7 2005), p. 072002. DOI: 10.1103/PhysRevD.71.072002.

[53] W. D. Apel et al. "New method to measure the attenuation of hadrons in extensive air showers". In: *Phys. Rev. D* 80 (2 2009), p. 022002. DOI: 10.1103/PhysRevD.80.022002.

[54] W.D. Apel et al. "Lateral distributions of EAS muons (800 MeV) measured with the KASCADE-Grande Muon Tracking Detector in the primary energy range – eV". In: *Astropart. Phys.* 65 (2015), pp. 55–63. DOI: 10.1016/j.astropartphys.2014.12.001.

[55] T. Antoni et al. "Muon Density Measurements with the KASCADE Central Detector". arXiv:astro-ph/0103363. 2001. URL: http://arxiv.org/abs/astro-ph/0103363.

[56] J. C. Arteaga-Velazquez et al. "Testing hadronic interaction models with the attenuation length of muons in KASCADE-Grande". In: *PoS* ICRC2015.314 (2015).

[57] T. Antoni et al. "The cosmic-ray experiment KASCADE". In: *Nucl. Instrum. Methods Phys. Res., Sect. A* 513.3 (2003), pp. 490–510. DOI: 10.1016/S0168-9002(03)02076-X.

[58] W.D. Apel et al. "Comparison of measured and simulated lateral distributions for electrons and muons with KASCADE". In: *Astropart. Phys.* 24.6 (2006), pp. 467–483. DOI: 10.1016/j.astropartphys.2005.10.001.

[59] K. Kamata and J. Nishimura. "The Lateral and the Angular Structure Functions of Electron Showers". In: *Prog. Theor. Phys. Suppl.* 6 (1958), pp. 93–155. DOI: 10.1143/PTPS.6.93.

[60] K. Greisen. "Cosmic Ray Showers". In: *Ann. Rev. Nucl. Sci.* 10 (1960), pp. 63–108. DOI: 10.1146/annurev.ns.10.120160.000431.

[61] N. N. Kalmykov and S. S. Ostapchenko. "The Nucleus-nucleus interaction, nuclear fragmentation, and fluctuations of extensive air showers". In: *Phys. Atom. Nucl.* 56 (1993), pp. 346–353.

[62] *KASCADE*. Last visited on 06.05.2016. URL: https://web.ikp.kit.edu/KASCADE/.

[63] W.D. Apel et al. "The KASCADE-Grande experiment". In: *Nucl. Instrum. Methods Phys. Res., Sect. A* 620 (2010), pp. 202–216. DOI: 10.1016/j.nima.2010.03.147.

[64] A.A. Lagutin and R.I. Raikin. "Lateral distribution of electrons in EAS at superhigh energies: predictions and experimental data". In: *Nucl. Phys. B, Proc. Suppl.* 97.1-3 (2001), pp. 274–277. ISSN: 0920-5632. DOI: 10.1016/S0920-5632(01)01282-8.

[65] W.D. Apel et al. "The KASCADE-Grande energy spectrum of cosmic rays and the role of hadronic interaction models". In: *Adv. Space Res.* 53.10 (2013). COSPAR 2012, Mysore, pp. 1456–1469. DOI: 10.1016/j.asr.2013.05.008.

[66] W.D. Apel et al. "KASCADE-Grande measurements of energy spectra for elemental groups of cosmic rays". In: *Astropart. Phys.* 47 (2013), pp. 54–66. DOI: 10.1016/j.astropartphys.2013.06.004.

[67] M. Finger. "Reconstruction of energy spectra for different mass groups of high-energy cosmic rays". PhD thesis. Karlsruher Institut für Technologie, 2011. URL: https://web.ikp.kit.edu/KASCADE/publication/PhD_Theses/M-Finger-Thesis.pdf.

[68] W.D. Apel et al. "The spectrum of high-energy cosmic rays measured with KASCADE-Grande". In: *Astropart. Phys.* 36.1 (2012), pp. 183–194. DOI: 10.1016/j.astropartphys.2012.05.023.

[69] *ANKA - Synchroton Radiation Facility*. URL: http://www.anka.kit.edu.

[70] M. Stümpert. "Suche nach Anisotropie in der kosmischen Strahlung mit KASCADE-Grande". PHD-Thesis. Universität Karlsruhe, 2007. URL: http://digbib.ubka.uni-karlsruhe.de/volltexte/1000007953.

[71] D.G. Watts and D.W. Bacon. "Using an Hyperbola as a Transition Model to Fit Two-Regime Straight-Line Data". In: *Technometrics* 16.3 (1974), pp. 369–373. DOI: 10.2307/1267666.

[72] G.V. Kulikov and G.B. Khristiansen. "On the size spectrum of extensive air showers". In: *J. Exptl. Theoret. Phys. (U.S.S.R.)* 35 (1958), p. 635. URL: http://www.jetp.ac.ru/cgi-bin/dn/e_008_03_0441.pdf.

[73] M. G. Aartsen et al. "Measurement of the cosmic ray energy spectrum with IceTop-73". In: *Phys. Rev. D* 88 (4 Aug. 2013), p. 042004. DOI: 10.1103/PhysRevD.88.042004.

[74] V.V. Prosin et al. “Primary CR energy spectrum and mass composition by the data of Tunka-133 array”. In: *EPJ web conf.* 99 (2015), p. 04002. DOI: 10.1051/epjconf/20159904002.

[75] *ROOT*. Last visited on 06.05.2016. URL: https://root.cern.ch.

[76] Y.S.Yoon et al. “Cosmic-Ray Proton and Helium Spectra from the first CREAM Flight”. In: *ApJ* 728.2 (2011), p. 122. DOI: 10.1088/0004-637X/728/2/122.

[77] T. Pierog. private conversation. Apr. 2016.

[78] A. Haungs. “Cosmic Rays from the Knee to the Ankle”. In: *Phys. Proc.* 61 (2015), p. 425. ISSN: 1875-3892. DOI: 10.1016/j.phpro.2014.12.094.

[79] R. Engel. “Air Shower Calculations With the New Version of SIBYLL”. In: *International Cosmic Ray Conference*. Vol. 1. International Cosmic Ray Conference. 1999, p. 415.

[80] *scikit-learn*. Last visited on 06.05.2016. URL: http://scikit-learn.org/stable/index.html.

Part II.

KASCADE Cosmic Ray Data Centre (KCDC)

1. Motivation

There are several reasons to motivate the development of the KASCADE Cosmic ray Data Centre (KCDC) [1–3]. The main purpose, of course, is to make the valuable data collected by KASCADE [4] available to colleagues as well as to the general public. This is common practice in astronomy, however, it is a novelty in astroparticle physics or in high-energy physics. Only recently data collected by the LHC experiments was made available publicly at the CERN Open Data Portal [5].

The publication of the data solves and avoids several problems that many experiments are facing right now or will be facing at some time in the future. One is connected to an important aspect of scientific work, which is the reproducibility of results. This can only be achieved if the data used for the analyses is kept available in addition to a thorough documentation of the detector itself and of the reconstruction procedures employed including software codes and meta data. For many past experiments the data are either lost or available only in a format that makes it hard or even impossible to read. In that sense KCDC is an archive for the data ensuring readability by supporting several commonly used output formats and it is a provider for the documentation needed to actually work with the data.

Being implemented as a web portal, KCDC addresses a key point of the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities" [6]:

"*Our mission of disseminating knowledge is only half complete if the information is not made widely and readily available to society. New possibilities of knowledge dissemination not only through the classical form but also and increasingly through the open access paradigm via the Internet have to be supported.*"

In addition to the publication of the data, it has always been a goal of KCDC to be used by pupils, students and - more general - the interested non-physicist. For this purpose, KCDC provides a section for educational tutorials.

When making data publicly available, the terms of usage have to be specified. For KCDC an "End User License Agreement (EULA)" has been developed (`https://kcdc.ikp.kit.edu/lawnorder`), which regulates the usage of both, the data and the web portal. This is necessary for example to clarify that the redistribution - especially of modified data - is not allowed and that publications based on the data

provided should follow good scientific practice. The latter is a main requirement for a meaningful discussion, should the result obtained by a user of KCDC disagree with the original work of the KASCADE collaboration.

The following description of KCDC (for the release named MERIDIAN) is split into three parts. The first part will cover the history of KCDC and the reasons to develop a new version. Next, the usage of the web portal by those who want to retrieve data or information is described and the usage by those who want to administrate the portal, i.e. the provider of the data will be briefly presented. The last part will focus on the implementation of the software, which was developed as part of this thesis.

2. MERIDIAN

The development of KCDC started in 2012 with a clear list of short-term goals in mind. The first topics focused to

- grant open access for the general public to a set of – initially 15 – fully calibrated quantities of high-quality air-showers measured with the KASCADE scintillator array,
- add a detailed description of the experiment and the data,
- develop analysis tutorials for teachers, pupils, and early stage students.

The first version of KCDC was released on 05/11/2013 after six months of intensive development. The status at that time was that roughly 160 million events have been released, stored in a MySQL database. Since the idea was to provide the means for the user to select a subset of this data, a "Data Shop" was developed together with the necessary processing structures (both hardware and software wise). This led to the introduction of user accounts, which are needed to manage the data requests and to provide the downloads.

Maybe underestimated at first, providing a detailed description of the data and the experiment turned out to be quite cumbersome and time consuming. When the experiment was operated and the reconstruction procedures were developed, no one had in mind that one day the data would be published. Therefore, with almost no detailed documentation at hand, it was a lot of work to gather the needed information.

The second large release of KCDC took place roughly a year after the first. The changes to the portal itself were mainly of cosmetic nature and extended documentation of the experiment was included, while the process of submitting a data request stayed largely the same. The reason it took so long to finalize the second version was the transition to a mongoDB to store the data. Since the data was stored in a MySQL database before, extensive changes to the data processing code have been necessary. In addition to the new database, three quantities were added to the list of calibrated quantities. Two of these have been recorded by the hadron calorimeter of

the central detector. The number of air-shower events provided to the public stayed the same.

Developed as a part of this thesis and described in detail in the next chapters, the third version – named MERIDIAN – was released on 22/03/2016. While the second version was largely based on the first one, MERIDIAN has been developed completely from scratch. There were several reasons to do so, all of them being connected to the long-term goals of KCDC. The most relevant goals in that regard were the inclusion of the remaining quantities of the KASCADE and KASCADE-Grande components and the publication of the software of KCDC as open-source. Especially having the latter in mind, a modular, plugin based design was chosen. Therefore, it is easy to extend and modify even key aspects of KCDC in order to adjust the installation to suit the needs of other experiments. In addition, the array-like quantities, such as particle densities at each station, have been added to the MERIDIAN release and the means to provide data for individual detector components has been included. Handling these aspects was not possible with the old “Data Shop”, hence a complete rewrite of the previous version was necessary.

3. The Web Portal

The entry page of the current KCDC version - called MERIDIAN - is shown in figure 3.1. It features two menus, the head menu on the top and the main menu on the left side, both being fully customizable within the administrator interface. The first two pages cover the motivation and legal regulations concerning the use of the portal as well as the data.

Figure 3.1.: The index page of the KCDC web portal (`https://kcdc.ikp.kit.edu/`).

The information pages contain documentation on KCDC, KASCADE and the data that is being published. The documentation on these pages is still under development, however, there is a manual [7] that can be downloaded in PDF format, which provides detailed descriptions of KCDC, KASCADE and the data.

There are pages for announcements, a section displaying frequently asked questions (FAQs), pages for the users to manage their account, for reporting a bug and - new in this release - for publications related to KCDC.

The main focus of this chapter is the description of the data retrieval process, which will be discussed in the next section.

3.1. The Data Shop

After an account has been created, the user gets access to the data shop, which can be used to either download preselected data or to create individual selections. When a request for a subset of the available data has been submitted, it is processed asynchronously in the background. The user will get a notification once the processing of the job has finished and the download of the data is kept available for two weeks. A detailed description of how individual selections are created is given in the following.

During its lifetime, KASCADE has accumulated several terabytes of data, therefore, it is not feasible to simply create one large file that contains all the data available and make it accessible on the website. KCDC provides two ways to reduce the total amount of data. The first one is to limit the number of returned observables by excluding single quantities or even entire detector components. In addition, the user can reduce the number of events by defining selection criteria on observables. This selection is done in the “Data Shop”. Its entry page in its current version is shown in figure 3.2.

In the left column, available detector components are listed, if they have not been selected yet. The selected components are listed in the second column. There are three pre-selection modes for a detector. Selected by default, deselected by default, and cannot be deselected (which means it is also selected by default as is the case for “Event Info”). The info box on the right side displays its default content, a description of the “Data Shop” and its usage. Its content will change if the user hovers over a detector or quantity, displaying a text shortly describing the detector or quantity.

On loading the page, only the available detectors are fetched from the server. The descriptions are fetched when needed. The same applies to the quantities. They are only requested once a detector has been selected, which happens automatically in case the detector is selected by default.

The third column contains a list of quantities belonging to the currently selected detector or detector-component. As for the detectors, a quantity can be set to be selected or deselected by default or as always selected. The latter will always be shipped to the user. This is the case for the run and event number, which uniquely identify an event.

In addition to the selection of quantities, the third column is also used by the user to add selection criteria that are to be applied during the processing of the request. In the case displayed in figure 3.2 there are four such criteria.

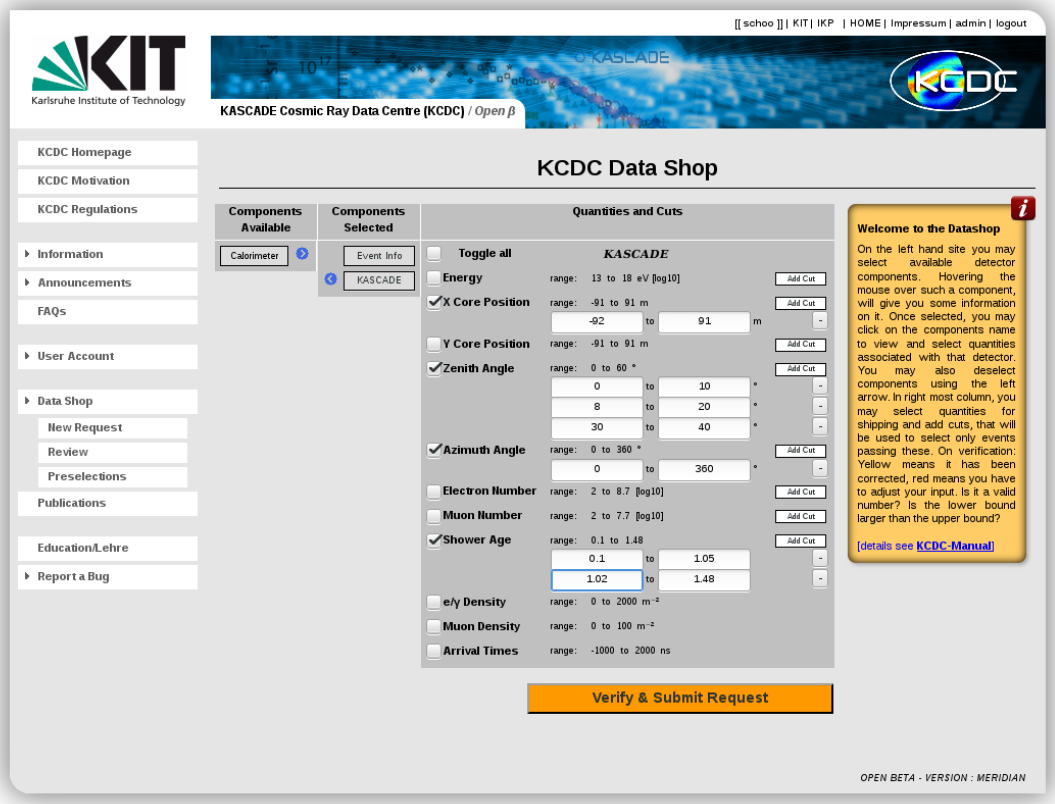

Figure 3.2.: The entry page of the KCDC “Data Shop” pages.

The user supplied cut for the X-coordinate is out of range (see Fig. 3.2). The available range is displayed above the added cut. Taking a look at the summary in figure 3.3, this has been corrected and the user is informed that the cut on the X-coordinate is now obsolete. In addition to providing a summary of the selected quantities and defined selection criteria, the possibility to choose the file format that will be used to store the selected data is provided. Three formats are currently supported, ASCII, ROOT, and HDF5. While ASCII files do not require the use of special libraries, ROOT files are commonly used in high-energy physics and HDF5 files, which are often used by astronomers, are becoming popular also in cosmic ray physics. Details on the formats of ROOT and HDF5 files can be found in the corresponding documentations (see [8] and [9], respectively).

As a next step, the user may then either submit the request or go one step back to the first page of the “Data Shop” in order to change the selection of quantities or criteria.

The review page shown in figure 3.4 is used to display and manage previously submitted requests. For the request that was submitted in the example above, the entry shows that there are no selection criteria applied on X. Taking a look at the selection criteria applied to the zenith angle shows that overlapping ranges are accepted separately as is shown in the summary. However, the review entry shows that they have been merged in the backend (which will be discussed in chapter 4) prior to being processed. The criteria on the azimuthal angle is obsolete and the full range is used, while the selection on the shower age has been accepted, merged again in

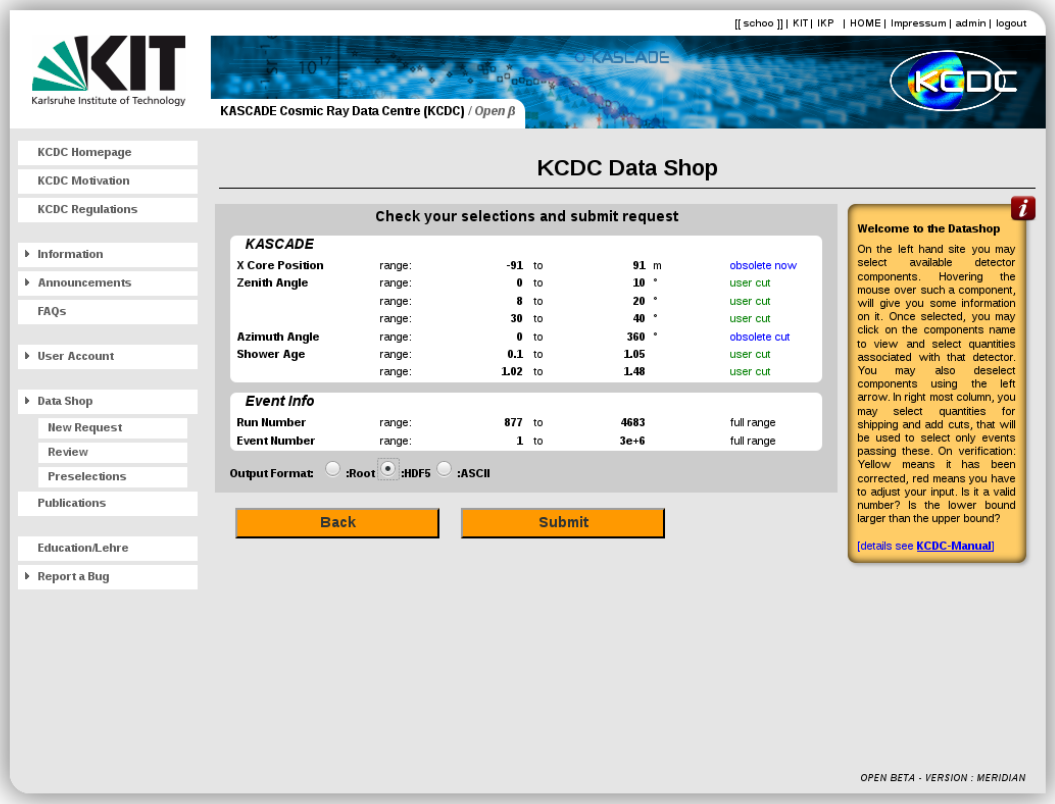

Figure 3.3.: The confirmation page of the KCDC “Data Shop” pages.

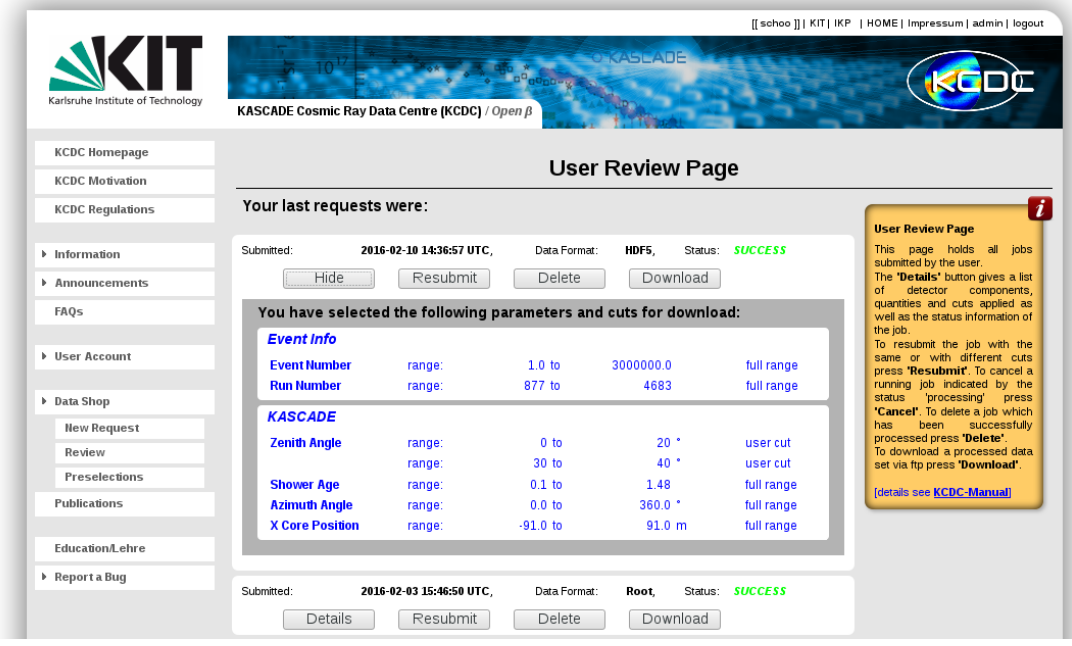

Figure 3.4.: The review page of the KCDC “Data Shop” pages.

the backend and found to be identical to the full range.

Although only the review entry of one request is shown in figure 3.4, the complete history of requests submitted previously by the user can be reviewed. Each entry consists of the following information:

- the date of submission
- the output format selected
- the current status of the job
- the summary of selected quantities and cuts

The status may be either of:

- "PENDING", if the job has not started yet
- "STARTED", if it has started and is currently being processed
- "SUCCESS", if the data is ready for download
- "REVOKED", if the request has been canceled by the user
- "FAILED", if the task has failed for some reason
- "DLEXPIRED", if the download has expired (results are not kept indefinitely)

In addition, the user has the opportunity to resubmit a request. A reason may be that it has expired, failed or been canceled. Another, more common, reason would be to use a previous request as a template, because on a resubmit the user is redirected to the first page of the data shop, which is pre-filled according to the resubmitted request.

Regardless of the chosen output format, the created zip-file contains a copy of the license agreement and an info.txt file which contains a summary of the request; example:

```
Written number of events: 33967
Your requested observables:
Calorimeter (calorimeter)
    Nhad [Number Hadrons]: type=Numerical [Int32] unit=None
General Data (general)
    Datetime [Date]: type=Date unit=None
    Ev [Event Number]: type=Numerical [Int32] unit=None
    R [Run Number]: type=Numerical [Int32] unit=None
KASCADE-array (array)
    Arrival [Arrival Times]: type=Array [Int32] unit=ns
    Az [Azimuth]: type=Numerical [Float64] unit=°
    E [Energy]: type=Numerical [Float64] unit=eV [log10]
    EDensity [Electron Density]: type=Array [Float64] unit=m^-2
    MDensity [Muon Density]: type=Array [Float64] unit=m^-2
    Ne [Number Electrons]: type=Numerical [Float64] unit=None
        Cut: 5.0 to 8.0
    Nmu [Number muons]: type=Numerical [Float64] unit=None
    Xc [X-Coordinate]: type=Numerical [Float64] unit=m
```

```
Yc [Y-Coordinate]: type=Numerical [Float64] unit=m
Ze [Zenith]: type=Numerical [Float64] unit=°
    Cut: 0.0 to 40.0
```

In case of ASCII, there is one file for each component, the ordering of the columns being the same as shown in the info.txt. It is possible that one or more components are missing from an event. To take this into account, there is another file shipped: "row_mapping.txt". The columns are ordered in the same way as the components are listed in the info.txt file. They contain the row number for each components' file that corresponds to the data recorded during the same event. In case a component was inactive, the column is set to –1.

In case of a ROOT file, it will contain one TTree object for each component plus a tree that maps their entries belonging to the same event. Again, the entry number of an inactive component will be set to –1.

The solution for HDF5 files is similar to the one for ROOT files. The file will contain one table for each component plus one table for the map.

While it may not be most convenient to have to manage several files or tables and trees, the alternative to store dummy values for missing components is not chosen, since it has several drawbacks. One is the increased file size, for example if KASCADE-Grande is released as currently planned, one would have to ship dummy entries for several years worth of data, if KASCADE is also selected and the user would still have to check if the entries are valid or only placeholders.

3.2. Administrating the Data Shop

The available components and the associated quantities can be easily configured from the administrator interface. This is demonstrated in figure 3.5 using the component "Event Info" as an example. The component itself has a unique name used internally ("general" in this case), the name to be displayed in the data shop and the selection mode, which has already been discussed. In addition a description may be given, which will be displayed in the information box. It may be supplied in HTML or reStructuredText.

The quantities associated with the detector are displayed below the detector description. Each quantity has an internal name, which has to be unique in combination with the detector. Since the recontructions for KASCADE and for KASCADE-Grande have both provided the number of electrons, the quantities may both be named Ne, however, there can only be one quantity named Ne for a single component.

In addition to the displayed name and selection mode, a quantity may have a unit defined. The displayed unit is the degree sign in unicode. The quantity may have one of three types: Numerical, Date or Array-like. In case of a Date, the sub-type is ignored, otherwise it specifies the exact type of the quantity, a double precision floating-point number in this case.

The fields in the "Cuts" group define whether the user can add cuts, how dates are to be displayed (e.g. yyyy-mm-dd) and the possible range of the quantity. It is important to note that these values are only meant as a guideline for the user. They

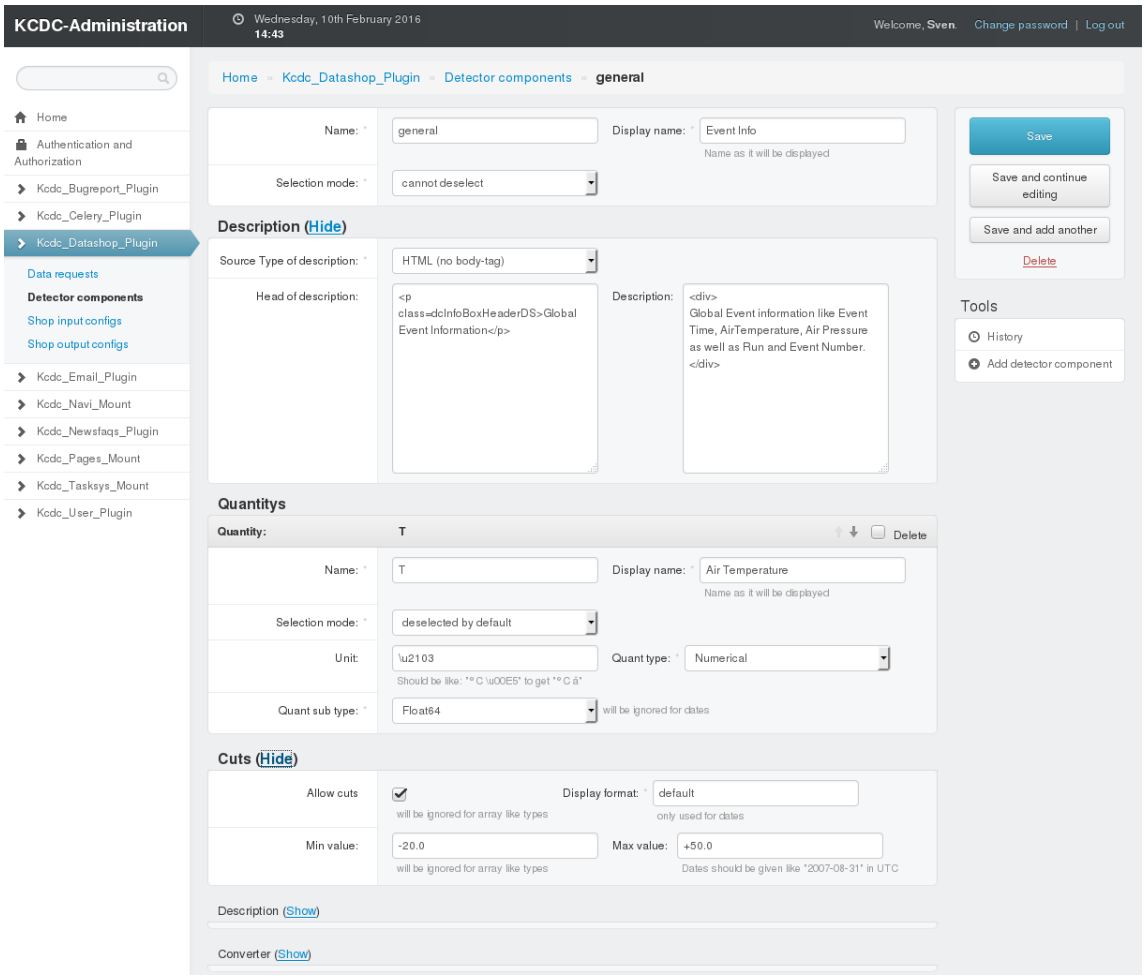

Figure 3.5.: The administrator page of the KCDC data shop.

will be used to ensure the user does not define cuts outside of this interval. However, they will not be enforced during the processing. This means that they will not be included in the query and, therefore, if there is data outside of this range, it will be shipped, if the user did not define a cut on that quantity. The reason for doing it this way is to keep the query times as low as possible, assuming that there is no data in the database which is outside the specified range.

A quantity may have a description that will be shown in the information box. In addition there may be converters defined, which are used during the processing to help retrieving compound types, because it is currently not possible to map compound types directly in the administrator interface.

This concludes the brief presentation of the administrator interface, however, as can be seen in the same figure, almost every plugin can be configured via the administrator interface.

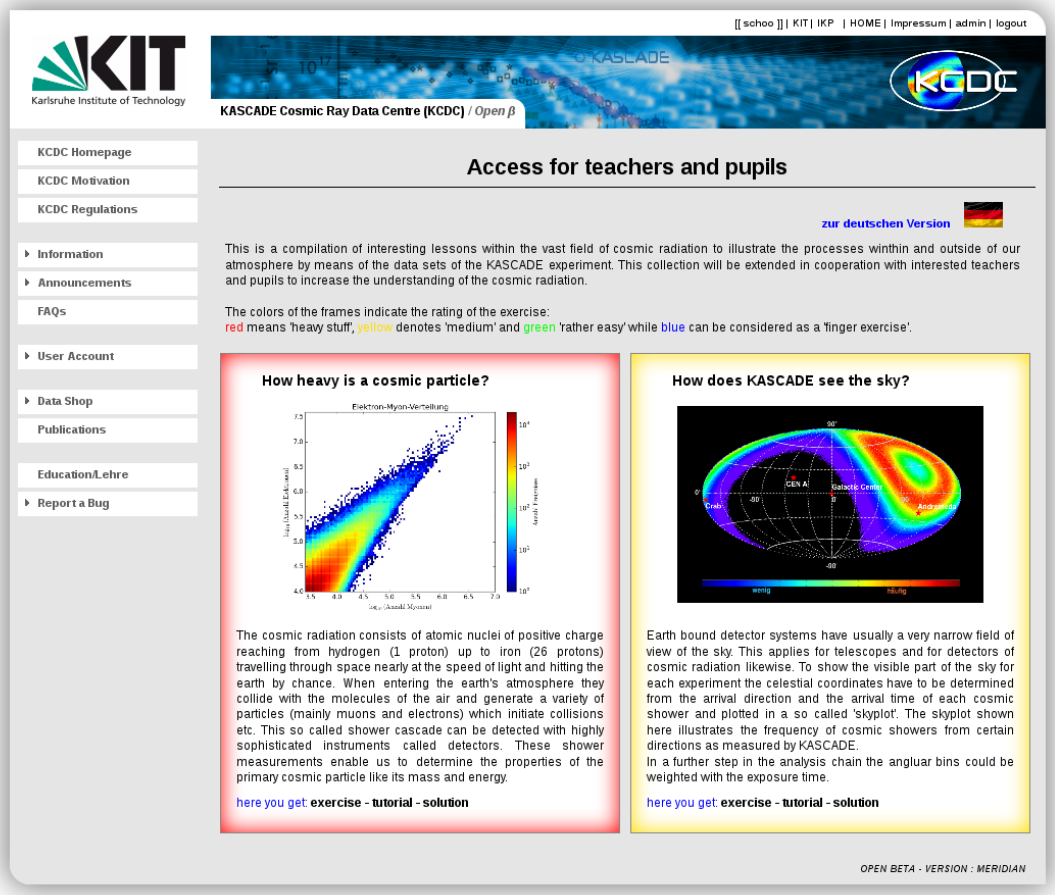

Figure 3.6.: The entry page of the KCDC educational pages.

3.3. Education

An important goal of data publication is to make it available for educational purposes. The entry page of the educational section of KCDC is shown in figure 3.6.

The main addressees of the tutorials provided here are pupils and students interested in astroparticle physics. Currently, there are only two examples available in both English and German. The topics covered by the two examples are “How heavy is a cosmic particle?” and “How does KASCADE see the sky”. Both examples contain information on the needed software, the physics background, the question to be solved and its solution, which is provided as example code together with a discussion of the results.

The tutorial on the two-dimensional shower-size spectrum is written using Python, while the map of the arrival direction of the cosmic particles is produced using C++. In the near future, additional tutorials will be developed by bachelor students as well as by the KCDC-Team, preferably in collaboration with teachers to ensure that there are tutorials that do not overwhelm the pupils. It might even be necessary to provide the same task in different difficulty levels in order to keep it challenging and by that interesting.

4. The Software

This chapter is about the design and implementation of the software of KCDC.

4.1. Setup

The backbone of the web portal is written in Python (version 3.4) [10]. It uses several third party packages, the most important one being Django [11] which the web portal part of the software is built upon. While the choice of a task system is kept flexible, KCDC uses a Celery [12] based task-queue, backed by RabbitMQ [13] as a message broker for the processing of the data requests as well as for some periodic maintenance tasks [1]. A PostgreSQL database has been chosen for the storage of the web portal data, such as the user data and the management/configuration data of the web portal itself. The actually published data is stored in a mongoDB [14]. Due to its schemaless design, it suits our need to store the data for several separate detector components together for each event, taking into account that some components might have been offline during the recording of an event and that some observables are arrays of varying size. Static files, such as plain HTML, JS, and CSS format, are served by a Nginx [15] server, which also acts as a proxy server for a Gunicorn [16] instance, which is a Python-based WSGI [2] HTTP server.

All these choices are completely up to the user of the software as will be discussed in the next section.

4.2. Implementation

KCDC has been designed with the release of its software in mind. Therefore, it has to be relatively easy to adjust, replace, or add components to ensure that it is usable for other projects as well. This is achieved by a modular layout, which is shown schematically in figure 4.1.

[1] Tasks are not executed synchronously on the server, but asynchronously on a dedicated machine. Celery is used to add the tasks to the task queue, which is maintained by RabbitMQ nodes. Celery also provides the framework for the worker nodes, i.e. the processes that check the queue for open tasks and then process these tasks.

[2] WSGI is short for "Web Server Gateway Interface". It is an interface between the server (e.g. Gunicorn) and the web application or framework (e.g. Django).

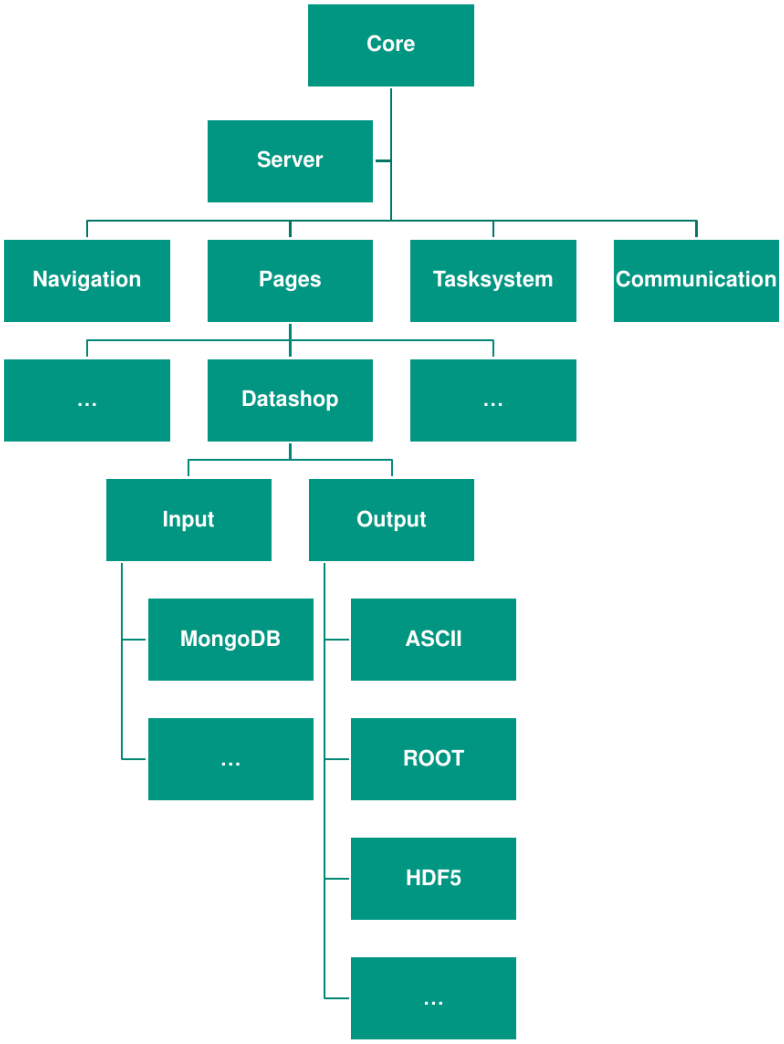

Figure 4.1.: The modular layout of the KCDC software.

The Core of KCDC consists of two packages. The first package - the plugin-core package - provides the basic functionality needed for the plugin system. With its help, a simple python class can be turned either into a plugin or a mount-point, i.e. a class that manages and utilizes its plugins. The features added to these classes are heavily customizable and extensible. The only needed component is the basic plugin system, which creates the plugin-system's structure (based on tree-like parent-children relations) and adds the capability to iterate over the children using pre-, post- or level-order tree traversal [3]. In addition, path based access [4] to a specific child is possible as well as invoking or applying functions to specific children or the whole system or sub-system.

Furthermore an optional component for managing settings is available, which may

[3]Tree traversal is a method used to visit each child of the tree exactly once. Pre-, post- and level-order specify in which order the children are accessed.

[4]If the path to a certain child is known, a child may be accessed directly instead of traversing all the child nodes until the correct one is found. For example, if a certain node has three children and a child (grandchild) of the second child has to be accessed, then instead of checking all the children of e.g. child1 first, one may specify the path directly, i.e. parent->child2->grandchild.

be used to manage plugin, mount or module level settings. It is customizable in several ways, the main ones being the way the settings are collected (e.g. at plugin or module level etc.) and possibly merged, and the way the setting is accessed. The access of the setting is based on retrievers, which are functions or function objects (functors) that implement strategies on how a specific setting is to be found and returned, i.e. retrieved. Several retrievers may be used, the order in which they have been added is important, however, because they will be tried one after another. The first successful retrieval will abort the iteration over the available retrievers. For KCDC, three retrievers are available. The first one is a general path based retriever. For example a setting "SETT1" of plugin "plugin1" in module "mod1" may be accessed from any mount or plugin by "settings.mod1.plugin1.SETT1". Using "settings.SETT1", this first retriever will fail, however, "plugin1" may be defined as the default location of the next retriever, which simply looks for SETT1 in its default location. In case of KCDC, this is the settings class of Django itself.

For KCDC, the setting system is set up to collect plugin or mount level settings at the module level. Therefore, in order to retrieve SETT1, we would use settings.mod1.SETT1 as the direct path. However, it is also possible to access it via settings.SETT1. This way the first two retrievers will fail and a third one will retrieve SETT1 for us. This retriever analyses the Python call stack to retrieve the correct module. This makes it more expensive than the direct path method and is in practice seldomly used.

The second core package "kcdc-core" utilizes the plugin-core package, which could be used in other contexts as well, and adjusts it to work with the Django web-framework. Its main purpose is to load, initialize, and manage the plugin structure of KCDC. The initialization of the kcdc core system is split in two phases, one phase running before the Django setup has run and one after Django has been fully setup. During the first phase, the "INSTALLED_APPS", which is a Django setting that lists all apps (modules) to be loaded, is analyzed to check for the presence of plugins or mounts. If one is found, the configuration of its loggers [5] is processed and possibly defined module level settings are loaded into the settings system. After this step, all plugins are known to the core system and possible plugin/mount level settings are processed. In addition the method for a pre-Django setup is invoked for each member of the plugin-system.

This concludes phase one, phase two starts directly after the Django setup has finished. It gives each plugin/mount the opportunity to execute setup code that needs Django to have already been fully initialized.

Taking a look again at figure 4.1, another distinguished component is the Server. This is the base Django project with its configuration files for Django and the plugin system. Each plugin, mount or module may define default settings within its packages, however, during setup, the Server is checked for user-overrides of these settings. This way, a user simply needs to install a plugin being able to configure it in one place without the need to access or even modify the plugins code base. This

[5] Normally the logging configuration of Django is used directly, however, it is not directly accessible to the plugin before the log-system has been set up by Django. Hence, a plugin may specify its logging configuration within its own module and the configuration will be injected into the Django log-system automatically. The result will be identical to a directly edited logging configuration, however, this way no changes have to be applied by the user of the plugin.

is also the only package that has to be supplied by the user himself, since it contains the KCDC-specific setup, such as the configuration of the Gunicorn server, the PostgreSQL database etc. Everything else is plugin based and may be replaced, added, or extended at will. Therefore, each plugin can be developed as a self-contained unit.

The third level in figure 4.1 shows the mount-points provided by KCDC. The Navigation mount may be used by any app to request a HTML representation of a menu of the available web-pages. Given a menu type, the Navigation mount will ask its children one after another if they are able to produce a menu of this type. The first child that confirms to be able to produce the menu is used to create the menu. In return, the mount provides all information on the available pages, which may be grouped together (e.g account specific pages are grouped together). The menu tree generated will take any access restrictions into account as well as the user status, such as is logged in, is a superuser, has permissions etc. In addition the mount manages the navigation caches to minimize database accesses. The layout of the menu tree and the access restrictions can be configured using the Django administration pages. There is no need to modify configuration files, however, there are mechanisms to auto-generate the menu tree from URLs provided by the plugins.

The Pages mount is used to manage the page-level requirements such as static files. In addition, it provides an administration interface for an easy method to quickly change the look-and-feel of the website. It is also responsible for adding the URLs defined for each page to the Django URLs registry. This way a page can be added or removed by simply installing or removing its package without the need to modify the global URLs registry yourself. This ensures the possibility for an easy "install and forget" way to plug together the web portal.

There are several page-plugins implemented within KCDC, one for user management, one for bug reports, another one for news, FAQs and changelogs, one for educational purposes and the data shop. All of these are in principle optional and do not depend on each other.

The data shop is as much a plugin as it is a mount-point. It is a plugin to the Pages mount, but it provides itself two mount points, one for the input format, i.e. the kind of source where the measured data is stored and one for the output formats. The only input plugin provided by KCDC is one for a mongoDB. There are only very few requirements for the input plugin, which is in principle the format in which the data is expected to be returned in. In short, there is no base class enforcing a layout or interface of an input plugin. As long as it looks and behaves like an input plugin it is considered to be one (which is called duck-typing), no matter how it works internally.

As output formats, KCDC provides three types: ASCII, HDF5, and ROOT. All the data shop has to know about these plugins is whether they can process array-like data. For example the plugin for the ASCII format does not support array-like types. If at least one array-like observable has been selected, ASCII will no longer be given as a choice to the user. Other than that, the data shop does not need to know about the implementation of the plugins as long as they behave properly. Therefore, it is easy to extend the data shop to offer additional formats.

The task or job that processes the data retrieval is submitted to the "Tasksystem"

mount, which acts as a central point used to submit any tasks defined by the plugins to the task queue (One time jobs as well as periodically repeated ones.). It may have many plugins attached to it. Submitting a job can be done in a "fire and forget" way if one does not care how the job is being processed. The mount will just pick a default system if it is defined, otherwise it will use the first of the attached systems. Of course, it keeps track of which system has been used for a specific task, so it can be queried for the result and the current status of the task, or a task may be canceled or revoked if it is executed asynchronously. It is easy to switch the way tasks are being processed, however if one relies on a specific system to be used, this information can be passed to the mount too.

In case of KCDC the only system is implemented using Celery and RabbitMQ to provide an asynchronous task-queue.

The last mount is the Communication mount. It can be used to send messages to users or admins. Currently, the only plugin provided by KCDC is an email plugin, however, e.g. a twitter plugin would also be possible. Sending messages is done by passing the message together with its type (in this case it is an email) and type specific information (such as authentication credentials, addressee etc.) to the mount, which will "ask" its children (one after another) whether they are able to send messages of this type. The first one to accept this type of message, will be used to send the message, except if the use of a specific plugin is requested.

Using the plugin based design, KCDC is highly extensible and configurable and can, therefore, be easily adjusted to serve the needs of other experiments. In principle, it could also be used for something completely different, because the data shop, being implemented as a plugin, is optional, too.

The above description of the implementation has been simplified and should not be considered as complete. A complete and extensive description is beyond the scope of this text, however, a detailed documentation on the usage and development of the software will be included in its release.

4.3. Future Development of KCDC

The basic features for a data providing web portal are already implemented in KCDC, however, there is always room for new features or improved ones. Especially the educational pages have to be extended by more example analyses.

Regarding the published data, adding measurements of e.g. KASCADE-Grande and the other components of the experiment is planned for the future. In addition, the inclusion of air-shower simulations and/or the reconstruction software used to reconstruct the published quantities is currently being discussed. The capability for easily adding more components is already given, adding these would not require any changes to the code base.

Support for a permission based access to parts of the web portal is already included. While KASCADE has finished its data taking, other experiments might want to release only parts of their data to the public, but the entire data set for members of the collaboration. The basis for such a permission based restriction of the data access is given, but not fully included in the data shop yet. This would be a feature that will certainly be added before the software is released.

Since it is currently not needed for KCDC, selecting separate input sources for different components is not possible yet. Also the ability to more closely represent the data structure as it is stored in the source could be useful. For example, for the core position, Xc and Yc have to be selected individually, although it does not make sense to select only one of them. Quantities such as the particle densities are only useful with additional information like the distance in shower coordinates at which the density has been measured. Currently this is not represented in the admin interface, but has to be enforced during the processing of the data. Combined cuts, especially on Xc and Yc or in the N_e-N_μ-plane, could also be helpful.

Another feature planned for the future is the ability to provide the data points for published results e.g. the composition and energy spectrum of cosmic rays in case of KCDC for KASCADE and KASCADE-Grande. While such a central database is available for high energy physics, there is none available for cosmic ray physics yet. Additionally, the used reconstruction algorithms could be added.

A very basic requirement for open-source software is the provision of extensive documentation, clean code and a good coverage of unit- and functional tests. There is still room for improvements regarding all three categories.

To conclude, while there is still a lot of work to be done, KCDC has already become a highly extensible and adjustable implementation of a web portal for data publication. It is already in use and can be accessed via `https://kcdc.ikp.kit.edu`.

5. Some final words

The development of KCDC was - of course - not done by a single person alone. During the development of the first version of KCDC we have been six people on the core-team, namely Andreas Haungs, Benjamin Fuchs, Donghwa Kang, Doris Wochele, Jürgen Wochele, and myself. Without any of these and the support and input of Johannes Blümer (who chose the name of the project and developed the idea of KCDC together with Andreas), KCDC would not have had its first release in Nov. 2013, if it would have been released at all.

It is worth to note that KCDC is the product of a bunch of interested people who have had no prior experience in developing a web portal and who had other duties next to KCDC as well. It would have probably not worked so well if not every one of us had been as eager to learn something new and had been as personally committed to the project as we turned out to be.

I am happy that I was given the opportunity to be part of the team and I want to thank the team for their hard work and their support for new ideas as they often meant a lot of extra work for all of us, especially after Ben had left after the first release and someone came up with the crazy idea to completely change the software design to be plugin based, although we were aiming for a new release. It took quite some time, but I am still convinced it was the right decision ;).

Thanks again and all the best for future releases.

Sven

References

[1] *KCDC*. Last visited on 06.05.2016. URL: https://kcdc.ikp.kit.edu.

[2] S. Schoo et al. "The KASCADE Cosmic Ray Data Centre (KCDC)". In: *PoS* ICRC2015.262 (2015).

[3] A. Haungs et al. "KCDC - The KASCADE Cosmic-ray Data Centre". In: *J. Phys.: Conf. Ser.* 632.1 (2015), p. 012011. DOI: 10.1088/1742-6596/632/1/012011.

[4] T. Antoni et al. "The cosmic-ray experiment KASCADE". In: *Nucl. Instrum. Methods Phys. Res., Sect. A* 513.3 (2003), pp. 490–510. DOI: 10.1016/S0168-9002(03)02076-X.

[5] *CERN Open Data Portal*. Last visited on 06.05.2016. URL: http://opendata.cern.ch/.

[6] *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. Last visited on 06.05.2016. URL: http://openaccess.mpg.de/Berlin-Declaration.

[7] Wochele, Jürgen et al. *KCDC User Manual*. Last retrieved on 06.05.2016. URL: https://kcdc.ikp.kit.edu/static/pdf/base/kcdc-Manual.pdf.

[8] *ROOT*. Last visited on 06.05.2016. URL: https://root.cern.ch.

[9] *HDF5*. Last visited on 06.05.2016. URL: http://www.hdf5.com/.

[10] *Python 3.4*. Last visited on 06.05.2016. URL: www.python.org.

[11] *Django*. Last visited on 06.05.2016. URL: www.djangoproject.com.

[12] *Celery*. Last visited on 06.05.2016. URL: www.celeryproject.org.

[13] *RabbitMQ*. Last visited on 06.05.2016. URL: www.rabbitmq.com.

[14] *MongoDB*. Last visited on 06.05.2016. URL: www.mongodb.org.

[15] *Nginx*. Last visited on 06.05.2016. URL: www.nginx.com.

[16] *Gunicorn*. Last visited on 06.05.2016. URL: gunicorn.org.

Final Conclusions

KASCADE and KASCADE-Grande have been built to study the energy spectrum and mass composition of cosmic rays by measuring extensive air-showers induced by the primary particles. This information can be used to test astrophysical models on the origin and propagation of cosmic rays.

At the beginning of this work there was a clear set of three main goals to be reached:

1. Demonstrate the new level of accuracy reached by the simultaneous use of the KASCADE and KASCADE-Grande detectors.
2. Utilize the combined reconstruction to infer the energy spectrum and mass composition of cosmic rays with a better precision than before.
3. Acknowledging the fact that the KASCADE facility has finally been dismantled, develop the means to archive and publish the scientific data taken by the detector setups.

The first goal has clearly been reached. Especially in case of KASCADE-Grande a significantly improved reconstruction of the shower observables was made possible by the coherent treatment and combination of the two detectors.

The second task has been fulfilled to a large extent. It was shown that the combined array can be used to reconstruct the energy spectrum and mass composition in far greater detail than what was possible using the standalone reconstructions. The necessary reconstruction procedure has been developed and is ready to be used. Also, the features discovered beforehand (knees and ankles) of the light and heavy mass group spectra have been confirmed. The remaining limitation, which presently prevents a consistent reconstruction for measured events located in the former KASCADE array and for those located in KASCADE-Grande, is based on differences between measured data and simulations. It should be remarked that this differences were only unveiled due to the better accuracy reached with the present work. Once the hadronic interaction models have reached better agreement with measured data, the developed analysis is ready to be applied again.

Although not part of the initial goals, an important discovery was made regarding the differences between EposLHC and QGSJetII4, the models of the newest generation that take already into account the latest data of the Large Hadron Collider. These models have converged to a consistent description of the light mass group of cosmic ray primaries. However, their different extrapolation from proton-proton interactions to proton-nucleus and nucleus-nucleus interactions still results in spectra of heavy primaries that differ from one another. This discrepancy increases with energy. However, as the composition becomes heavier towards higher energies, the underlying

dependency might well be the mass of the primary particles instead of their energy. Nevertheless, the present analysis exhibited hints that a possible source for the mismatch between measurement and simulation lies in the lateral distribution of muons. The slope of the lateral density function used to infer the total number of muons is a fixed parameter and has been derived from simulated data. It turns out that the predicted lateral distribution is too flat compared to measured data resulting in an overestimation of the number of muons for events located in KASCADE and an underestimation for events located in KASCADE-Grande.

The further development of KCDC was very successful, where an important part was performed within this thesis. The software is already in use by the KASCADE-Grande collaboration with the additional prospect of being released as an open-source software package for other experiments to be used as a basis to start publishing their data as well. The web portal is actively used by the community and it is gradually extended in terms of the number of published events, as well as in terms of the available reconstructed observables.

There is always room for improvements, however, both, the combined analysis and KCDC provide a good basis for future endeavors.

Acknowledgments

First of all I want to thank Hans Jürgen Mayer who has developed the new KRETA version on which the present analysis is based upon. I want to thank Prof. Dr. Johannes Blümer and Prof. Dr. Ulrich Husemann as my reviewers. I am grateful for the discussions within the KASCADE-Grande collaboration and in particular with Andreas Haungs (my advisor), Juan Carlos Arteaga Velazquez, Donghwa Kang, Hans Jürgen Mayer, Tanguy Pierog, and Jürgen Wochele.

I want to thank the KCDC-Team, namely Benjamin Fuchs (off duty), Andreas Haungs, Donghwa Kang, Doris Wochele and Jürgen Wochele. In addition our main "KCDC-tester" Jürgen Oehlschläger provided valuable input.