

# Measurement of the flux of ultra high energy cosmic rays using data from very inclined air showers at the Pierre Auger Observatory

Der Fakultät für Mathematik, Informatik und Naturwissenschaften der  
RWTH Aachen University vorgelegte Dissertation zur Erlangung des  
akademischen Grades eines Doktors der Naturwissenschaften

von

Diplom-Physiker

**Hans Peter Dembinski**

aus Rostock



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Conventions</b>	<b>3</b>
<b>3</b>	<b>Ultra-high energy cosmic rays</b>	<b>9</b>
3.1	Cosmic rays . . . . .	9
3.1.1	Cosmic rays up to 100 TeV . . . . .	10
3.1.2	Cosmic rays above 100 TeV . . . . .	11
3.2	Extensive air showers . . . . .	15
3.2.1	Heitler-model of the hadronic cascade . . . . .	16
3.3	Very inclined air showers . . . . .	24
3.3.1	Muon component . . . . .	24
3.3.2	Lateral shower profile at ground level . . . . .	26
3.3.3	Time structure of the shower front . . . . .	30
3.3.4	Muon generated electromagnetic particles . . . . .	31
<b>4</b>	<b>Pierre Auger Observatory</b>	<b>33</b>
4.1	Surface detector . . . . .	35
4.1.1	Signal calibration . . . . .	35
4.1.2	Trigger system and data acquisition . . . . .	39
4.2	Fluorescence detector . . . . .	42
4.2.1	Signal calibration . . . . .	43
4.2.2	Trigger system and data acquisition . . . . .	45
4.3	Extensions and new developments . . . . .	46
<b>5</b>	<b>Air shower and event simulation</b>	<b>47</b>
5.1	Simulation of air showers . . . . .	47
5.1.1	Hadronic interaction models at low energies . . . . .	48
5.1.2	Hadronic interaction models at high energies . . . . .	49
5.1.3	Thinning . . . . .	52
5.2	Simulation of surface detector events . . . . .	53
5.2.1	Un-thinning . . . . .	53
5.3	Mass production of very inclined air showers . . . . .	55
5.3.1	Simulation setup . . . . .	56
5.3.2	Comparison with older productions . . . . .	61
5.4	Modeling very inclined air showers . . . . .	62
5.4.1	Number of muons on the ground . . . . .	63
5.4.2	Universality of the normalized muon density profile . . . . .	69
5.4.3	Parameterisation of the muon density profile . . . . .	69

---

5.4.4	Comparison with another model . . . . .	77
5.5	Surface detector response to very inclined air showers . . . . .	81
5.5.1	Properties of the Auger Water-Cherenkov detector . . . . .	81
5.5.2	Modeling the signal response . . . . .	82
5.5.3	Analysis of the electromagnetic signal component . . . . .	86
5.5.4	Comparison with another model . . . . .	89
<b>6</b>	<b>Reconstruction of cosmic ray properties</b>	<b>91</b>
6.1	Vertical and very inclined showers . . . . .	92
6.2	SD event reconstruction of very inclined showers . . . . .	93
6.2.1	Idealised reconstruction . . . . .	95
6.2.2	Realistic SD reconstruction . . . . .	100
6.2.3	Reconstruction resolution and bias . . . . .	111
6.2.4	Comparison with references . . . . .	117
6.2.5	Constant intensity analysis . . . . .	124
6.2.6	Bias correction of the energy estimator . . . . .	127
6.2.7	Summary . . . . .	128
6.3	FD event reconstruction . . . . .	129
6.3.1	Pulse finding and rejection of random pixels . . . . .	129
6.3.2	Reconstruction of the shower axis . . . . .	130
6.3.3	Reconstruction of the shower energy . . . . .	131
6.3.4	Systematic uncertainty of the energy . . . . .	134
<b>7</b>	<b>Energy calibration of the Surface Detector</b>	<b>137</b>
7.1	Event selection . . . . .	138
7.1.1	SD event selection . . . . .	139
7.1.2	FD event selection . . . . .	141
7.1.3	Cut optimisation . . . . .	145
7.2	Calibration method . . . . .	149
7.2.1	Model components . . . . .	151
7.2.2	Statistical and numerical bias . . . . .	164
7.3	Application to data . . . . .	167
7.4	Summary . . . . .	174
<b>8</b>	<b>Flux of ultra-high energy cosmic rays</b>	<b>175</b>
8.1	Exposure of the surface detector . . . . .	176
8.2	Unfolding of the cosmic ray flux . . . . .	179
8.2.1	Unfolding problem and solutions . . . . .	179
8.2.2	RUN-algorithm . . . . .	182
8.2.3	Application to data . . . . .	190
8.3	Summary . . . . .	196
<b>9</b>	<b>Acknowledgements</b>	<b>199</b>
<b>A</b>	<b>Analytical and numerical calculations</b>	<b>201</b>
A.1	Point of shower maximum and particle interaction lengths . . . . .	201
A.2	Exp-normal distribution . . . . .	202
A.3	Effective area of a SD station and average muon signal . . . . .	203
A.4	Projection of surface detector stations . . . . .	205
A.5	Calculation of the data model $f_{\text{tot}}$ . . . . .	207
<b>B</b>	<b>Air shower simulation: Technical summary</b>	<b>209</b>
<b>C</b>	<b>SD station: signal response model</b>	<b>211</b>

## CONTENTS

---

<b>D Reconstruction of SD and FD events: technical details</b>	<b>213</b>
<b>E SD event reconstruction: background rejection</b>	<b>217</b>
<b>F Event selection: examples of rejected events</b>	<b>221</b>
<b>G Cosmic ray flux: tables and consistency checks</b>	<b>227</b>

---

**CONTENTS**

# Chapter 1

## Introduction

When Victor Hess discovered the cosmic radiation in 1912 [1], he laid a foundation for the upcoming field of high energy particle physics. For about 40 years, the cosmic radiation was the only source of particles with energies of several GeV. The analysis of cosmic rays and their reactions with matter led to several discoveries, for instance the positron in 1933 [2], the muons in 1936 [3], and the pions in 1947 [4].

In the 1950's, the main focus of particle physicists shifted towards man-made particle accelerators, which offered several advantages over the cosmic radiation. The key development was the synchrotron as a scalable accelerator for particles up to multi-GeV energies. The Cosmotron, completed in 1953, was one of the first [5]. Artificial accelerators provide almost monochromatic beams of high intensity with a defined particle content, in contrast to the cosmic radiation, and overall much more control over the particle interactions.

Cosmic rays remained of considerable interest for astrophysicists as messenger particles of very powerful astrophysical processes. In 1939, Pierre Auger discovered cosmic ray induced particle showers in the atmosphere, called extensive air showers [6]. He analysed coincident particle counts in several spatially separated detectors and estimated from the number of registered particles, that cosmic rays exist with energies up to  $10^{16}$  eV. His concept of measuring cosmic rays through their extensive air showers in an array of particle detectors turned out to be extremely scalable and is used till today. In 1963, Linsley found the first cosmic rays around  $10^{20}$  eV with a large surface detector [7].

It is a downside of the surface detector measurement, that the reconstruction of the energy and mass of the cosmic ray from the measured signals depend strongly on model calculations of the air shower development, which still have quite large theoretical uncertainties. A new detection method, the fluorescence method, was established with the Fly's Eye/HiRes experiment in the 1980's [8], which reduces this model dependency considerably. The fluorescence method allows to follow the full air shower development in the atmosphere. This is achieved by observing fluorescence light with appropriate telescopes which was emitted by nitrogen molecules after an excitation through the passing air shower.

It was realised eventually that fluorescence and surface detectors are complementary in many ways so that a combination, a hybrid detector, would be even more versatile. The Pierre Auger Observatory is such a hybrid detector and currently the world's largest cosmic ray observatory. Its design focuses on the detection of ultra-high energy cosmic rays from  $10^{18}$  eV to  $10^{20}$  eV and above.

The center of mass energy in the first interaction of the most energetic cosmic rays in the atmosphere reaches almost  $10^3$  TeV. This is two orders of magnitude larger than the center of mass energy in the largest artificial accelerator, the Large Hadron Collider [9]. These large interaction energies and the precise measurement of air showers at the highest energies with the observatory raises new interest in particle physicists for cosmic rays and air showers.

Like many of its predecessors, the Pierre Auger Observatory focuses on the analysis of air showers with zenith angles up to  $60^\circ$  although it is sensitive up to  $90^\circ$ . This restricted class of events is called vertical showers by convention, the remain are called very inclined. Most of the collected events are vertical in this sense: about 75 % of the events in the surface detector and about 87 % of the events in the fluorescence

---

detector fall into the vertical category.

Very inclined air showers are more difficult to simulate and model. The curvature and the geomagnetic field of the Earth may be neglected for vertical showers but not for very inclined showers. The muon component of a very inclined shower dominates the signals measured in the surface detector. The signals are weaker and have a more complex pattern due to geomagnetic deflections of the muons. Nevertheless, there is an ongoing effort to reliably reconstruct these events [10–17] so that they may be used in cosmic ray studies. A reliable reconstruction of very inclined showers would increase the number of detected cosmic rays by about 30 %, as well as the overall sky coverage of the Pierre Auger Observatory.

There are also other gains. The signals in the surface detector generated by very inclined showers are dominated by the muons in the shower. A proper reconstruction of such events therefore challenges the understanding of this shower component. Searches for cosmic neutrinos in the surface detector data of the Pierre Auger Observatory benefit from the understanding of hadron induced very inclined air showers, which form the main background [18].

This work contributes to the understanding of hadron induced very inclined air showers measured with the surface detector of the Pierre Auger Observatory. This is done by presenting a full analysis chain from the simulation of such events, over their reconstruction from the collected data, towards the final goal of a measurement of the flux of ultra-high energy cosmic rays between  $10^{18}$  eV and  $10^{20}$  eV. A precise energy calibration of the surface detector measurement with the fluorescence detector is a key step in this chain.

This thesis is structured as follows. Chapter 2 lists the basic conventions used throughout the work. Chapter 3 provides an overview over cosmic rays, with a focus on cosmic rays at ultra-high energies and extensive air showers.

The simulation of a large library of very inclined air showers was a major part of the work and is discussed in Chapter 5. The simulations are used to study and model the muon component of very inclined air showers and the response of the surface detector of the Pierre Auger Observatory to such showers.

The reconstruction of cosmic ray properties from data of the Pierre Auger Observatory is discussed in Chapter 6, whereas the discussion focuses on the reconstruction of surface detector events generated by very inclined air showers. The reconstruction will make heavy use of the models derived in the simulation chapter and will be thoroughly tested with simulated events.

The cosmic ray energy reconstructed from the data of the surface detector has large systematic uncertainties, if it is based only on theoretical models of the shower development. These uncertainties can be greatly reduced by calibrating the surface detector with the fluorescence detector. The latter is able to perform an almost calorimetric measurement of the cosmic ray energy. The calibration is performed with a small sample of events, which are observed in both detectors simultaneously. A significantly improved method for this kind of calibration is introduced and applied in Chapter 7.

The last chapter, Chapter 8, discusses the calculation of the cosmic ray flux from the reconstructed cosmic ray events. An unfolding technique is used to obtain the true flux from the measured flux. The result is compared with the most recent result obtained from the analysis of vertical showers at the Pierre Auger Observatory.

# Chapter 2

## Conventions

This chapter summarises the coordinate system and naming conventions used throughout this work. Most of the conventions mentioned here are taken from the official conventions guide of the Pierre Auger Observatory [19].

### Particle notation

The following chapters deal with extensive air showers, particle showers generated through high energy interactions of a cosmic ray in the atmosphere. During the development of this air shower, particles and anti-particles are generated in equal amounts. For almost all purposes, the distinction between particles and anti-particles is irrelevant in this study.

Therefore, terms like “electrons” and “muons” will generally refer to both the particle and anti-particles, unless explicitly stated otherwise. Another commonly used term in the context of air showers is “electromagnetic particles”, which refers to electrons, positrons, and photons (but not muons).

## Coordinate systems

### Ground coordinate system

Fig. 2.1 shows the main coordinate system to describe the shower. It is a local cartesian coordinate system, defined at the impact point of the shower axis on the ground. The impact point is called the *shower core*.

It is not feasible to define a single global rectangular coordinate system for the Pierre Auger Observatory, because Earth’s curvature is relevant over the size of the surface array, as demonstrated in Fig. 2.2.

The zenith angle  $\theta$  has the usual astronomical definition. The definition of the azimuth angle  $\phi$  in the coordinate system of the Pierre Auger Observatory differs from the astronomical one. The direction defined by  $(\theta, \phi)$  points to the origin of the cosmic ray in the sky, not in the direction of shower propagation.

### Lateral coordinate system

It is useful to introduce a special coordinate system to describe the lateral profile of an air shower in the context of this work: the *shower front plane* coordinate system. It is depicted in Fig. 2.3. The shower front plane contains the point where the shower axis intersects the ground plane, which is called the *shower core*. It is perpendicular to the shower direction and oriented so, that the y-axis is parallel to the geomagnetic field projected into the plane.

It should be emphasized, that observations and predictions are always done in the ground plane, and only projected in the shower front plane. The coordinate system has the advantage, that it preserves some of the principal symmetries of the shower profile. It helps to separate geometrical effects from physical effects.

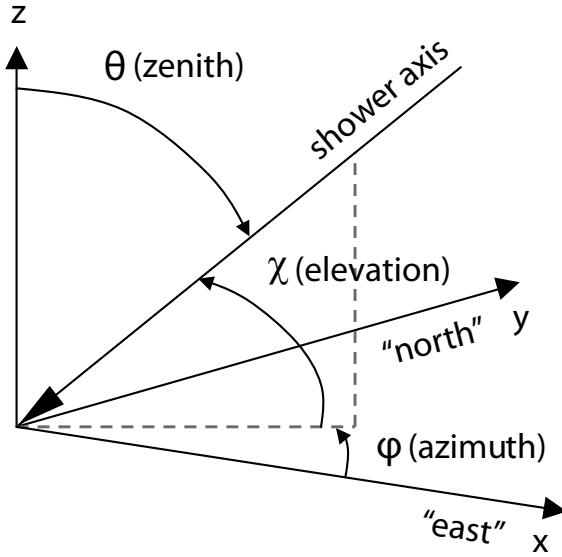


Figure 2.1: The drawing shows the local cartesian coordinate system. The center of the system is the impact point of the shower axis on the ground (shower core). The x-axis points into the geographic east, the y-axis points into the geographic north, the z-axis points upwards (adapted from [19]). The zenith angle  $\theta$  is counted from the vertical direction. The azimuth angle  $\phi$  starts in the geographic east and is counted counter-clockwise. The elevation angle  $\Omega = \pi/2 - \theta$  is less common, but sometimes used in the context of Fluorescence telescope measurements.

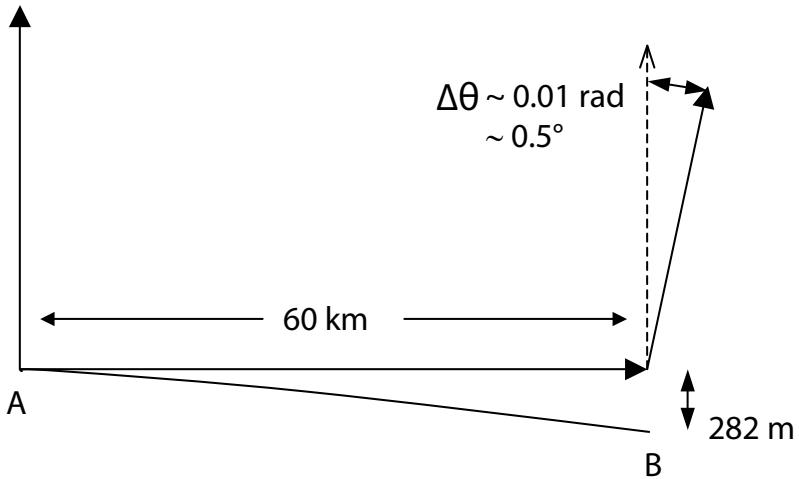


Figure 2.2: The drawing illustrates the effect of Earth's curvature over the size of the surface array of the Pierre Auger Observatory. A shower with  $\theta_A = 0^\circ$  in the local rectangular coordinate system  $A$  at the west end of the array, has a zenith angle  $\theta_B \approx 0.5^\circ$  in the local coordinate system  $B$  at the east end. Also, the center of  $B$  appears by about 280 m deeper in  $A$ . Distances and angles are not to scale (from [19]).

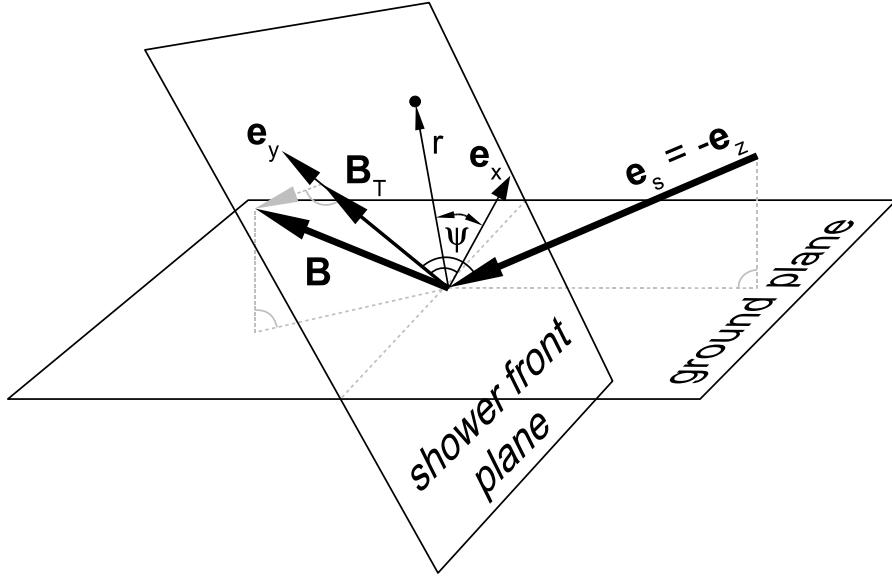


Figure 2.3: A drawing of the lateral shower coordinate system, after [20]. Arcs without arrows indicate right angles. The center of the system is the impact point of the shower axis on the ground. The x-y-plane is perpendicular to the shower axis. The z-axis is anti-parallel to the shower direction vector  $e_s$ . The orientation of this plane is chosen such that the y-axis is parallel to the projected vector of the geomagnetic field  $\mathbf{B}$  into the plane.

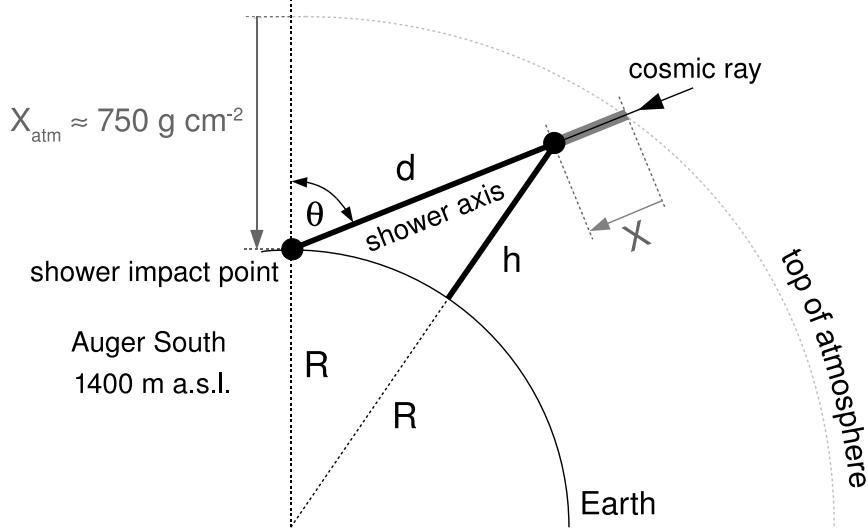


Figure 2.4: The drawing shows the coordinates, which are used to describe a point in the longitudinal profile of the shower. The distances  $R$ ,  $d$ , and  $h$  are geometrical lengths, while  $X$  is a slant depth, counted from the top of the atmosphere. To describe a particular point on the shower axis, the distance  $d$  is the distance of a particular point on the shower axis from the impact point on the ground, the height  $h$  is the altitude of this point above the ground level. Another way to describe the point is via the accumulated slant depth  $X$  along the shower axis. The total atmospheric depth  $X_{\text{atm}}$  depends on the zenith angle  $\theta$  of the trajectory and the ground altitude, e.g.  $X_{\text{atm}} \approx 750 \text{ g cm}^{-2}$  for  $\theta = 0^\circ$  and the site of the southern Pierre Auger Observatory.

---

## Longitudinal coordinate system

Fig. 2.4 shows the coordinates used to describe the longitudinal shower development. The most common quantity to describe the longitudinal profile is the *slant depth*  $X$ , defined by the integral

$$X = \int ds \rho(s), \quad (2.0.1)$$

where  $\rho(s)$  is the local air density, and  $s$  is a path length counted from the top of the atmosphere along the shower axis. The slant depth is good to describe the shower in terms of interaction and attenuation lengths, which are independent of the material if expressed in this unit.

In this work, a point in the longitudinal development will also be characterised by its altitude above the ground  $h$ , and its distance  $d$  to the shower impact point on the ground (= shower core), as indicated in Fig. 2.4.

## Other coordinate systems

Landmarks and geographic maps of the Earth are usually expressed in the **UTM coordinate system**. Points are expressed in northing, easting, and altitude above sea level within a particular reference zone. The system takes the curvature of the Earth into account and is therefore not cartesian: the northing and easting coordinates measure distances on the surface of a reference ellipsoid.

The **geomagnetic field** has a coordinate system based on the magnitude and direction of the field vector. The direction is expressed with the inclination angle  $\theta$  and the declination angle  $\delta$ . The inclination is the vertical angle. It is zero, if the field is parallel to the horizon, and  $-90^\circ$  ( $90^\circ$ ) if the field points vertically downwards (upwards). The declination is the horizontal angle. It is zero, if the field points to the geographic north, and is counted clock-wise from there.

## Acronyms

<b>ADC</b>	Analog Digital Converter
<b>ADST</b>	Advanced Data Summary Tree (event data format)
<b>AGN</b>	Active Galactic Nuclei
<b>AIRES</b>	AIR-shower Extended Simulations (air shower simulator)
<b>AMIGA</b>	Auger Muon-detectors and Infill for the Ground Array
<b>APF</b>	Aerosol Phase Function (Monitor)
<b>a.s.l.</b>	above sea level
<b>QCD</b>	Quantum ChromoDynamics
<b>CDAS</b>	Central Data Aquisition System
<b>QED</b>	Quantum ElectroDynamics
<b>CLF</b>	Central Laser Facility
<b>CMB</b>	Cosmic Microwave Background
<b>CORSIKA</b>	COsmic Ray SImulations for KAskade (air shower simulator)
<b>DAQ</b>	Data AQuisition
<b>EPOS</b>	Energy-conserving quantum mechanical multiple scattering approach, based on Partons (parton ladders), Off-shell remnants, and Splitting of parton ladders
<b>FADC</b>	Flash Analog Digital Converter
<b>FD</b>	Fluorescence detector
<b>FLUKA</b>	FLUktuierende KAskade (german for fluctuating cascade)
<b>FoV</b>	Field of View
<b>FRAM</b>	Fotometric Robotic Atmospheric Monitor
<b>GHEISHA</b>	Gamma-Hadron-Electron-Interaction SH(A)ower code
<b>GPS</b>	Global Positioning System
<b>HAM</b>	Horizontal Attenuation Monitor

<b>HEAT</b>	High Elevation Auger Telescope
<b>IGRF</b>	International Geomagnetic Reference Field
<b>LED</b>	Light Emitting Diode
<b>LIDAR</b>	Light detection and ranging
<b>p.d.f.</b>	probability density function
<b>PLD</b>	Programmable Logic Device
<b>PMT</b>	Photomultiplier tube
<b>QGSJet-II</b>	Quark-Gluon-String model with JETs
<b>RUN</b>	Regularised UNfolding
<b>SD</b>	Surface detector
<b>SDP</b>	Shower Detector Plane
<b>UTM</b>	Universal transverse mercator
<b>UV</b>	Ultra Violet
<b>VEM</b>	Vertical Equivalent Muon
<b>XLF</b>	eXtreme Laser Facility

## Mathematical notation

$\langle x \rangle$	average of $x$
$\sigma[x]$	statistical uncertainty of $x$
$\sigma_{\text{sys}}[x]$	systematic uncertainty of $x$
$x$	vector (small letter)
$M$	matrix (capital letter)
$x^T$	transpose of $x$
$\text{erf}(x)$	error function $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x dt e^{-t^2}$

---

# Chapter 3

## Ultra-high energy cosmic rays

This chapter provides a general overview over cosmic rays. The focus is laid on ultra-high energy cosmic rays above  $10^{18}$  eV and the extensive particle showers which they initiate in the atmosphere of the Earth. Understanding the properties of such air showers is fundamental for the Pierre Auger Observatory and similar instruments which rely on the indirect observation of cosmic rays via this phenomenon.

The development of an air shower in the atmosphere is best described with full Monte-Carlo simulations of the elementary particle processes, which are covered in Chapter 5. This chapter tries to provide simple model calculations for many air shower properties which are very approximate, but allow to understand many basic air shower features.

### 3.1 Cosmic rays

The following general discussion of cosmic rays is based on ref. [21–24].

The term *cosmic rays* generally refers to stable and charged particles which travel through interstellar or even intergalactic space. The total flux of cosmic rays at the top of the atmosphere is about  $1000 \text{ m}^{-2} \text{ sr}^{-1} \text{ s}^{-1}$  and dominated by protons of a few GeV. The differential flux  $J(E) \propto dN/dE$  is approximately a power law  $J(E) \propto E^\alpha$  with a spectral index  $-2.6 \lesssim \alpha \lesssim -3.2$ . The range of cosmic ray energies is huge. The cosmic rays with the highest energies detected so far slightly exceed  $10^{20}$  eV, but are extremely rare: only about one such particle per  $\text{km}^2$  and century arrives at the Earth. The measured flux  $J(E)$  above 1 TeV is shown in Fig. 3.1.

Experimentally, two energy ranges need to be distinguished. Up to 100 TeV, the flux  $J$  is large enough for balloon and outer space experiments, which measure the cosmic rays directly with appropriate particle detectors close to or above the top of the atmosphere. Such experiments are able to measure the cosmic ray energies and the mass composition directly, see *e.g.* ref. [25, 26]. Therefore, the flux  $J(E)$  and the relative abundance of different cosmic ray nuclei is well known in this energy range. The latter can be used to learn something about the propagation of the rays.

At energies above 100 TeV, the differential flux eventually becomes so low that only ground based experiments can provide the necessary exposure to collect enough events in a reasonable time frame. These experiments cannot observe the cosmic ray directly. Instead, they sample the extensive air shower generated by interactions of the cosmic ray with atmospheric matter. An extensive air shower is equivalent to a particle shower in a hadronic calorimeter. Therefore, one can say that ground based experiments use the atmosphere as a calorimeter for the cosmic rays.

Soft hadronic interactions are most important for the development of an extensive air shower. These interactions cannot be described in the perturbative approach to quantum chromodynamics. Therefore, systematic uncertainties arise in the reconstruction of the cosmic ray properties from the air shower data. The reconstruction of the cosmic ray mass  $A$  is particularly difficult. The relative abundances of different cosmic ray nuclei are generally not known at these energies. In most cases only a measure of the average

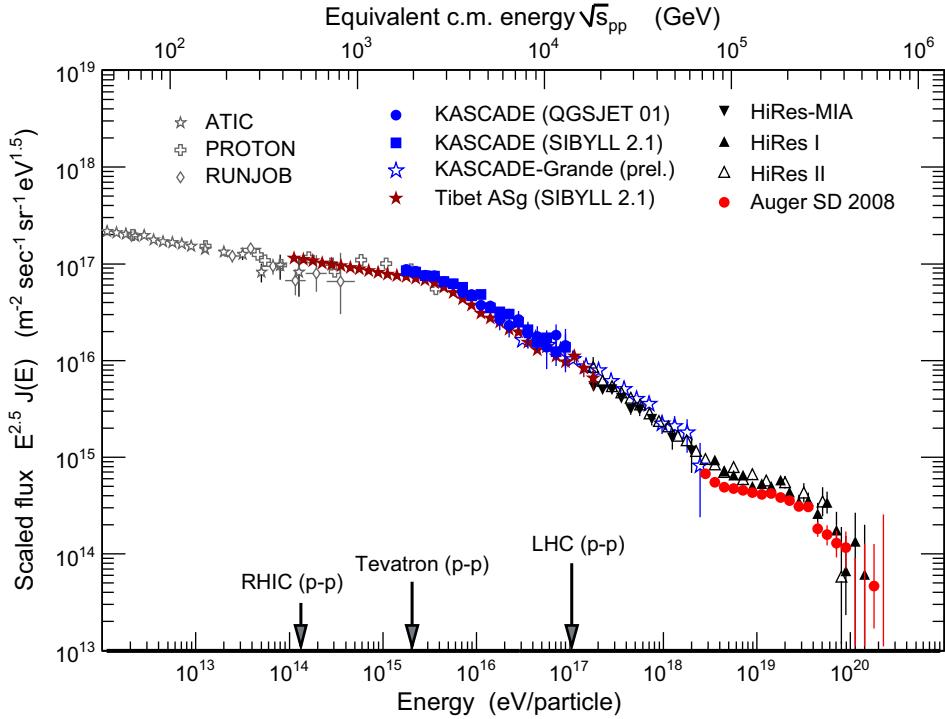


Figure 3.1: Shown is a compilation of measurements of the cosmic ray flux as a function of the cosmic ray energy, done by several air shower experiments. Note, the scaling of the cosmic ray flux with  $E^{2.5}$ , which emphasizes the features. The second horizontal axis shows the center of mass energy in the first interaction of the cosmic ray, if it was a proton. This interaction energy is compared with some collider experiments (see ref. [24] and references therein).

logarithm of the cosmic ray mass  $\langle \ln A \rangle$  can be derived. The differential flux  $J(E)$  has comparably large systematic uncertainties as well.

### 3.1.1 Cosmic rays up to 100 TeV

The relative abundance of nuclei with the charge number  $Z > 1$  in cosmic rays is similar to the interstellar medium [27]. This is a strong indication that cosmic rays are initially normal interstellar matter which is accelerated in an astrophysical process.

It is generally accepted that the majority of cosmic rays from a few GeV up to 100 TeV are accelerated in supernova blast waves. Turbulent magnetic fields in the shock front of these waves are able to accelerate charged particles from the interstellar medium. The general process is called first-order Fermi acceleration, see *e.g.* ref. [21].

The relative abundance of two groups of elements (Li, Be, B) and (Sc, Ti, V, Cr, Mn) is many orders of magnitude larger in cosmic rays than in the interstellar medium. These elements are essentially absent as end products of stellar nucleosynthesis. The discrepancy can be explained by spallation processes during the propagation of the cosmic rays. Spallation can occur when a cosmic ray collides with a proton from the interstellar medium. The cosmic ray nucleus is transformed or fragmented in the process. As a consequence, some of the more abundant elements are converted and fill up the rare elements. The Lithium-group is mostly generated by spallation of carbon and oxygen, while the Scandium-group is mostly generated by spallation of iron.

With knowledge about the spallation cross-sections at these energies, it is possible to derive the traversed slant depth  $X$  of the cosmic rays from the abundances of the spallation products. The slant depth is

of the order of  $10 \text{ g cm}^{-2}$  for cosmic rays of a few TeV and has to be compared with the average density of the interstellar medium of about proton per  $\text{cm}^3$ . The corresponding path length of cosmic rays is then of the order of 1000 kpc, much larger than the extensions of our galaxy which is 50 kpc at most.

This large path length can be understood by considering the gyro radius  $r_g$  of cosmic rays in this energy range. A handy formula for the gyro radius is

$$r_g = \frac{p c}{Z e B} \approx (1.1 \times 10^{-7} \text{ pc}) \frac{p/(\text{GeV}/c)}{Z(B/\text{nT})}, \quad (3.1.1)$$

whereas  $p$  and  $Z$  are the momentum and charge number of the cosmic ray. The galactic magnetic field is of the order of 0.1 nT and therefore a 100 TeV proton has a gyro radius  $r_g \approx 0.1 \text{ pc}$ . This is only a small fraction of the thickness of the galactic disk of about 300 pc.

Cosmic rays up to 100 TeV are apparently confined in our galaxy by magnetic fields and travel for a long time on complex trajectories until they eventually escape the galaxy. This interpretation is fortified by a more direct measurements of their average time of residence which can be obtained from an analysis of the relative abundance of long-living radioactive isotopes in the cosmic radiation (see ref. [22] and references therein).

The confinement into the galaxy is less efficient at larger energies, which makes the observed energy spectrum of cosmic rays steeper than the input spectrum at the source.

### 3.1.2 Cosmic rays above 100 TeV

Much less is certain about cosmic rays above 100 TeV. To a large degree, this is a consequence of the less precise or even controversial data delivered by experiments, which leave more room for interpretation. The acceleration mechanism which works so well at lower energies also runs into problems. The shock acceleration in supernova blast waves is not powerful enough to explain cosmic rays with energies much larger than 100 TeV and has to be replaced by something else.

A comparison of the gyro radius  $r_g$  of a cosmic ray with the thickness of the galactic shows, that protons up to about  $10^{17} \text{ eV}$  and iron nuclei up to about  $10^{18} \text{ eV}$  can be confined in our galaxy by magnetic fields. Cosmic rays up to these energies are generally believed to be of galactic origin. Above, they are considered to be extragalactic.

A rapid change in the spectral index  $\alpha$  of the differential flux  $J(E)$  is observed around  $4 \times 10^{15} \text{ eV}$ . The feature is called the *knee*. The spectral index  $\alpha$  changes from about  $-2.7$  to about  $-3.2$ . Furthermore, an increase of the average logarithm of the cosmic ray mass  $\langle \ln A \rangle$  with the energy  $E$  is observed above the knee which continues up to about  $10^{17} \text{ eV}$ .

There are two concurrent classes of theories for the knee. The first assumes that galactic accelerators reach their energy limit around the knee, so that cosmic ray components of different mass  $A$  successively vanish. In general, nuclei with a larger charge number  $Z$  are better confined via magnetic fields in the source and can be accelerated to higher energies.

The second class of theories assumes powerful accelerators, able to accelerate cosmic rays up to about  $10^{18} \text{ eV}$  or larger. In this scenario, the cosmic ray components of different mass  $A$  successively vanish, because the galactic magnetic fields are not able to confine them. The maximum energy of a cosmic ray in the galaxy depends again on its charge number  $Z$ .

Another feature is observed around  $4 \times 10^{18} \text{ eV}$ , the so called *ankle*. The spectral index  $\alpha$  of the differential flux  $J(E)$  changes again from about  $-3.2$  to about  $-2.6$ . There are two theories for the ankle. The first assumes that the galactic component of the cosmic ray flux extends up to the ankle and is taken over by a harder intergalactic component.

The second theory is the so-called dip model [28]. Here, the take-over of the extra-galactic component happens earlier, somewhere between the knee and the ankle. The ankle is explained as a consequence of the propagation of extragalactic cosmic ray protons in the cosmic microwave background (CMB). During their propagation, the protons loose energy by  $e^+e^-$  production

$$p\gamma \rightarrow p e^+ e^-, \quad (3.1.2)$$

which starts around  $10^{18} \text{ eV}$  and produces the ankle feature in the dip model.

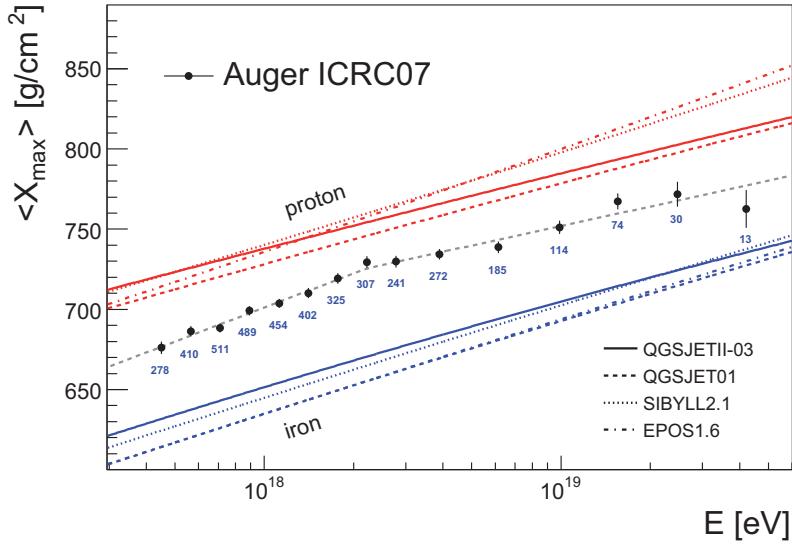


Figure 3.2: Shown is the depth of the electromagnetic shower maximum  $X_{\max}$  extracted from data of the Pierre Auger Observatory as a function of cosmic ray energy. The data are shown as black dots, the gray dashed line is a fit. The straight lines above and below the data are predictions from several hadronic interaction models (from [29]).

The dip model assumes that the cosmic rays between  $10^{17}$  eV and  $10^{19}$  eV are mostly protons while the takeover model expects an iron dominated composition up to  $10^{18}$  eV and then a fast transition to the composition of extragalactic cosmic rays. Measurements of the cosmic ray composition help to decide between these models, but do not clearly favor one or the other. The depth of the electromagnetic shower maximum  $X_{\max}$  is an example of a mass-sensitive air shower observable with the property  $\langle X_{\max} \rangle \propto \langle \ln A \rangle$ . A recent measurement from the Pierre Auger Observatory is shown in Fig. 3.2.

The last feature in the cosmic ray flux is the cut-off, experimentally established around  $7 \times 10^{19}$  eV. The spectral index  $\alpha$  changes from about  $-2.6$  to about  $-4.3$  [30]. The position of the cut-off matches a 40 years old prediction, the Greisen-Zatsepin-Kuzmin limit (GZK limit) [31, 32]. The GZK limit is also a consequence of the high density of CMB photons. Cosmic ray protons above the limit loose energy through pion production

$$p \gamma(\text{CMB}) \rightarrow p \pi^0 \rightarrow p 2\gamma \quad (3.1.3)$$

$$\rightarrow n \pi^+ \quad (3.1.4)$$

while heavier cosmic rays nuclei are destroyed by photo-disintegration. Fig. 3.3a) shows the attenuation lengths for several nuclei at ultra-high energies. Only nearby sources can possibly contribute to the observed flux on the Earth above the GZK limit and therefore the flux should be suppressed.

The GZK limit is a reasonable explanation based on well-established physics, like Lorentz-invariance and cross-sections measured in the laboratory. Nevertheless, there is also another possible explanation for the cut-off. Fig. 3.3b) shows a compilation of possible intergalactic accelerators for cosmic rays. In first approximation, the maximum energy provided by a source is proportional to its typical extension  $L$  and typical magnetic field strength  $B$ , times the charge number  $Z$  of the cosmic ray. If the acceleration process is a scaled variant of the shock front acceleration in supernova blasts, it is also proportional to the speed of the shock  $\beta_s$ , measured in units of the speed of light. Put together, this gives a magnitude estimate [33]:

$$E_{\max} \simeq (10^{18} \text{ eV}) Z \beta_s \left( \frac{L}{\text{kpc}} \right) \left( \frac{B}{\mu\text{G}} \right), \quad (3.1.5)$$

with  $1 \text{ G(auss)} = 10^{-5} \text{ T}$ . The estimate shows that the known astrophysical objects are barely able to reach an energy level of  $10^{20}$  eV. It is therefore also possible that the observed cut-off is caused by the energy limit of the accelerators.

Conventional acceleration scenarios for ultra-high energy cosmic rays are called *bottom-up* approaches in contrast to the more exotic *top-down* approaches. The latter explain the extragalactic flux by rare decays of super heavy relics from the big bang. The relics may be new particles, topological defects, or primordial black holes. A general feature of top-down models is an enhanced photon and neutrino flux at the Earth, and a cosmic composition made entirely of (anti-)protons or neutrons.

The composition measurements performed by the Pierre Auger Observatory makes top-down scenarios unlikely, since also heavy cosmic rays seem to be present in the cosmic ray flux. Recent limits on the neutrino and photon fluxes obtained from data of the Pierre Auger Observatory add further constraints [34–36]. In particular, the photon fraction above  $10^{19}$  eV is smaller than 2% at 95% confidence, which excludes most top-down models.

Finally, the experimentally observed anisotropy of cosmic rays above  $5.7 \times 10^{19}$  eV favors the explanation of the flux suppression with the GZK limit. At large distances, the arrival directions of cosmic rays are made isotropic by random intergalactic magnetic fields. Thus, an anisotropy can only be observed, if only nearby sources contribute to the flux observed at the Earth.

Fig. 3.4 shows the anisotropic distribution of the 27 highest energy events from ref. [37]. The data allows to reject the hypothesis that the distribution is isotropic with a confidence of 99 % [37, 38]. The result is obtained from a comparison of the arrival directions of the highest energy cosmic rays with position of active galactic nuclei (AGNs) in the sky. Still, the observed correlation does not prove a general connection of extragalactic cosmic rays with AGNs, which may simply be tracers of the true sources.

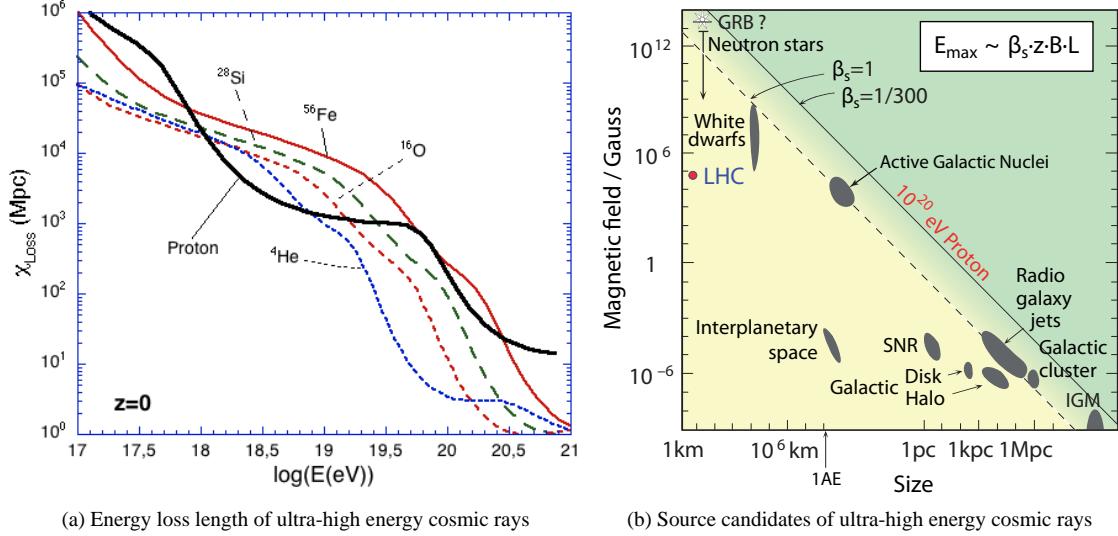


Figure 3.3: a) The graphs show the attenuation lengths for several nuclei due to interactions with the cosmic microwave background, calculated at a red-shift  $z = 0$  (from ref. [39]). b) The figure shows source candidates of cosmic rays in relation to the typical magnetic field strength and the size of their acceleration regions. The maximum energy increases from the bottom left to the top right. The Large Hadron Collider is shown for comparison (from ref. [24], original: ref. [33]).

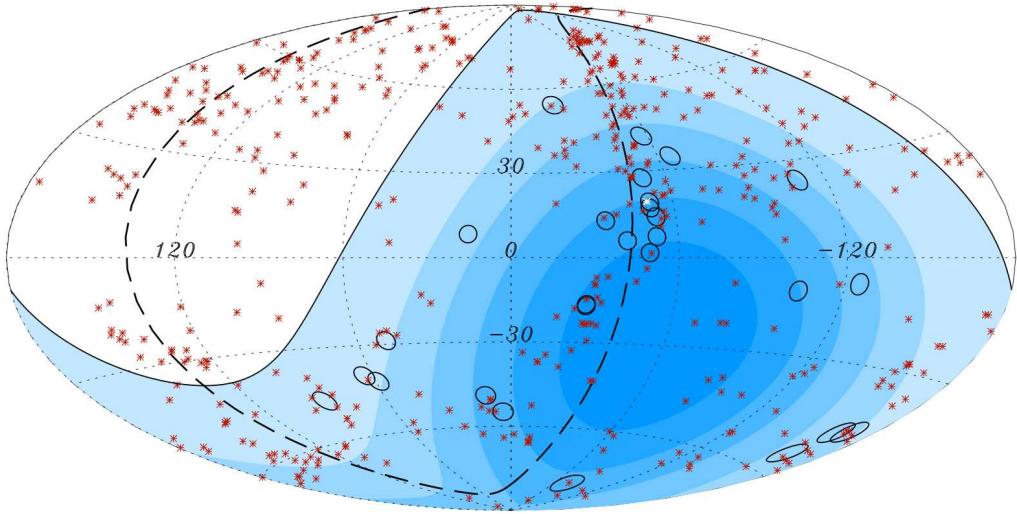


Figure 3.4: Shown is a Aitoff projection of the celestial sphere in galactic coordinates. Circles with a radius of  $3.1^\circ$  represent the arrival directions of the 27 cosmic rays with the energies larger than  $5.6 \times 10^{19}$  eV detected by the Pierre Auger Observatory up to 26<sup>th</sup> May 2006. Red asterisks indicate the positions of 472 of galaxies with active nuclei from the 12<sup>th</sup> edition of the catalog of quasars and active nuclei [40] within 75 Mpc distance. The solid line shows the border of the field of view of the Southern Pierre Auger Observatory, if zenith angles up to  $60^\circ$  are taken into account. The shades of blue indicate the relative exposure of the observatory in these coordinates, each band has equal integrated exposure. The dashed line is the supergalactic plane. Centaurus A, one of the closest active galactic nuclei, is marked in white (from ref. [37]).

## 3.2 Extensive air showers

A cosmic ray proton or nucleus with an energy larger than a few GeV initiates an *air shower* in the atmosphere of the Earth. An air shower is a cascade of secondary particles generated by the successive interactions of the primary cosmic ray with the electrons and nuclei in the atmosphere. As the cosmic ray energy increases, the term *extensive air shower* is used, referring to an air shower with a lateral size of at least several hundred meters. The particles in an extensive air shower form a slightly curved front, which moves with the speed of light. The phenomenon is well described in the literature, see *e.g.* ref. [21–23, 44].

Fig. 3.5 illustrates the main components in an air shower and Fig. 3.6 shows a simulation of two very inclined showers. The secondary particles in an air shower can be grouped into four basic components: the electromagnetic, muonic, hadronic, and the neutrino component.

In each hadronic interaction, the primary cosmic ray loses about half its kinetic energy, which serves to produce hadrons. Light mesons are preferred, but also baryons are produced in rare cases. Pions are most frequent (about 90 %), followed by kaons (about 10 %) [44]. The pion multiplicity per interaction increases slowly with the beam energy, it is of the order of 10 (at typical energies after a few cascading steps) to a few 100 (in the first interaction) for cosmic rays between  $10^{18}$  eV and  $10^{20}$  eV [22].

The unstable mesons decay before making another interaction if their time-dilated decay length  $\gamma\beta c\tau$  is shorter than their hadronic interaction length  $l_{\text{int}}$ . An overview of the decay constants  $c\tau$  and hadronic interaction lengths  $l_{\text{int}}$  of typical shower particles is given in Table 3.1 and Fig. 3.7b).

About 1/3 of the pions in the first interaction are neutral. Through their short life-time, they decay almost immediately into two high energy photons. The probability for another hadronic interaction remains slim even for neutral pions generated in the first interaction of a  $10^{19}$  eV cosmic ray, and it can be safely neglected at lower energies. The  $\pi^0$ -decay feeds the electromagnetic cascade, which dilutes the initial photon energy further into numerous electromagnetic particles of low energy by pair production and bremsstrahlung processes.

The charged pions and the kaons except for  $K_S$  have much longer life-times and usually produce more particles in successive hadronic interactions with air nuclei. This forms the hadronic component. In successive cascading steps, the kinetic energy of each pion is used up to generate more particles.

Again, about 1/3 of the energy after each step of the hadronic cascade goes into the electromagnetic component and forms new electromagnetic sub-showers. There are feed-back processes like electromagnetic interactions with air nuclei and direct pair production of muons, but their effect is negligible in hadronic showers. In the end, most of the cosmic ray energy is carried by electromagnetic particles.

The hadronic cascade comes to an end, when the time delayed decay length  $\gamma_\pi\beta_\pi c\tau_\pi$  of the charged pions becomes smaller than their hadronic interaction length  $l_{\text{int}}$ . It is possible to define a *critical energy*  $\xi_c^\pi$  for charged pions from this decay condition

$$\gamma_\pi \beta_\pi c \tau_\pi \stackrel{!}{=} l_{\text{int}}(h_{\text{max}}^\pi) \quad \Rightarrow \quad \xi_c^\pi \approx \frac{l_{\text{int}}(h_{\text{max}}^\pi)}{c\tau_\pi} m_\pi c^2, \quad (3.2.1)$$

whereas  $l_{\text{int}}(h_{\text{max}}^\pi)$  is the typical distance between two interactions at the altitude  $h_{\text{max}}^\pi$  of the hadronic shower maximum and  $m_\pi$  the mass of the pion. The critical energy marks the energy at which most charged pions decay. The critical energy  $\xi_c^\pi$  ranges between 10 GeV and several 100 GeV, depending on the zenith angle  $\theta$  of the shower.

A charged pion decays almost always in a muon and a muon neutrino, feeding the last two shower components. The muon inherits about 80 % of the pion energy in this two-body decay. The typical muon energy at the production point is therefore of the same order as the critical energy  $\xi_c^\pi$ .

The electromagnetic cascades come to an end, when the electrons reach their critical energy  $\xi_c^e \approx 87$  MeV in air. Electrons at this energy start to lose more energy in collisions faster than in radiative processes and get quickly absorbed in the atmosphere. A part of the energy dumped by electromagnetic particles into the air is released through the isotropic emission of fluorescence light. Relativistic electrons and muons also produce Cherenkov light in air, which is collimated in the forward direction.

The point where the number of electromagnetic particles reaches a maximum can be measured with telescopes, that detect the emitted fluorescence light. Fig. 3.2 shows measurements of the depth of the electromagnetic shower maximum  $X_{\text{max}}$  from the Pierre Auger Observatory. The atmosphere above the

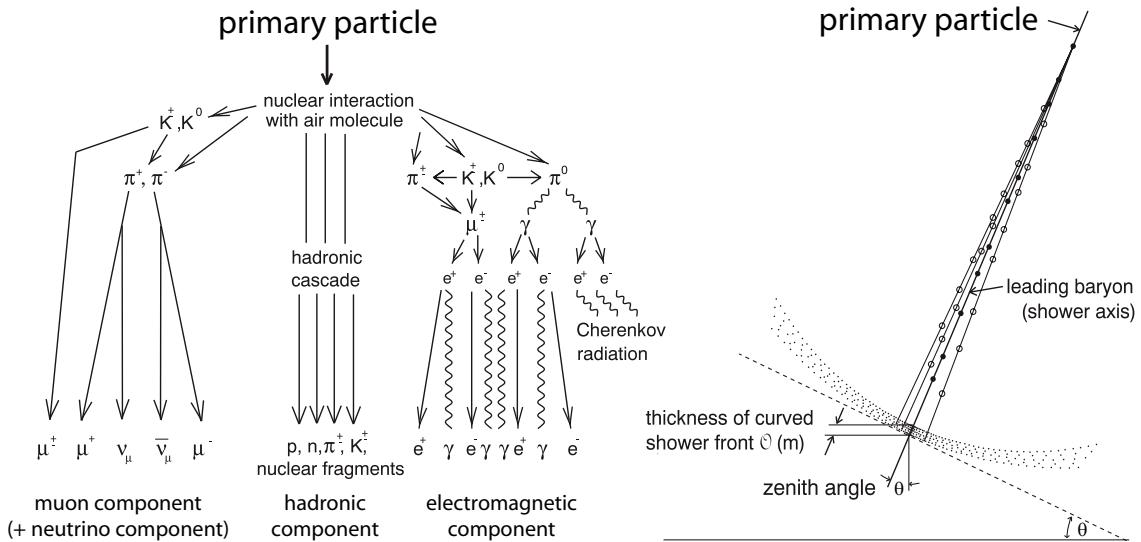


Figure 3.5: Left: Shown are the main processes in an extensive air shower, which form the hadronic, electromagnetic, muonic, and neutrino component (see text). Right: Drawing of a lateral section through the shower with its shower front of finite width (see text). The drawings are not to scale (adapted from [41]).

Table 3.1: The table shows an overview of the decay constants  $c\tau$  of the most common unstable particles in extensive air showers, taken from ref. [23]. Experimental uncertainties are not shown.

Particle	$\pi^{\pm}$	$\pi^0$	$K^{\pm}$	$K_S$	$K_L$	$\mu^{\pm}$	$n$
$c\tau/m$	7.8	$2.5 \times 10^{-9}$	3.7	$2.7 \times 10^{-2}$	5.1	659	$2.6 \times 10^{10}$

observatory is thick enough to contain the air shower development at least up to the maximum. The total atmospheric slant depth at the site as a function of the shower inclination  $\theta$  is shown in Fig. 3.7a).

The electromagnetic cascade is continuously fed by the hadronic cascade. The electromagnetic shower maximum is therefore coupled with the end of the hadronic cascade, and should be reached shorter after the latter. In first approximation, both are coincident. The majority of the muons are produced at end of the hadronic cascade.

At the electromagnetic shower maximum, the number of electromagnetic particles is much larger than the number of muons. At larger shower inclinations, the total atmospheric depth is large enough to fully absorb the electromagnetic component. After this transition, the shower is dominated by muons. Ground based detectors observe this transition at zenith angle between  $60^\circ$  and  $70^\circ$ , depending on their altitude.

### 3.2.1 Heitler-model of the hadronic cascade

A quantitative treatment of extensive air showers is quite complex and either involves solving coupled cascade equations or full Monte-Carlo simulations of the involved microscopic processes. Both approaches are described in Chapter 5.

To gain a qualitative understanding of an air shower, it is instructive to analyse a simplified view of the air shower cascade, which allows analytical calculations. Before the era of high-speed computing, Heitler presented such a model for the development of a photon induced shower [46]. An adaption of Heitler's approach to the case of hadronic cascades is presented in ref. [22, 45], which are the basis of the following discussion.

Fig. 3.8 illustrates the model approach. The hadronic cascade is approximated by a discrete sequence of simultaneous pion splittings. Each splitting occurs after a constant splitting length  $\ln 2 \lambda_\pi$ , which is the slant depth interval after which the pion has a 50 % chance for an inelastic interaction. At each splitting

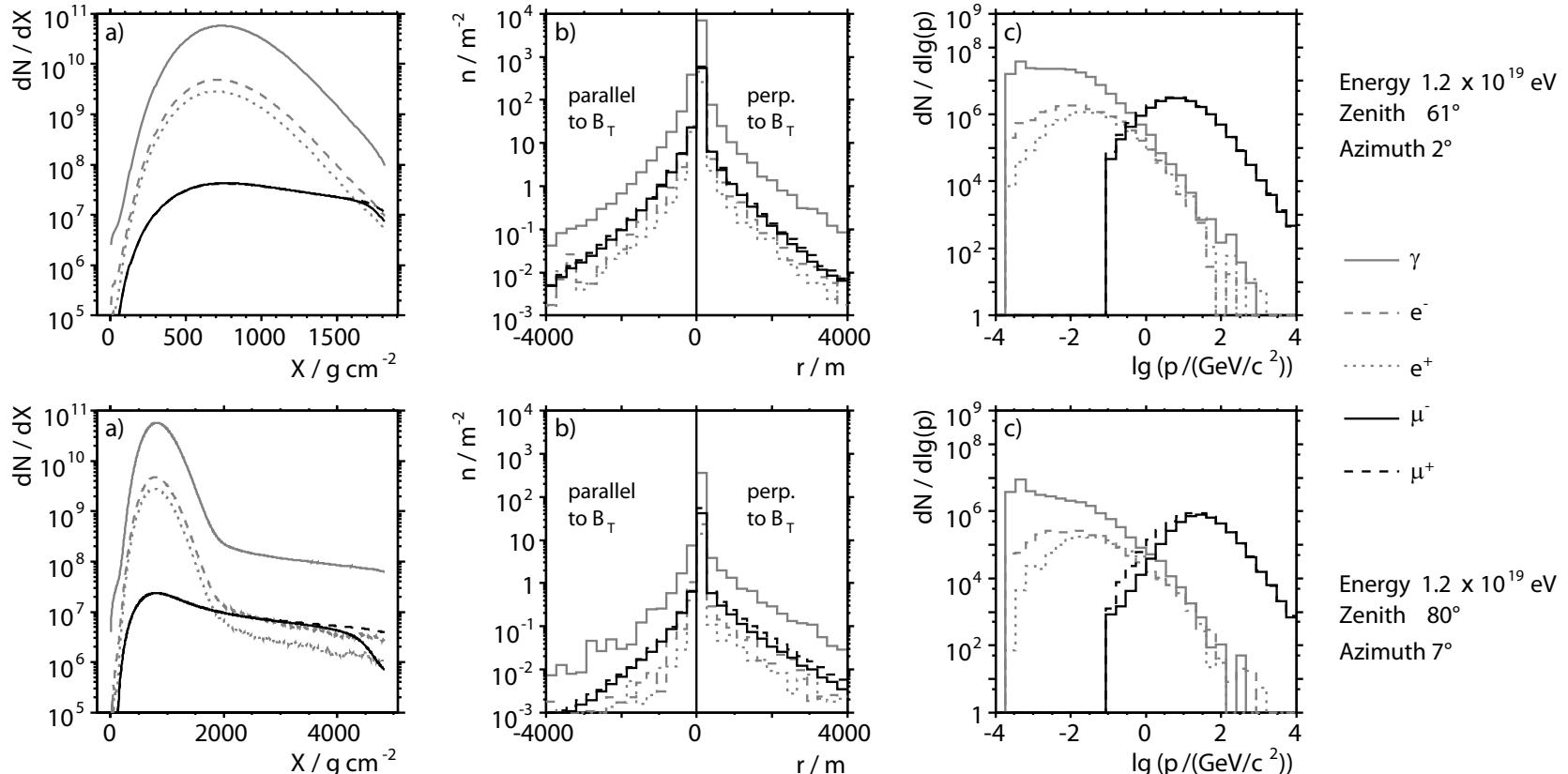


Figure 3.6: The graphs show two simulated proton showers from Chapter 5: a) number of particles vs. slant depth, b) particle number density on the ground vs. radial distance from the shower axis, c) energy spectrum on the ground. The right side of figure b) shows the radial distribution in slice perpendicular to the projected geomagnetic field  $B_T$ , while the left side shows a slice parallel to  $B_T$  (see Fig. 2.3). The particle depletion along  $B_T$  in b) for  $\theta = 80^\circ$  is caused by geomagnetic deflections. The peak in the photon energy distribution in c) is caused by  $e^+/e^-$  annihilation. The energy thresholds in all plots are 250 keV for electromagnetic particles and 0.1 GeV for muons.

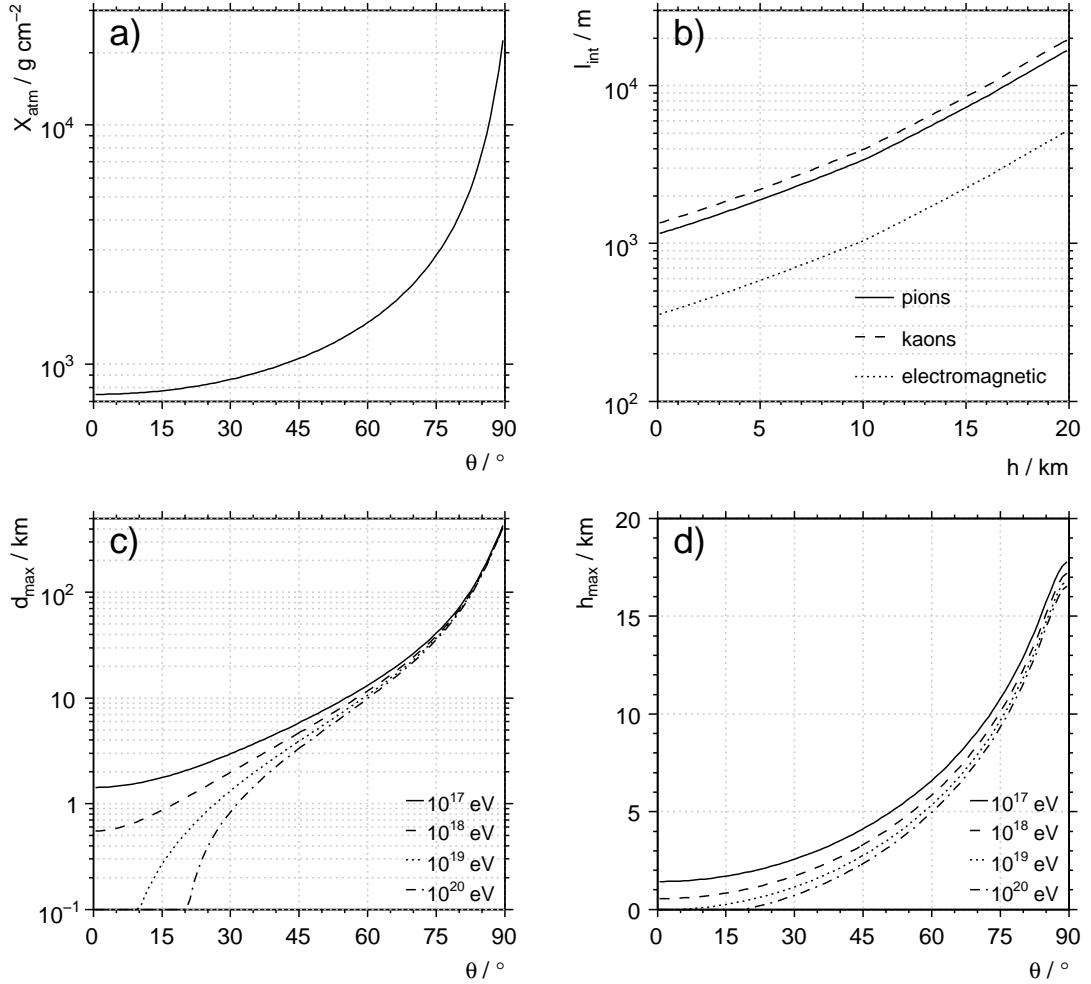


Figure 3.7: The figures show for an observer at an altitude of 1400 m: a) the total slant depth  $X_{\text{atm}}$  of the atmosphere, which an air shower has to pass to reach the ground [42, 43], b) the interaction length of some shower particles as a function of the altitude  $h$  (approximately valid between 10 GeV and 1000 GeV) [21, 42, 43], c) the distance  $d_{\text{max}}$  between the electromagnetic shower maximum and the impact point on the ground plane, and d) the altitude of said maximum above the ground  $h_{\text{max}}$  [29, 42, 43]. The calculations are covered in Appendix A.1.

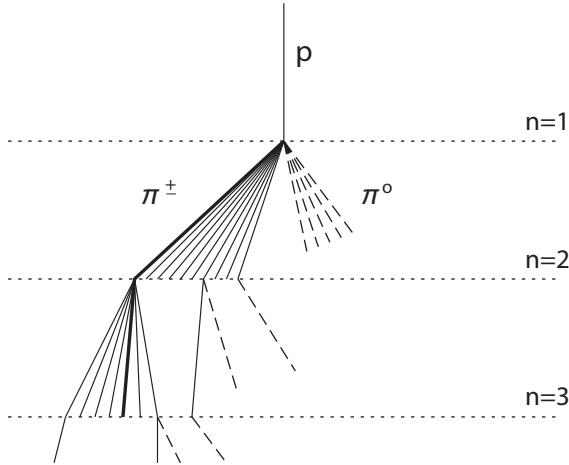


Figure 3.8: Schematic view of a hadronic shower, according to Heitler's approach. Dashed lines indicate neutral pions, which decay immediately and yield electromagnetic showers (not shown). The thick line indicates the leading particle, which carries a significant fraction of the total energy in the interaction. The residual energy is divided equally between the other pions. Not all pion lines are shown after the  $n = 2$  level (adapted from [45]).

level, a constant number of  $N_{\text{ch}}$  charged pions and  $\frac{1}{2}N_{\text{ch}}$  neutral pions are generated from each initial particle. Neutral pions decay immediately and feed the electromagnetic cascade.

One particle after each interaction is the leading particle, which gets a constant fraction  $(1 - \kappa)$  of the energy of the parent. In a real shower, the leading particle is usually a baryon. In the simplified model, it is usually treated like a charged pion and in rare cases like a neutral pion. The remaining energy fraction  $\kappa$  is divided equally on the other particles. Neutral pions decay immediately. A charged pion enters the next cascading step, until its energy falls below the critical energy  $\xi_c^\pi$ . If this happens, it decays into a muon before making another interaction. Neutrinos are neglected in the picture.

Many approximations are done to get simple analytical formulas. The most important ones are described in the following. The overall idea is to model the shower in terms of average interactions.

- **Proton cosmic ray.** The air shower in the simplified model is always initiated by a proton. The model will be generalised to the case of heavy cosmic rays in a later step.
- **Pion shower.** The air shower is approximated as a pure pion shower. As 90 % of the generated particles after each interaction are pions, this seems like a reasonable first approximation.
- **Constant values for  $\lambda_\pi$  and  $N_{\text{ch}}$ .** The pion interaction length  $\lambda_\pi$  and the  $\pi^\pm$  multiplicity  $N_{\text{ch}}$  are both weak functions of the beam energy  $E_\pi$ . Both are set constant.
- **Leading particle approximations.** When two hadrons interact, a significant fraction of the total energy is carried away by a single leading particle. The *inelasticity* of the interaction can be described by the parameter  $\kappa$ , which is the remaining fraction of the total energy available for immediate particle production. The inelasticity parameter  $\kappa$  is a function of the primary energy, and not well known at ultra-high energies. The predictions vary in the range 0.4 to 0.9 [21]. A constant value of 0.6 is used here.

In a real shower, the leading particle is usually a baryon, so that the energy of the leading particle remains in the hadronic cascade. However, there is also a small probability  $\epsilon$  for the leading particle to be a neutral pion. The neutral pion would decay immediately, taking a lot of energy out of the hadronic cascade. A constant probability  $\epsilon = 0.05$  is used here.

In the following, some basic formulas are derived from the simplified shower model. For comparisons with full simulations and data, also some numerical values are calculated. It shall be emphasized, that

the numerical values have to be considered with care and are by no means quantitative predictions. They mainly serve to show that the simplified shower model is indeed able to produce the right magnitudes. Any quantitative prediction of the model needs to be confirmed by realistic air shower simulations to be of any value.

The following gives a summary of the parameter values used in these calculations:

$E$	$= 10^{19}$ eV	energy of the initial cosmic ray
$N_{\text{ch}}$	$= 10$	average charged pion multiplicity [45]
$\xi_c^\pi$	$= 10$ GeV	critical energy for pions from the last section
$\lambda_\pi$	$= 100$ g cm $^{-2}$	average pion interaction length [21]
$\kappa$	$= 0.6$	total energy fraction available for pion production
$\epsilon$	$= 0.05$	probability that the leading particle is a $\pi^0$

### Energy calibration function

According to Fig. 3.8, the energy  $E$  of the cosmic ray is divided in the first interaction into three groups: the energy  $(1 - \kappa)E$  is carried by the leading pion,  $\frac{2}{3}\kappa E$  is carried by the other charged pions, and  $\frac{1}{3}\kappa E$  by neutral pions. The neutral pions decay immediately. The energy fraction  $E_\pi^{(1)}/E$  that remains in the hadronic cascade depends on whether or not the leading particle is a neutral or charged pion. There are two cases:

$$\begin{aligned} \text{leading particle } \pi^\pm: \quad P &= (1 - \epsilon), & \frac{E_\pi^{(1)}}{E} &= (1 - \kappa) + \frac{2}{3}\kappa \\ \text{leading particle } \pi^0: \quad P &= \epsilon, & \frac{E_\pi^{(1)}}{E} &= \frac{2}{3}\kappa. \end{aligned}$$

On average, the energy fraction

$$\frac{E_\pi^{(1)}}{E} = \kappa \left( \epsilon - \frac{1}{3} \right) + 1 - \epsilon \approx 0.78 \quad (3.2.2)$$

remains in the hadronic cascade.

This situation repeats in every step, so that in the  $n$ -th step, the remaining energy  $E_\pi^{(n)}$  in the hadronic cascade is

$$E_\pi^{(n)} = \left[ \kappa \left( \epsilon - \frac{1}{3} \right) + 1 - \epsilon \right]^n E. \quad (3.2.3)$$

The total number of charged pions in the  $n$ -th step is  $N_{\text{ch}}^n$ . The pion cascade stops, when the average energy per pion reaches the critical energy  $\xi_c^\pi$ :

$$\frac{E_\pi^{(n_c)}}{N_{\text{ch}}^{n_c}} \stackrel{!}{=} \xi_c^\pi \quad \Rightarrow \quad n_c = \frac{\ln[E/\xi_c^\pi]}{\ln[N_{\text{ch}}] - \ln[\kappa(\epsilon - \frac{1}{3}) + 1 - \epsilon]} \approx 8. \quad (3.2.4)$$

The charged pions decay into muons at  $n = n_c$ , which allows to derive the total number of muons  $N_\mu$  in the shower:

$$\ln N_\mu = \ln N_\pi = n_c \ln N_{\text{ch}} = \underbrace{\frac{\ln[N_{\text{ch}}]}{\ln[N_{\text{ch}}] - \ln[\kappa(\epsilon - \frac{1}{3}) + 1 - \epsilon]}}_{1/\gamma} \ln \left[ \frac{E}{\xi_c^\pi} \right]. \quad (3.2.5)$$

This can be rewritten in a form of a simple power law:

$$E = \xi_c^\pi \times N_\mu^\gamma. \quad (3.2.6)$$

The power law is a very important result, because it allows to reconstruct the primary energy of the cosmic ray by counting the total number of muons  $N_\mu$  in the shower. Eq. (3.2.6) will be generalised later in this section.

The parameters  $\rho_c^\pi$  and  $\gamma$  are constants in the simplified shower model. In reality, they are very weak functions of the cosmic ray energy  $E$ , but it will be confirmed with full air shower simulations in Chapter 5, that the approximation  $\gamma \approx \text{const.}$  and  $\xi_c^\pi \approx \text{const.}$  is excellent for  $10^{18} \text{ eV} < E < 10^{20} \text{ eV}$ . The numerical value  $\gamma \approx 1.11$  obtained in the simple approach is close to the simulation results.

### Invisible energy

Not all the energy of the cosmic ray is released into the atmosphere. In a real shower, most of the energy not carried by electromagnetic particles is carried by muons and neutrinos. The neutrino energy is lost entirely and muons loose only a small fraction of their energy in the atmosphere.

Fluorescence detectors observe the fluorescence and Cherenkov light, which is proportional to the energy deposit of the air shower in the atmosphere. The energy  $E_{\text{inv}}$  carried by the muons and neutrinos is invisible for these detectors. The energy measurement of the fluorescence detectors is corrected for this invisible energy, the correction is calculated with full air shower simulations.

The invisible energy fraction  $E_{\text{inv}}/E$  can be calculated with Eq. (3.2.3) within the simplified model:

$$E_{\text{inv}}/E = \frac{E_{\pi}^{n_c}}{E} = \left( \frac{E}{\xi_c^\pi} \right)^{1/\gamma-1} \approx \left( \frac{E}{\xi_c^\pi} \right)^{-0.10} \approx 0.14. \quad (3.2.7)$$

This result is of the same order as full simulations [47].

The equation shows, that the correction decreases slowly with the cosmic ray energy  $E$ , which is one of the reasons why the fluorescence method works best at ultra-high energies. The model also shows, that the correction has a sensitive dependency on  $\xi_c^\pi$  and  $\gamma$ , which depend on the details of the hadronic interactions in the shower. It will be shown later, that it also has a sensitive dependency on the mass  $A$  of the cosmic ray.

### Depth of hadronic shower maximum

With Eq. (3.2.4) and the splitting length  $\ln 2 \lambda_\pi$ , it is possible to calculate the depth of the pion shower maximum  $X_{\max}^\pi$ :

$$\begin{aligned} X_{\max}^\pi &= X_0 + n_c \ln 2 \lambda_\pi = X_0 + \frac{\ln 2 \lambda_\pi}{\ln[N_{\text{ch}}] - \ln[\kappa(\epsilon - \frac{1}{3}) + 1 - \epsilon]} \ln \left[ \frac{E}{\xi_c^\pi} \right] \\ &\approx 30 \text{ g cm}^{-2} + 19 \text{ g cm}^{-2} \ln \left[ \frac{E}{\xi_c^\pi} \right] \approx 590 \text{ g cm}^{-2}, \end{aligned} \quad (3.2.8)$$

whereas  $X_0 = \ln 2 \lambda_p(10^{19} \text{ eV}) \approx 30 \text{ g cm}^{-2}$  is the extrapolated proton-air interaction length at  $10^{19} \text{ eV}$  from ref. [21]. As discussed before, the  $X_{\max}^\pi$  of the hadronic cascade should occur shortly before the  $X_{\max}$  of the electromagnetic cascade. The model result is of the same order as the maximum of the muon shower in Fig. 3.6, which marks the end of the hadronic cascade.

A formula for the depth  $X_{\max}$  of the electromagnetic shower maximum may be derived with a similar approach [45]. One obtains the same functional form:

$$X_{\max} = X_0 + C \ln E \quad (3.2.9)$$

whereas  $C$  is a constant. If the average  $X_{\max}$  is plotted against the logarithm of the cosmic ray energy, one should obtain a straight line. Fig. 3.2 shows some simulations of  $X_{\max}$ , which are indeed straight lines in this representation.

### Shower-to-shower fluctuations

There are physical fluctuations in the air shower development, that even occur if identical cosmic rays are injected into the atmosphere. These shower-to-shower fluctuations affect the distribution of energy between muons and electromagnetic particles in an air shower and the depth of the electromagnetic and hadronic shower maximum.

The shower-to-shower fluctuations can be traced back more or less directly to statistical fluctuations in the first few interactions of the cosmic ray with the atmosphere. Statistical fluctuations in the interactions of the second particle generation already average out, because of the large number of particles  $N_{\text{ch}}^{(1)} > 100$  produced in the first interaction.

Only because of the leading particle effect, the second and third interactions matter at all. If there was no leading particle, the fluctuations caused by the second particle generation were already by an order of magnitude  $(N_{\text{ch}}^{(1)})^{1/2} \lesssim 10$  smaller.

The shower-to-shower fluctuations of the total number of muons  $N_\mu$  on the ground turn out to be important in the context of this work, and shall be estimated in the context of the simplified shower model. Only the effect of the first interaction is considered. Two sources for fluctuations of  $N_\mu$  can be identified: the division of energy in the first hadronic interaction and fluctuations in the altitude of the muon production.

The energy division is discussed first. A considerable amount of the total energy  $E$  of the cosmic ray goes directly into the electromagnetic component in the first interaction. The energy  $E_\pi^{(1)}$  remains in the hadronic cascade. Only this energy can serve to produce muons in the end. Thus, one can approximately write:

$$\frac{\sigma[N_\mu]}{N_\mu} \approx \frac{\sigma[E_\pi^{(1)}]}{E_\pi^{(1)}}. \quad (3.2.10)$$

To estimate the fluctuations of  $\sigma[E_\pi^{(1)}]/E_\pi^{(1)}$ , it is instructive to turn back to the initial considerations, which led to Eq. (3.2.2). If the leading particle always was a charged pion, fluctuations of  $E_\pi^{(1)}$  could only come from the branching ratio  $p = 2/3$  of charged to neutral pions, produced in the hadronic interaction. The variance of  $p$  can be approximated by binomial statistics:

$$\frac{\sigma[p]}{p} = \frac{1}{p} \sqrt{\frac{p(1-p)}{\frac{3}{2}N_{\text{ch}}^{(1)}}} \lesssim 6\%, \quad (3.2.11)$$

whereas  $\frac{3}{2}N_{\text{ch}}^{(1)}$  corresponds to the sum charged and neutral pions. The value  $N_{\text{ch}}^{(1)} \gtrsim 100$  is inserted to obtain the numerical result. The energy fraction not carried by the leading particle is  $\kappa$ , so the fluctuation caused by this effect is

$$\frac{\sigma_A[N_\mu]}{N_\mu} \approx \kappa \frac{\sigma[p]}{p} \lesssim 3\%. \quad (3.2.12)$$

Much larger is the fluctuation caused by the leading particle itself. There is only a small chance  $\epsilon$  for the leading particle to be neutral pion, but the energy fraction  $(1 - \kappa)$  immediately lost is large. The variance of the two cases serves as an estimate of the fluctuations:

$$\frac{\sigma_B[N_\mu]}{N_\mu} \approx \frac{\sigma[E_\pi^{(1)}]}{E_\pi^{(1)}} = \frac{(\epsilon(1-\epsilon))^{1/2}(1-\kappa)}{\kappa(\epsilon - \frac{1}{3}) + 1 - \epsilon} \approx 12\%. \quad (3.2.13)$$

This type of fluctuation is interesting, because it has a strong asymmetry. In a realistic shower, this effect should lead to a tail towards low numbers of muons. Such a tail is indeed observed in full air shower simulations, as will be shown in Chapter 5.

The second source of fluctuations is the altitude of the muon production, which is also to a large degree determined by the first interaction. The basis of this discussion is Eq. (3.2.5), which shows how the total number of muons  $N_\mu$  depends on parameters of the simplified model.

The calibration constant  $\gamma$  depends only weakly on the hadronic interaction parameters. It is assumed, that the fluctuations of  $\gamma$  are negligible. This leaves fluctuations of the critical energy  $\xi_c^\pi$  to consider. Eq. (3.2.1) shows, that the critical energy  $\xi_c^\pi$  is proportional to interaction length  $l_{\text{int}}$  at the altitude of the muon production, thus

$$\frac{\sigma_C[N_\mu]}{N_\mu} \approx \gamma \frac{\sigma[\xi_c^\pi]}{\xi_c^\pi} = \frac{\sigma[l_{\text{int}}]}{l_{\text{int}}}. \quad (3.2.14)$$

The interaction length  $l_{\text{int}}$  depends on the altitude of the muon production  $h_{\text{max}}^\pi$ , which varies with the depth of the pion shower maximum  $X_{\text{max}}^\pi$ . Most of the fluctuation of  $X_{\text{max}}^\pi$  is generated in the first interaction. The depth  $X_0$  of the first interaction in Eq. (3.2.8) has an exponential distribution. The interaction

length at  $10^{19}$  eV is  $\lambda_p(10^{19} \text{ eV}) \approx 40 \text{ g cm}^{-2}$  which is then also the size of the fluctuation of  $X_0$ . The value  $\sigma[X_{\max}^\pi] \approx 50 \text{ g cm}^{-2}$  for the whole cascade seems like a good estimate.

This corresponds to a fluctuation of about 1 km in the muon production altitude  $h_{\max}^\pi$ . This has an estimated effect of about 10 % on  $l_{\text{int}}$ , so that

$$\frac{\sigma_C[N_\mu]}{N_\mu} \approx 10 \%. \quad (3.2.15)$$

The combination of all effects gives 15 %, in good accord with the simulation results in Chapter 5. The analysis shows, that  $\sigma[N_\mu]/N_\mu$  is approximately independent of the cosmic ray energy  $E$  and direction, at least in very inclined air showers with  $60^\circ < \theta < 90^\circ$ . This feature is also confirmed in Chapter 5.

### Heavy cosmic rays

So far, only proton induced air showers were regarded in the simplified shower model. The formulas derived so far can be generalised to heavy cosmic rays with  $A$  nucleons by using the *superposition approximation* [21, 22, 48]. In this approximation, a heavy cosmic ray is regarded as a superposition of  $A$  simultaneous and independent proton showers, each carrying an energy  $E/A$ .

This approximation describes a particular limit of the first interaction of a heavy nucleus, in which the cosmic ray suffers a full fragmentation very high in the atmosphere. That the first interaction is high in the atmosphere is quite realistic, because the hadronic interaction length of a heavy nuclei is approximately by a factor  $A^{2/3}$  smaller than for a proton. In reality, however, it is more probable for the heavy cosmic ray nucleus to successively fragment into smaller pieces. Still, for the simple calculations considered here, the superposition approximation is good enough.

A particular important consequence of this picture also holds in a realistic scenario: if a variable  $x$  in an air shower depends only weakly on the primary energy of the cosmic ray, then it will also depend only weakly on the mass  $A$  of the cosmic ray:

$$\frac{dx}{dE} \approx 0 \quad \Rightarrow \quad \frac{dx}{dA} \approx 0. \quad (3.2.16)$$

In the simplified shower model, it is possible to derive generalised formulas for heavy cosmic rays from the proton results by substituting  $E \rightarrow E/A$  and summing over  $A$  showers, wherever appropriate. It is instructive to express the new generalised quantities in terms of the corresponding quantities of a proton shower with the same energy  $E$ :

$$[N_\mu]_A / [N_\mu]_p = A^{1-1/\gamma} \approx A^{0.10} \approx 1.4 \quad (3.2.17)$$

$$\left[ \frac{E_\pi}{E} \right]_A / \left[ \frac{E_\pi}{E} \right]_p = A^{1-1/\gamma} \approx A^{0.10} \approx 1.4 \quad (3.2.18)$$

$$\left[ \frac{\sigma[N_\mu]}{N_\mu} \right]_A / \left[ \frac{\sigma[N_\mu]}{N_\mu} \right]_p = \frac{1}{\sqrt{A}} \approx 0.13 \quad (3.2.19)$$

$$[X_{\max}^\pi]_A - [X_{\max}^\pi]_p = \frac{\ln 2 \lambda_\pi}{\ln[N_{\text{ch}}/(1 - \frac{1}{3}\kappa)]} \ln A \approx 19 \text{ g cm}^{-2} \ln A \approx 76 \text{ g cm}^{-2}, \quad (3.2.20)$$

where  $A = 56$  for an iron shower is used for the numerical value in each line.

Apparently, air showers initiated by heavy cosmic rays develop higher in the atmosphere, have more muons, and smaller shower-to-shower fluctuations. The numerical results for iron are again in of the same order as the full simulations in Chapter 5.

The generalised energy calibration function for heavy cosmic rays is

$$E = A^{1-\gamma} \xi_c^\pi \times N_\mu^\gamma, \quad (3.2.21)$$

which means that the general power law structure is preserved, if  $A$  is constant. In particular, the exponent  $\gamma$  is unchanged, which is in good accord with full simulations. In reality,  $\gamma$  is a very weak function of  $E$ , and therefore also a very weak function of  $A$ .

The last important insight is, that the power law is even preserved, if  $A$  is related to a power law of the primary energy  $E$  itself. The following energy dependence for the cosmic ray mass shall be analysed

$$A(E) = A_0 \times E^\beta, \quad (3.2.22)$$

whereas  $A_0$  and  $\beta$  are constants. In this case, the generalised energy calibration function turns to

$$E = A_0^{(1-\gamma)/(1+\beta(\gamma-1))} \xi_c^{\pi/(1+\beta(\gamma-1))} \times N_\mu^{\gamma/(1+\beta(\gamma-1))} \quad (3.2.23)$$

and the power law index  $\gamma$  is replaced by  $\gamma_A = \gamma/(1 + \beta(\gamma - 1))$ . If the composition of cosmic rays becomes heavier (lighter) as the energy increases, then  $\gamma_A < \gamma$  ( $\gamma_A > \gamma$ ), because  $\gamma > 1$  and  $\beta > 0$  ( $\beta < 0$ ).

This special case is not just academic. It turns out, that Eq. (3.2.22) is at least a good approximation to real data between  $2 \times 10^{18}$  eV and  $2 \times 10^{19}$  eV. This can be shown with Fig. 3.2. The measurements of the depth  $X_{\max}$  of the electromagnetic shower maximum are well fitted by a (broken) line in this graphical representation so that experimentally the following relation is observed

$$(X_{\max} - \text{const.}) \propto \ln E. \quad (3.2.24)$$

It is not clear whether Fig. 3.2 shows a change in the cosmic ray composition, because its interpretation depends on air shower simulations, which apparently have quite a large spread. However, it shall be assumed, that there is a composition change. The simplified air shower model predicts that

$$(X_{\max} - \text{const.}) \propto \ln A, \quad (3.2.25)$$

see Eq. (3.2.20), which was derived for the hadronic shower maximum, but an analog result is obtained for the electromagnetic shower maximum. Eq. (3.2.24) and Eq. (3.2.25) can only be fulfilled simultaneously, if  $A$  is either constant or changing like Eq. (3.2.22).

This is an important result, which will be picked up again in Chapter 7.

### 3.3 Very inclined air showers

It was discussed in the previous section, that a transition exists between  $\theta = 60^\circ$  and  $70^\circ$ , where the electromagnetic component becomes absorbed in the atmosphere, and the shower starts to be dominated by the muon component. The so called *very inclined air showers* with zenith angles between  $60^\circ$  and  $90^\circ$  are the main topic of this work, and will be focused on in the following.

#### 3.3.1 Muon component

The average muon energy immediately after their production is of the order of the critical energy  $\xi_c^\pi$  of the pions. The pions still have large Lorentz factors  $\gamma > 100$  at the end of the hadronic cascade, which are inherited by the muons. The muons therefore form a collimated beam in the forward direction. Positive and negative charges are produced in equal numbers. After their production, the muons are only affected by Coulomb scattering, energy losses, decay, and geomagnetic deflections.

##### Coulomb scattering

The charged particles in the shower are scattered predominantly in the electric Coulomb fields of air nuclei. Multiple scattering is well covered by Molière's theory [51, 52]. In a normal approximation [53] to this theory, the average squared deflection angle after a given slant depth  $X$  is [49]

$$\langle \theta_s^2 \rangle = \frac{X}{\lambda_{\text{em}}} \left( \frac{E_s}{m \gamma \beta^2} \right)^2, \quad (3.3.1)$$

where  $\lambda_{\text{em}} \approx 37.7 \text{ g cm}^{-2}$  is the radiation length in dry air,  $E_s \approx 0.021 \text{ GeV}$  is a scattering constant, and  $m$ ,  $\gamma$ , and  $\beta = v/c$  are the mass, Lorentz factor and velocity of the scattering particle in the laboratory frame. This effect is important for electrons at all zenith angles. It is sub-dominant for muons at  $\theta \lesssim 80^\circ$ . At larger zenith angles, Coulomb scattering and geomagnetic deflections are the main sources of the angular divergence of the muons from the shower axis which is observed on the ground [84].

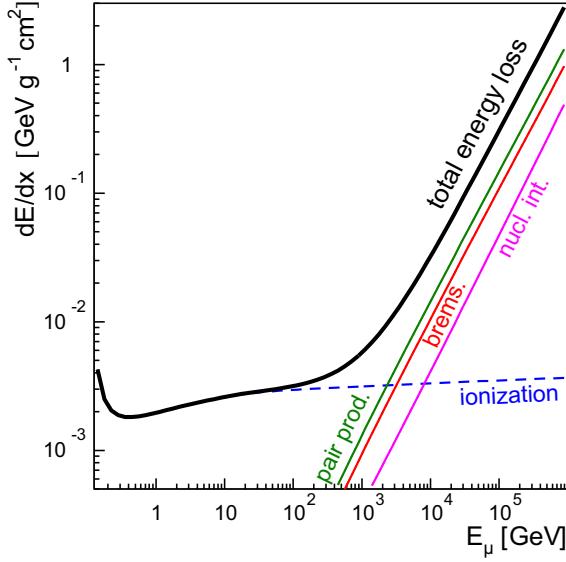


Figure 3.9: The graphs show the total energy loss of muons in dry air as calculated in CORSIKA [49], and the individual contributions due to ionisation loss, direct  $e^+e^-$  production in the electric field of a nucleus, bremsstrahlung, and nuclear interaction (adapted from [50]).

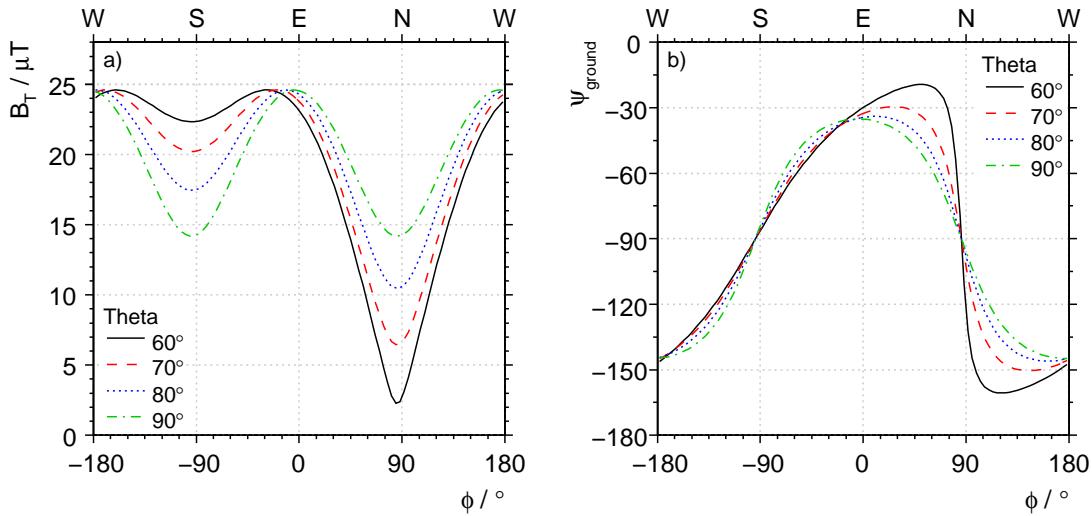


Figure 3.10: a) The graph shows the magnitude of the perpendicular component of the geomagnetic field with respect to the shower axis as a function of the shower direction. b) The plot shows the orientation of the ground plane in the shower plane coordinate system, which depends on the orientation of the shower axis and the geomagnetic field vector (see Chapter 2). The angle  $\psi_{\text{ground}}$  points to the early arriving part of the shower in this coordinate system. For both plots, a magnetic field of  $24.6 \mu\text{T}$  with an inclination of  $-35.2^\circ$  and a declination of  $4.2^\circ$  is used.

### Energy losses

Fig. 3.9 shows the energy loss of muons as a function of the energy. The moderate ionisation energy loss of about  $3 \text{ MeV/g cm}^{-2}$  is the only energy dissipation process for most muons. The total energy loss can be estimated from Fig. 3.2 and Fig. 3.7a), by assuming that all muons are produced at the depth  $X_{\max}$  of the electromagnetic shower maximum. The energy loss depends on the zenith angle  $\theta$  of the shower. It is only about 3 GeV at  $\theta = 60^\circ$ , but increases up to 70 GeV for as the zenith angle  $\theta$  approaches  $90^\circ$ . Because the muon energy  $\propto \xi_c^\pi$  also increases with  $\theta$ , the muon attenuation caused by energy loss remains small up to about  $80^\circ$ .

### Decay

The decay length  $\beta\gamma c\tau$  of a 10 GeV muon is 66 km, and thus much larger than the distance  $d_{\max}$  between the muon production point and the ground up to zenith angles  $\theta$  of about  $80^\circ$ , as shown in Fig. 3.7c). At  $\theta \gtrsim 80^\circ$ , a transition happens and muon decays become significant. In addition, muon attenuation due to energy loss processes becomes important.

The muon energy spectrum observed at the ground level is wide, as shown in Fig. 3.6. The decay and the muon attenuation effects above  $80^\circ$  make the spectrum harder by removing muons in the low energy tail. As a consequence, the average muon energy increases. The average energy at the muon production point increases, too, since the critical energy  $\xi_c^\pi$  for pions also increases with the zenith angle  $\theta$ . Together, all these effects lead to much higher average muon energies at the ground level as the zenith angle approaches  $90^\circ$ .

### Geomagnetic deflections

In the zenith angle range  $60^\circ \lesssim \theta \lesssim 70^\circ$ , where the electromagnetic component becomes extinct, also lateral deflections  $\delta x$  of muons caused by the geomagnetic field  $\mathbf{B}$  become relevant. The deflections can be approximated as [10, 54]

$$\delta x \simeq \frac{eB_T d^2}{2E_\mu/c} \quad \text{for } E_\mu \gg m_\mu c^2 \text{ and } d \gg \delta x, \quad (3.3.2)$$

whereas  $e$  is the elementary charge,  $d$  the distance between the muon production point and the ground along the shower axis,  $E_\mu$  is the muon energy, and  $B_T$  is the perpendicular component of the geomagnetic field  $\mathbf{B}$  with respect to the shower direction.

Fig. 3.10a) shows  $B_T$  as a function of the shower direction relative to the geomagnetic field. In the extreme case, a muon propagates in a  $25 \mu\text{T}$  geomagnetic field at the Auger South site. For a 10 GeV muon at a zenith angle of  $\theta = 60^\circ$  ( $80^\circ$ ), this leads to about 40 m (1600 m) lateral displacement after 10 km (66 km) travelled distance to the ground level. Geomagnetic deflections may therefore be neglected for  $\theta < 60^\circ$ , but are important in very inclined showers. They introduce an asymmetry in the lateral profile of the muon density, which is shown in Fig. 3.11.

The deflection also causes an increase of the total path length of the muons. Compared to the muon decay length, this increase remains negligible, so that the total number of muons on the ground is independent of the shower azimuth angle  $\phi$ :

$$\frac{dN_\mu}{d\phi} \approx 0. \quad (3.3.3)$$

### 3.3.2 Lateral shower profile at ground level

The lateral profile of an extensive air shower consists of photons, electrons, muons, and hadrons at all zenith angles. The density of these particles decreases very fast with increasing radial distance  $r$  to the shower axis, approximately like  $r^{-\alpha}$  for a certain  $r > r_{\min}$ . The values of  $\alpha$  and  $r_{\min}$  are in general different for the four components.

The lateral divergence in the shower is generated through the Fermi-motion of the partons in hadronic interactions, through the conservation of the transverse momentum in the two or more products of a decay, and to some degree by the geomagnetic field.

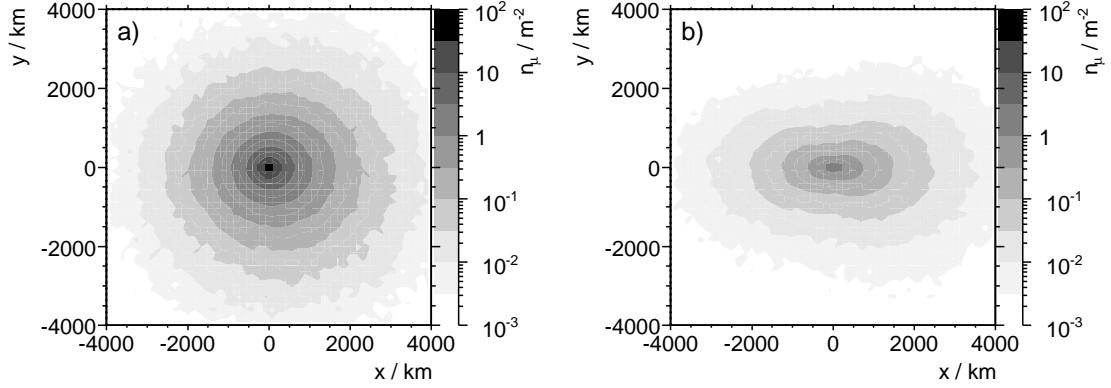
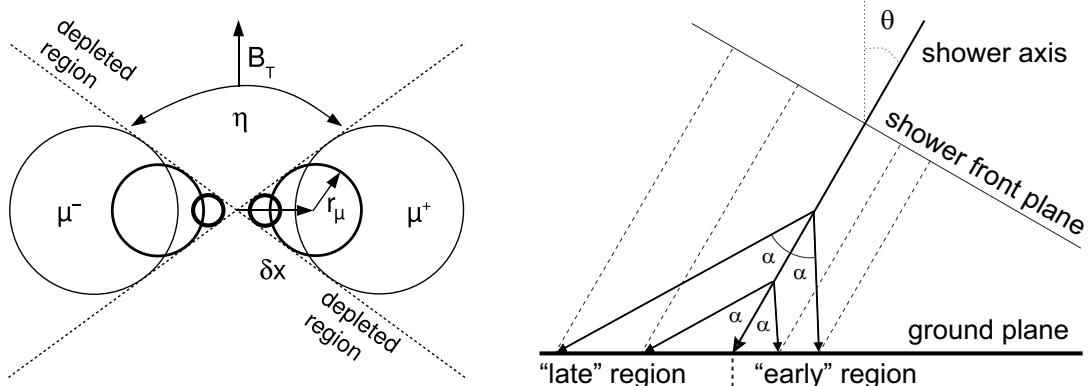


Figure 3.11: The profiles show the simulated muon density  $n_\mu$  on the ground at  $E \approx 10^{19}$  eV and  $\phi \approx 0^\circ$  in the lateral coordinate system. The showers are the same as in Fig. 3.6. a) The radial symmetry around the shower axis is still largely intact. b) The radial symmetry is broken by geomagnetic deflections.



(a) Simplified picture of the deflection of muons through the geomagnetic field in the shower front plane. For  $\eta > 1$ , depleted regions form (after [10]).

(b) Asymmetry in the density observed in the “early” and “late” regions of the shower on the ground. Distances are not to scale (adapted from [55]).

Figure 3.12: The geometrical drawings illustrate a) the geomagnetic and b) the early-late asymmetries in the lateral muon density  $n_\mu$  on the ground in very inclined air showers, which are explained in the text. a) The drawing shows rings formed by muons of equal energy  $E_\mu$  around the shower axis in the shower front plane. The line thickness indicates the magnitude of  $E_\mu$ . The geomagnetic field  $B_T$  separates the muon charges, shifting the rings by an offset  $\pm\delta x$ . If the field is strong enough, two depleted regions with an angular size  $\eta$  form around the magnetic field direction. b) The drawing in shows the shower from the side. Muons generated with equal angles  $\alpha$  to the shower axis at different points on the shower axis appear more dense in the early region of the shower on the ground than in the late region. Muons in the late region travel longer before reaching the ground level and get more attenuation than muons in the early region.

The lateral density<sup>1</sup> profile of the muons  $n_\mu$  is the most important one for the signal profile measured in an array of Water-Cherenkov detectors on the ground at zenith angles larger than  $60^\circ$ . Electrons and photons are more numerous than muons at all zenith angles, but their energy is usually so low that they make only short tracks in water, and thus they produce much less Cherenkov light than muons per particle. The same arguments holds for the left-over hadrons. While the electrons still contribute to the total signal, the contribution of the hadrons can safely be neglected [56].

Fig. 3.11 shows  $n_\mu$  for two example showers and the loss of radial symmetry at large zenith angles. There are two kinds of asymmetry observed in the lateral muon profile  $n_\mu$ : geomagnetic asymmetries and so called *early-late* asymmetries. The asymmetries are discussed separately.

### Geomagnetic asymmetries

An analytical model of the muon density  $n_\mu$  in very inclined showers is derived in ref. [10, 54]. It is valid in the range  $10^{18} \text{ eV} < E < 10^{20} \text{ eV}$  and  $60^\circ < \theta < 90^\circ$ .

The muon density  $n_\mu$  is calculated in the shower front plane at the ground level. Coulomb multiple scattering, muon decay, and energy losses are all neglected. This is a good first approximation and removes all sources of early-late asymmetries. Furthermore, the decay angle  $\theta_\mu$  between the muon direction and the direction of its mother particle, the pion, is neglected. Initially, also the geomagnetic field is set to zero.

In this case, a muon inherits its direction from its mother particle, the charged pion. The pions have a certain transverse momentum  $p_T$  with the respect to the shower axis, which they obtain in hadronic interactions. This leads to the following ansatz

$$r_\mu = \frac{p_T}{p_\mu} d(\theta) + r_\pi \simeq \frac{c p_T}{E_\mu} d(\theta) + r_\pi \quad \Leftrightarrow \quad E_\mu = \frac{c p_T d(\theta)}{r_\mu - r_\pi}, \quad (3.3.4)$$

which relates the radial distance  $r_\mu$  of the muon from the shower axis to its energy  $E_\mu$ , the distance  $d$  from the production point to the ground, and the radial distance  $r_\pi$  of the parental pion.

Simulations show [57, 58], that the  $p_T$  distribution of pions is approximately independent of the energy  $E$  and mass  $A$  and direction of the cosmic ray. The radial distance  $r_\pi$  of the pions is of the order of 10 m to 100 m [57]. The distance  $d \approx d_{\max}$  depends only on the zenith angle  $\theta$  in good approximation for  $\theta > 60^\circ$ . Fig. 3.7c) shows, that the dependency of  $d$  on the cosmic ray energy  $E$  can be neglected at large inclinations. It follows from Eq. (3.2.16), that the dependency on the cosmic ray mass  $A$  can be neglected, too.

The angle  $\tilde{\theta}_\mu$  between the muon and the shower axis due to the transverse momentum  $p_T$  of the pion is

$$\tilde{\theta}_\mu \simeq \frac{p_T}{p_\mu}. \quad (3.3.5)$$

An average transverse momentum  $p_T \approx 0.3 \text{ GeV}$  yields an angle  $\tilde{\theta}_\mu \approx 1.7^\circ$ .

Eq. (3.3.4) relates the muon density  $n_\mu$  with the muon energy distribution  $dN/dE_\mu$ :

$$n_\mu = \frac{dN}{dx dy} = \frac{1}{2\pi r_\mu} \frac{dN}{dr_\mu} = \frac{1}{2\pi r_\mu} \left| \frac{dE_\mu}{dr_\mu} \right| \frac{dN}{dE_\mu} = \frac{c p_T d(\theta)}{2\pi r_\mu (r_\mu - r_\pi)^2} \frac{dN}{dE_\mu}, \quad (3.3.6)$$

where  $dN/d\psi = \text{const.}$  is used (no geomagnetic field). If a power law  $dN/dE_\mu \propto E_\mu^{-\alpha}$  is assumed for the muon energy spectrum (compare with Fig. 3.6, which shows muon energy spectra observed on the ground), one obtains

$$n_\mu \propto r^{-1} (r - r_\pi)^{-2+\alpha} \simeq r^{-3+\alpha} \text{ for } r \gg r_\pi, \quad (3.3.7)$$

which is in good accord with full simulations.

The geomagnetic field is now included. Inserting Eq. (3.3.4) into Eq. (3.3.2) from the previous section leads to an offset  $\delta x$  in the shower front plane

$$\delta x = \frac{eB_T(\theta, \phi)d(\theta)}{2p_T} (r_\mu - r_\pi) \approx \eta(\theta, \phi) r_\mu \quad \text{for } r_\mu \gg r_\pi, \quad (3.3.8)$$

<sup>1</sup>Strictly speaking, this “density” is actually the integrated muon flux through a particular ground area element over the arrival time interval of the shower front.

whereas  $r_\mu$  is the radial distance of the muon in absence of the geomagnetic field. The factor  $\eta(\theta, \phi)$  depends only on the shower direction in this approximation. Eq. (3.3.8) has geometrical implications that are illustrated in Fig. 3.12a). If the geomagnetic field is off, muons with equal energies form rings around the shower axis. If the geomagnetic field is turned on, positive and negative charges are separated and each ring separates into two, having an offset  $\pm\delta x$  from the shower axis.

If  $\eta > 1$ , which is the case for  $\theta \gtrsim 80^\circ$ , depleted regions form, which contain no particles in the simplified model. This allows to understand the butterfly shape of  $n_\mu$  at these angles, see Fig. 3.11, although the effect is not so distinct in a real shower.

Eq. (3.3.6) allows another important conclusion. The shape of the muon density  $n_\mu$  essentially depends only on the shape of the muon energy spectrum  $dN/dE_\mu$ . The latter is approximately independent of the cosmic ray energy  $E$  and mass  $A$ , and therefore the same is true for the shape of the muon density  $n_\mu$ .

The universality of the muon energy spectrum  $dN/dE_\mu$  can be understood with Fig. 3.7d). The altitude of the muon production does not change significantly with the cosmic ray energy  $E$ , and thus the critical energy  $\xi_c^\pi$  of the pions remains approximately constant. The insensitivity to the cosmic ray mass  $A$  then follows again from the insensitivity to the energy  $E$ , see Eq. (3.2.16).

These results imply, that the lateral muon density profile  $n_\mu$  on the ground approximately factorises into the total number of produced muons  $N_\mu^0$ , which depends on energy  $E$  and mass  $A$  of the cosmic ray, a zenith angle  $\theta$  dependent muon attenuation factor  $a$ , and a normalised lateral density profile  $p_\mu$ , which depends only on the shower direction [11, 13, 54]:

$$n_\mu \approx N_\mu^0(E, A) \times a(\theta) \times p_\mu(r, \psi; \theta, \phi), \quad (3.3.9)$$

The muon attenuation factor  $a(\theta)$  takes the so far neglected absorption effects into account, which reduce the total number of muons  $N_\mu$  on the ground with respect to the total number of produced muons  $N_\mu^0$

$$N_\mu \simeq a(\theta) \times N_\mu^0. \quad (3.3.10)$$

The attenuation factor  $a$  depends only on the shape of the muon energy spectrum  $dN/dE_\mu$  and the distance  $d(\theta)$  to the ground. Since the former is approximately constant, the muon attenuation  $a(\theta)$  is only a function of the zenith angle  $\theta$ . The factorisation will be confirmed quantitatively in Chapter 5 and used in Chapter 6.

A final remark on the derivation of  $n_\mu$  shall be made. To get a simple analytical model, the decay angle  $\theta_\mu$  between the direction of the muon and its parental pion and the Coulomb scattering of muons were neglected, like in the original derivation [10, 54]. These approximations shall be challenged now.

The fixed angle  $\theta_\mu$  between the muon direction and the direction of its parent, the charged pion, is [49]

$$\cos \theta_\mu = \frac{\gamma_\pi \gamma_\mu - \gamma'_\mu}{\gamma_\pi \beta_\pi \sqrt{\gamma_\mu^2 - 1}} \quad (3.3.11)$$

where  $\gamma_\pi$  and  $\gamma_\mu$  are the Lorentz factors of the charged pion and the muon,  $\beta_\pi = v_\pi/c$  is the pion velocity, and  $\gamma'_\mu \approx 1.039$  is the fixed Lorentz factor of the muon in the center-of-mass frame of the pion.

If the Lorentz factors are large  $\gamma_\pi \approx \gamma_\mu \gg 1$ , the muon angle  $\theta_\mu$  is approximately given by

$$\theta_\mu \simeq \frac{\sqrt{2}}{\gamma_\pi}. \quad (3.3.12)$$

For a Lorentz factor of 100, one obtains a decay angle  $\theta_\mu \approx 0.8^\circ$ . This is the same magnitude as the earlier derived muon angle  $\tilde{\theta}_\mu \approx 1.7^\circ$  generated by the transverse momentum. One can further calculate with Eq. (3.3.1) that the average angular divergence generated by Coulomb scattering becomes larger than the angular divergence inherited by the pions at  $\theta \approx 80^\circ$ . Both effects therefore should be regarded in a full analytical discussion.

However, it turns out that all these effects lead to  $E_\mu \propto r_\mu^{-1}$  and  $n_\mu \propto r^{-3+\alpha}$  in first approximation. Thus the qualitative results derived so far remain valid.

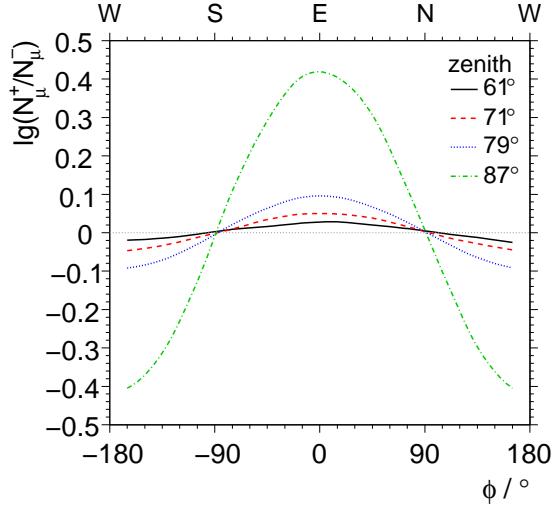


Figure 3.13: The plot shows the simulated ratio  $N_\mu^+ / N_\mu^-$  of positively and negatively charged muons, which arrive at the ground as a function of the shower azimuth  $\phi$  for different zenith angles. Positive muons are bent downwards for showers coming from the east. In this case, the ratio  $N_\mu^+ / N_\mu^-$  is larger than one, because negative muons have to travel longer paths and suffer more attenuation. At  $\theta \gtrsim 85^\circ$ , some muons are even bent so that they never reach the ground. The situation is reversed for showers coming from the west. The ratio is derived from air shower simulations described in Chapter 5 and approximately independent of the cosmic ray energy  $E$  and mass  $A$ .

### Early-late asymmetries

There are other effects that destroy the radial symmetry of the muon density  $n_\mu$  on the ground in very inclined showers beside the geomagnetic field. They are summarised under the term early-late asymmetries [16, 55].

The early-late asymmetries arise, because the lateral particle profile is not measured in the shower front plane perpendicular to the shower axis, but in the ground plane. This leads to differences in the muon density  $n_\mu$  in the early and late part of the shower, which are partly caused by a geometrical effect and partly caused by muon attenuation.

Fig. 3.12b) illustrates this. Muons in the early arriving part of the shower appear less inclined and more dense to an observer on the ground than muons in the late arriving part. This is the geometrical effect.

Furthermore, muons in the late part of the shower travel longer to the ground than muons in the early part. Muons in the late part suffer more energy losses and have more time to decay before reaching the ground level. The differences are relevant, which is shown by the simulated muon charge ratio in Fig. 3.13. This ratio is a function of the azimuth angle  $\phi$ , mostly because of the attenuation effects.

In general, the early-late asymmetries are not aligned with the geomagnetic field asymmetry. The angle between both in the shower plane coordinate system is shown in Fig. 3.10b).

### 3.3.3 Time structure of the shower front

The time structure of very inclined air showers is investigated in ref. [57, 58]. A general discussion of the shower front at all zenith angles can be found *e.g.* in ref. [59]. Muons travel nearly in straight lines and are the leading particles in the shower front at all zenith angles. In first approximation, the shower front is well described by a sphere expanding with the speed of light. The center of the sphere is approximately the point, where the hadronic cascade ends and the majority of the muons are produced.

The shower front has a finite width of a few meters, equivalent to a few tens of nanoseconds, see Fig. 3.5. At  $\theta \lesssim 60^\circ$ , this width is dominated by the delay of the electromagnetic particles. Electromagnetic

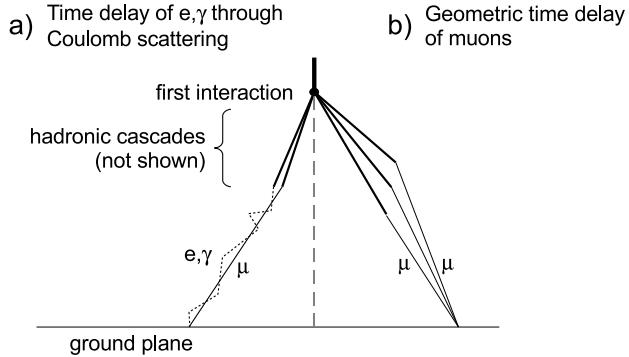


Figure 3.14: The drawing shows how a time delay is generated, even if all particles in the shower would move with the speed of light. Muons are the first arriving particles at all angles. Electromagnetic particles generated in the hadronic are delayed with respect to the muons because of Coulomb scattering. The time delay of the muons is mostly due to small differences in the individual velocities, at distances  $r > 1000$  m, also differences in the geometric paths play a role [57, 58]. All effects in the drawing are greatly exaggerated, the distances are not to scale.

particles get spread out in time by multiple scattering in contrast to muons, as illustrated in Fig. 3.14a).

At  $\theta \gtrsim 60^\circ$ , the electromagnetic component is extinct and the residual width is caused by the small differences in the velocity  $\beta$  of the relativistic muons and geometrical delays, which are shown in Fig. 3.14b). The muon delays are smaller than the delays of the electromagnetic particles and the shower front is better defined in very inclined air showers.

### 3.3.4 Muon generated electromagnetic particles

A small but non-negligible electromagnetic component is detectable on the ground at all angles, which is produced by the muons themselves. A discussion of this component can be found in ref. [11, 13, 16]. The muons produce photons and electrons mainly through decay.

A muon decays into an electron and two neutrinos. On average, the electron inherits about 1/3 of the muon energy, which is at least 106 MeV. This is many times the rest mass of an electron, but also quite close to the critical energy  $\xi_c^e$ . Thus, the electrons usually only generate short electromagnetic subshowers, before ionisation energy losses become dominant and they are quickly absorbed. On the other hand, because of the broad energy spectrum of the muons and their continuous energy losses, some muons always decay and refill the remaining electromagnetic particle spectrum.

The muons are therefore accompanied by a halo of electromagnetic particles of moderate lateral extension, typically a few tens of meters. If the density of muons is larger than  $0.1 \text{ m}^{-2}$ , the halos overlap and form a continuous electromagnetic particle front.

In general, the electromagnetic halo front arrives in coincidence with the muon front on the ground, because the electromagnetic particles do not survive long enough to build up a noticeable time delay. The particle density of the electromagnetic halo front follows the muon density closely, so that the number ratio  $n_{\text{em}}/n_\mu$  of electrons and muons remains nearly constant.

At larger zenith angles of  $\theta \gtrsim 85^\circ$ , a significant fraction of the muons can even reach energies up to a few TeV, where bremsstrahlung, direct pair production, and muon-nucleon interactions become important, see Fig. 3.9. These muon interactions favor the production of secondary particles which carry a significant part of the muon momentum, which are then able to initiate small electromagnetic sub-showers along the muon path. Therefore,  $n_{\text{em}}/n_\mu$  increases again after a plateau between  $65^\circ$  and  $85^\circ$  (see Chapter 5).

### 3.3. VERY INCLINED AIR SHOWERS

# Chapter 4

## Pierre Auger Observatory

The Pierre Auger Observatory is the world's largest instrument for measuring cosmic rays of ultra-high energies. It consists of two parts, one on the northern and one on the southern hemisphere, called *Auger North* and *Auger South*.

Auger North is currently in its planning stage [60] and will be built in the south-eastern part of Colorado, USA, near the city of Lamar. The northern site has a latitude of  $38^\circ$  north and an average altitude of 1100 m above sea level. The current design envisions an instrumentation of a ground area of 20 000 km<sup>2</sup>.

Auger South was completed in 2008 and instruments a ground area of 3000 km<sup>2</sup>. The observatory is located on a plateau north-east of the city of Malargüe, Argentina. The site has a latitude of  $35^\circ$  south and an average altitude of 1400 m above sea level.

Auger South already allows a good survey of the galactic center and the southern sky. Together with Auger North, the Pierre Auger Observatory will achieve almost uniform sky coverage, which is motivated by detailed anisotropy studies. The huge exposure of the observatory allows detailed studies of cosmic rays at the highest energies. Already now, Auger South collects about 200 events above  $10^{19}$  eV and about one event above  $6 \times 10^{19}$  eV each month. Auger North will triple these rates.

The design of the Pierre Auger Observatory incorporates two well-established measurement techniques, which are both based on the indirect observation of cosmic rays through air showers. The first method employs telescopes, which detect the fluorescence light generated by collisions of electrons and positrons in the shower with nitrogen molecules in the air. The second method samples the lateral density of the shower front on the ground with an array of particle detectors.

Fig. 4.1 gives an overview over Auger South. The southern part of the observatory has four fluorescence detector (FD) buildings, located at four corners of the surface detector (SD) array. Each FD building houses six telescopes and overlooks  $180^\circ$  in azimuth and about  $30^\circ$  of elevation above the horizon. The four FD buildings together completely cover the area instrumented by the 1630 particle detector stations of the surface detector<sup>1</sup>, which are placed in a triangular grid with a neighbor-to-neighbor distance of 1.5 km. The geography of the site is shown in Fig. 4.2. The SD array is very flat, the mean ground slope is smaller than 1 %.

The SD and the different FD buildings operate independently. The FD buildings operate in dark moonless nights and have an effective duty cycle of about 13 %. They observe the longitudinal shower profile directly, and perform an almost calorimetric measurement of the cosmic ray energy. The longitudinal shower profile allows sensitive studies of the nature of the cosmic rays, whether they are predominantly protons or iron nuclei, or something completely different.

The surface detector with a duty cycle of almost 100 %. It is the work horse to obtain the event statistic, but it observes only the part of the fraction of the air shower, which arrives at the ground level. The direction of the cosmic ray can be well reconstructed from this information. In principle, the reconstruction of the energy and mass from the surface detector signals has large theoretical uncertainties, but these uncertainties can be avoided by calibrating the SD with the FD.

<sup>1</sup>Current state at August, 19<sup>th</sup>, 2009.

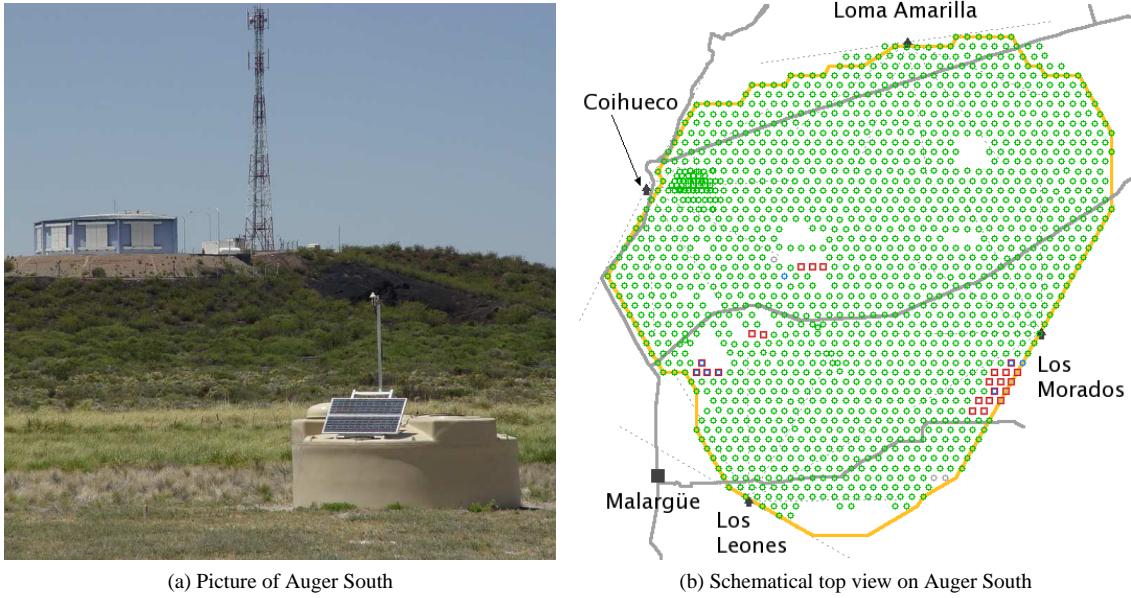


Figure 4.1: a) The picture shows one of the four FD buildings, a communication tower, and a SD station in the foreground. b) The drawing depicts a top view on Auger South. Green circles represent active SD stations, red and blue squares those which are just installed. Black dots represent the four FD buildings. Solid gray lines are roads, thin dashed lines indicate the field of views of the fluorescence telescopes. The yellow solid line indicates the design size of Auger South.

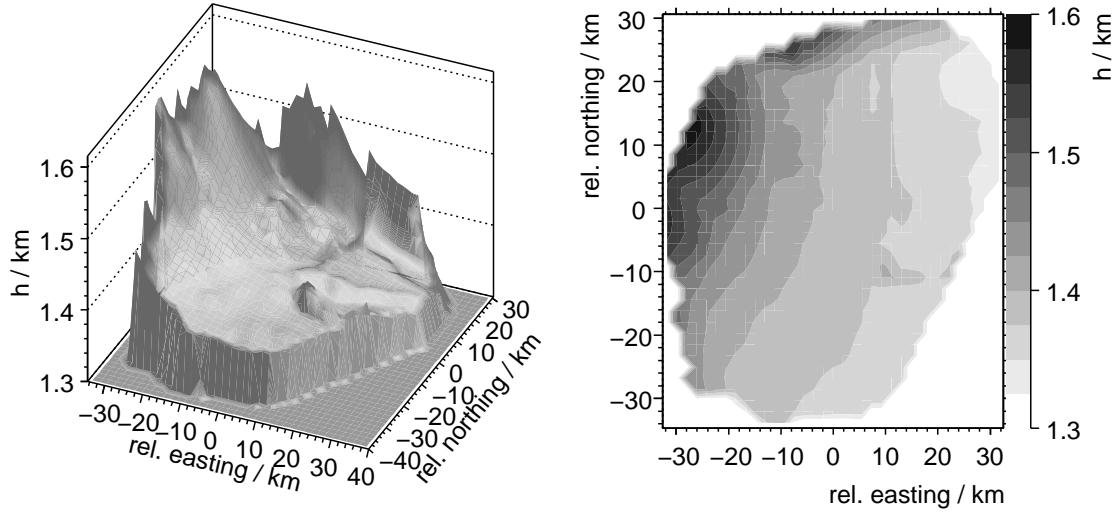


Figure 4.2: Altitude maps generated from the positions of all deployed SD stations of Auger South are shown in two different presentations. The positions are in UTM coordinates relative to the approximate center of the SD array. The large parts of the array are very flat. An increase in the altitude of about 250m is observed in the north-western direction, which marks the beginning of the Andes. The geographic features appear somewhat exaggerated due to the plot scales.

This is just one example for a genuine advantage of the Pierre Auger Observatory, which is its ability to observe air showers in several FD buildings or in the FD and the SD simultaneously. These simultaneous measurements can be used to cross-check the detectors and to quantify detection uncertainties.

The southern Pierre Auger Observatory was designed for cosmic rays with energies above  $10^{18}$  eV. An extension of the southern observatory to efficiently detect cosmic rays in the energy range of  $10^{17}$  eV to  $10^{18}$  eV is currently being built in the north-west corner of the array. This low energy extension will also serve as a test site for new detector developments, which benefit from the existing infrastructure and the possibility to compare new measurements with a well-understood reference. The low energy extensions and the new developments are shortly discussed at the end of this chapter.

## 4.1 Surface detector

The surface detector (SD) of the southern observatory is documented in ref. [61]. The cell unit of the array is the Water-Cherenkov-detector station [62] shown in Fig. 4.3. The Water-Cherenkov detectors of the Pierre Auger Observatory are based on those of the Haverah Park experiment [63].

Each SD station works completely autonomous. It has two solar panels and a battery, that provide a constant power of 10 W. The station is connected with the central data acquisition system (CDAS) in Malargüe via a radio link to the nearest communication tower. A common time reference over the array is established with the Global Positioning System (GPS) [64]. Each station is equipped with a commercial GPS receiver (Motorola OnCore UT), providing a pulse per second, which synchronises an internal 100 MHz clock. A micro-controller (80 MHz 403GCX PowerPC) runs the local station software being responsible for data acquisition, monitoring and communication. The latest data is always stored in memory and available upon request to the CDAS.

The exterior shell of the station is a cylindric polyethylene tank with a radius of 1.8 m and a height of 1.55 m. It encloses a Tyvek<sup>TM</sup> liner filled with  $12 \text{ m}^3$  of purified water up to a level of 1.2 m. The water has a resistivity of  $5 - 15 \text{ M}\Omega\text{cm}$  and a refraction index of 1.33.

Fast particles from an air shower produce Cherenkov light in this medium, which is detected by three semi-hemispheric 9 inch photomultipliers (Photonis XP1805PA/1). The photomultipliers (PMTs) are optically coupled to the water with a special compound, which guides about 90 % of the incoming light to the photocathode. The base of each PMT has two outputs, connected to the anode and to the last dynode, respectively. The nominal end-to-end gain factor of the PMTs is  $2 \times 10^5$ . The dynode signal is additionally amplified by a factor of 32.

The anode and dynode signals are filtered and digitized at 40 MHz using 10 bit Flash Analog Digital Converters (FADCs). Both signals together provide a high dynamic range from a few to about  $10^5$  photoelectrons, encoded effective in 15 bits with 5 bits overlap.

The FADC traces have a bin size of 25 ns, each bins contains 0 – 1023 channels. A pedestal of 50 channels is added to the signal to observe possible fluctuations of the baseline. The digitized signals are processed by a fast programmable logic device (PLD), which forms trigger decisions. In case of a trigger, 768 bins of all six FADC traces are read out and stored together with the time stamp of the trigger, whereas 100 bins are stored before and 668 bins after the trigger. The first 100 bins are in general signal free and used to analyse the signal baseline.

The station automatically collects monitoring data, which is send to the CDAS every 10 minutes. Among such data are the current PMT, battery and CPU board voltages and the water temperature in the tank. Finally, each station is equipped with an on-board LED flasher, which can be used to study the PMT linearity.

### 4.1.1 Signal calibration

The signal calibration of the SD stations is documented in full detail in ref. [65], which is the basis of the following summary.

Cherenkov pulses generated by charged particles within the water volume show exponential decays with a typical decay time of 60 ns. The decay time depends on the liner reflectivity and the water quality. An average pulse shape generated by a vertically and centrally through-going muon is shown in Fig. 4.4b).

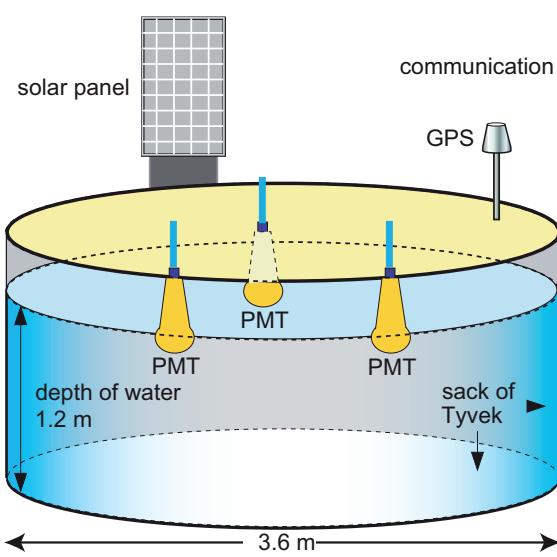


Figure 4.3: The drawing shows a station of the surface detector of the Pierre Auger Observatory, as described in the text (adapted from ref. [41]).

The both the integral and the height of the pulse above the baseline are proportional to the Cherenkov light generated by the particle. The pulse integral  $Q$  is equivalent to the charge collected at the PMT anode, the pulse height  $I$  is proportional to the peak current at the anode. The pulse integral  $Q$  is more precise and therefore used as the signal reference in the final event analysis. On the contrary, the station level triggers are implemented as thresholds for the pulse height and therefore based on  $I$ .

The average values of  $Q$  and  $I$  generated by identical particles varies from station to station and from PMT to PMT. The same is true for their ratio. Each PMT has slight differences in its amplification properties and its optical coupling to the water. Each station has slight variations in the quality of the water and the liner reflectivity. The PMT amplification also has a temperature dependence, which gives rise to daily and seasonal variations.

The trigger sensitivity of individual SD stations and the analysis of SD data should not dependent on such individual and varying properties. Therefore, the SD stations continuously perform a self-calibration. Nature provides an excellent calibration source: a uniform background flux of atmospheric muons is constantly generated by cosmic rays of a few GeV, which produces a high rate of muon hits of roughly 2.5 kHz in each station.

This flux can be used to express both  $Q$  and  $I$  in terms of a physical reference: the *vertical equivalent muon* = VEM. A VEM has a corresponding pulse height  $I_{\text{VEM}}$  and charge  $Q_{\text{VEM}}$ , which are used as base units for  $I$  and  $Q$ . Signals and trigger thresholds expressed in VEM are independent of individual station or PMT characteristics.

A single station cannot measure the directions or impact points of individual background muons. It therefore cannot measure the VEM-signal directly, but there are two indirect ways to derive the VEM signal from the background flux.

The first method is more precise, but computationally too expensive for the limited capabilities of a single station and therefore done offline (offline calibration). The second method is less precise, but simpler and can performed online every minute by the station itself (online calibration). Both kinds of calibrations are performed individually for each PMT and are robust against failures of individual PMTs in a station.

### Offline calibration

The integrals  $Q_i$  and heights  $I_i$  of Cherenkov pulses generated by the background muons are collected over some time to generate histograms. A clear peak turns up in these histograms, which is closely related to

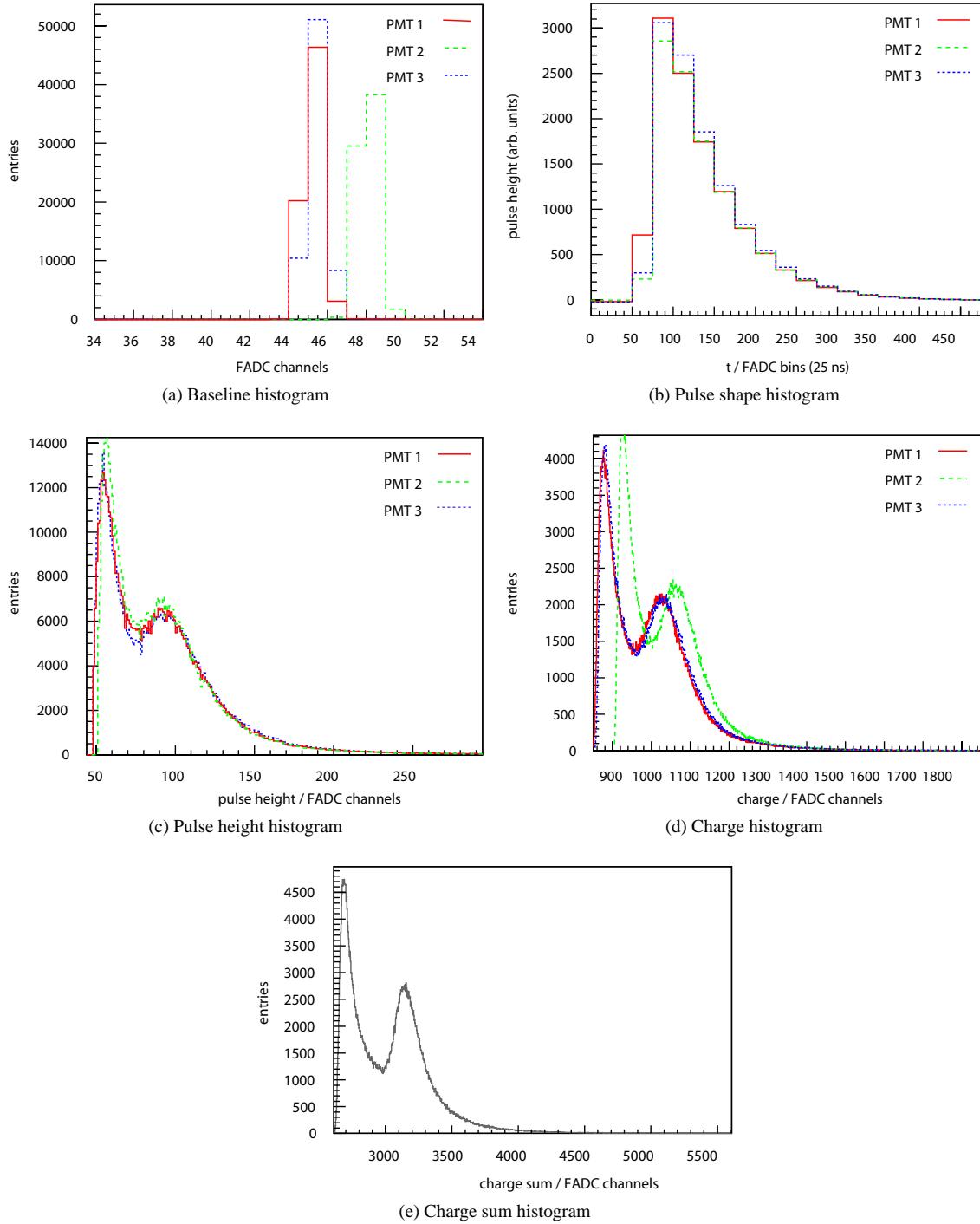


Figure 4.4: The plots show a sample of the 13 histograms of calibration data, which each surface detector station re-generates every minute. This data is send together with the signal trace upon a readout request by the central data acquisition system (adapted from ref. [65]).

the signal generated by a VEM. Examples are as shown in Fig. 4.4c)–e). The peak positions shall be called  $Q_{\text{VEM}}^{\text{peak}}$  and  $I_{\text{VEM}}^{\text{peak}}$ .

The VEM peak stands out in these histograms, because the track length for a vertically through-going muon is independent of the impact point on the surface of a cylinder. This makes the VEM signal more frequent than other signals. A small additional correction factor has to be applied, because the peak position  $\bar{Q}_{\text{VEM}}^{\text{peak}}$  in the histogram of the summed charge of all three PMTs actually corresponds to  $1.09 \pm 0.02$  VEM and the peak position  $Q_{\text{VEM}}^{\text{peak}}$  in the charge histogram of individual PMTs to  $1.03 \pm 0.02$  VEM.

These correction factors were found in dedicated experiments with a reference station [66], in which scintillators on top and underneath the station allowed to identify vertically and centrally through-going muons via coincidences. The origins of these shifts are well understood [67].

The histogram-based calibration of the signal has a precision of 2 %, which is much better than the typical statistical resolution of the signal. The station re-generates the required histograms of  $Q$  and  $I$  every minute and sends them together with the signal trace to the CDAS upon a station readout. The actual signal calibration is then performed offline.

The signals used in the event reconstruction is based on the calibrated pulse charge  $Q$ :  $S \text{ [VEM]} = Q/Q_{\text{VEM}}^{\text{peak}}$ .

### Online calibration

Above a certain threshold  $I_{\text{thr}}$ , the PMT trigger rate is dominated by real muon hits from the background flux over other forms of noise. Because this background flux is almost constant, the rate of such a threshold trigger corresponds to a physical threshold value in units of VEM.

For example, a threshold  $I_{\text{thr}} \approx 3$  VEM for a single PMT corresponds to a rate of about 100 Hz. The number of channels  $I_{\text{VEM}}^{\text{est}}$  above the baseline which corresponds to 1 VEM could therefore be derived by adjusting  $I_{\text{VEM}}^{\text{est}}$  until the rate of triggers with  $I > 3 I_{\text{VEM}}^{\text{est}}$  is 100Hz. The actual procedure is a more complex, but uses the same principle.

The relation between the trigger rate and the corresponding threshold value in VEM only has to be established once for a reference station, which itself may be calibrated manually with the histogram-based method.

The main purpose of this rate-based calibration is to retain a uniform trigger performance over the whole SD array. This is achieved by re-calibrating  $I_{\text{VEM}}^{\text{est}}$  with this procedure every minute. The thresholds of the proper station trigger algorithms are expressed in terms of VEM, and thus independent of individual PMT characteristics.

The rate-based calibration is also used to initially set up the end-to-end gain of each PMT by adjusting the high voltage upon the startup of a station. The high voltage is generally not re-adjusted during normal operation. The average gain per PMT set up in this way is  $3.4 \times 10^5$ , with an average of 94 photoelectrons per VEM. The initial calibration of the gain ensures a balanced peak response for all PMTs in a station, despite differences in the light collection efficiency and the optical coupling to the water. It also ensures a proper dynamic range for the electronics.

The discussion shows, that the station level triggers and the final signals used in an event analysis are based on different calibration methods. An absolute offset between  $I_{\text{VEM}}^{\text{est}}$  and  $I_{\text{VEM}}^{\text{peak}}$  is observed at the level of 5 %, with a seasonal variation of 3 % [65]. The seasonal variation is generated by a physical variation of the intensity of the background flux of muons, which in turn is caused by the differences in the atmospheric density profiles of summer and winter. The absolute offset is not important for the event analysis, but the seasonal variation of 3 % leads to a corresponding variation of the lowest level event trigger rate.

Also part of the online calibration is the monitoring of the dynode to anode signal ratio  $R$ . The nominal gain factor of the last dynode to the anode is 32, but the true value is continuously monitored and derived from real signals. About 100 large pulses are collected over a time span of three minutes. The pulses are selected as such that they stand out above the noise level in both the anode and the dynode traces, but do not saturate the dynode.

Simply calculating the ratio  $R$  from the heights of these pulse is not accurate enough, because the dynode and the anode signals have a phase delay  $\epsilon$  of about 5 ns. This phase delay is introduced by the additional amplification step of the dynode signal. The correct mathematical model of the relation is fitted

to the collected data instead:

$$A_i = \frac{1}{R}((1 - \epsilon)D_i + \epsilon D_{i+1}), \quad (4.1.1)$$

whereas  $A_i$  is the anode signal in time bin  $i$  in FADC channels above the baseline, and  $D_i$  and  $D_{i+1}$  are the corresponding dynode signals above the baseline in the same and the following bin.

### 4.1.2 Trigger system and data acquisition

The trigger system of the surface detector is documented in ref. [68]. The system has three levels. Trigger events are called T1 to T3, according to the trigger level. T1 and T2 events are generated locally in each station. The station stores the corresponding signal traces in a short time buffer. T2 events are automatically reported together with the station ID and the trigger time to the CDAS.

The CDAS continuously monitors the stream of T2s and searches for compact patterns in time and space, which are the signature of an air shower. If such a configuration is found, the CDAS emits a T3 and the array is read out. The third level trigger is therefore also called event level trigger or simply central trigger.

Local stations with a T1 or T2 are regarded as *candidate* stations in the event. They may be part of an air shower or just random coincidences, generated by the background flux of atmospheric muons. It is not the purpose of the third level trigger to distinguish between real showers and noise. More complex methods are better suited for this task and are applied offline as part of the reconstruction of an event candidate.

A significant number of candidate events are rejected by the offline selection, too. Therefore, the whole event reconstruction chain has to be regarded in an analysis of the detection efficiency of the surface detector.

The CDAS also counts the number of T2 events generated by each station every second. This data is written into special files and used offline to determine the exposure of the surface detector to cosmic rays. It allows to see whether a station was in data acquisition mode at a given second. The exposure calculation is discussed in Chapter 6 and Chapter 8.

#### Station level triggers

The FADC traces of all PMTs in a station are continuously monitored in hardware by a programmable logic device (PLD). Three kinds of triggers are implemented in this device. All work on the high-gain dynode signal and are applied to the calibrated pulse height  $I/I_{\text{VEM}}^{\text{est}}$ .

- **Thr1 (T1).** This a simple threshold trigger and a first level trigger. It requires a coincident crossing of a threshold of 1.75 VEM in all three PMTs. In the rare case, that only two (one) PMTs are operational, the threshold is 2 VEM (2.8 VEM). The trigger produces a rate of about 100 Hz.
- **Thr2 (T2).** This trigger is a stricter version of the Thr1 and a second level trigger. It requires a coincident crossing of a threshold of 3.2 VEM in all three PMTs. In the rare case, that only two (one) PMTs are operational, the threshold is 3.8 VEM (4.5 VEM). The trigger produces a rate of about 20 Hz.
- **Tot (T1, T2).** This time-over-threshold trigger is a first level trigger, but automatically promoted to the second level. It requires at least two PMTs to have at least 13 bins with more than 0.2 VEM in a sliding time window of 120 bins = 3  $\mu\text{s}$ . If only one PMT is operational, it is already regarded as sufficient, if only the remaining PMT satisfies the trigger condition. The trigger produces a rate of about 1 – 5 Hz.

The signal calibration ensures, that the dominating Thr1 and Thr2 rates remain constant. The Thr1 and Thr2 typically selects short signals with a length of 6 – 8 bins, equivalent to 150 – 200 ns, which are typically generated by muons. Example traces for each trigger condition are shown in Fig. 4.5.

The ToT rate varies slightly from station to station and is not as well balanced by the station calibration as the simple threshold triggers. It is much more pure than the other triggers, but still dominated by coincident hits of two muons in the sliding time window. It is typically generated in stations close to the shower core, where the dominant contribution to the signal comes from electromagnetic particles in the shower front, which are spread out in time.

#### 4.1. SURFACE DETECTOR

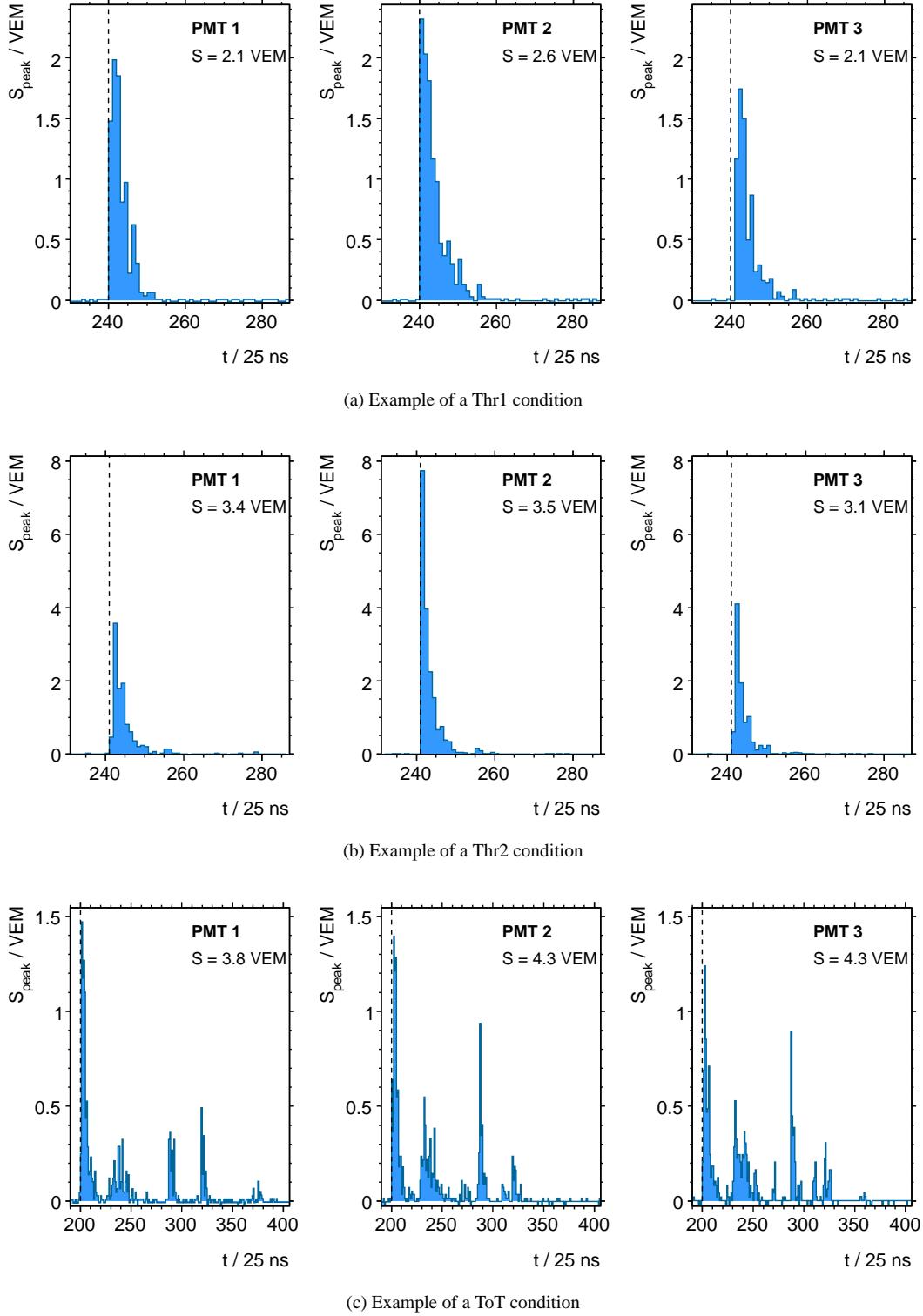


Figure 4.5: The graphs show examples of signal traces, which generated first or second level trigger events. The triggers are based on the calibrated pulse height  $S_{\text{peak}} \propto I$  in VEM of these traces. Also shown in each plot is the corresponding signal  $S$  based on the charge  $Q$  of the pulse. The dashed line marks the start time of the traces.

### Event level triggers

The CDAS monitors the stream of incoming T2 events with a software on a central computer. The T2 events are stored in a temporary buffer and sorted according to their trigger time. This buffer is then scanned with a sliding window of  $50 \mu\text{s}$  in search of a compact spatial configuration of the T2-emitting stations in the window.

The trigger algorithms applied in this search use the concept of *crowns* to measure the distance between two stations. Each station in a regular triangular grid is surrounded by six closest neighbors, which form the first crown. The second crown is the ring of the second-closest neighbors and so one. Fig. 4.6 visualises the concept, the three valid trigger conditions are listed below.

- **3ToT.** This trigger condition asks for a compact configuration of three ToT triggers. It requires an arbitrary central station to have one partner in the first crown and a second partner in the first or the second crown. Each station participates in this trigger pattern about three times a day. About 90 % of the trigger events are generated by real air showers, most of them fall in the zenith range  $0^\circ < \theta < 60^\circ$ .
- **4C1.** This trigger condition asks for a compact configuration of four T2 triggers, regardless of the type. It requires an arbitrary central station to have at least two partners within the first two crowns and another one within the first four crowns. Each station participates in this trigger pattern about two times a day. About 10 % of the trigger events are generated by real air showers, most of them fall in the zenith range  $60^\circ < \theta < 90^\circ$ .
- **BIG.** This trigger condition is automatically fulfilled, if 30 T2 stations or more are found in a single time window of  $50 \mu\text{s}$ . This limits the computing time of the other patterns and is also a catch-all condition for anything highly unusual. BIG triggers are extremely rare and occur only a few times per year. All event candidates triggered with this condition so far were either created by lightning strikes in the array or by very inclined conventional air showers with energies close to  $10^{20} \text{ eV}$ . The rejection of lightning events is discussed in Chapter 6.

If any of these conditions is fulfilled, the central computer emits a T3. Another way to get a T3 is upon external request by the fluorescence detector. Fig. 4.6 shows an example configuration for the 3ToT and the 4C1 trigger.

Up to January 2007, the entire surface detector array was read out upon a T3 event. The SD array was still growing at that time up to a point, where a read-out of the whole array produced an unacceptably high load on the radio communication system. The CDAS was therefore changed and since then reads out only the first six crowns around each T2 [69].

Stations with a T1 are only considered as candidates, if their trigger time is not further apart than  $(6 + 5n) \mu\text{s}$  from that of the nearest T2-station, with  $n$  being the crown number. Otherwise, such stations are regarded as having no signal.

Upon a read-out request, every candidate station sends its station ID, signal traces, trigger information, position, calibration histograms, and an error code concerning the communication to the CDAS. The latter is used as part of the monitoring of the communication systems.

The task of the CDAS ends after the event is written into a data storage system and the third level trigger is therefore the last real trigger of the SD. The event candidates are further processed offline and have to pass two more selection levels. Because of a similarity to a trigger level concept, these levels are often referred to as the “Physics trigger” T4 and the “Quality trigger” T5. They are not actual triggers, since they operate on already recorded data. The T4 rejects signals generated by background muons. The T5 rejects events close to the border of the array, which are likely to have a large systematic reconstruction uncertainty. These selections are discussed in Chapter 6.

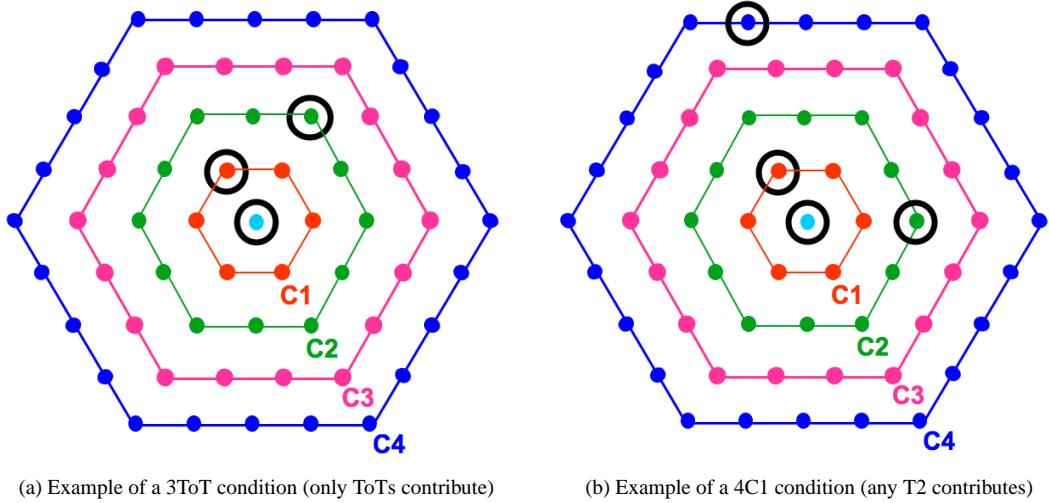


Figure 4.6: The plots give a schematical example for the two main third level trigger conditions discussed in the text. The SD array is shown from the top. The central station is surrounded by crowns, C1 to C4, of neighboring stations. Encircled points represent stations with a second level trigger.

## 4.2 Fluorescence detector

The fluorescence detector (FD) of the southern observatory is documented in ref. [70]. It uses the same detection method as the successful Fly's Eye/HiRes experiment [8].

The atmosphere and the fluorescence detector together form a calorimeter. The calorimeter is sufficiently large to absorb cosmic rays even at ultra-high energies: a shower has to overcome at least 8 hadronic interaction lengths and 20 radiation lengths until it reaches the ground level. Charged particles in an air shower excite nitrogen molecules along their path, which partly de-excite by isotropically emitting fluorescence light. The number of emitted photons is proportional to the energy loss of the charged particle in the atmosphere. The conversion factor is called *fluorescence yield*.

The wavelength-dependent fluorescence yield can be measured in the laboratory. Due to quenching effects, the yield depends on the temperature, pressure, and humidity of the air [71, 72]. Since these and other important atmospheric parameters are constantly changing over time, the atmosphere has to be monitored carefully over the site. Under standard conditions ( $20^{\circ}\text{C}$ , pressure of 1013 hPa), about 5 photons per MeV are emitted between 300 nm and 400 nm in dry air [73].

With the fluorescence yield it is possible to reconstruct the total energy loss of the electromagnetic shower component in the air by collecting the emitted fluorescence light along the shower path. The electromagnetic component carries about 90 % of the total energy of the cosmic ray, as discussed in Chapter 3, and therefore allows an almost calorimetric energy measurement.

The fluorescence light is detected with 24 Schmidt telescopes, which are housed in four separate FD buildings – Los Leones, Los Morados, Loma Amarilla, and Coihueco – located at the border of the SD array. Each telescope has a field of view of  $30^{\circ} \times 28.1^{\circ}$ . Each FD building covers an azimuth angle of  $180^{\circ}$ .

The telescope design is shown in Fig. 4.7. The optical system consists of a diaphragm with an aperture of  $3.8\text{ m}^2$  and a spherical mirror, which has an area of  $12\text{ m}^2$  and a radius of curvature of 3.4 m. To improve the signal to noise ratio a UV transmitting filter (Schott MUG-6) is integrated into the aperture. A ring of corrector lenses behind the filter reduces aberrations in the optical system and increases the collection area by almost a factor of two. The optical system keeps the angular size of the image of a point source on the camera smaller than  $0.5^{\circ}$ .

The camera is installed in the focal plane of the mirror and consists of 440 hexagonal PMTs (Photonis XP3062) arranged in a matrix of 22 rows and 20 columns. Each PMT forms a pixel of the camera and has

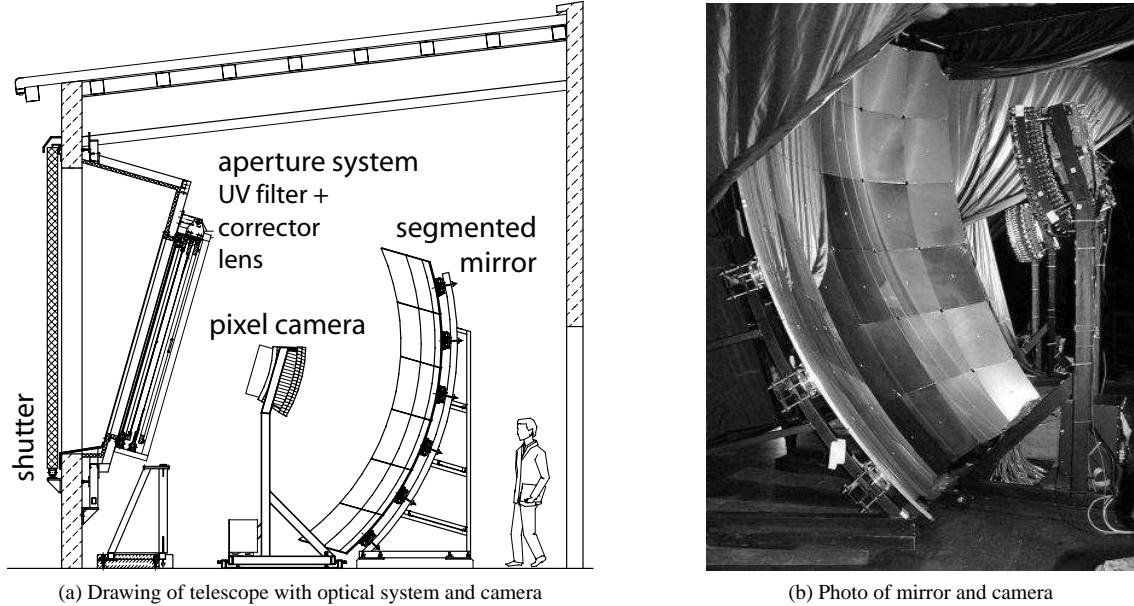


Figure 4.7: The drawing on the left shows the optical components and the camera of a telescope of the fluorescence detector of the Pierre Auger Observatory. The shutter on the outside provides protection from bad weather and dust and from light of the sun and the moon. The photo on the right shows the camera and the mirror, the camera of another telescope is visible in the background. The photo also shows the curtains in the front and the back of the telescope, which were lifted for the photo, but normally hang down to catch stray light.

a field of view of  $1.5^\circ$ .

The light collection efficiency over the surface of a PMT decreases towards its border and small gaps between adjacent PMTs are unavoidable due to the mechanical structure of the camera. These effects would decrease the collection efficiency over the entire camera to 70 %, but are avoided by placing small triangular mirror segments around each PMT. These segments, called Mercedes stars, increase the collection efficiency to 94 %.

The filtered and amplified pixel signal is digitized every 100ns with a 10MHz Analog Digital Converter (ADC) with a dynamic range of 12 bits. Each PMT actually has to cover a range of 3 to  $10^5$  photoelectrons which is equivalent to 15 bits. A compression scheme is used to cover this range with only 12 bits. The scheme exploits, that only one pixels of the camera needs the full dynamic range at a given point in time, while the other pixels only need a fraction of their 12 bit range. Each pixel has its own circular buffer which stores the last 100  $\mu$ s of the signal trace.

In order to be cost-efficient, not every PMT has its own high voltage unit. Instead, PMTs with similar gain characteristics are grouped during the installation of the camera and are connected to one of several high voltage channels per camera. The high voltages for the PMTs in the same channel are fine-tuned with programmable potentiometers. The gain is balanced over the entire camera to a level of 1 %.

Each FD building uses a commercial GPS receiver (Motorola OnCore UT) to synchronise itself to a common reference time. The same type of receiver is used in each SD station.

### 4.2.1 Signal calibration

The fluorescence detector reconstructs the total number of photons generated by an air shower from a detected fraction of the isotropically emitted light in a certain distance. In order to derive the number of photons at the aperture of the telescope from the integrated pixel signals, the telescopes need to be calibrated. Similarly, the absorption and scattering properties of the atmosphere have to be measured to

derive the total number of photons emitted by the shower from the sampled number of photons at the aperture.

The signal calibration of the telescopes is done in two steps. An absolute calibration is done three to four times per year, which relates the integrated pixel signals to a defined number of photons at the aperture. A relative calibration is done every night of operation before and after the data taking. The relative calibration allows to keep track of changes between two absolute calibrations.

### Absolute calibration

The absolute calibration is a full end-to-end calibration of each telescope. It accounts for the optical system, the camera, and the electronic readout. In order to perform the calibration, a diffuse light source with a diameter of 2.5 m is mounted at the aperture of a telescope. The light source resembles a drum. It is constructed to have an almost uniformly illuminated front, which covers the aperture of the telescope completely.

The drum is illuminated by a calibrated UV LED diode and provides a known and uniform photon flux at each PMT of the camera. The absolute calibration has an uncertainty of 6 %. The relative wavelength dependency of the FD signal response is also measured with the drum, but using a Xenon flasher with a filter wheel to select individual frequency bands.

### Relative calibration

The relative calibration is performed with three diffusive light sources, mounted at different positions of the optical system. They allow to monitor the short and long term changes in different groups of detector components. This is done by illuminating the camera with these light sources and recording the measured signals.

The first light source is mounted at the mirror and illuminates the camera directly. It is used to track relative changes in the calibration of the PMTs and the camera electronics.

The second light source is mounted at the center of the camera itself. Its light is reflected at the mirror and falls back into the camera. The second light source is used to track relative changes in the camera and the mirror reflectivity, for example due to the accumulation of dust on the mirror.

The third light source is mounted in front of the aperture of the telescope and illuminates the aperture. Its light passes the whole optical system. This light source is used to monitor relative changes in the aperture, the mirror reflectivity, and the camera.

The measurements of all three light sources can be disentangled offline and then also allow to keep track of changes in individual detector components.

### Atmospheric monitoring

The Pierre Auger Observatory uses an extensive monitoring program to keep track of atmospheric properties. A summary of these activities can be found in ref. [74, 75]. The main task of the atmospheric monitoring is to derive all parameters for a proper calculation of the fluorescence yield along the shower track and the Mie scattering of UV light on aerosols in the atmosphere.

The vertical profiles of the density, temperature, and relative humidity of the atmosphere at the site were measured with a program of weather balloon launches, which rise up to an altitude of 25 km above sea level [41]. Several of launches were conducted every month over the course of a year, to measure both the daily and seasonal variations of the atmospheric conditions.

The measured profiles are used to calculate the fluorescence yield and the atmospheric slant depth of the atmosphere above the observatory. The atmospheric density profile varies by about 5 % around an average value, with a corresponding variation in the atmospheric slant depth profile of about  $7 \text{ g cm}^{-2}$ . Variations of the density profile have a potential small impact on the energy estimator of the surface detector, which has to be derived from air shower and detector simulations. Variations in the slant depth profile affect the reconstruction of the depth  $X_{\max}$  of the electromagnetic shower maximum.

The aerosol content of the air is measured with elastic backscatter LIDAR (Light Detection And Ranging) units [76]. A LIDAR unit consists of a steerable UV laser and an optical system with a PMT to

detect the backscattered light. Each FD building is equipped with one unit. The time resolved signal of backscattered light after a UV shot allows to reconstruct the aerosol extinction coefficient at any given spot along the path of the laser beam. The aerosol extinction coefficient is used to correct for the attenuation of UV light on its way from the shower axis to the telescope due to Mie scattering.

The LIDAR units also detect clouds in the field of view of the telescopes. Clouds can distort the measured light curve obtained from a shower, either by adding reflected light to the curve or by absorption of light, depending on whether the cloud was behind or in front of the shower. In addition to the LIDAR units, there are infrared cameras installed at each FD building, which monitor the cloud coverage of the sky.

A versatile instrument is the Central Laser Facility (CLF) in the approximate center of the SD array [77]. The solar powered facility houses a steerable 355 nm UV laser, which fires laser shots of up to 7 mJ into the air. The laser beam can take almost any inclination from vertical shots into the air down to almost horizontal shots. The scattered light from the beam is detected with the FD telescopes. At the maximum power level, the amount of scattered light is roughly equivalent to the fluorescence light emitted by a  $10^{20}$  eV air shower.

The CLF provides a test beam for the fluorescence detectors, which is used to analyse the energy and angular resolution of the telescopes, the aerosol extinction coefficient of the atmosphere. It can also be used to derive the angular and energy resolution of the FD telescopes. The CLF is also used to measure an absolute timing offset between the SD and FD by feeding a small fraction of the laser light into a specially prepared nearby SD station, which triggers the station. A second facility, called XLF, is currently being built.

The measurements of the aerosol extinction coefficient by LIDARs and the CLF are extended by Horizontal Attenuation Monitors (HAMs), Aerosol Phase Function Monitors (APFs), and Fotometric Robotic Atmospheric Monitor (FRAM) on the site. The HAMs measure the wavelength dependency of the aerosol extinction coefficient. The APFs measure the aerosol differential scattering cross-section [78]. The FRAM measures the wavelength dependent integrated aerosol extinction coefficient through the atmosphere by tracking several bright stars [79]. The last method is non-invasive in contrast to the others and used to cross-check the other instruments.

Finally, every FD building and the CLF are equipped with standard sensors for barometric pressure, air temperature, relative humidity, wind speed, and wind direction.

#### 4.2.2 Trigger system and data acquisition

The trigger system of the fluorescence detector has three levels. The first level trigger operates on individual pixels and keeps the overall noise rate constant for the higher levels. The second and third level triggers operate on individual telescopes and select tracks in the camera with a certain spatial and time ordering, which are the signature of real air showers.

- (1) **Pixel trigger.** The first level trigger is implemented in the logic of a Field-Programmable Gate Array (FPGAs). The signal trace from each pixel is integrated over a sliding window of the last 16 bins ( $= 1.6 \mu\text{s}$ ). The integral is compared to a threshold value to form the trigger. The threshold value is not fixed, but continuously adjusted to maintain a trigger rate of 100 Hz for each pixel.
- (2) **Spatial structure.** The second level trigger is also implemented in the logic of a FPGA. The trigger searches for a combination of five adjacent pixel triggers, which form a rough line in the camera. One out of the five pixels may be silent, which makes the trigger more tolerant to pixel failures and weak showers. The trigger produces a rate of about 0.1 – 100 Hz per telescope.
- (3) **Time structure.** The third level trigger is implemented in software and rejects lightning events, muon impacts in the camera, and plain random pixel triggers. The trigger produces a rate of about 0.01 Hz per telescope.
  - (a) **Lightning rejection.** Lightning can cause hundreds of pixels to trigger in bursts of 100 Hz. It has to be rejected fast so that it does not congest the signal buffers and causes dead time. Cuts on the time development of the pixel trigger multiplicity in steps of 100 ns and the total number of triggered pixel reject 99 % of the lightning events while removing less than 1 % of real showers [80].

- (b) **Muon and random trigger rejection.** Random triggers and muon impacts are rejected by looking at the time structure of the peaks in the ADC traces of individual pixels. The peak time search is used for this part of the trigger is comparably slow, but manageable after the basic lightning rejection.

The local software in each FD building combines coincident third level triggers and starts the readout of the participating telescopes. The collected event data is send to the CDAS, together with a request to trigger the SD array readout. In order to set a proper time window for the SD trigger, a rudimentary online reconstruction of the shower development is performed.

The FD reconstruction algorithm uses the measured arrival time in at least one SD station to improve the accuracy of the reconstructed shower geometry, as will be discussed in Chapter 6. If the cosmic ray energy is around or below  $10^{18}$  eV, the shower may only trigger one or two stations, which is not enough for an independent SD reconstruction, but enough to improve the FD reconstruction.

The data acquisition of the fluorescence detector is remotely operated by human shifters from the central campus of the observatory in Malargüe, who open and close telescopes according to the environmental conditions and monitor the data acquisition. In addition, automatic safety protocols are implemented in the local slow control PC of each FD building, which performs orderly shutdowns in case of power failures, communication problems, or bad environmental conditions like rain or heavy storms.

Data is collected, if the moon fraction is smaller than 60 %, which is the case in about 16 nights each month. The data aquisition starts after and ends before the phase of astronomical twilight, yielding an average time of 10 hours of operation. Individual telescopes remain closed, if they would get direct moon light or if they get too much reflected moon light from clouds.

On average, the FD reaches at a duty cycle of 13 %. Since the thresholds of the pixel triggers depend on the amount of background light, its sensitivity is a function of time and reaches a maximum during the darkest nights of each monthly data taking period.

### 4.3 Extensions and new developments

The southern Pierre Auger Observatory works efficiently in the cosmic ray energy range of  $10^{18}$  eV to  $10^{20}$  eV. Two extensions are currently being built to increase this range down to about  $10^{17}$  eV. Air showers develop higher in the atmosphere and have a smaller lateral extension as their energy decreases. In order to detect these showers efficiently, fluorescence detector telescopes are needed, which look up higher into the sky, and a surface detector array with a smaller grid size.

The High Elevation Auger Telescope (HEAT) extension consists of three new telescopes, which are located close to the FD building Coihueco [81]. They are almost identical to the standard telescopes, but cover the required higher elevation angle from  $30^\circ$  to  $60^\circ$ .

The Auger Muon-detectors and Infill for the Ground Array (AMIGA) extension fills up the standard SD array near the Coihueco FD building with additional stations to make a more dense grid [82]. The extension is partly visible already in the top left corner of Fig. 4.1a). Through the infill, the grid size is reduced to 750 m in a fraction of the SD array and partly even to 433 m.

In addition, scintillation detectors are buried underground next to some of the SD stations in the denser grid. The scintillation detectors are a new development for the Pierre Auger Observatory. They are shielded from electromagnetic particles by the earth and therefore only detect the muon component of the shower. Used in coincidence with the standard SD stations the buried detectors will help to separate the electromagnetic and the muonic component of the lateral shower profile.

The area covered by the low-energy extensions HEAT and AMIGA is an ideal testbed for new air shower detectors. Air showers generate radio waves, which are detectable in the frequency range of 1 MHz to 100 MHz with suitable antennas [83]. A prototype array of about 100 autonomous radio detectors is planned, which will be embedded in the AMIGA array, and record air shower data alongside the established detectors.

# Chapter 5

## Air shower and event simulation

This chapter deals with air shower simulations and the insights that are gained from the simulations in the context of this study. A large library of very inclined air showers and corresponding surface detector events was generated as a part of this work, which is described here.

The library is used to model the muon component of very inclined air showers (see also ref. [84]) and the signal response of the surface detector to such air showers in the second part of this chapter. The derived models are used in the reconstruction of surface detector events measured with the Pierre Auger Observatory in Chapter 6.

### 5.1 Simulation of air showers

It was discussed in Chapter 3, that an extensive air shower initiated by a cosmic ray nucleon forms a hadronic cascade, which feeds an electromagnetic component along the way. The hadronic cascade mostly consists of unstable mesons, which eventually decay into a penetrating muon component and into neutrinos.

Ground based experiments observe the cosmic ray only indirectly through its air shower. Therefore, they usually rely strongly on quantitative predictions of air shower observables as a function of the properties of the primary cosmic ray.

To predict observables, two main approaches were developed in the past, and one unified approach is currently emerging.

- **Cascade equations.** The development of an air shower can be treated with a coupled set of cascade equations [85, 86]. These are coupled differential equations that describe the change in the flux of a particle species analytically with loss and gain terms (see *e.g.* [21]). The continuous and average treatment of the particle flux is a good approximation after a few interaction lengths, because of the huge number of secondary particles. The treatment of shower-to-shower fluctuations can be included into the approach [87].

The equations are solved numerically and the necessary computing time is very moderate. The solutions are 1-dimensional profiles of the number density of various particle species in the shower as a function of the slant depth  $X$ , which allow valuable predictions for fluorescence telescope measurements. Not yet achieved is a full 3-dimensional calculation of the shower, so that the cascade equation approach is currently unable to predict observables for surface detector arrays.

The program CONEX [87] implements the pure cascade equation method and has found several applications in the Pierre Auger Collaboration.

- **Monte-Carlo simulation.** Another way to calculate an air shower is the full Monte-Carlo simulation, which explicitly follows every single particle and treat all relevant microscopic processes according to their cross-sections and probabilities. It is currently the only way to predict the lateral shower profile, which is the essential information for surface detector arrays.

A full Monte-Carlo simulation of an air shower with more than  $10^{11}$  particles above  $10^{18}$  eV needs huge amounts of computing resources. Therefore, a statistical sampling algorithm is applied to cut down the complexity to an acceptable level.

Two full Monte-Carlo simulation programs of widespread use in the Pierre Auger Collaboration are AIRES [88] and CORSIKA [49, 89].

- **Hybrid simulation.** A comparably new approach tries to combine the calculation speed of the cascade equation method with the full Monte-Carlo treatment of the lateral shower profile. This is achieved by calculating the first and last steps of the shower with the full Monte-Carlo simulation, and the intermediate steps with cascade equations.

The first program which implemented this idea was SENECA [90], which is also used in the Pierre Auger Collaboration. Recently, there is an effort to merge the CONEX and CORSIKA programs [91], which will enable this feature in CORSIKA in the future.

A discussion of the air shower simulation programs and hadronic interaction models can only be a snapshot of the current situation, as there is vivid motion in the field [92, 93]. A summary of past and recent programs that use the Monte-Carlo simulation approach is given in ref. [94].

The simulation programs typically use external well-established codes to implement hadronic and electromagnetic interactions. The electromagnetic interactions are well determined at all energies by the perturbative theory of quantum electrodynamics (QED). The EGS4 code [95], originally developed for collider experiments, implements most of the necessary electromagnetic interactions in air showers and is utilized in all modern air shower simulation programs. It is usually extended by some processes, that specially appear in the context of air shower simulations at ultra-high energies, like the LPM-effect [89, 96, 97], direct muon pair production in the electric field of a nucleus, and electromagnetic interactions of photons, electrons, and muons with nuclei.

The treatment of hadronic interactions is not as straight-forward, because most interactions in air showers happen at low momentum transfer, where the strong coupling constant  $\alpha_s$  is too large for the standard perturbative approach to quantum chromodynamics (QCD). Phenomenological models and effective theories are used instead to calculate the interaction cross-sections.

The treatment of hadronic interactions is generally split into a high energy and a low energy part, with a threshold between the two at about 80 GeV. The low energy part is usually treated by a model, which is based directly on collider data. The high energy models are based on theoretically motivated extrapolations of such low energy data.

The low energy models have a considerable impact on the shape of the lateral profile of the particle density in an air shower, but almost no impact on the longitudinal profile [98]. For the high energy interaction model, it is the other way round [93]. This observation can be understood in context of the analytical discussion of air showers in Chapter 3. It was shown there, that only low energy muons are able to diverge far away from the shower axis, which in turn are generated by pions of low energies in the last steps of the hadronic cascade. In contrast, the interactions relevant for development of the longitudinal profile occur at high energies.

An overview of contemporary hadronic interaction models is given in the following. Another important feature of modern air shower simulations is the statistical weight-sampling. The sampling may introduce biases to air shower observables and its impact therefore has to be analyzed. It discussed at the end of this section.

### 5.1.1 Hadronic interaction models at low energies

The most commonly used hadronic interaction models at low energies are GHEISHA [99], URQMD [100, 101], and FLUKA [102, 103].

FLUKA combines an array of models for hadronic interactions for different energy ranges in itself. The proclaimed aim of FLUKA is to rely on parametrized collider data as little as possible. Instead, the authors use QCD motivated models wherever possible. The code and the theoretical framework are continuously improved.

URQMD is based on a QCD motivated approach called Quantum Molecule Dynamics. It allows to compute general hadronic interactions, but the approach specially aims at collisions of two heavy ions. This specialization may be disadvantageous in the treatment of the low energy hadronic cascade in air showers, where most interactions are pion+nucleus and nucleon+nucleus.

GHEISHA is a well tested and mature model in the context of collider experiments. It was the standard for almost two decades, but is no longer actively maintained at present. It uses phenomenological fits to tabulated experimental data in many cases.

Partial comparisons of predictions from these and other models with data can be found in ref. [98, 104]. Data from fixed target experiments cover only a small fraction of the total interaction phase space of interest in air showers. New experiments are desirable to extend this data. The analysis in ref. [104] remains inconclusive, while ref. [98] favors FLUKA and URQMD over GHEISHA.

### 5.1.2 Hadronic interaction models at high energies

Many high energy hadronic interaction models are on the market. They extrapolate hadronic interactions over several orders of magnitude beyond the data of current colliders and their predictions for cross-sections at ultra-high energies vary significantly. Instead of a listing of many similar models, a general overview of the theoretical approaches shall be given, based in parts on ref. [21, 22, 94].

Contemporary models usually have three theoretical components [105]: (a) a component to calculate the “soft” part of the interaction of individual hadrons where the exchanged transverse momentum  $p_T$  is low, (b) a component to calculate the “hard” part of said interaction where  $p_T$  is high, and (c) a component to calculate hadron+nucleus and nucleus+nucleus cross-sections from the basic hadron+hadron cross-sections.

The idea for the splitting of the soft and hard regime is obtained from the measured total inelastic  $pp(\bar{p})$  cross-section. It features some resonances up to laboratory energies of about 10 GeV, then a nearly flat plateau over nearly two orders of magnitude in the energy, and finally a slow increase [23]. A similar plateau is observed in the  $Kp$  and  $\pi p$  cross-sections.

This gave rise to the *minijet* model [106, 107] for interactions above laboratory energies of about 10 GeV, which splits the inelastic cross-section  $\sigma_{\text{tot}}^{hh}$  of two hadrons into a “soft” and “hard” part

$$\sigma_{\text{tot}}^{hh}(\sqrt{s}) \approx \sigma_{\text{soft}}^{hh}(\sqrt{s}) + \sigma_{\text{hard}}^{hh}(p_T^{\min}, \sqrt{s}), \quad (5.1.1)$$

where  $\sqrt{s}$  is the center-of-mass energy, and  $p_T^{\min}$  is a lower threshold of transverse momentum exchange.

The soft part  $\sigma_{\text{soft}}^{hh}$  is thought to be dominated by complex peripheral QCD interactions between the hadrons, in which the constituent quarks quarks participate collectively. It is assumed to vary slowly with the center-of-mass energy  $\sqrt{s}$ .

The hard part  $\sigma_{\text{hard}}^{hh}$  then is responsible for the bulk increase of the total cross-section with  $\sqrt{s}$ , and caused by interactions of individual partons, see Fig. 5.1b). These multiple interactions give rise to small hadronic jets. The hard part of the total cross-section can be calculated in perturbative QCD with the help of the QCD factorization theorems [108] and experimentally measured parton structure and fragmentation functions.

Although the first ansatz of the minijet model was Eq. (5.1.1), this picture is actually too simple and has since evolved. The basic separation remains a guiding idea. The SIBYLL model [105, 109] is a pure minijet model, which treats  $\sigma_{\text{soft}}^{hh}$  as a constant.

Most contemporary hadronic interaction models are based on the Gribov-Regge theory [110–114], for example the QGSJet model [115–120]. Gribov-Regge theory is an effective field theory that works only in the limit of low momentum transfer and multiple-scattering of partons. The hadronic interactions are described by the exchange of hypothetical particles, called pomerons, see Fig. 5.1a). A pomeron can be seen as a QCD cascade of quarks and gluons built between the interacting partons. It is charge and color neutral.

There are soft and hard pomerons, which correspond to the two terms in Eq. (5.1.1). In Gribov-Regge models, also  $\sigma_{\text{soft}}$  varies with the center-of-mass energy  $\sqrt{s}$ . The model QGSJet became a common choice in the cosmic ray community because it showed good agreement with cosmic ray data around  $10^{15}$  eV, for example in a comparative analysis of data from the KASCADE air shower experiment with several hadronic interaction models [121, 122].

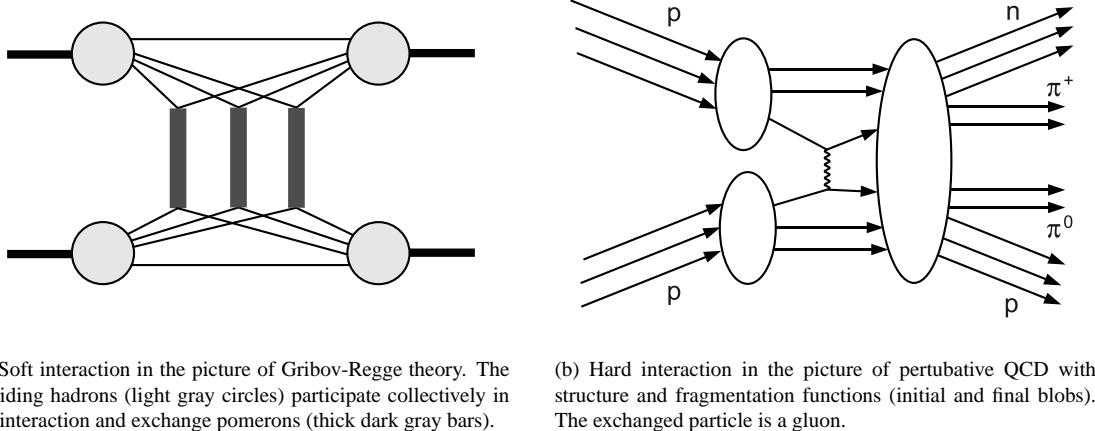


Figure 5.1: Two diagrams show one out of many configurations, that contribute to a “soft” and “hard” cross-section of two colliding protons.

The hadronic interaction models predict not only the total inelastic cross-sections, but also the differential cross-sections. In case of the hard scatterings, perturbative QCD together with the factorization theorem yields the differential cross-sections. Soft interactions need a phenomenological treatment, based on a microscopic Monte-Carlo simulation of the interaction.

A variant of the dual parton model (DPM) of QCD string production and fragmentation [123, 124] is usually used. It is based on the idea, that strings of “color field lines” should span between interacting partons. The string picture is an analogy to classical fields in electromagnetism, with the difference that the field lines of the strong force attract each other, because the force mediating gluon carries color charges itself. A string gains potential energy as the interacting partons move apart, and then fragments into color neutral objects.

The final component of an interaction model is a theory to calculate hadron+nucleus and nucleus+nucleus interactions. The Gribov-Regge theory can be extended to such cases and the interactions are treated in a similar way as the interaction of individual hadrons. The minijet model SIBYLL on the other hand is built on the multiple scattering theory of Glauber [125] and an advanced superposition approach [48], which approximates a nucleus as a superposition of free nucleons.

QGSJet has been succeeded by QGSJet-II [126–128], which includes interference terms of soft and semi-hard pomeron amplitudes. QGSJet-II started to combine the soft and hard treatment into a common theoretical framework, which further improves the agreement of this model with collider data.

The latest generation of models are based on the so called Parton-based Gribov-Regge theory [129], which removes the separation between soft and hard interactions and treats them consistently. In particular, energy conservation is strictly assured for the first time in the particle production.

The EPOS model [130, 131] is a representative of the latest model generation. The special feature of EPOS is its enhanced cross-section for baryon production compared to other models, which the authors derived from comparisons of EPOS predictions with data of heavy ion collisions. As a consequence, EPOS currently generates more muons in air showers than any other model.

The author hopes to point out with this overview, that there is still a lot of theoretical uncertainty in the description of hadronic interactions at ultra-high energies and that even now the field is rapidly progressing. In general, the models co-evolve with the available data. Several experiments at the LHC at CERN are dedicated to heavy ion collisions and sampling of the forward direction, which will greatly help the model builders in the future.

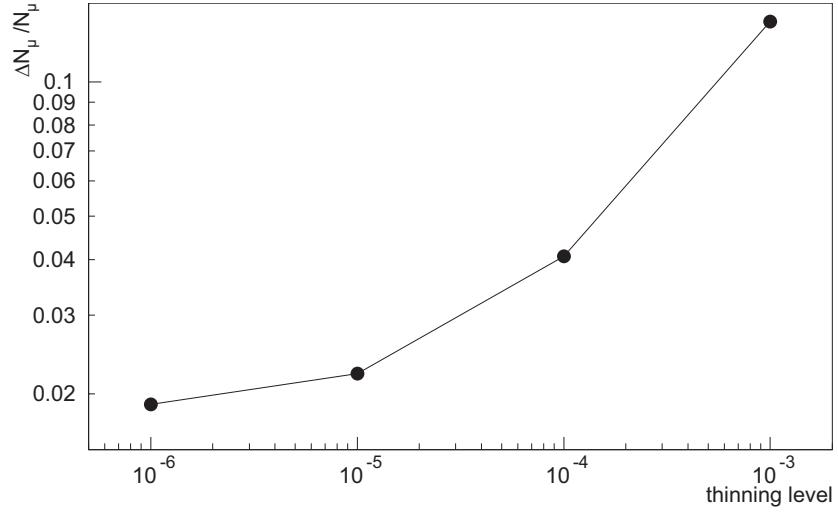


Figure 5.2: The graph shows the relative artificial fluctuation of the total muon number  $N_\mu$  in CORSIKA simulations as a function of the thinning level  $\epsilon$  (adapted from ref. [89]).

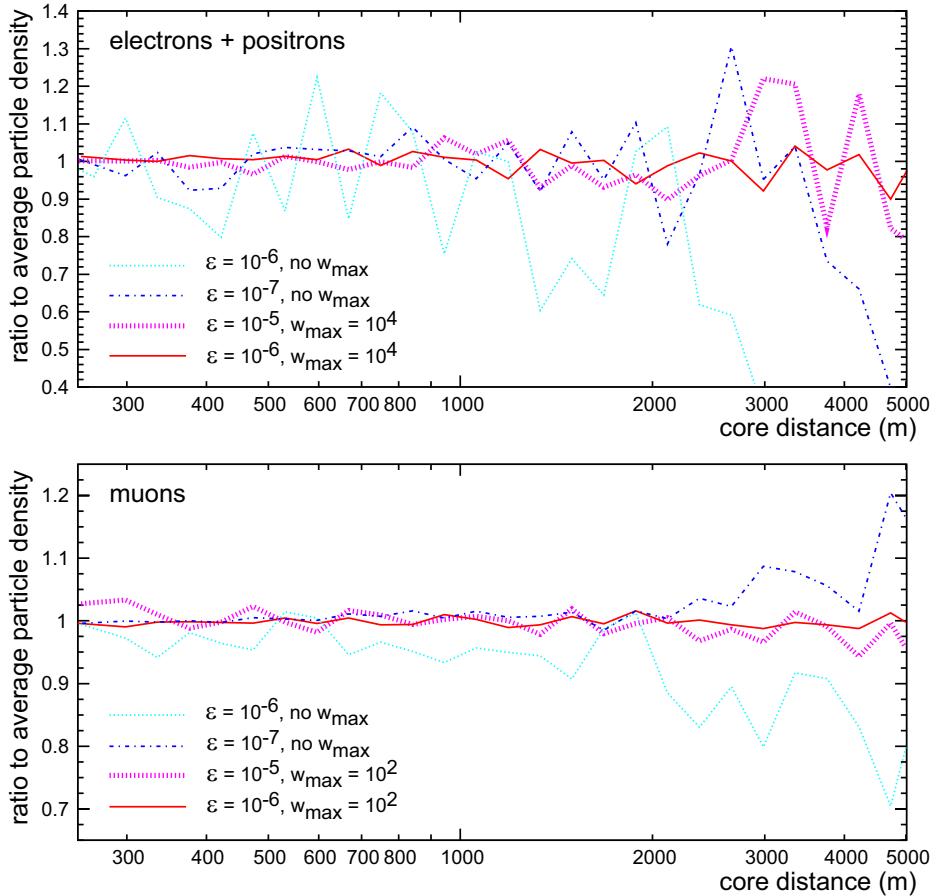


Figure 5.3: The graphs compare biases and fluctuations in the lateral density profile of proton showers simulated with the thinning levels  $\epsilon$  and weight limits  $w_{\text{max}}$ . The simulations are done with the air shower program CORSIKA at an energy  $E = 10^{19}$  eV and a zenith angle  $\theta = 0^\circ$  (adapted from ref. [132]).

### 5.1.3 Thinning

In principle, the full Monte-Carlo approach requires to follow every single particle in the air shower explicitly. This approach is not feasible even with modern computing resources at energies above  $10^{18}$  eV, where the shower consists of more than  $10^{10}$  particles.

It was demonstrated recently in the Pierre Auger Collaboration, that the full simulation of a single air shower at  $10^{19}$  eV [133] takes about 1.5 years on an AMD Opteron™ CPU. The required storage space is enormous, too: a few 100 GB for a single shower. On the other hand, data analyses which rely on knowledge of the distribution of an observable need at least 10 times more simulated than real events, and thus, huge libraries of simulated air showers are desired.

To reduce the computational complexity to a reasonable level, a weight-sampling of the shower is introduced, which is called *thinning*. The thinning discards secondary particles of low energy during the shower development, which have a negligible impact on the global development of the shower in the best case [134, 135]. This reduces the amount of particles to explicitly follow by orders of magnitude.

The actual implementation of a thinning algorithm has to handle some special cases, see *e.g.* [89], but the basic idea is to follow all particles down to a certain energy fraction  $\epsilon$  of the cosmic ray and then start to follow only one secondary particle  $i$  from each multiple particle production process at random, with a probability  $p_i$  equal to the fraction of the total energy carried by the  $i$ -th particle after the process:

$$p_i = \frac{E_i}{\sum_{i=0}^n E_i}. \quad (5.1.2)$$

The surviving particle gets a weight  $w_i = 1/p_i$  to conserve the energy in the shower. The survivor may interact again, but only weight is accumulated in the subsequent shower development. The number of actually followed particles remains constant. The energy fraction  $\epsilon$  is also called the *thinning level* of the shower. The thinning is turned off for particles, which have accumulated a preset maximum weight  $w_{\max}$  [136]. All secondaries particles are followed again in such a case.

The thinning algorithm solves the computational and storage issues, but also introduces artificial fluctuations and biases in shower observables. These need to be kept at a minimum.

Firstly, the thinning may only start after the global shower-to-shower fluctuations are established. As discussed in Chapter 3, the first interactions are the most important for the observed global shower-to-shower fluctuations. Any thinning at this stage would increase the natural shower-to-shower fluctuations, which has to be avoided. The thinning level  $\epsilon$  therefore needs to be small.

Fig. 5.2 shows the effect of the thinning level  $\epsilon$  on the global fluctuation of the total number of muons  $N_\mu$  in the shower. For  $\epsilon = 10^{-6}$ , the effect is negligible. With the simplified shower model from Chapter 3 and an assumed average multiplicity of 100 in the first few interactions, it is possible to estimate that the thinning starts in the third to forth generation of secondary hadrons at this thinning level, which is on the safe side.

Secondly, the weight limit  $w_{\max}$  may not be too large. As discussed in Chapter 3, the lateral distance  $r_i$  of a particle from the shower axis is strongly correlated with the particle energy  $E_i$ . The thinning tends to remove low energy particles and therefore makes the lateral density profile artificially steep, as shown in Fig. 5.3. The weight limit  $w_{\max}$  helps to reduce this bias.

While an impact on global shower features can be mostly avoided by careful choices  $\epsilon$  and  $w_{\max}$ , additional the artificial fluctuations in local shower features cannot and have to be dealt with. The effect shall be illustrated for a simple counting observable: the number of particles  $N$  which fall into an area on the ground, which is small compared to the structure of the shower front.

If  $n$  is the number of weighted particles that arrive in this area, which carry individual weights  $w_i$ , then  $N$  is the sum of the weights

$$N = \sum_i^n w_i. \quad (5.1.3)$$

In the limit  $w_i \rightarrow 1$ , the fluctuations of  $N$  follow a Poisson distribution, so that  $\sigma[N] \approx \sqrt{N}$ .

In local areas of the shower front, the individual weights tend to be of similar size, so that it is possible to write approximately

$$N \approx \langle w \rangle n, \quad (5.1.4)$$

whereas  $\langle w \rangle \approx w_i$  is the average weight. In this case,  $n$  follows a Poisson distribution, so that the fluctuation of  $N$  can be calculated as

$$\sigma[N] \approx \langle w \rangle \sigma[n] = \langle w \rangle \sqrt{n} = \sqrt{\langle w \rangle} \sqrt{N}. \quad (5.1.5)$$

The example shows, that the thinning increases the natural Poisson fluctuations in observables like the lateral muon density profile  $n_\mu$  roughly by a factor equal to the square-root of the average local particle weight at a particular point on the ground.

## 5.2 Simulation of surface detector events

The full Monte-Carlo simulation of an air shower produces a set of weighted particles on the ground level. Each particles has a detailed space and time coordinates, a momentum and an energy. The particles can be analyzed directly to learn something about the global features of an air shower or they can be used to further simulate the detector response to such an air shower. Simulated events provide the connections between air shower observables and detector signals. They are also used to calculate the detection efficiency of the detector, which is often difficult to obtain experimentally.

The simulation of the surface detector (SD) of the Pierre Auger Observatory is done in the software framework `Offline` of the Pierre Auger Observatory [137]. The default is a full Monte-Carlo simulation of the SD, which shall be summarized here. The simulation only treats photons, electrons, and muons from the air shower. All heavier particles barely produce Cherenkov light in an SD station [56] and are neglected.

The site of the southern Pierre Auger Observatory is almost flat, as shown in Chapter 4, and thus the impact point of the simulated air shower may be placed at an arbitrary position in the SD array. Once the impact point is set, the simulation of surface detector events may be roughly divided into three steps.

- (1) **Shower un-thinning.** A set of unweighted particles is recovered from the weight-sampled shower around each station with a so called *un-thinning algorithm*. Some particles out of this set hit the station. They are picked at random with a probability given by the effective area of the station with respect to the shower front. Particles which hit the station are placed at random positions on its hull.
- (2) **Signal response of a SD station.** The particles are individually tracked through the hull and the whole detector in a full Monte-Carlo simulation of all particle processes based on GEANT4 [138]. The optical photons of the generated Cherenkov-light are individually tracked with a custom code [139], which treats all relevant attenuation and scattering processes. The optical photons are either absorbed or generate a photoelectron in one of the PMTs. The output of the station simulation is a discrete time sequence of photoelectron counts, which is converted into a standard VEM trace with noise in the subsequent simulation of the station electronics.
- (3) **Simulation of Online-triggers.** The calibrated signal trace is checked against the station level triggers T1 and T2, as discussed in Chapter 4. Finally, the T3 trigger condition is checked against the stations with a T2.

The complete event is written into the same data structures as real events, which makes them indistinguishable from real events in the point of view of subsequent analyses steps.

The simulation fully imitates the signal calibration of the SD stations a simulated flux of background muons, as described in Chapter 4. This procedure makes the simulated signal very robust, because small errors in the parameters of the tank simulation, like the PMT efficiencies or the liner reflectivity, are counterbalanced. The simulation of the signal response of a SD station to muons is compared with experimental data in Fig. 5.4. The agreement is very good.

### 5.2.1 Un-thinning

The other simulation steps are straight forward, but the un-thinning of the weight-sampled air shower needs some discussion. The full procedure is described ref. [140].

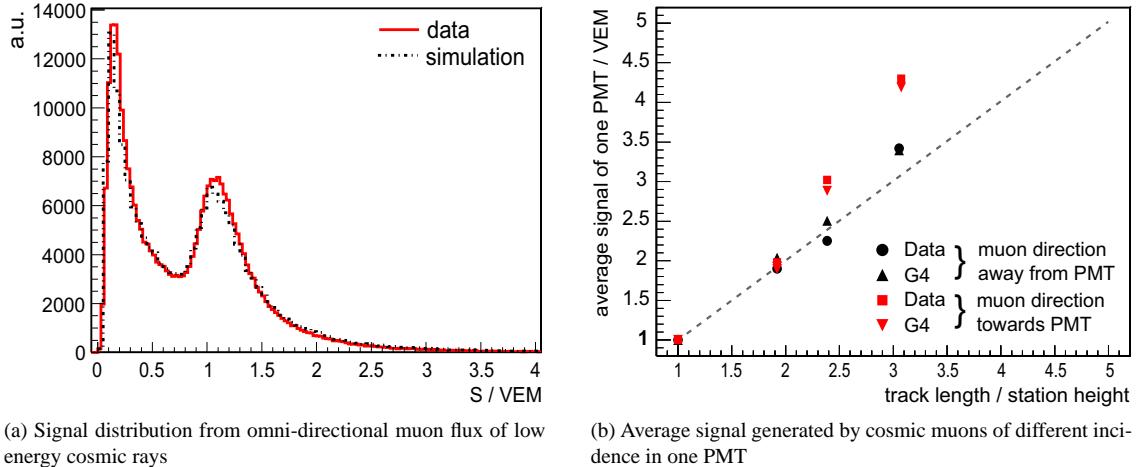


Figure 5.4: a) The signal distribution obtained from the flux of atmospheric muons is compared with a simulation of this flux and the signal response. The peak close to zero is generated by noise. The peak around 1 VEM is generated by vertically and centrally through-going muons. The small shift of this peak is well understood, see Chapter 4. b) Signals of selected muon tracks from the background flux are compared with the simulation. The tracks are selected in the experiment via pairs of scintillators, placed around the station and used as triggers for the desired geometrical configuration. Muons produce larger signals if they face a PMT, because a significant fraction of the Cherenkov light falls directly into the PMT (plots adapted from [67]).

The un-thinning is essential to get the correct time structure of the signal trace and the correct signal and time fluctuations. One way to see this is to regard the example of a weighted electron in the SD station, which shall represent 100 electrons. If such a weighted electron was directly processed, the signal in the SD station would be the random signal of a single electron, but scaled by a factor of 100. In reality, the 100 electrons would arrive over some time, making a wide signal, while the weighted electron would generate just a single peak. Furthermore, the start time and the signal size uncertainties would be too large by factor of 10.

The un-thinning method tries to avoid this effect by recovering a set of unweighted particles from the weight-sampled ones with a minimum bias. The method is based on the observation that the shower front is featureless and particles are uncorrelated over small scales in good approximation. This means that particles in a small area of the shower front may arbitrarily switch positions without significantly affecting the output of the event simulation, if the timing relative to the shower front is preserved.

Thus, it is possible to define a sampling area  $A_{\text{sample}}$  around each SD station, which is larger than the actual effective area of the station  $A_{\text{station}}$  by a factor  $\tilde{w}$

$$A_{\text{sample}} = \tilde{w} A_{\text{station}}. \quad (5.2.1)$$

Instead of the weighted particles which would normally fall directly into the effective area  $A_{\text{station}}$  of the station, now all weighted particles from the sampling area  $A_{\text{sample}}$  are regarded for a possible insertion into the station.

To conserve the energy, a particle with weight  $w_i$  in the sampling area has to be re-weighted with the factor  $1/\tilde{w}$ . The re-weighted particles can be regarded as average representatives of identical clones. The number  $k$  of such clones which need to be generated from a particle with weight  $w_i/\tilde{w}$  is given by the Poisson distribution

$$f(k) = \frac{\tilde{w}}{w_i} e^{-k\tilde{w}/w_i} \quad (5.2.2)$$

The unweighted clones are then placed randomly on the surface of the station. The arrival time of each clone is corrected after the spatial shift, so that the time delay  $\tau$  with respect to the shower front is conserved. Then, the time delay is artificially randomized to roughly imitate the original time structure of the

shower front. The randomized time delay  $\tau'$  is modeled with a log-normal distribution, which is a good approximation of the true arrival time distribution. The formula used in praxis is

$$\tau' = \tau \exp(G) \quad (5.2.3)$$

whereas  $G$  is random number from a normal distribution with mean 0 and variance 0.1. Ideally,  $A_{\text{sample}}$  is large enough, so that  $w_i/\tilde{w} \ll 1$  and the production of identical clones is very unlikely. The thinning effects are then essentially avoided.

On the other hand,  $A_{\text{sample}}$  still has to remain small compared to the large scale structure of the air shower. The default size of  $A_{\text{sample}}$  in the SD simulation of the Pierre Auger Observatory is defined relative to the distance  $r$  from the shower axis:

$$\Delta r = 0.1 r \quad \Delta\psi = 0.15 \quad (5.2.4)$$

whereas  $\Delta r$  and  $\psi$  are the radial and angular size of the sampling area  $A_{\text{sample}}$

$$A_{\text{sample}} = \Delta\psi ((r + \Delta r/2)^2 - (r - \Delta r/2)^2) \quad (5.2.5)$$

in the lateral coordinate system. The achievable weight-reduction factor  $\tilde{w}$  depends on the radial distance  $r$  to the shower axis and the zenith angle  $\theta$  of the shower. The magnitude of  $\tilde{w}$  at the typical distance  $r = 1000$  m is:

$$\tilde{w}(1000 \text{ m}) \approx \begin{cases} 3 \times 10^3, & \theta = 0^\circ \\ 6 \times 10^3, & \theta = 60^\circ \\ 7 \times 10^3, & \theta = 80^\circ. \end{cases} \quad (5.2.6)$$

The calculation of the effective area  $A_{\text{station}}$  of a SD station is carried out in Appendix A.3

### 5.3 Mass production of very inclined air showers

A library of 6480 simulated air showers was produced in the course of this work [141] and an equal number of SD events. The library contains cosmic rays with energies between  $10^{18}$  eV to  $10^{20}$  eV to cover the range of interest of the Pierre Auger Observatory. The zenith angle range  $60^\circ < \theta < 88^\circ$  includes only very inclined air showers. Many simulation options are chosen as such, that the library may be used as an extension of the existing production LD1\_GPF00q2fl [142] in the zenith angle range  $0^\circ < \theta < 60^\circ$ .

About a quarter of the air showers were computed at the Ohio Supercomputer Center [143], the remaining air showers were computed at the Lyon supercomputing center CC-IN2P3 [144]. The production consumed approximately 5500 days on an AMD Opteron™ CPU with 2 GHz or equivalent processor. Air showers need 631 GByte of storage. The simulation of the SD events needs only a small fraction of these resources and was performed at the local institute. The SD events need 64 GByte of storage. The air showers and the SD events are freely available for the Pierre Auger Collaboration. Download instructions can be found online [142].

The mass production of air showers on the computing clusters was managed with a set of self-made software tools. The management software [141] is based on a *client-server* concept. A MySQL database as a central server stores a table with the defining properties of each air shower in the library. A set of identical clients, which may run independently on many machines, communicate with the server. Each client independently follows a simple algorithm:

- (1) Find a shower in the table that is not already processed. Mark the shower as being processed.
- (2) Run the air shower simulation and wait for its termination.
- (3) Check the simulation output for a simulation abort. If the output is fine, mark the shower as finished in the table of the central database. Otherwise, mark the shower as broken.

The system is designed for minimal manual intervention of the user. The only maintenance task are periodic checks for broken simulations and finding the cause of such errors.

The setup and the features of the library are discussed in the following. The production was divided in two distinct simulation runs, the technical and physical aspects of these runs will be pointed out. A technical summary of all features can be found in Appendix B.

### 5.3.1 Simulation setup

The work on the library started in the year 2005 and it was decided then to use the full Monte-Carlo simulation program CORSIKA for the production. The air shower simulation needed to predict the lateral shower profile, so that a cascade equation approach like CONEX could not be used. Other possibilities included the hybrid air shower program SENECA and the full Monte-Carlo program AIRES. Since SENECA was just emerging at that time, it was not considered. CORSIKA and AIRES were both already well established simulation codes, so the choice was between the two.

CORSIKA and AIRES share many similarities, but an important difference is the treatment of low energy interactions. AIRES employs a simplified treatment of the hadronic cascade at low energies, which is an enhanced form of Hillas' splitting algorithm [86, 134]. CORSIKA uses a range of detailed models, which are well established in other parts of high energy physics. The enhanced splitting algorithm in AIRES is very fast, but it is a rough approximation to true hadronic interactions.

It was shown in the first section of this chapter, the low energy hadronic interaction model has a strong impact on the lateral density profile of muons. This study is largely based on this profile and thus CORSIKA was favored.

Based on private communications that were later published in ref. [98], FLUKA was chosen as the low energy hadronic interaction model in CORSIKA. Because of the considerable theoretical uncertainties in the hadronic interaction models at high energies, half of the air showers were simulated with high energy hadronic interaction model QGSJet-II and the other half with EPOS. Differences between both sets are used to estimate the current systematic uncertainties of the simulation. The transition energy between the low energy and high energy interaction models is the default choice of 80 GeV.

The correct simulation of very inclined showers has to be activated in CORSIKA by a choice of certain code options. The options basically turn off some optimizations intended for vertical showers. By default, CORSIKA neglects upward going particles, as the probability that such particles generate downward-going particles in following interactions is very small in most cases. This is no longer true in near horizontal air showers, and CORSIKA can be forced to follow upward going particles with the UPWARD option.

Also, the curvature of Earth's atmosphere cannot be neglected in very inclined showers, as it is possible for  $\theta < 60^\circ$ . The CURVED option enables a curved atmospheric model in CORSIKA.

Finally, the calculation of the longitudinal shower profile needs special care. By default, CORSIKA calculates the longitudinal shower profile  $dN/dX_{\text{vert}}$  as a function of the vertical slant depth  $X_{\text{vert}}$  instead of the slant depth along the shower axis  $X$ . This is sufficient, as long as the curvature of the Earth can be neglected, because then  $X_{\text{vert}} \approx X \cos \theta$ . With the SLANT option, CORSIKA can be convinced to bin the longitudinal profiles correctly in  $X$ .

Unfortunately, it was not possible to use the last option together with the other two when the production of the library started: this feature was added to CORSIKA later. Thus, the first production run is barely usable for FD simulations and other analyses that rely on the longitudinal shower profile. However, all simulated showers are suitable for SD simulations and studies of the lateral profile on the ground.

#### Parameter distribution

A central features of a library of air showers is the distribution of the continuous cosmic ray parameters  $(E, \theta, \phi)$  within the range of the library. There are two possible choices:

- (A) The parameter space  $(E, \theta, \phi)$  is divided into a regular grid and air showers are simulated at discrete points  $(E_i, \theta_i, \phi_i)$  in this grid.
- (B) Some distribution function is chosen for the parameter space  $(E, \theta, \phi)$  and air shower parameters are picked at random from the distribution function.

The discrete case (A) is optimal, if the library is primarily used to parameterize air shower observables. The continuous case (B) is optimal, if the library is primarily used to compare the distribution of air shower observables with the according distribution in real data.

For example, it is simpler to parameterize the depth  $X_{\text{max}}$  of the electromagnetic shower maximum as a function of the cosmic ray energy  $E$  with discrete library, because then  $X_{\text{max}}$  can be averaged at

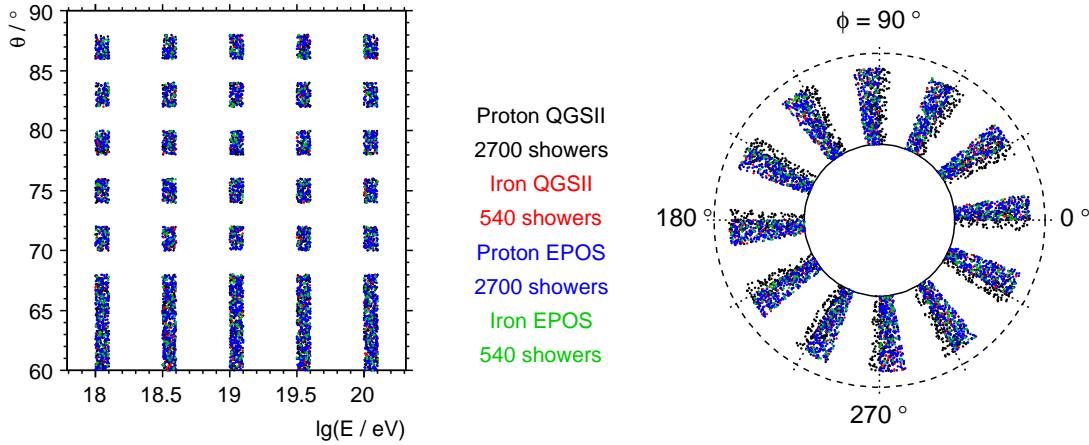


Figure 5.5: The plots illustrate the distribution of the simulated air showers. The scatter plot on the left shows the distribution of the zenith angle  $\theta$  and cosmic ray energy  $E$ , the custom plot on the right the distribution of the azimuth angle  $\phi$ . The radius of each point in the right plot is chosen at random and the polar angle according to the azimuth  $\phi$ . The first 1800 showers followed a different azimuth angle convention and are shifted with respect of the others shows by an azimuth angle of  $4.2^\circ$ .

discrete points in  $E$ . To compare the distribution of the measured  $X_{\max}$  with the expected distribution from simulations, it is necessary to use a continuous library. The continuous library can be re-weighted to reflect the distribution of  $(E, \theta, \phi)$  in real events, which is not possible as such with a discrete library.

When the library was planned, it was tried to comply with both requirements to get a general purpose library. The showers are randomly distributed in small finite bins in the  $(E, \theta, \phi)$  space and therefore some areas are covered with a dense continuous distribution, while others are left empty. This compromise allows a future extension of the library into a fully continuous form. Still, studies of shower properties in local regions in the  $(E, \theta, \phi)$  space are possible with reasonable statistic. The parameter distribution is illustrated in Fig. 5.5, the technical details can be found in Appendix B.

### Thinning and energy thresholds

For the air shower thinning in the simulation, it was decided to use a thinning level  $\epsilon = 10^{-6}$  and the weight limits

$$w_{\max}(\text{hadrons}, \mu) = \frac{E}{10^{15} \text{ eV}} \quad (5.3.1)$$

$$w_{\max}(e, \gamma) = \frac{E}{10^{13} \text{ eV}}, \quad (5.3.2)$$

which are a compromise between the computational requirements and the statistical quality of the lateral profile, with  $E$  being the cosmic ray energy. The latter limit is for electromagnetic particles, which are more numerous and therefore allowed to accumulate larger weights.

The maximum weight  $w_{\max}$  is made proportional to the cosmic ray energy to keep the number of actually followed particles in the air shower simulation roughly constant as the cosmic ray energy  $E$  increases. The computing time is proportional to the number of secondary particles in the shower, which in turn is roughly proportional to the cosmic ray energy  $E$ . If the maximum weight  $w_{\max}$  is also proportional to the cosmic ray energy  $E$ , the increase is effectively suppressed and showers of all energies take similar times to compute.

In addition to the basic thinning, CORSIKA also has a radial thinning algorithm. The radial thinning is applied when the simulated particles are written to the disk and therefore does not influence the shower development. It allows to reduce the size of the output file by removing particles near the shower

axis and re-weighting the remaining ones. Particles are removed with a probability following a  $(r_{\text{thin}}/r)^4$  distribution, whereas  $r_{\text{thin}}$  is a maximum radius to be defined by the user.

The radial thinning is supposed to reduce the statistical quality where detectors saturate anyway because of the huge number of particles. However, it turned out in the course of this analysis that there is not much to gain from the radial thinning in very inclined showers, because the number of particles close to the axis is already reduced through the atmospheric attenuation. The radial thinning was turned off in the second production run to avoid the unnecessarily large artificial fluctuations near the shower axis.

CORSIKA follows particles down to certain minimum momentum. The momentum thresholds also significantly affect the computing time. For photons, a momentum threshold is strictly necessary, because the differential cross-section for ionization energy loss has an infrared divergence. Other particles below a certain momentum may not be of interest, because they fall below a detection threshold. The following thresholds are chosen:

$$p_{\text{thr}}(\text{hadrons}, \mu) = 0.1 \text{ GeV} \quad (5.3.3)$$

$$p_{\text{thr}}(\text{e}, \gamma) = 250 \text{ keV}. \quad (5.3.4)$$

The threshold  $p_{\text{thr}}(\text{hadrons}, \mu)$  is used for the electromagnetic particles and approximately equal to the threshold of Cherenkov light production in water for electrons. Similarly, the threshold  $p_{\text{thr}}(\text{hadrons}, \mu)$  is the effective energy threshold for muons in water [16]. Electromagnetic particles which fall below the threshold are added to the profile of the longitudinal energy loss of the shower. Hadrons and muons are dropped from the simulation.

### Ground altitude, geomagnetic field, atmosphere

The development of an air shower depends on the local conditions of the observation site: the altitude of the ground, the profile of the atmospheric density, and the local geomagnetic field.

The geography of the southern observatory was shown in Chapter 4. The site is very flat. The altitude varies only by less than 300 m over a distance of about 50 km. The corresponding maximum difference in slant depth is negligible, less than  $1 \text{ g cm}^{-2}$  at  $\theta = 60^\circ$ , for example. The ground plane altitude in CORSIKA is set to 1425 m.

Values for the geomagnetic field  $\mathbf{B}$  can be obtained from the IGRF-10 model [145]. The geomagnetic field varies slowly with time and the position of the observer, which is shown in Fig. 5.6. CORSIKA uses a fixed geomagnetic field. The geomagnetic declination  $\delta_B$  has no influence on the air shower simulation, as the internal coordinate system of CORSIKA is always oriented so, that the geomagnetic field is parallel to the  $x - z$ -plane. The following values are used for the remaining parameters, the field strength  $B$  and the inclination  $\theta_B$ :

$$B = 24.6 \mu\text{T} \quad \theta_B = -35.2^\circ.$$

The geomagnetic declination  $\delta_B$  comes into play, when CORSIKA's coordinate system is transformed back into the site coordinate system. This transformation is done as part of the SD simulation by the Auger Offline framework. A constant value of

$$\delta_B = 4.2^\circ$$

is used here. This value is apparently not in best agreement with the IGRF-10 prediction and should be made time dependent in a future update of the analysis software.

So far, the time dependency of the geomagnetic field can be neglected. Fig. 5.6 shows, that the variation of the magnitude between 2005 and 2010 is at the level of 1 %, the variation of the inclination less than  $0.5^\circ$ , and the variation of the declination less than  $0.8^\circ$ . The long-term evolution of Earth's magnetic field cannot be predicted. Over the life-time of the experiment of 20 years, the changes may become relevant.

CORSIKA uses a homogeneous geomagnetic field, although the true field shows a spatial variation. The muons in near horizontal showers travel up to 400 km through the atmosphere until they reach the ground, which is a significant distance compared to the variation scale of the geomagnetic field. Fig. 5.6 shows, how the geomagnetic field changes in the vicinity of the surface detector array. The effects turn out to be of the same order as the time evolution effects for showers with zenith angles  $\theta \lesssim 85^\circ$ .

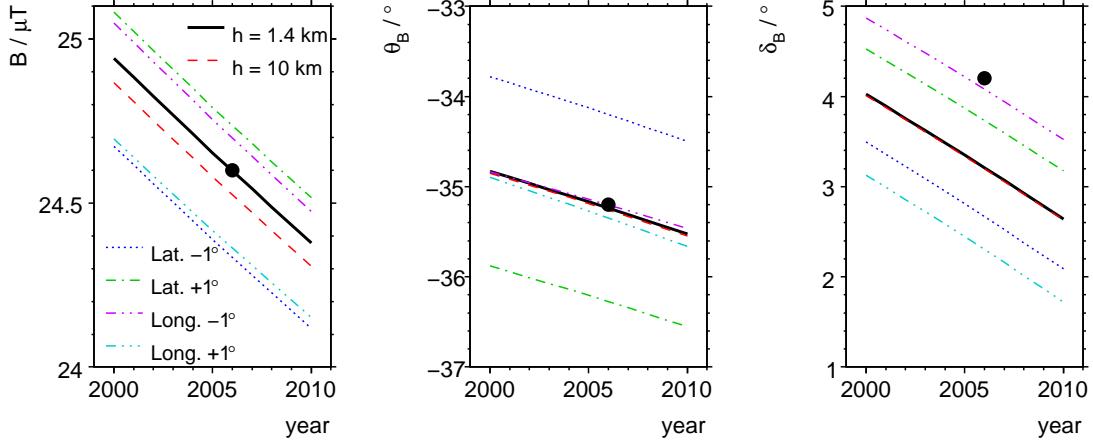


Figure 5.6: The graphs show the evolution of the geomagnetic field from 2000 to 2010 at the center of the surface detector at the Southern Pierre Auger Observatory, as predicted by the IGRF model [145]. Depicted from left to right is the total magnitude of the geomagnetic field, the inclination, and declination. The solid line shows the geomagnetic field at the center of the Southern SD, at a ground altitude of 1400 m. The dashed line shows the field in an altitude of 10 km above sea level. The other lines show how the field varies, if the longitude and latitude of the latter point is changed by  $\pm 1^\circ$ , which is equivalent to moving the observation point by about 100 km to the north, south, east, and west. The large black dot represents the fixed geomagnetic field in the simulation.

The atmospheric density profiles over Malargüe [41–43] differ slightly from the global reference model, the US standard atmosphere [146], as shown in Fig. 5.7. Furthermore, the density profiles show variations on the scale of months and days.

The total atmospheric depth  $X_{\text{atm}}$  shows negligible variation, but a considerable difference to the standard model of 10 % with a variation of about 5 % is observed in the air density  $\rho_{\text{air}}$  at an altitude of 17 km. These differences should have an effect on the produced number of muons  $N_\mu$ , if the altitude  $h_{\max}^\pi$  of the maximum of the hadronic shower cascade is close to this altitude. The simplified air shower model derived in Chapter 5 predicts  $N_\mu \propto \rho_{\text{air}}(h_{\max}^\pi)$ .

For zenith angles smaller than  $70^\circ$ , the hadronic shower maximum is below the critical altitude range and the differences to the U.S. standard atmosphere can be neglected. At larger zenith angles, simulations with the correct atmosphere could show up to 10% more muons on average and a seasonal variation of up to 5 %. The systematic shift is small compared to other theoretical uncertainties in the total number of muons  $N_\mu$  and therefore not critical for this study. The seasonal variation of up to 5 % leads to a corresponding variation in the SD energy scale, since the total number of muons  $N_\mu$  on the ground is to derive the cosmic ray energy. This energy variation could distort the measured cosmic ray flux in Chapter 8, but the impact evaluated with real data turns out to be negligible.

The first production run used the U.S. standard atmosphere. The second production run used the spring atmosphere shown in Fig. 5.7, which is close to an average atmospheric profile over Malargüe.

### Event simulation

The simulation of surface detector events was done with the `Offline`-framework [137], using a developer version of 2.5-Godot with the internal revision number 7732.

The SD array in the simulation emulates the completely constructed detector, without holes or gaps. Every air shower was used to produce one simulated surface detector event. The core position is randomly set in a square of  $2 \text{ km} \times 2 \text{ km}$  in the approximate center of the array.

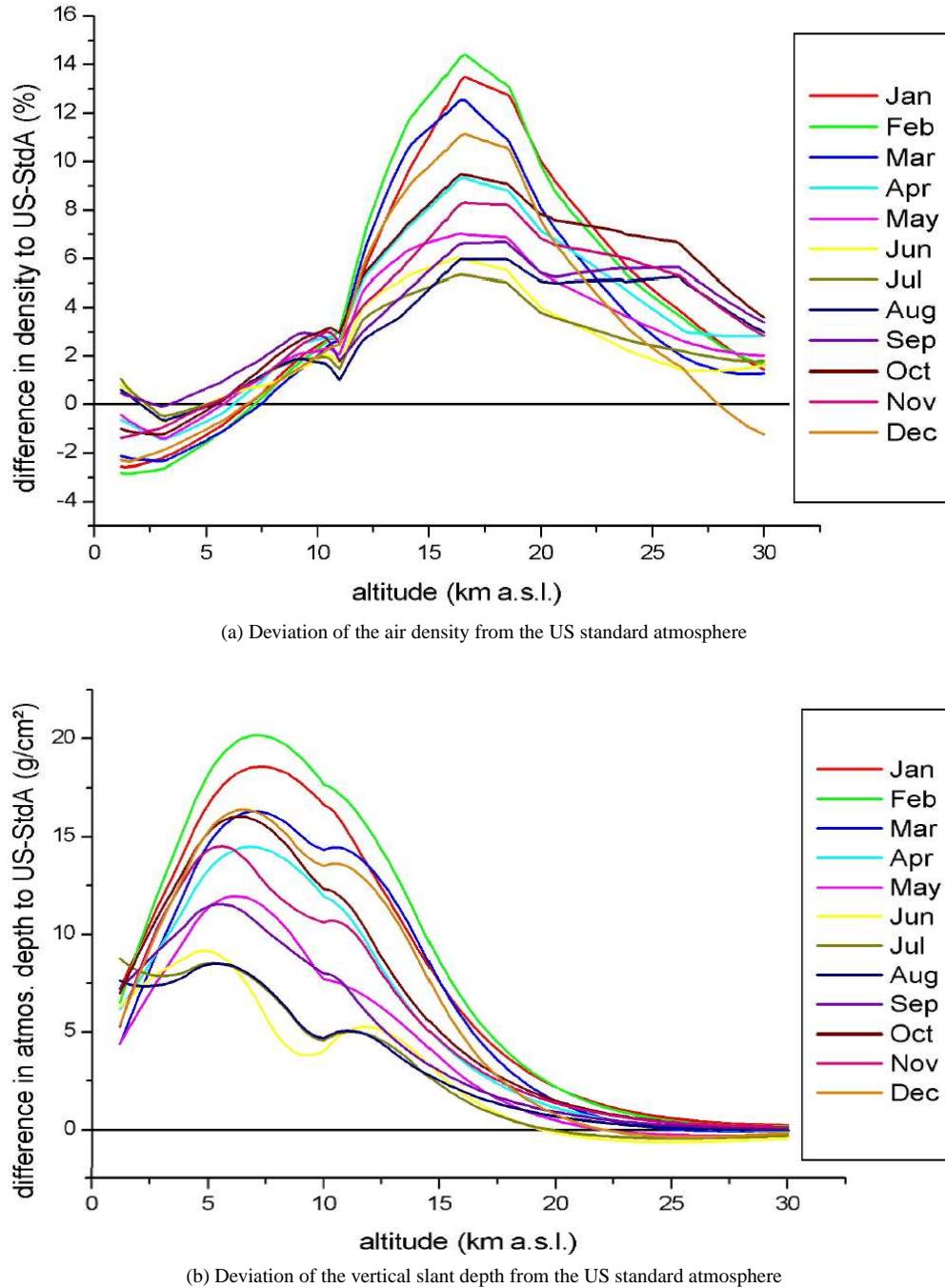


Figure 5.7: The graphs show the difference between monthly models of the Malargüe atmosphere and the US standard atmosphere [146] (from [43]).

### 5.3.2 Comparison with older productions

Large libraries of air showers exist which were generated with former versions of CORSIKA and other hadronic interaction models. It is worthwhile to compare this library with the former productions in order to estimate the impact of the changes.

Fig. 5.8 shows a comparison of the combinations

- CORSIKA-6.0 + QGSJet01c + GHEISHA2002d (configuration A), and
- CORSIKA-6.5 + QGSJet-II-3 + FLUKA2006 (configuration B, this work),

for an example shower close to a zenith angle of  $60^\circ$ . Configuration A was a common choice before QGSJet-II succeeded QGSJet and FLUKA succeeded GHEISHA.

The electron density profiles in both configurations seem to agree, but the large fluctuations do not allow any quantitative conclusions. The muon density profiles  $n_\mu$  on the other hand show significant deviations. The total number of muons  $N_\mu$  in configuration B is about 10 % lower than in configuration A and the muon density profile  $n_\mu$  is significantly steeper at radial distances  $r > 2000$  m.

The differences of the muon component are analyzed further through individual variations of the CORSIKA version and the hadronic interaction models. The change of the CORSIKA version has no apparent effect. The global offset of about 10 % turns out to be caused by the change of the high energy hadronic interaction model, while the steeper decline of the lateral muon density  $n_\mu$  is due to the change from the low energy hadronic interaction model. The latter effect is confirmed by another analysis [98]. The sensitivity of  $n_\mu$  at large radii to the low energy hadronic model is apparent.

It may surprise that the decrease in the number of muons is not accompanied by an according increase in the number of electrons due to energy conservation. There is corresponding increase, but it is smaller by an order of magnitude because the muons carry only a tenth of the total cosmic ray energy, as explained in Chapter 3.

The comparison allows to conclude that the choice of both the high and the low energy hadronic interaction model significantly affects the muon profile  $n_\mu$  on the ground. Comparisons of low energy hadronic interaction models favor FLUKA over GHEISHA [98], so that the newer simulations made with FLUKA should be preferred in analyses of very inclined air showers.

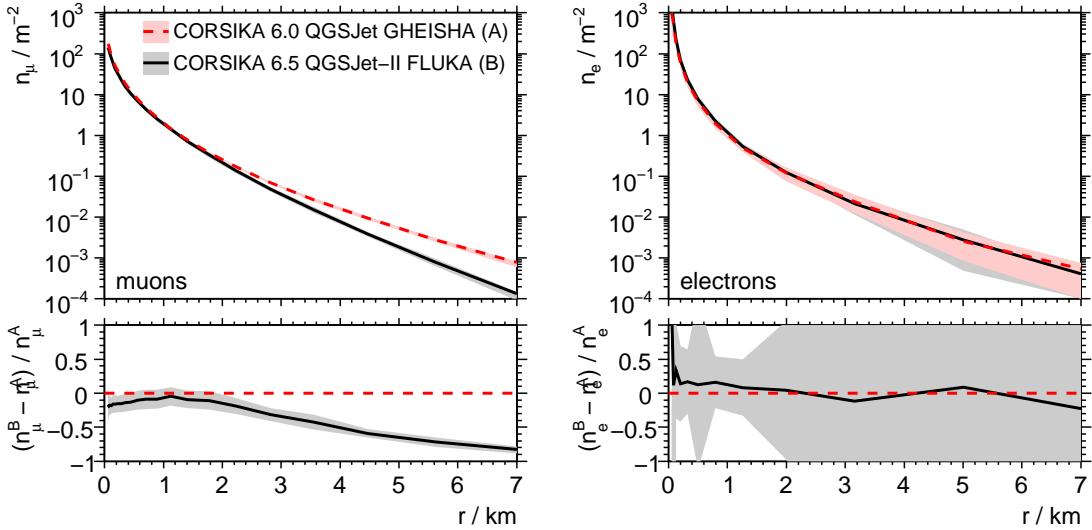


Figure 5.8: The plots show the average lateral density of muons and electrons of an example shower at  $\theta \approx 60^\circ$  and  $E \approx 2 \times 10^{19}$  eV for the two combinations CORSIKA-6.0 + QGSJet01c + GHEISHA2002d (configuration A) and CORSIKA-6.5 + QGSJet-II-3 + FLUKA2006 (this work). The averages are built from 10 showers each, the altitude of the first interaction is fixed in all showers to 33 km. The colored bands indicate the variation of the densities in the individual showers. All other settings are the same as for the showers with the run IDs 4681-6480 in Table B.2.

## 5.4 Modeling very inclined air showers

It was argued in Chapter 3, that the lateral density profile of the muons  $n_\mu$  in very inclined air showers should factorize into three parts: the total number of produced muons  $N_\mu^0$ , the attenuation factor  $a$ , and the normalized density profile  $p_\mu$ :

$$n_\mu(r, \psi; E, A, \theta, \phi) \simeq N_\mu(E, A, \theta) p_\mu(r, \psi; \theta, \phi), \quad (5.4.1)$$

whereas  $N_\mu$  depends on the energy  $E$ , atomic mass  $A$ , and zenith angle  $\theta$  of the cosmic ray and  $p_\mu$  on the coordinates  $(r, \psi)$  in the lateral coordinate system and the orientation  $(\theta, \phi)$  of the shower axis relative to the geomagnetic field vector (compare with Eq. (3.3.9)).

It was further argued that the total number of muons  $N_\mu$  on the ground is related to the cosmic ray energy via a power law:

$$N_\mu(E, A) \simeq a(\theta) C(A) E^{1/\gamma}, \quad (5.4.2)$$

whereas  $a(\theta)$  describes the muon attenuation in the atmosphere and  $C(A)$  is a factor that depends on the atomic mass  $A$  and  $\gamma$  is a constant (compare Eq. (3.2.21)).

Eq. (5.4.1) and Eq. (5.4.2) are approximations, which need to be confirmed over the energy range observed at the Pierre Auger Observatory of  $10^{18}$  eV to  $10^{20}$  eV with full air shower simulations. This will be done in the following with the help of the air shower library from the previous section. Similar analyses have been done before [11, 13, 54], based on simulations with the air shower program AIRES and the hadronic interaction model SIBYLL. They are confirmed in this study on the basis of another simulation program and other hadronic interaction models.

In a second step, a full parameterization of the muon density  $n_\mu$  on the ground will be derived from the air shower library on the basis of Eq. (5.4.1) and Eq. (5.4.2). The parameterization of  $n_\mu$  will be an important input for the event reconstruction in Chapter 6.

### 5.4.1 Number of muons on the ground

The total number of muons  $N_\mu$  on the ground is regarded first. Integrating Eq. (5.4.1) over the ground surface and inserting Eq. (5.4.2) yields

$$N_\mu(E, A, \theta) \simeq a(\theta) N_\mu^0(E, A) \simeq a(\theta) C(A) E^{1/\gamma}. \quad (5.4.3)$$

The factorization in Eq. (5.4.1) and the validity of Eq. (5.4.2) are confirmed by fitting Eq. (5.4.3) to simulated showers. An analysis of the bias of the fit shows the quality of the approximations. The fits are done separately for the hadronic interaction models QGSJet-II and EPOS.

In order to perform the fit, functional forms for  $C(A)$  and  $a(\theta)$  need to be found. As the air shower library only contains two types of cosmic rays, protons and iron nuclei, the functional form of  $C(A)$  cannot be derived. The fit is done separately for proton and iron showers, which reduces  $C(A)$  to a constant in each fit. The factorization of  $C(A)$  is confirmed by showing that the independent parameterizations divided by  $C(A)$  approximately agree for both nuclei.

The parameterization of the attenuation function  $a(\theta)$  turns out to be surprisingly simple. It is fitted very well by a power law of the distance  $d_{\max}$  between the electromagnetic shower maximum and the shower impact point on the ground along the shower axis, as shown in Chapter 3:

$$a(\theta) \propto d_{\max}(\theta)^\beta, \quad (5.4.4)$$

whereas  $\beta$  is a constant.

There are two points about this equation, which need some discussion. The first is, that the muon attenuation should in general depend on the distance of the hadronic shower maximum to the impact point on the ground, not on  $d_{\max}$ , which is the distance of the *electromagnetic* shower maximum to the impact point on the ground.

At  $\theta > 60^\circ$ , however, the difference between the two distances is small compared to  $d_{\max}$ , as explained in Chapter 3. The use of  $d_{\max}$  then is just more convenient, as it can be calculated numerically without any simulation input. The derivation is described in Appendix A.1. The fit uses the  $d_{\max}(\theta)$  curve for a cosmic ray energy of  $10^{19}$  eV.

The second point is the functional form of a power law. Intuitively, one might expect an exponential form

$$a(\theta) \propto e^{-d_{\max}(\theta)/(\gamma c\tau)}, \quad (5.4.5)$$

where  $\gamma$  is an average Lorentz factor of the muons in the shower and  $c\tau$  is their decay constant. On the other hand, the energy distribution of muons is wide and energy loss processes also play a role. Together, these effects apparently turn into an effective power law in the considered zenith angle range.

The number of muons  $N_\mu$  on the ground is now a function of independent power laws terms. To fit the model to the simulated data, it is convenient to rewrite Eq. (5.4.3) as

$$\begin{aligned} N_\mu &= 10^\alpha [d_{\max}(\theta)/\text{km}]^\beta [E/\text{EeV}]^{1/\gamma} \\ \Leftrightarrow \lg N_\mu &= \alpha + \beta \lg[d_{\max}(\theta)/\text{km}] + \frac{1}{\gamma} \lg[E/\text{EeV}], \end{aligned} \quad (5.4.6)$$

whereas  $\alpha$ ,  $\beta$ , and  $\gamma$  are the free parameters in the fit.

The fit of Eq. (5.4.6) turns out to be non-trivial. The usual least-squares fit expects Gaussian fluctuations around the mean, see *e.g.* ref. [147]. The data shows asymmetric fluctuations and therefore the least-squares fit yields biased results. The bias refers to the mean of the distribution of the residuals  $(N_\mu - \langle N_\mu \rangle)/\langle N_\mu \rangle$ , whereas  $\langle N_\mu \rangle$  is calculated by the fit and  $N_\mu$  the random data.

The general procedure in case of non-Gaussian fluctuations is to model the probability density function of the fluctuations and use the likelihood method to obtain the best parameter estimates. It turns out, that the fluctuations are not described well by standard distributions. A special distribution, which shall be named *exp-normal* distribution, yields a good description, but is mathematically too complex to be of any use. The exp-normal distribution is derived and discussed in Appendix A.2.

Therefore, another approach is used, based on a modification of the standard least-squared method. The unmodified least-squares method is based on the numerical minimization of the sum of the squared residuals around a fit

$$\chi^2 = \sum_i \left( \frac{N_\mu^i - \langle N_\mu \rangle}{\sigma_i} \right)^2, \quad (5.4.7)$$

whereas  $\langle N_\mu \rangle$  is calculated with the model and  $\sigma_i$  is the expected uncertainty of data point  $N_\mu^i$ . The uncertainties  $\sigma_i$  are not known *a priori*, but it turns out that the relative fluctuations are approximately constant, so that it is possible to use a constant  $\sigma_{\text{rel}} = \sigma[N_\mu]/N_\mu \approx \sigma_i/N_\mu^i$  in the fit. The sum of the residuals can be rewritten as

$$\chi^2 = \sum_i \left( \frac{N_\mu^i/\langle N_\mu \rangle - 1}{\sigma_{\text{rel}}} \right)^2 \quad (5.4.8)$$

To enforce unbiasedness, a penalty term is added to the sum:

$$\tilde{\chi}^2 = \chi^2 + \lambda \left( \sum_i \frac{N_\mu^i/\langle N_\mu \rangle - 1}{\sigma_{\text{rel}}} \right)^2. \quad (5.4.9)$$

The new term is the squared average of the residual distribution. The penalty factor  $\lambda$  can be used to tune the importance of the penalty term for the minimization. The simple choice  $\lambda = 1$  is used here. The numerical minimization of  $\tilde{\chi}^2$  is done with the MINUIT package [148].

The best estimate of  $\sigma_{\text{rel}}$  is obtained *a posteriori* from the width of the distribution of the residuals  $(N_\mu - \langle N_\mu \rangle)/\langle N_\mu \rangle$ . The size of  $\sigma_{\text{rel}}$  does not affect the position of the minimum of  $\chi^2$  and may therefore be set to 1 in a first iteration of the fit. The first iteration is then used to derive  $\sigma_{\text{rel}}$  from the residual distribution. The second iteration of the fit then uses the best estimate of  $\sigma_{\text{rel}}$  to obtain meaningful uncertainty estimates for the parameters  $\alpha$ ,  $\beta$ , and  $\gamma$ .

The parameter uncertainties derived from the modified fit are nevertheless only approximate. MINUIT calculates the parameter uncertainties based on the assumption that the squares of the residuals  $(N_\mu - \langle N_\mu \rangle)/\langle N_\mu \rangle$  follows a  $\chi^2$  distribution. The fluctuations of the residuals  $(N_\mu - \langle N_\mu \rangle)/\langle N_\mu \rangle$  are only approximately Gaussian and thus the sum in Eq. (5.4.8) follows only approximately the  $\chi^2$  distribution. Adding a penalty term like in Eq. (5.4.9) also biases the uncertainty estimates. The impact of the penalty term on the uncertainties can be analyzed by varying the penalty factor  $\lambda$ . It turns out that the impact of the penalty term is negligible in this case.

The fits and the analyses of the distribution of the residuals  $(N_\mu - \langle N_\mu \rangle)/\langle N_\mu \rangle$  for the different hadronic interaction models and cosmic ray masses are shown in Fig. 5.9. The residual distribution has mean at zero as guaranteed by the fit. The distribution is compared with a Gaussian and the exp-normal distribution. Both models work well in case of iron showers, but proton showers have a tail towards low number of muons and are better described by the exp-normal distribution. A possible explanation for this tail was given in Chapter 3.

Fig. 5.10 and Fig. 5.11 show detailed analyses of the model bias and of  $\sigma_{\text{rel}}$  as a function the cosmic ray energy and direction. The model bias turns out to be smaller than 2 %, which is an excellent result. The relative size  $\sigma_{\text{rel}}$  of shower-to-shower fluctuations is a constant in very good approximation. The uncertainty of  $\sigma_{\text{rel}}$  is calculated according to ref. [149]:

$$\sigma^2[\sigma^2] = \frac{1}{N} \left( \mu_4 - \frac{N-3}{N-1} \sigma^4 \right), \quad (5.4.10)$$

whereas  $N$  is the number and  $\mu_4$  the forth central moment of the measurements.

The fits already confirmed the basic structure of Eq. (5.4.1) and Eq. (5.4.2). It is still necessary to show that the cosmic ray mass  $A$  only scales  $N_\mu$ , but leaves the other dependencies of  $N_\mu$  unaffected. This is the case, if the fit parameters  $\beta$  and  $\gamma$  are approximately independent of  $A$ . Fig. 5.12 shows the comparison. The parameter  $\beta$  varies at the level of 2 %,  $\gamma$  at the level of 1 %. Propagated to  $N_\mu$  this corresponds to a bias of up to 10 % in the range  $10^{18} \text{ eV} < E < 10^{20} \text{ eV}$  and  $60^\circ < \theta < 88^\circ$ .

Table 5.1 summarizes the final results. The following conclusions are derived for cosmic rays in the range  $10^{18} \text{ eV} < E < 10^{20} \text{ eV}$  and  $60^\circ < \theta < 88^\circ$ .

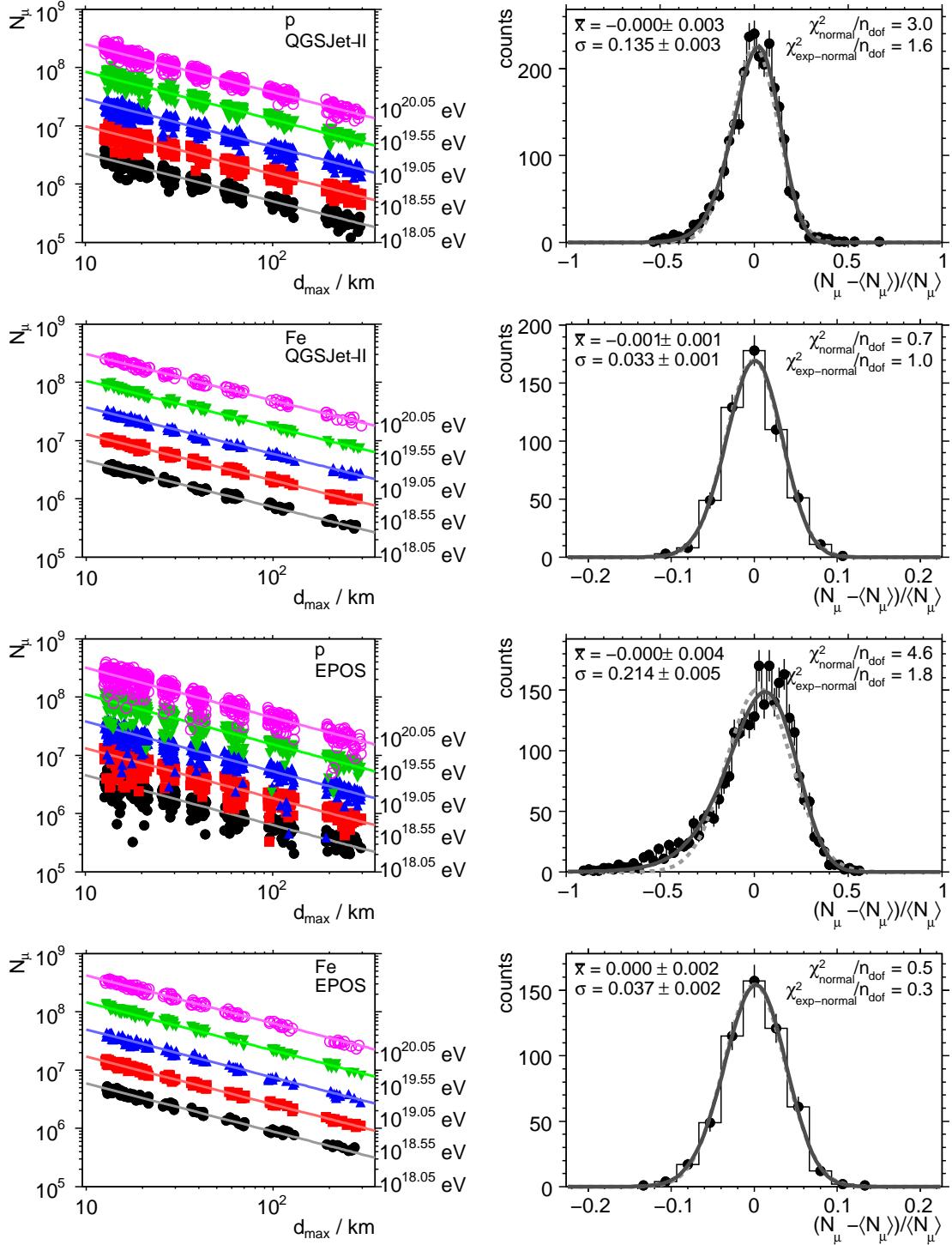


Figure 5.9: Left: The points show the simulated muon number  $N_\mu$  on the ground as function of  $d_{\max}(\theta)$  for proton and iron showers, simulated with QGSJet-II and EPOS. The five different energy intervals from Fig. 5.5 are indicated by different markers and colors. The solid lines are energy slices of the two-dimensional model  $N_\mu(E, \theta)$ . Right: The histograms show the distribution of the residual around the fit shown on the left side. The normal and the exp-normal distributions are shown in comparison. The reduced  $\chi^2$  values quantify the agreement of these fluctuation models with the observed distribution.

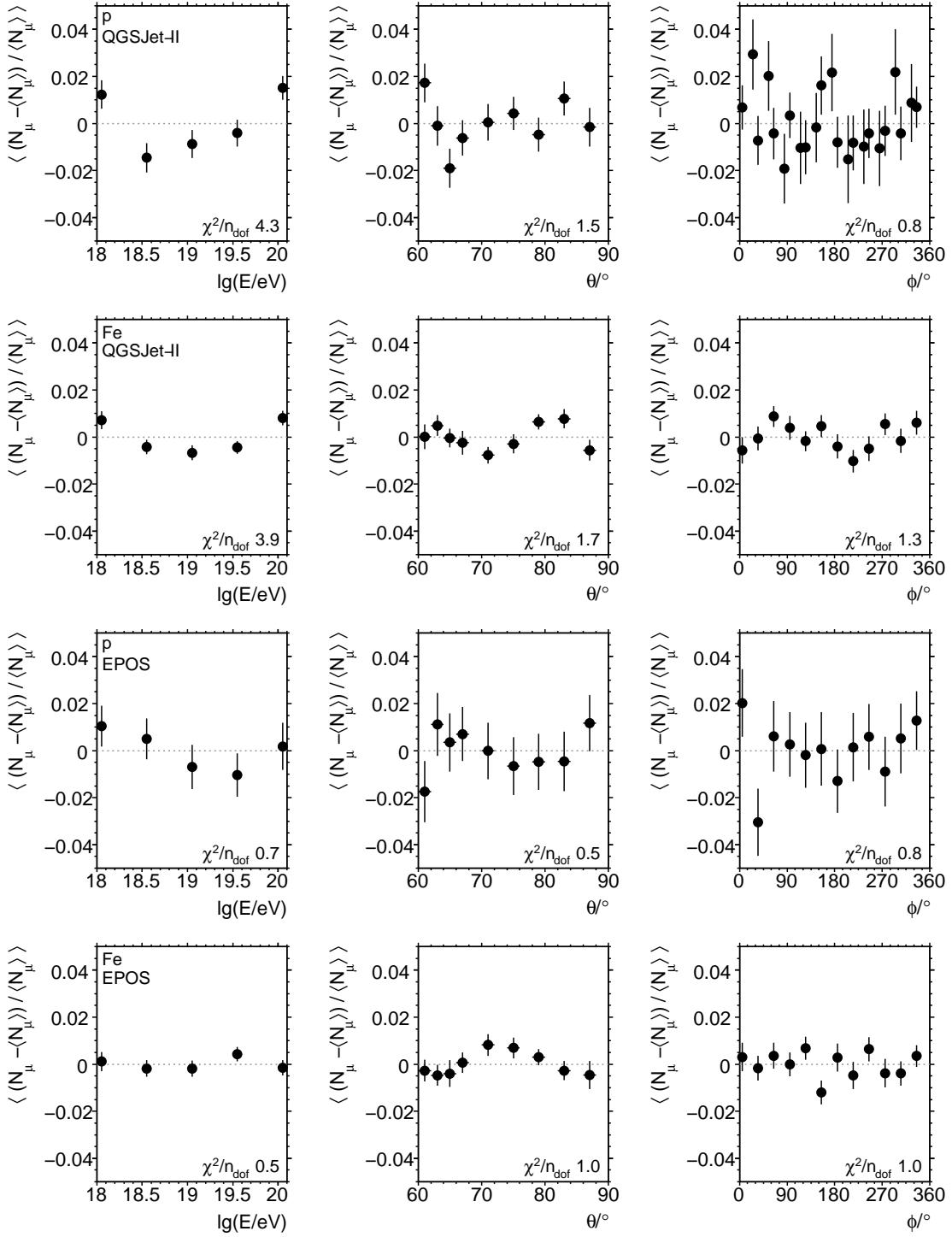


Figure 5.10: The plots show the bias  $\langle (N_\mu - \langle N_\mu \rangle) / \langle N_\mu \rangle \rangle$  of the  $N_\mu$  model as a function of the cosmic ray energy  $E$ , the zenith angle  $\theta$ , and the azimuth angle  $\phi$  for all combinations of cosmic ray masses and hadronic interaction models. The reduced  $\chi^2$  values quantify the agreement with the zero expectation.

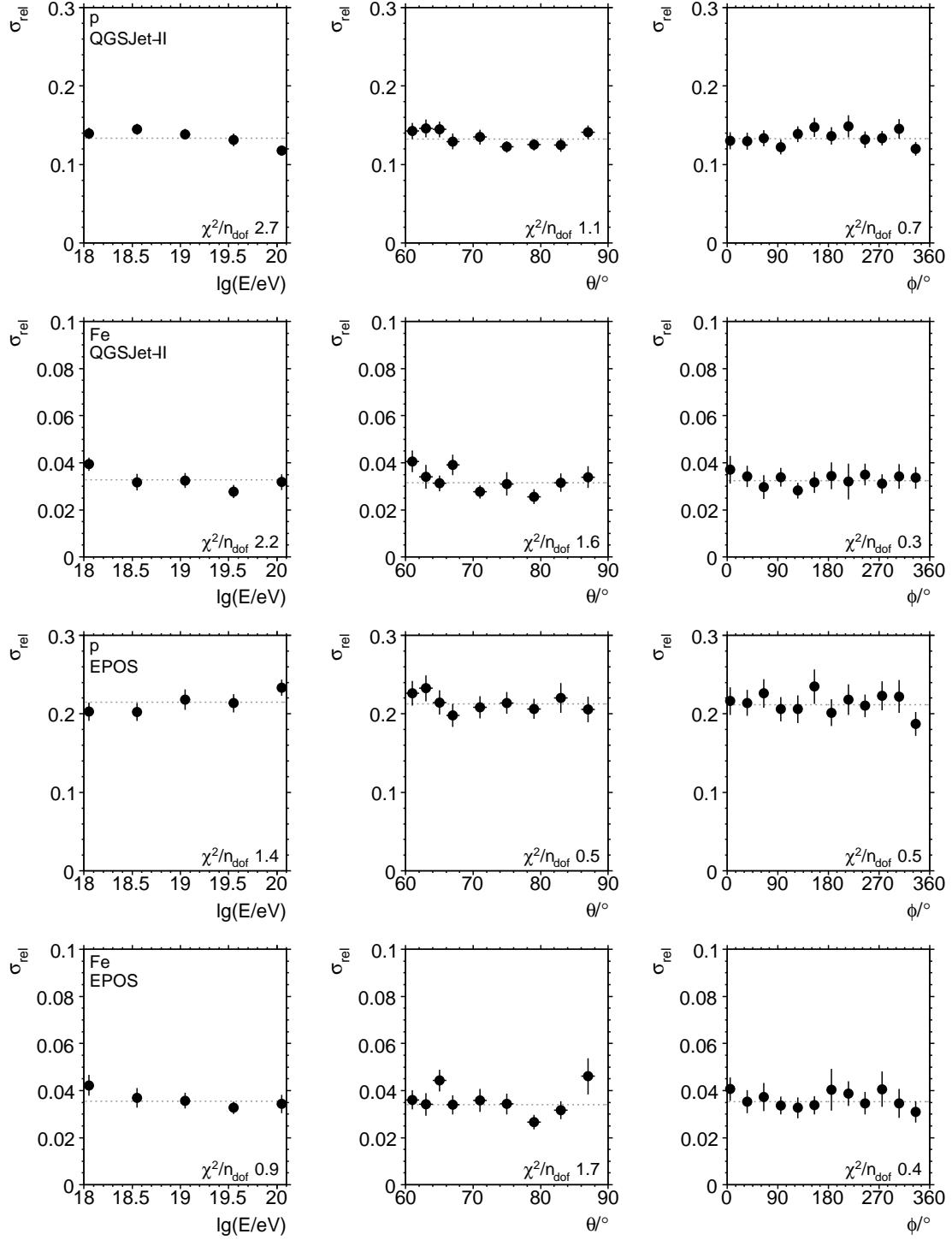


Figure 5.11: The plots show the width  $\sigma_{\text{rel}}$  of the residual distribution around the  $N_\mu$  model as a function of the cosmic ray energy  $E$ , the zenith angle  $\theta$ , and the azimuth angle  $\phi$  for all combinations of cosmic ray masses and hadronic interaction models. The reduced  $\chi^2$  values quantify the agreement with a constant.

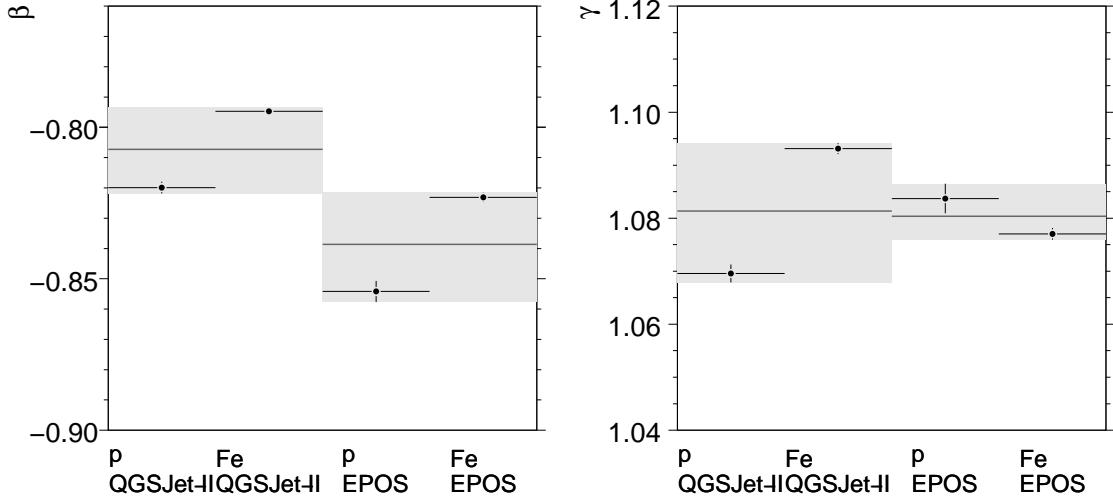


Figure 5.12: The plots show the variation of the parameters  $\beta$  (left) and  $\gamma$  (right) from Eq. (5.4.6) with the cosmic ray mass and the hadronic interaction model. The horizontal lines show the average values of  $\beta$  and  $\gamma$  for each hadronic interaction model, the gray bands estimate the systematic uncertainties.

Table 5.1: The table shows the fitted parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  of Eq. (5.4.6). The ratio  $N_\mu/n_\mu^{\text{ref}}$  shows the total number of muons on the ground relative to proton showers simulated with QGSJet-II. The ratio depends weakly on the zenith angle  $\theta$  and the cosmic ray energy  $E$ . The given value is derived at  $E = 10^{19}$  eV and  $\theta = 70^\circ$ , the systematic uncertainty due to this dependency in the range  $10^{18}$  eV  $< E < 10^{20}$  eV and  $60^\circ < \theta < 88^\circ$  is shown in braces. The last column shows the relative shower-to-shower fluctuation  $\sigma_{\text{rel}}$  of the total number of muons  $N_\mu$  on the ground. The systematic uncertainty shown in braces is estimated from Fig. 5.11.

Hadronic model	QGSJet-II		EPOS		
	Cosmic ray	proton	iron	proton	iron
$\alpha$		$7.300 \pm 0.003$	$7.403 \pm 0.002$	$7.468 \pm 0.006$	$7.546 \pm 0.003$
$\beta$		$-0.820 \pm 0.002$	$-0.795 \pm 0.001$	$-0.854 \pm 0.003$	$-0.823 \pm 0.002$
$\gamma$		$1.070 \pm 0.001$	$1.093 \pm 0.001$	$1.084 \pm 0.002$	$1.077 \pm 0.001$
$N_\mu/N_\mu^{\text{ref}}$		$1.000 \pm 0.010$	$1.314 \pm 0.007$	$1.281 \pm 0.018$	$1.718 \pm 0.008$
		$(-0.082 + 0.120)$		$(-0.103 + 0.069)$	$(-0.034 + 0.030)$
$\sigma[N_\mu]/N_\mu$		$0.135 \pm 0.002$	$0.034 \pm 0.001$	$0.214 \pm 0.003$	$0.038 \pm 0.001$
		$(\pm 0.020)$		$(\pm 0.005)$	$(\pm 0.005)$

- Eq. (5.4.1) and Eq. (5.4.2) are valid at the level of 10 %. This is sufficiently small to model the total number of muons  $N_\mu$  on the ground with these approximations in the event reconstruction of Chapter 6. The  $\theta$ -dependency of  $N_\mu$  has the largest systematic uncertainty and may impose a systematic bias in the reconstruction of the energy  $E$  of air showers from SD data. It will be shown in Chapter 6 with an analysis of real events that there is no significant bias introduced by the  $\theta$ -dependency of  $N_\mu$ .
- The relative size  $\sigma_{\text{rel}}$  of the shower-to-shower fluctuations of  $N_\mu$  is constant for cosmic rays with a given mass  $A$  in very good approximation. The relative fluctuation  $\sigma_{\text{rel}}$  is sensitive to the cosmic ray mass  $A$  and sensitive to hadronic interaction models if  $A$  is small.

The predicted shower-to-shower fluctuations of  $N_\mu$  for proton showers range between 14% and 21%. The fluctuations depend mostly on the first hadronic interactions of the cosmic ray in the atmosphere, as discussed in Chapter 3. The hadronic interaction models are extrapolated to their extreme in these interactions, which may explain the large systematic variation in  $\sigma_{\text{rel}}$ .

The shower-to-shower fluctuations of iron showers do not show this sensitivity. The details of the first hadronic interactions average out to some degree in the interactions of many nucleons.

- The absolute scale of  $N_\mu$  is sensitive to the cosmic ray mass  $A$  and the hadronic interaction models. The disagreement between the models of about 35 % is of the same magnitude as the difference between proton and iron showers. Other comparisons find differences up to 50 % [11, 13, 131].

### 5.4.2 Universality of the normalized muon density profile

The first part of Eq. (5.4.1) is already confirmed. The next step is to show, that  $p_\mu$  is approximately independent of the cosmic ray energy  $E$  and mass  $A$ . This feature is often called *universality* in the context of extensive air showers.

In order to show the universality, simulated showers in a small  $(E, \theta, \phi)$  region are averaged, which yields an average profile  $p_\mu$  and a variance. The profiles of showers arriving from the east and west can be merged, since they show a basic mirror symmetry. This yields a statistic of ten showers per zenith angle and energy interval.

Fig. 5.13 shows the variation of  $p_\mu$  with the cosmic ray energy  $E$  for proton showers. Within a radius of 4 km, variations up to 10 % are found between  $10^{18}$  eV and  $10^{20}$  eV. The observed variation of  $p_\mu$  within a single energy interval seems to be dominated by shower-to-shower fluctuations.

Fig. 5.14 shows the variation of  $p_\mu$  with the atomic mass  $A$ . It was argued in Chapter 3, that an insensitivity of a shower observable to the cosmic ray energy  $E$  is connected with an insensitivity to the cosmic ray mass  $A$ . The principle is confirmed here. The variation of the normalized profile  $p_\mu$  with the atomic mass  $A$  is again at the level of 10 %.

The universality of  $p_\mu$  can be confirmed at a level of 10 %. This is the same level of systematic uncertainty as in the first part of Eq. (5.4.1). The impact of the approximations in Eq. (5.4.1) on the event reconstruction will be derived in Chapter 6 and turn out to be smaller than 10 %.

It is possible to conclude from the universality of  $p_\mu$  that the lateral profile of the muon energy  $E_\mu$  is also universal to a similar degree. This is because there is a strong correlation between the radial distance  $r_\mu$  of a muon from the shower axis and the energy  $E_\mu$  of the muon, as discussed in Chapter 3. This will be used in the next step.

The results confirm an earlier analysis, performed with another air shower simulation program and other hadronic interaction models [13].

### 5.4.3 Parameterisation of the muon density profile

The total number of muons  $N_\mu$  was already parameterized at the beginning of this section. The normalized lateral profile of the muon density  $p_\mu$  is parameterized in the following to obtain a full model of the muon density  $n_\mu$  on the ground with Eq. (5.4.1).

The parameterization [14, 84] used in this study is entirely phenomenological. The approach differs from the one presented in ref. [11, 54]. In the other approach, a model of  $p_\mu$  is derived by applying an

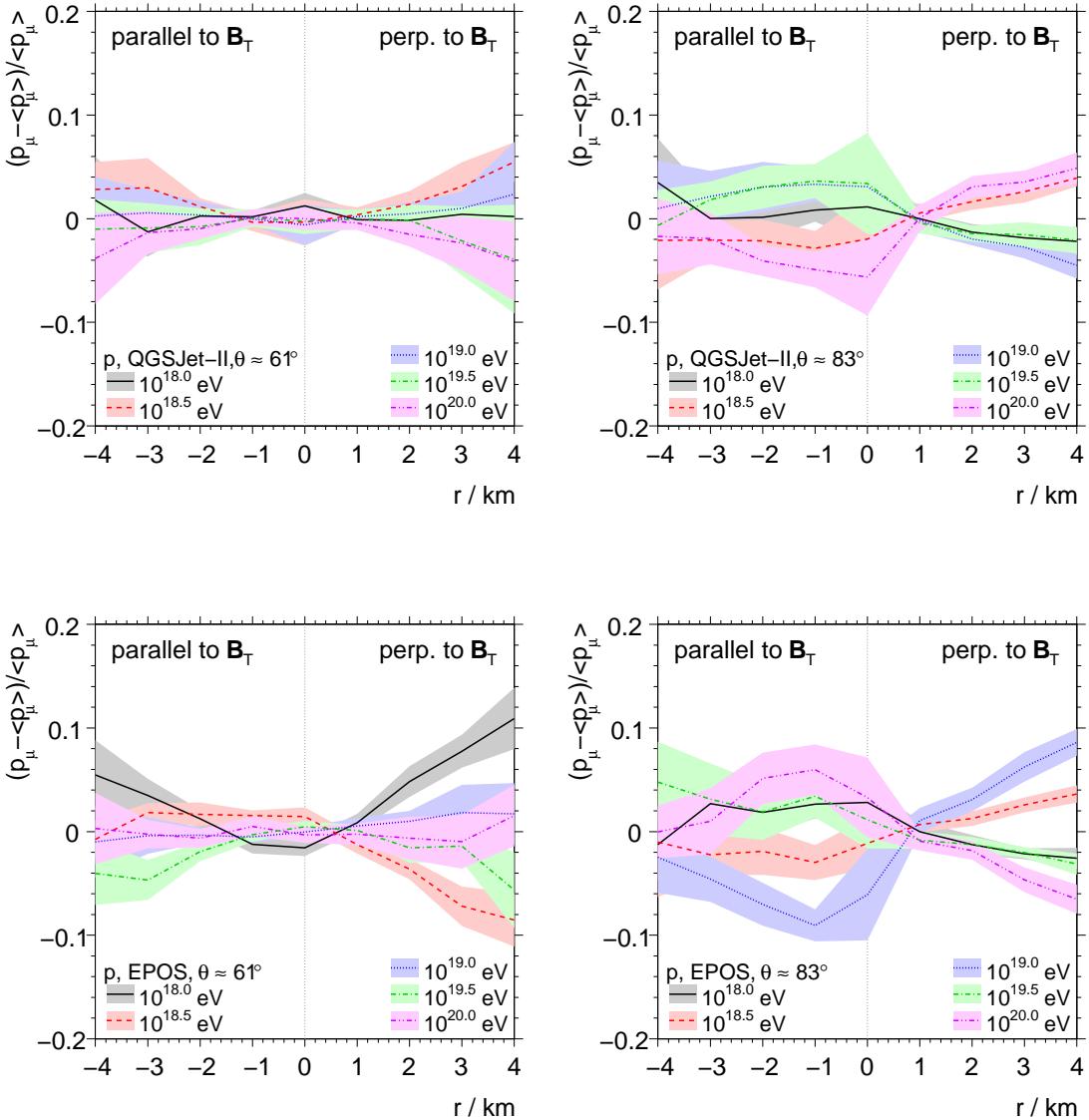


Figure 5.13: The plots show the variation of the normalized lateral density profile of muons  $p_\mu$  on the ground as a function of the cosmic ray energy  $E$  in proton showers. The plots in the top row show simulations done with QGSJet-II, the bottom row those with EPOS. The left and right sides show the result at different zenith angles. The profiles are obtained from showers arriving from the eastern and western direction, so that the geomagnetic field effect is at its maximum. The colored bands show the shower-to-shower fluctuation. The graph at  $r < 0$  ( $r > 0$ ) shows the variation parallel (perpendicular) to the geomagnetic field component  $\mathbf{B}_T$ , which is oriented perpendicular to the shower axis. All profiles are normalized to one at  $r = 1000$  m.

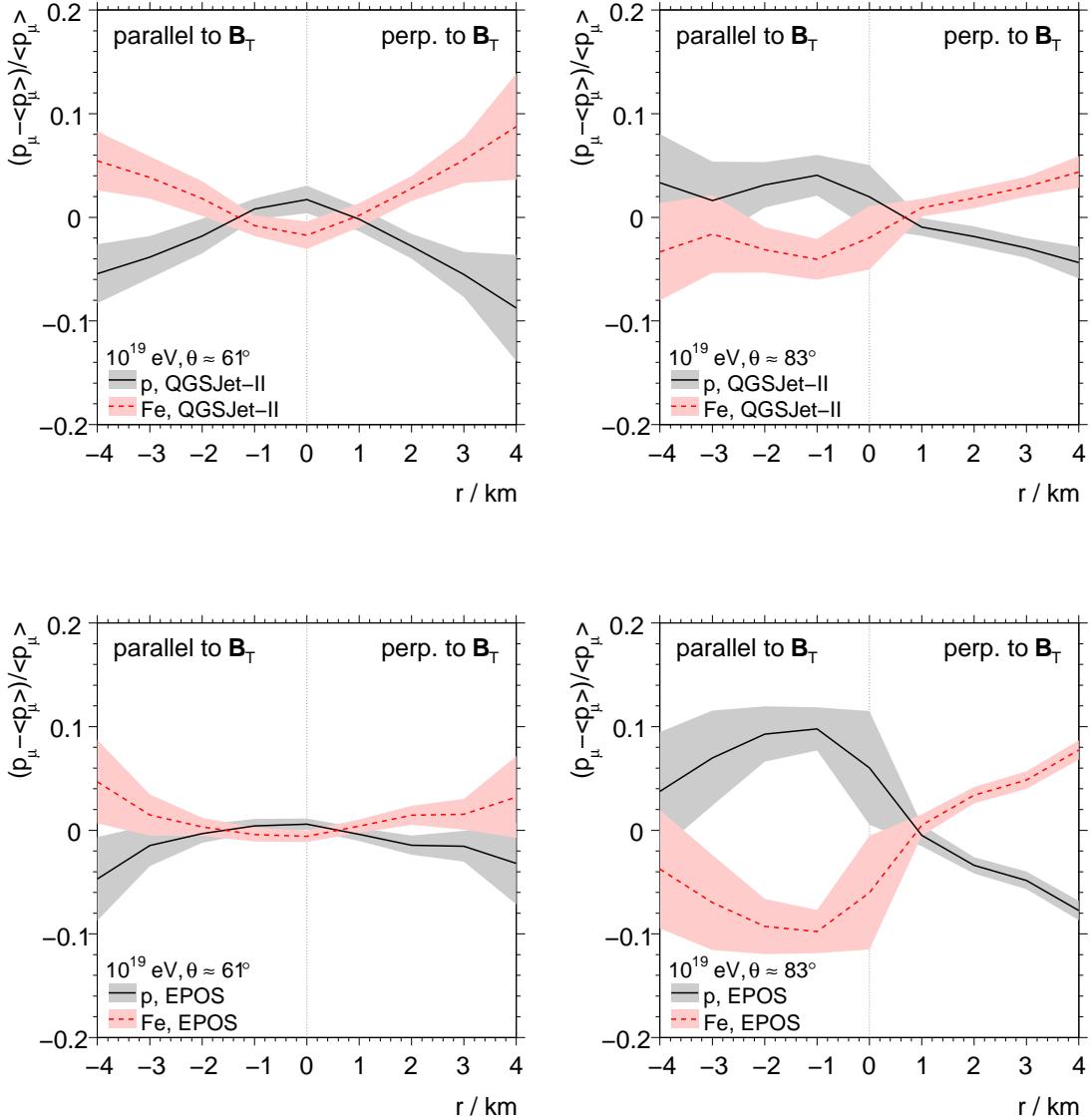


Figure 5.14: A similar analysis is shown as in Fig. 5.13, but this time the cosmic ray energy  $E$  is kept at  $10^{19} \text{ eV}$  and the cosmic ray mass  $A$  is varied.

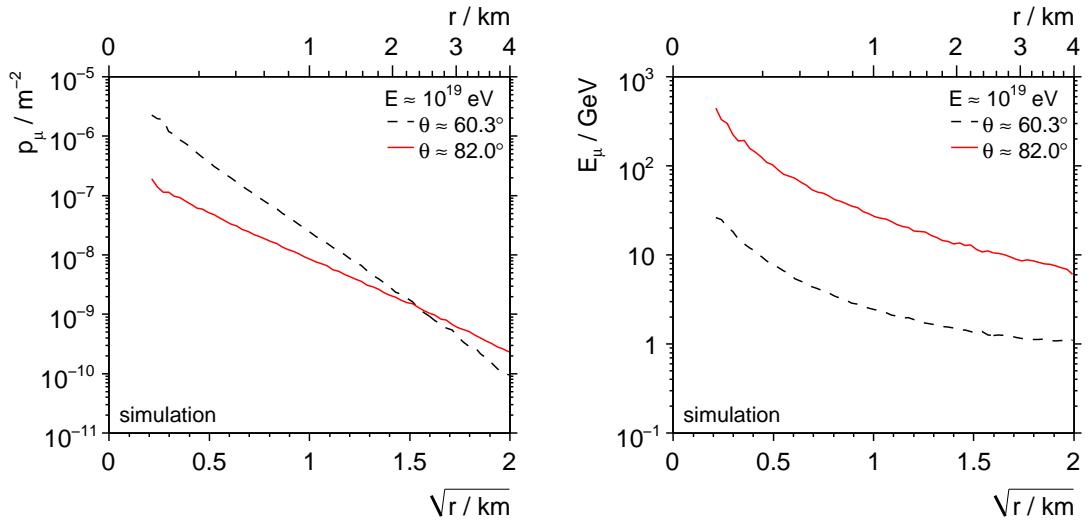


Figure 5.15: The plots show the normalized muon density (left) and the average muon energy (right) as a function of the square root of the radial distance  $r$  to the shower axis for two example proton showers simulated with QGSJet-II from Table B.1.

analytical model of the geomagnetic deflection to averaged simulated lateral profiles without geomagnetic deflections. The result is tabulated and interpolated to obtain a continuous model.

The approach presented here is not based on an analytical model of  $p_\mu$ . Instead, it is general parameterization of the simulation output that relies only on very general properties of the lateral profile, mainly that  $p_\mu$  is a smooth function with a negligible small scale structure. The parameters are derived from a fit which makes no special restrictions to the distribution of the shower parameters ( $E, \theta, \phi$ ) as long as they sufficiently cover the range of interest. Thus, the parameterization may be applied to any kind of general purpose air shower library.

The parameterization of  $p_\mu$  is done in two steps. In the first step, a parameterization based on a polynomial expansion in  $\sqrt{r}$  and a Fourier expansion in  $\psi$  is fitted to  $\lg p_\mu$ -profile of each individual shower, whereas  $(r, \psi)$  are the polar coordinates in the lateral coordinate system. It turns out, that  $\sqrt{r}$  is an ideal variable for the expansion, since the approximation  $\lg p_\mu \simeq \alpha \sqrt{r} + \text{const.}$  is already very good, as shown in Fig. 5.15. The first step leads to a set of  $k$  parameters  $\{c_k\}$  for every shower.

These parameters  $\{c_k\}$  obtained from each shower turn out to be smooth functions of the direction of the shower  $c_k = c_k(\theta, \phi)$ . Thus, in the second step, a parameterization based on a polynomial expansion in  $d_{\max}(\theta)$  and a Fourier expansion in  $\phi$  is fitted to every parameter  $c_k$ , which leads to  $l$  coefficients. It will turn out, that the distance  $d_{\max}$  between the shower maximum and the impact point of the shower on the ground is an ideal variable for this kind of expansion.

The final result is a  $k \times l$  matrix of coefficients, which gives an average continuous description of  $p_\mu(r, \psi; \theta, \phi)$ . The full procedure will be discussed in the following.

With only slight modifications, this procedure can also be used to obtain a parameterization of the average muon energy  $E_\mu(r, \psi; \theta, \phi)$ . The lateral profile of the muon energy  $E_\mu$  is interesting, because the signal  $S_\mu$  generated by a muon in a Water Cherenkov-detector has a weak energy dependency. This dependency is neglected in this study, but may be included in a future update. The discussion of the parameterization will focus on obtaining a high-quality model of  $p_\mu$ . The parameterization of  $E_\mu$  is discussed alongside to illustrate how the procedure can be generalized.

### First step: Local parameterization

The first step describes the parameterization on the level of a single simulated air shower. The simulation provides weighted particles in the ground plane. Each weighted muon carries a position and momentum

vector among other things. The algorithm is applied to each shower in the set.

The ground position of the muons are projected into the shower front plane. This plane is divided in 30 cells in  $\psi$ , and 30 cells in  $\sqrt{r}$ , ranging from 0 to  $\sqrt{4000}$  m. The normalized density in each cell  $(i, j)$  is calculated by adding the muon weight and dividing the sum by the corresponding ground area

$$A_{\text{cell},ij} = \frac{1}{2 \cos \theta} (\psi_{i+1} - \psi_i) (\sqrt{r_{j+1}^4} - \sqrt{r_j^4}), \quad (5.4.11)$$

and the total number of muons on the ground  $N_\mu$ .

The logarithm of the raw data can now be parameterized with the following expansion in  $\sqrt{r}$  and  $\psi$ :

$$\begin{aligned} \lg p_\mu &= \sum_{k=0}^3 \hat{r}^k \times \left( \sum_{j=0}^3 c_{kj}^g \cos(j\psi) + \sum_{j=1}^3 c_{kj}^u \sin(j\psi) \right) \\ &\text{with } \hat{r} = 2\sqrt{r/4000 \text{ m}} - 1. \end{aligned} \quad (5.4.12)$$

The use of the reduced variable  $\hat{r}$  optimizes the accuracy of the numerical computation of the fit, which is done with the linear least-squares method<sup>1</sup>, see *e.g.* [147]. The fit leads to 28 parameters per shower.

The use of the linear least-squares method is a key ingredient of the approach, since so many parameters need to be estimated from the data unambiguously and automatically for each shower. It shall be noted, that fitting the logarithm of a random variable instead of the variable itself introduces a bias to the fitted parameters, in analogy to the discussion in Section 5.4.1. The bias will be analyzed after the fit and eventually corrected.

The early-late asymmetry described in Chapter 3 is contained within the dipole ( $j = 1$ ) terms of the Fourier expansion, the geomagnetic deflections in the quadrupole terms ( $j = 2$ ). Both asymmetries contribute to the octopole terms ( $j = 3$ ).

The choice of the upper limits  $k_{\max} = 3$  and  $j_{\max} = 3$  of the expansions are based on the accuracy of the raw simulated data. Lower orders of the expansions do not reproduce the structure of the lateral profile with sufficient accuracy, but higher orders which correspond to short scale structures start to be dominated by statistical and artificial fluctuations of the raw data.

To obtain a parameterization of the profile of the muon energy, the same procedure is followed, but starting from a grid of average muon energies in the shower front plane. The muon energy in each cell of the grid is calculated from its content under regard of the particle weights. The logarithms of the energies are then fitted with the same parameterization.

### Second step: Global parameterization

Each parameter  $c_{kj}^{g,u}$  of the first step is now regarded as a function of  $(\theta, \phi)$ . Again, it is assumed that this dependency is slowly varying. The parameterization in  $\phi$  is done again with the first terms of a Fourier expansion. The parameterization in  $\theta$  is a polynomial expansion in  $d_{\max}(\theta)$ , the distance between the shower maximum and the shower impact point on the ground along the shower axis.

The distance  $d_{\max}$  is roughly the scale on which the geomagnetic deformations, the divergence and attenuation in the shower grow, as discussed in Chapter 3. It is therefore the ideal variable for an expansion of  $p_\mu$ .

$$\begin{aligned} c_{kj}^g &= \sum_{m=0}^5 \hat{d}(\theta)^m \left( \sum_{\ell=0}^5 \tilde{c}_{kjml}^{gg} \cos(\ell\phi) + \sum_{\ell=1}^6 \tilde{c}_{kjml}^{gu} \sin(\ell\phi) \right) \\ c_{kj}^u &= \sum_{m=0}^5 \hat{d}(\theta)^m \left( \sum_{\ell=0}^5 \tilde{c}_{kjml}^{ug} \cos(\ell\phi) + \sum_{\ell=1}^6 \tilde{c}_{kjml}^{uu} \sin(\ell\phi) \right) \\ &\text{with } \hat{d} = 2[d_{\max}(\theta) - d_{\max}(60^\circ)]/[d_{\max}(90^\circ) - d_{\max}(60^\circ)] - 1, \end{aligned} \quad (5.4.13)$$

<sup>1</sup>Numerically more efficient and more accurate methods exist to fit parameterizations, which consist entirely of orthogonal functions. However, it turns out that the chosen form is fitted very well with standard methods, and further optimization is not needed.

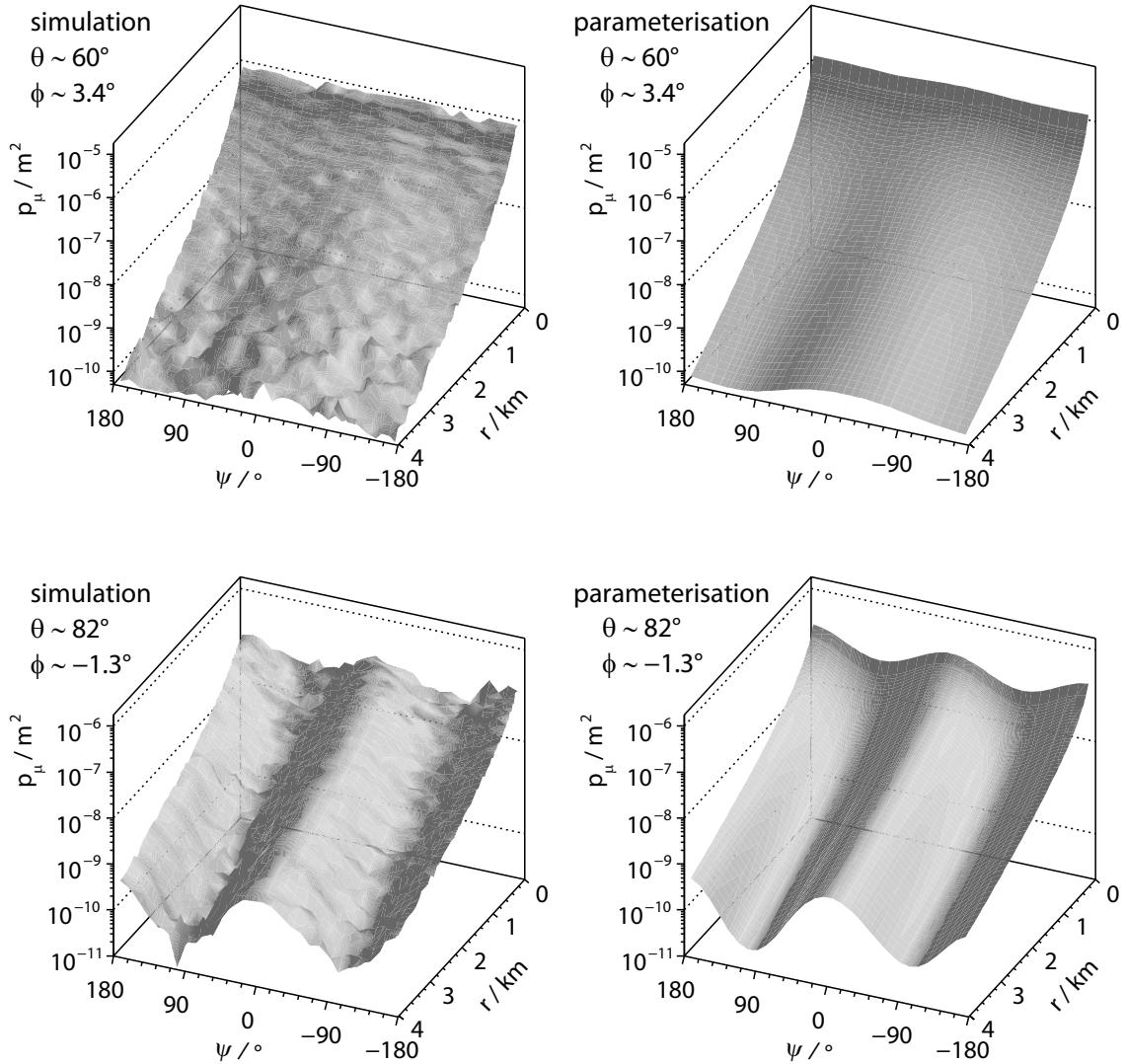


Figure 5.16: The plots show example profiles of the normalized muon density, obtained from simulated proton showers around  $10^{19}$  eV. The left side shows the raw profile from the simulation, while the right side shows the final parameterization. The geomagnetic field effect is at its maximum in both examples.

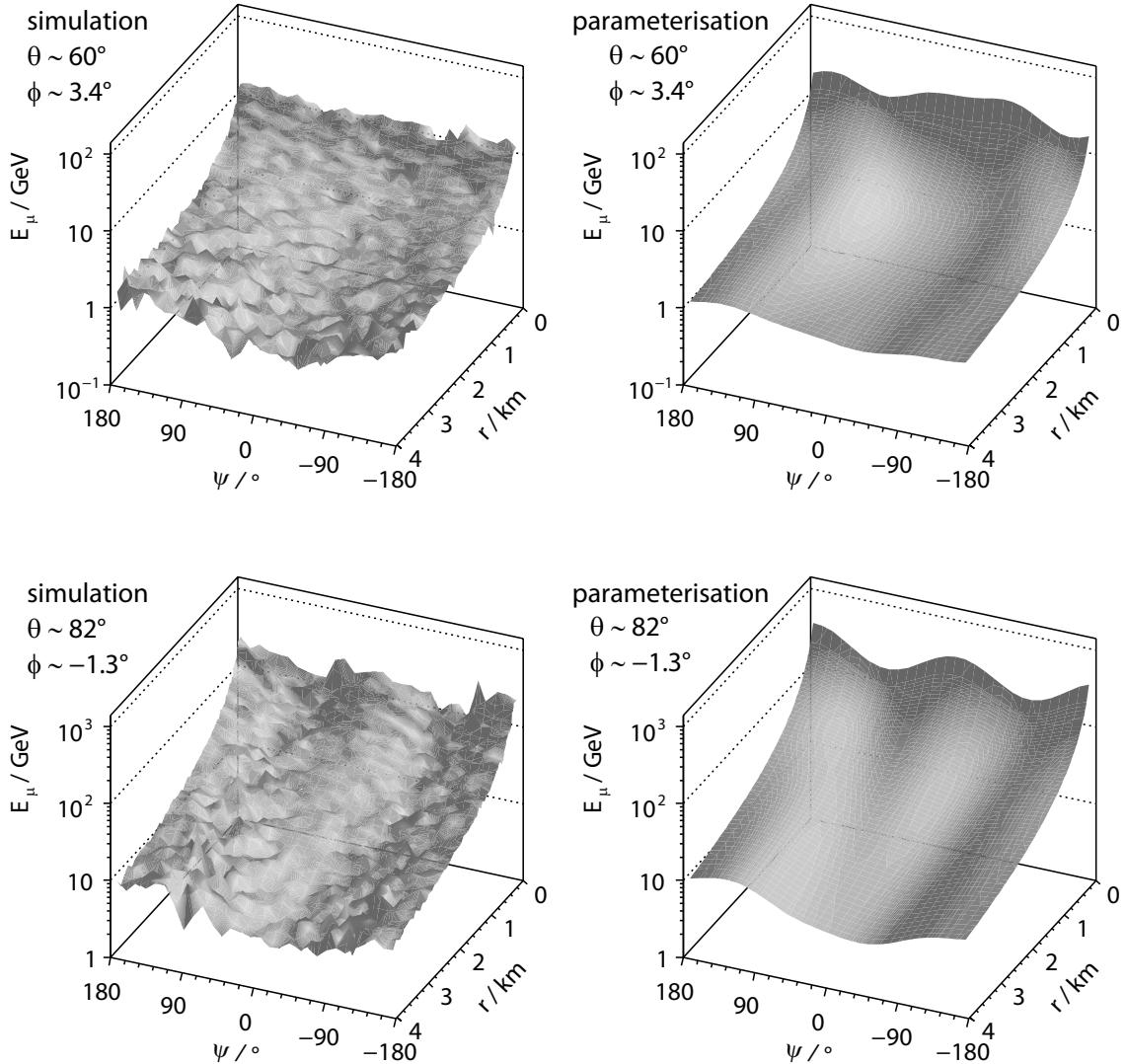


Figure 5.17: The plots show example profiles of the average muon energy, obtained from simulated proton showers around  $10^{19}$  eV. The left side shows the raw profile from the simulation, while the right side shows the final parameterization. The geomagnetic field effect is at its maximum in both examples.

whereas  $\hat{d}$  is the reduced variable. The fit is again done with a linear least-squares method, for the same reasons as in the first step.

The second parameterization of the lateral profile of the muon energy is done in the same way, but the reduced variable  $\hat{d}$  is exchanged with

$$\hat{\theta} = 2(\theta - 60^\circ)/(90^\circ - 60^\circ) - 1. \quad (5.4.14)$$

It turns out empirically that the parameterization of the energy profile works better in this variable.

The two sets of  $28 \times 66 = 1848$  coefficients

$$\tilde{c}_{kjml}^{gg}, \tilde{c}_{kjml}^{gu}, \tilde{c}_{kjml}^{ug}, \tilde{c}_{kjml}^{uu}$$

are the final result of the parameterizations. They provide a full description of the normalized lateral profile of the muon density  $p_\mu$ , and the muon energy  $E_\mu$ .

Some example profiles are shown in Fig. 5.16 and Fig. 5.17. The  $p_\mu$  profile examples show the dipole from the early-late asymmetries with a maximum at about  $\psi \approx -20^\circ$  well in the example at  $\theta \approx 60^\circ$ , while the quadrupole from the geomagnetic deflections has a clear signature in the example at  $\theta \approx 82^\circ$ .

The dipole from the early-late asymmetries is much weaker in the  $E_\mu$  profile, but it can be also spotted there: The  $E_\mu$  profile has a minimum where the  $p_\mu$  profile shows a maximum, as some low energy muons decay before reaching the ground in the late part of the shower, which in turn slightly increases the average muon energy in the late part.

The parameterization of  $p_\mu$  is more precise than the parameterization of  $E_\mu$ . Both are affected to some degree by large artificial fluctuations in the simulated profiles at small radial distances  $r < 250$  m. These fluctuations are a consequence of the radial thinning near the shower axis, which was turned on in the first production run. While the fit of the  $p_\mu$ -parameterization is quite robust against these fluctuations, but they distort the  $E_\mu$ -parameterization, as can be seen in Fig. 5.17 for the shower near  $60^\circ$ . The radial thinning should be avoided in very inclined air showers.

### Bias and distortion

The parameterization procedure is applied to proton showers simulated with QGSJet-II in the range  $60^\circ < \theta < 88^\circ$ . The result needs to be quantitatively checked against the input for possible biases, since the fit itself is not unbiased. This is done by analyzing the distribution of the residuals  $(p_\mu - \langle p_\mu \rangle)/\langle p_\mu \rangle$  of the data  $p_\mu$  around the parameterization  $\langle p_\mu \rangle$ . The average of this distribution estimates the bias, the width of the distribution the precision of the parameterization.

Getting a meaningful residual distribution is not trivial. The simulated profiles have shower-to-shower, statistical, and artificial fluctuations, which should to be separated from the distortion introduced through the parameterization. Shower-to-shower fluctuations are correlated variations over the full shower, statistical and artificial fluctuations are uncorrelated fluctuations from the counting of particles in a certain region. The artificial fluctuations are a consequence of the shower thinning during the simulation.

Multiple showers with a varying cosmic ray energy  $E$ , zenith angle  $\theta$  and azimuth angle  $\phi$  shall not averaged for the residual analysis. In this case, it is only the possible to reduce the statistical and artificial fluctuations in every individual lateral profiles. The width of the distribution of the residuals  $(p_\mu - \langle p_\mu \rangle)/\langle p_\mu \rangle$  will include the shower-to-shower fluctuations and the effect of the parameterization. If the combined result is not much larger than the observed shower-to-shower fluctuations of the normalized profile  $p_\mu$  in Fig. 5.13, the parameterization is acceptable.

The following algorithm is used to reduce the statistical and artificial fluctuations in the simulated  $p_\mu$ -profiles.

- (1) Parts of the first step of the parameterization are repeated. The particles in the shower front plane are sorted into 30 cells in  $\psi$  and 30 cells in  $\sqrt{r}$ , ranging from 0 to  $\sqrt{4000}$  m, and the  $p_\mu$ -profile is calculated. Instead of a fit, a second grid of the same form is filled with the prediction from the parameterization (it is evaluated at the bin centers).
- (2) Pairs of adjacent cells in  $\psi$  are averaged once, reducing the number of cells in  $\psi$  to 15. Then, for each direction in  $\psi$ , adjacent cells in  $\sqrt{r}$  are merged and the contents averaged, until the merged cell

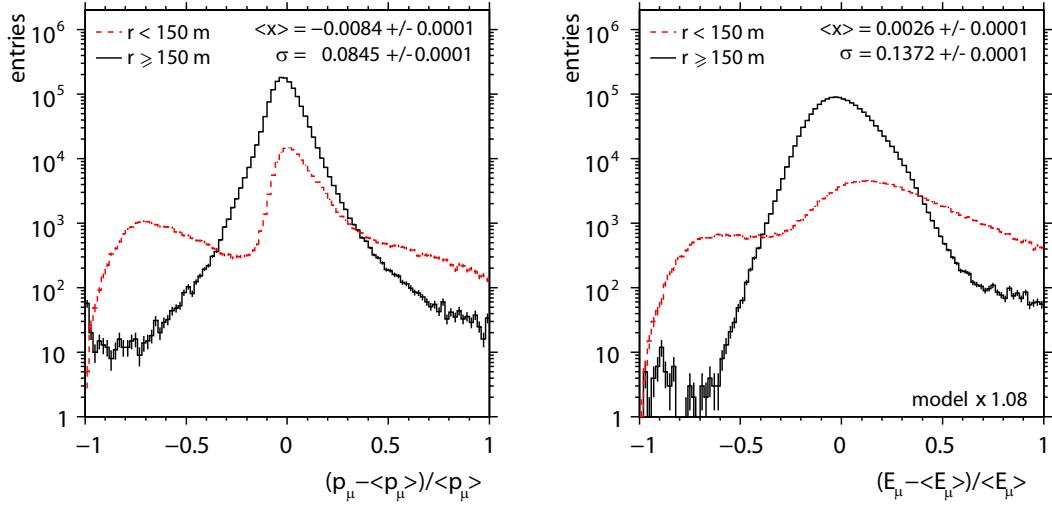


Figure 5.18: The left plot shows the residual distribution of the normalized lateral profile of the muon density  $p_\mu$  from proton showers simulated with QGSJet-II around the parameterization. The right plot shows the same for the lateral profile of the muon energy  $E_\mu$ , respectively. The parameterization of the latter was multiplied with the factor 1.08 to compensate its bias. Residuals from cells closer than 150 m to the shower axis are dominated by large artificial fluctuations and are thus excluded from the analysis.

contains more than 400 weighted particles. This is done starting from  $r = 0$  and working outwards. After this procedure, most remaining cells contain around 400 weighted particles, except for a few left-over cells at large  $r$ , which could not grow enough. The latter are not used for the comparison between the simulated showers and the parameterization.

- (3) The second grid that holds the parameterization is merged in the exact same way as the first one, so that the grids have pairs of cells of equal extensions. Finally, the residual is calculated for each pair of corresponding cells, and saved together the radius  $r$  at the center of the cell.

The algorithm assures that the statistical and artificial fluctuations are reduced to a level of about 5 %. The application to the profile of the muon energy  $E_\mu$  is analogue.

Fig. 5.18 shows the results of this analysis, after its application to the full input set of proton showers simulated with QGSJet-II. The residuals obtained from cells closer and farther than 150 m from the shower axis are separated. Only the latter are meaningful, since the former are dominated by the large artificial fluctuations through the radial thinning in most showers of the input set.

The analysis shows, that the parameterization of  $p_\mu$  shows a negligible bias smaller than 1 %, which is an excellent result. The parameterization of  $E_\mu$  on the hand is biased: the average of the residual distribution is off by about 8 %, if the parameterization is not modified. Fig. 5.18 shows the residual distribution with respect to the scaled parameterization of  $E_\mu$ , where the energies are multiplied by a factor of 1.08 to compensate the overall bias.

The residual distributions are not a Gaussian. Most residuals show a small variance, but there are also rare large deviations. The width of the distribution is estimated by the square root of the variance. In case of the  $p_\mu$ -parameterization, one obtains a width of 8 %, which is at the level of universality of  $p_\mu$ . This is an optimal result and shows that the parameterization approach introduces only negligible distortions.

In case of the  $E_\mu$ -profile, the square root of the variance is 14 %. This is not optimal, but still satisfactory for a side product.

#### 5.4.4 Comparison with another model

The model of the lateral profile of the muon density  $n_\mu$  developed in this work is based on a phenomenological parameterization of fully simulated profiles. This model, called model *B*, shall be compared with a

semi-analytical approach [10, 13, 54] from the literature, called model A.

Model A is based partially on air shower simulations and partially on an analytical model of the lateral muon profile. The analytical approach allows to understand the structure of the geomagnetic deformations of the profile and a part of the derivation of model A is repeated in Chapter 3. Model A reproduces most of the structure found in full air shower simulations remarkably well.

On the other hand, model A does not include some effects in its analytic approach that are present in the full simulation:

- The early-late asymmetries in the profile of the muon density are neglected.
- The muons are formed in two body decays of mesons. The distance of the mesons to the shower axis of some 100 m is neglected, instead the muons are assumed to originate from a point on the shower axis.
- The decay angle between the direction of the parent meson and the generated muon is neglected.

These effects were already discussed in Chapter 3. Model B automatically includes these effects and reproduces the full simulation to a level of 10 %. Model A does not achieve this level of agreement with the simulation input everywhere [10], because of these neglected effects.

Model A also uses another simulation setup than model B. Model A is based on the AIRES air shower program. QGSJet01 is used for hadronic interactions at high energy and the enhanced splitting algorithm of AIRES for low energies, which is tuned to GHEISHA. The simulation setups of model A and B are not too far from the two configurations A and B in the comparison in Section 5.3.2 and similar differences in the profiles of the muon density  $n_\mu$  can be expected.

Fig. 5.19 to Fig. 5.22 compare the two models. The global offset in the produced number of muons between QGSJet01 and QGSJet-II is of no interest for comparison and compensated by scaling model A down by a factor of 0.9. The profiles in a) and b) show the expected number of muons in a station of the surface detector as predicted by the models in the lateral coordinate system. The solid, dashed, and dotted contour lines represent density contours representing 100, 10, and 1 muon per surface station. Model A has a left-right symmetry in contrast to model B, because the early-late asymmetries are not included.

Shown in c) is the relative difference of the model predictions, in d) this difference is compared with the Poisson uncertainty expected for the average model. The difference in units of the expected Poisson uncertainty allows to estimate the severity of the difference in a fit of the profile to experimental data. For example, a relative difference of a factor of two is very significant if the average number of muons per station is 100, but not if the average number of muons is 0.1. Relative differences close to the shower axis are more emphasized in d) than in c), because of the small statistical uncertainties in this region.

Overall, differences up to +50 % and down to more than -100 % are observed. In terms of the Poisson statistic, the differences are larger than  $\pm 1\sigma$  in some places. Model A shows a steeper decline at a function of the radial distance from the shower axis than model B, which is caused by the different low energy hadronic interaction models used for the simulation inputs of both models. Model B has a dipole which is missing in model A and caused by the early-late asymmetries. Slight differences in the quadrupole structure of model A and B are visible at 84° and probably due to neglected effects in the modeling of the geomagnetic deflections in model A.

In conclusion, the differences of up to 100 % between model A and B are significant and will affect the event reconstruction of very inclined air showers in Chapter 6. The differences between model A and B are much larger than the differences between model B and its simulation input. The differences caused by the different approaches used to derive model A and B seem comparable to those caused by the differences in the simulation input.

A direct comparison of the models gives a first impression of the magnitude of the differences, but most meaningful is a comparison of the performance of both models in the reconstruction of simulated and real events. This comparison is done in Chapter 6.

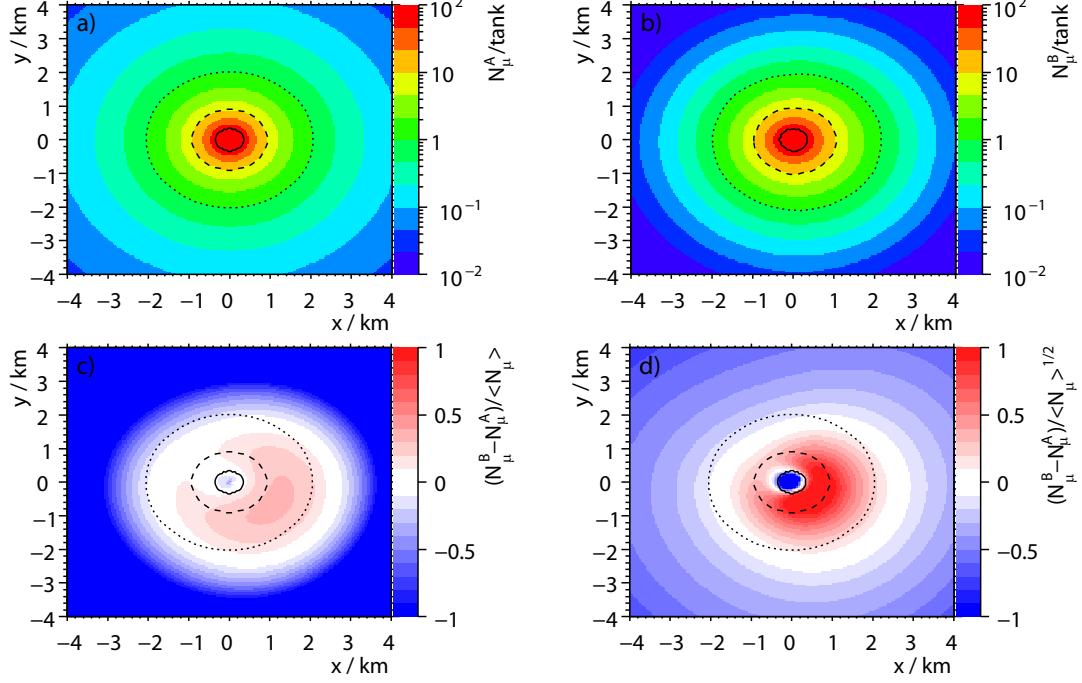


Figure 5.19: The profiles show the muon density predicted a) by model A [10, 13, 54] and b) by model B (this work) at  $10^{19}$  eV,  $\theta = 60^\circ$ , and  $\phi = 0^\circ$ . The profiles in c) and d) show the relative difference (see text). Model B is scaled down by a factor of 0.9.

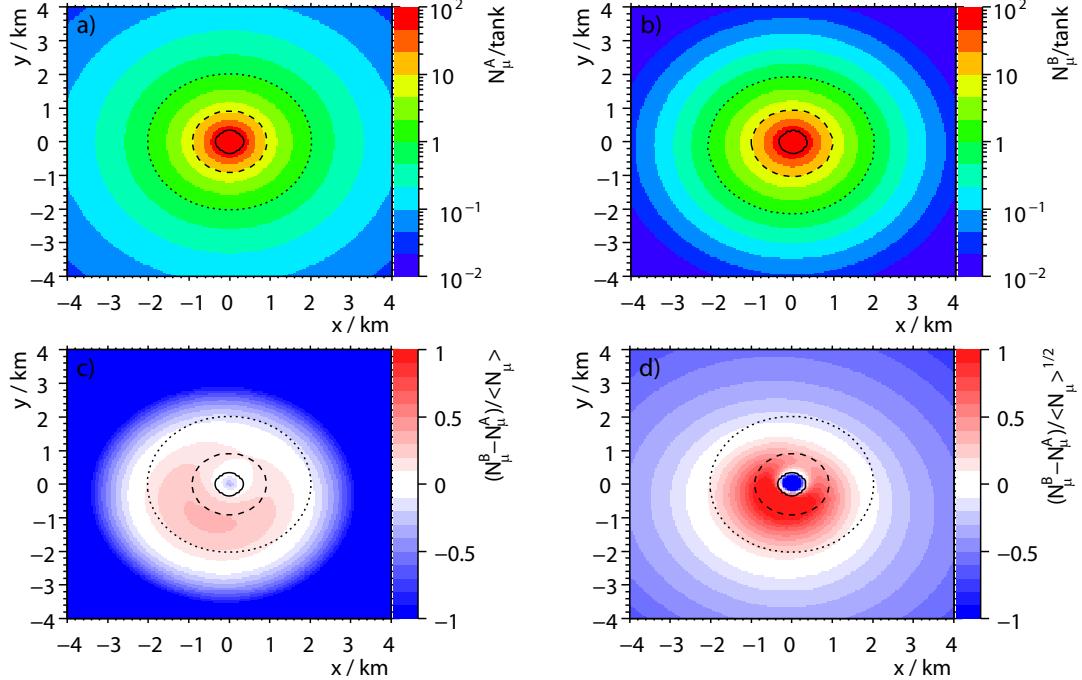


Figure 5.20: The profiles show the muon density predicted a) by model A [10, 13, 54] and b) by model B (this work) at  $10^{19}$  eV,  $\theta = 60^\circ$ , and  $\phi = 90^\circ$ . The profiles in c) and d) show the relative difference (see text). Model B is scaled down by a factor of 0.9.

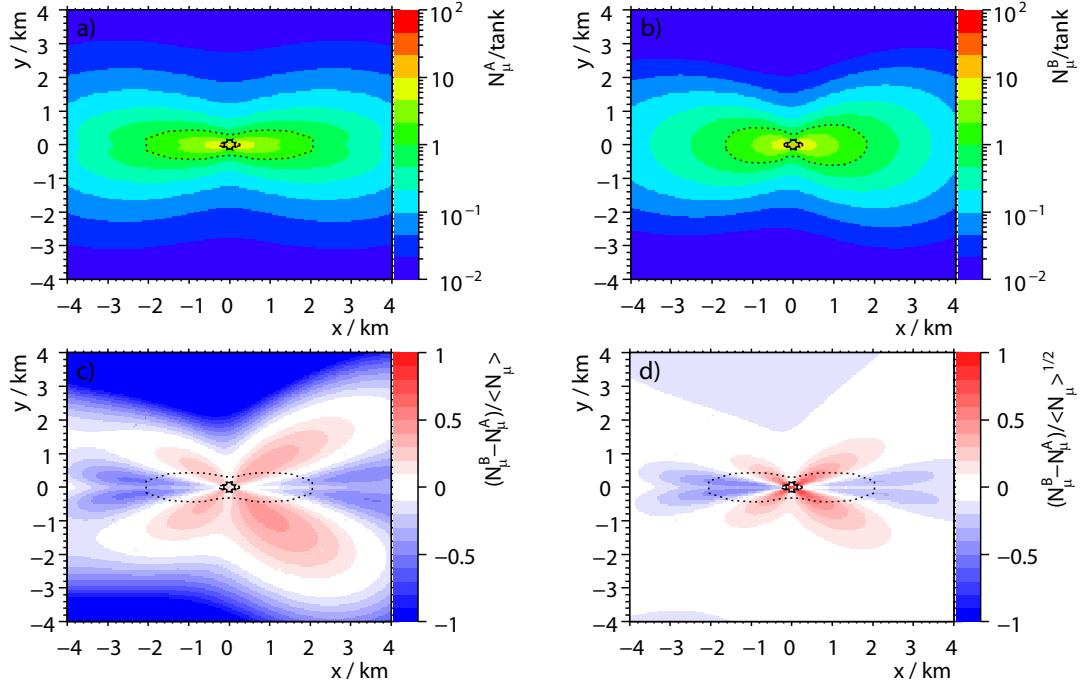


Figure 5.21: The profiles show the muon density predicted a) by model A [10, 13, 54] and b) by model B (this work) at  $10^{19}$  eV,  $\theta = 84^\circ$ , and  $\phi = 0^\circ$ . The profiles in c) and d) show the relative difference (see text). Model B is scaled down by a factor of 0.9.

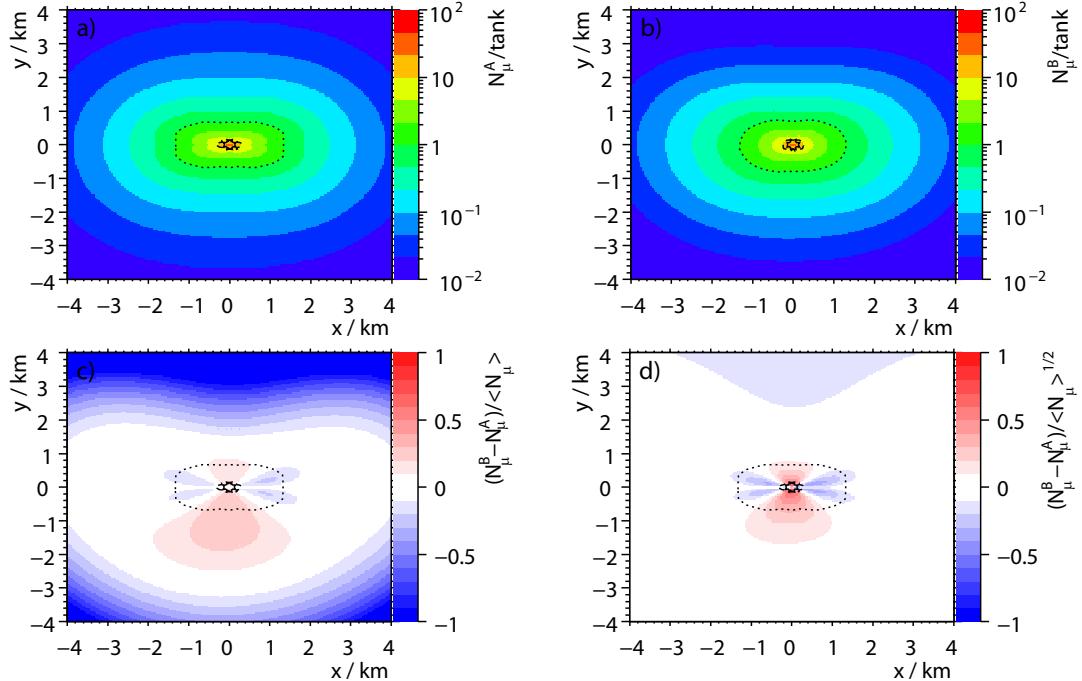


Figure 5.22: The profiles show the muon density predicted a) by model A [10, 13, 54] and b) by model B (this work) at  $10^{19}$  eV,  $\theta = 84^\circ$ , and  $\phi = 90^\circ$ . The profiles in c) and d) show the relative difference (see text). Model B is scaled down by a factor of 0.9.

## 5.5 Surface detector response to very inclined air showers

The surface detector samples the front of an air shower at several points on the ground. The SD signals in very inclined air shower are dominated by the contribution of muons. Therefore, the SD basically measures the lateral profile of the muon density  $N_\mu$  on several spots on the ground. With the model of the muon density  $n_\mu$  on the ground from the previous section, it is possible to reconstruct the total number of muons  $N_\mu$  on the ground from these measurements. The total number of muons  $N_\mu$  is used as an estimator of the cosmic ray energy  $E$  in Chapter 6.

In order to reconstruct the total number of muons  $N_\mu$  on the ground with this approach, it is necessary to understand and model the SD signal response to the particle mix in the front of a very inclined air shower. The model has to relate the measured signal  $S$  in a station with the number of muon hits  $k$  and regard the residual contributions of electrons and photons to the signal. The basic advantage in the reconstruction of very inclined air showers is the fact that the contributions from electromagnetic particles are small and can be treated approximately, while the main contribution to the measured signals stems from the muons. The signal response of the Auger Water-Cherenkov detectors to muons can be modeled very well.

In the following, a signal response model is derived from the library of simulated SD events which was presented earlier in this chapter. The derivation of the model follows the work of ref. [11, 13, 15–17] closely. A slight modification of the standard ansatz will improve the accuracy of the model in its application to the reconstruction of events in Chapter 6.

### 5.5.1 Properties of the Auger Water-Cherenkov detector

The signal response of a Water-Cherenkov detector to a fast charged particle is a function of the track length  $l$  that the particle travels in the water volume with a velocity larger than the speed of light in water. The track lengths  $l$  and therefore the generated signal varies from particle to particle, partly due to the geometry of the detection volume and partly due to the de-acceleration of particles in water.

The signals generated by two or more simultaneous particles in the detector add up. Cherenkov photons from different particles have a random phase relation and show no interference. The SD stations of the Pierre Auger Observatory have a linear response to Cherenkov light over a large dynamic range.

The signals generated in the surface detector by very inclined air showers are dominated by the contributions of muons, but also the response to electrons and photons needs to be regarded. The contribution of hadrons can be neglected [56].

- **Muons.** The muons arrive in parallel over the extension of a station. The average muon signal is proportional to the geometrical track length  $l_\mu$  of the muon in the detector in very good approximation. Muons in the energy range  $1.5 \text{ GeV} \lesssim E_\mu \lesssim 500 \text{ GeV}$  [16] fully penetrate the SD station and maintain 100 % production efficiency for Cherenkov light, but do not generate secondary particles via bremsstrahlung or direct  $e^+e^-$  pair production in the water. The average muon energy on the ground in very inclined air showers ranges between 1 GeV and 100 GeV, depending on the zenith angle  $\theta$  of the shower.
- **Electrons.** The average energy of electrons on the ground is at least by two orders of magnitude smaller than the muon energy. The production threshold for Cherenkov light is 0.8 MeV. The ionization energy loss in the water of about  $2 \text{ MeV cm}^{-1}$  stops electrons up to an energy of about 0.1 GeV. Cherenkov light is generated with 100 % efficiency over the full track in good approximation. Therefore, electrons up to about 0.1 GeV generate a signal proportional to their energy in good approximation. Electrons with energies larger than the critical energy in water of about 0.08 GeV [23] produce small sub-showers in the water and the relation between electron energy and signal becomes more complex, but in first approximation it is still a proportionality.
- **Photons.** A photon needs to be converted into an electron/positron pair to generate any Cherenkov light. The radiation length in water is 0.36 m [23] and a conversion therefore likely. The Cherenkov production threshold for photons is about 1.3 MeV [16], which is equal to the sum of the rest masses of two electrons and the kinetic energy for at least one electron to be above the Cherenkov threshold.

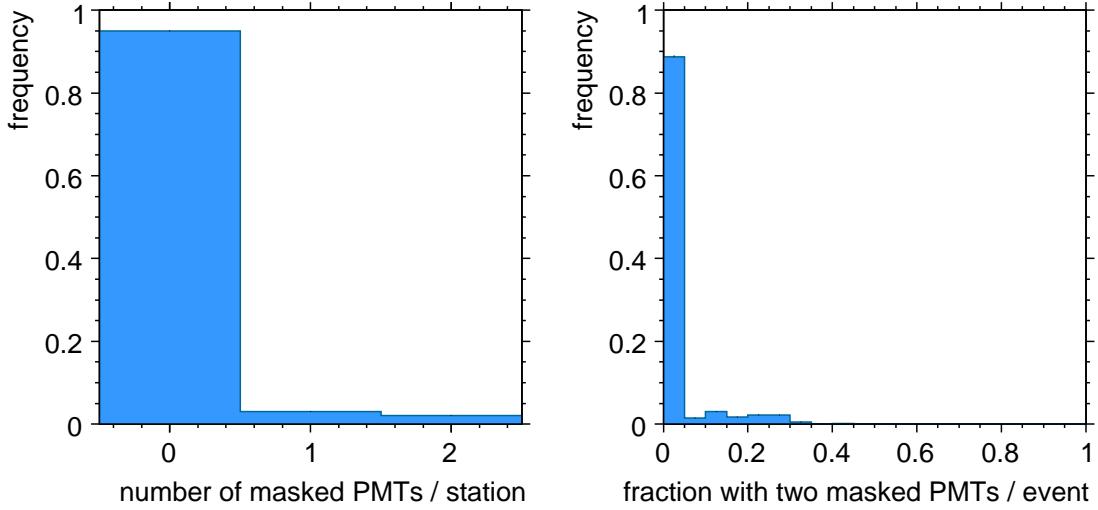


Figure 5.23: The left plot shows the frequency of masked PMTs in real events. A masked PMT is a faulty PMT. Its signal is not used for the station level triggers or in the event reconstruction.

The signal response of the Auger Water-Cherenkov is independent of the azimuth angle  $\phi$  of the incident particle in good approximation. The station design assures, that most Cherenkov photons suffer several reflections before hitting a PMT, so that the information about the initial direction of the photons is lost. This randomization is very efficient if the particles have a moderate inclination.

In very inclined air showers, a particle generated Cherenkov cone can touch a PMT directly, which amplifies the measured signal and introduces a dependency on the azimuth  $\phi$  of the particle on the level of a single PMT in the station. A  $\phi$ -variation  $\alpha \approx 15\%$  in the average signal response to muons is observed in simulations [15]. The total signal response of a station is the average of its three individual PMTs. The  $\phi$ -variation on the PMT level is well described by a cosine law, which cancels in the calculation of the average

$$\begin{aligned} S_\mu &\equiv \frac{1}{3}(S_{\mu,\text{PMT-1}} + S_{\mu,\text{PMT-2}} + S_{\mu,\text{PMT-3}}) \\ &= \frac{1}{3}((1 + \alpha \cos(\phi_\mu))S_\mu + (1 + \alpha \cos(\phi_\mu + 120^\circ))S_\mu + (1 + \alpha \cos(\phi_\mu + 240^\circ))S_\mu) = S_\mu \end{aligned}$$

If one of three PMTs is not operational, the  $\phi$ -variation of the signal response is still smaller than 5 %. Fig. 5.23 shows that most stations have all three PMTs in operation.

### 5.5.2 Modeling the signal response

The signal generated by the front of a very inclined air shower in an SD station now is modeled quantitatively, based on the previous discussion of the properties of the Auger Water-Cherenkov detector. The event reconstruction in Chapter 6 needs the full probability density function (p.d.f.) of the signal response to  $k$  muons.

In general, the signal  $S$  in an SD station is the sum of a muon contribution  $S_\mu$  and two kinds of contributions from electromagnetic particles

$$S = S_\mu + S_{\text{em}} + S_{\text{em-}\pi^0}, \quad (5.5.1)$$

whereas  $S_{\text{em}}$  is the contribution from electromagnetic particles generated in earlier muon decays and  $S_{\text{em-}\pi^0}$  the contribution from electromagnetic particles which originate from the hadronic cascade.

The muon-generated electromagnetic particles make the first contribution. They are called halo particles and were already discussed in Chapter 3. The energy spectrum of the halo particles and their density

relative to the muons are almost constant. Therefore, there is an almost fixed relation between the average muon signal  $\langle S_\mu \rangle$  and the average halo signal  $\langle S_{\text{em}} \rangle$

$$\frac{\langle S_{\text{em}} \rangle}{\langle S_\mu \rangle} \approx 0.2 \quad (5.5.2)$$

On average, the halo particles form a constant signal background for the muons.

While the initial energy of a halo particle is of the order of the energy of its parent, its energy is lost quickly in a small electromagnetic sub-shower. Therefore, the average energy of the halo particles is orders of magnitude lower than the average muon energy. The number of halo electrons is comparable to the number of muons. The number of photons above one MeV is even an order of magnitude larger. The signal ratio  $\langle S_{\text{em}} \rangle / \langle S_\mu \rangle$  is still small, because the muons with their larger energy generate Cherenkov light over their full geometrical track length in the detector volume, while the less energetic halo particles have shorter tracks.

The remaining electromagnetic particles from the hadronic cascade are not closely related to the muon density. Ideally, their contribution  $S_{\text{em-}\pi^0}$  to the total signal  $S$  is zero, which is the case at zenith angles  $\theta \gtrsim 65^\circ$ . At  $60^\circ < \theta < 65^\circ$ , the contribution is already small. The signal response model is based on the requirement, that  $S_{\text{em-}\pi^0}$  makes only a small contribution to the signal so that it is not necessary to treat this contribution in detail.

The signal components are random variables. Each has a corresponding p.d.f., which can be parameterized in general as

$$f_\mu(S_\mu | \theta_\mu, \phi_\mu, E_\mu, k) \quad (5.5.3)$$

$$f_{\text{em}}(S_{\text{em}} | E_\mu, \theta_\mu, n_\mu) \quad (5.5.4)$$

$$f_{\text{em-}\pi^0}(S_{\text{em-}\pi^0} | r, \psi, \theta, \phi, E, A), \quad (5.5.5)$$

whereas  $\theta_\mu, \phi_\mu, E_\mu$  are the average direction and energy of the  $k$  muons hitting the station;  $(r, \psi)$  is the position of the station relative to the shower axis in the lateral coordinate system;  $(\theta, \phi), E, A$  are the direction, energy and mass of the cosmic ray.

Only  $S_\mu$  depends directly on the number of muon hits  $k$ . The electromagnetic signal  $S_{\text{em}}$  is tightly correlated to  $k$ , but actually a function of  $n_\mu$ . Both  $S_\mu$  and  $S_{\text{em}}$  are completely determined by the properties of the muon front close to the ground, but the dependency of  $S_{\text{em-}\pi^0}$  needs to be traced back to the primary properties of the cosmic ray.

The general p.d.f. of the total signal response to  $k$  muons is the convolution of these p.d.f.s:

$$\begin{aligned} f(S) &= f(S | \theta_\mu, \phi_\mu, E_\mu, k; r, \psi, \theta, \phi, E, A) \\ &= \int_0^\infty d\tilde{S}_{\text{em}} \int_0^\infty dS_{\text{em-}\pi^0} f_\mu(S - \tilde{S}_{\text{em}}) f_{\text{em}}(\tilde{S}_{\text{em}} - S_{\text{em-}\pi^0}) f_{\text{em-}\pi^0}(S_{\text{em-}\pi^0}) \end{aligned} \quad (5.5.6)$$

Modeling the complete p.d.f. is a complex task and has so far not been done. Instead, several approximations are used to reduce the general model to practical proportions.

The p.d.f.  $f_\mu$  of the muon signal is reduced to a function of  $\theta_\mu$  and the number of muons hits  $k$ . The dependency on  $\phi_\mu$  cancels in very good approximation and the dependency on  $E_\mu$  is very weak, as discussed earlier. A study of simulated SD events finds a  $E_\mu$ -dependency of the muon signal  $S_\mu$  at the level of about 5 % [150] in zenith angle range  $60^\circ < \theta < 80^\circ$ , which is neglected in this study.

The remaining p.d.f.  $f_\mu(S_\mu | \theta_\mu, k)$  of  $k$  muon hits follows from the p.d.f.  $g(S_\mu | \theta_\mu) \equiv f(S_\mu | \theta_\mu, 1)$  of a single muon hit via auto-convolution:

$$\begin{aligned} S_\mu^k &= S_{\mu,1}^1 + S_{\mu,2}^1 + \cdots + S_{\mu,k-1}^1 + S_{\mu,k}^1 \\ \Rightarrow f_\mu(S_\mu | \theta_\mu, k) &= \int ds_{k-1} \int ds_{k-2} \cdots \int ds_2 \int ds_1 \times \\ &\quad g(S_\mu - s_{k-1} | \theta_\mu) g(s_{k-1} - s_{k-2} | \theta_\mu) \cdots g(s_2 - s_1 | \theta_\mu) g(s_1 | \theta_\mu). \end{aligned} \quad (5.5.7)$$

Thus, it is sufficient to model  $g(S_\mu | \theta_\mu)$ .

The approximations towards  $f_{\text{em}}$  and  $f_{\pi^0\text{-em}}$  are radical in the standard approach. They are replaced by delta-functions

$$\begin{aligned} f(S) &\simeq \int_0^\infty d\tilde{S}_{\text{em}} \int_0^\infty dS_{\text{em-}\pi^0} f_\mu(S - \tilde{S}_{\text{em}}) \delta(\tilde{S}_{\text{em}} - S_{\text{em-}\pi^0}, \langle S_{\text{em}} \rangle) \delta(S_{\text{em-}\pi^0}, \langle S_{\text{em-}\pi^0} \rangle) \\ &= f_\mu(S - \langle S_{\text{em}} + S_{\text{em-}\pi^0} \rangle) \simeq \frac{1}{1 + \langle \epsilon \rangle} f_\mu \left( \frac{S}{1 + \langle \epsilon \rangle} \right), \end{aligned} \quad (5.5.8)$$

whereas  $\langle \epsilon \rangle \equiv \langle S_{\text{em}} + S_{\text{em-}\pi^0} \rangle / \langle S_\mu \rangle$  is the average contribution from electromagnetic particles. The following approximation is used:

$$S_\mu \approx S - \langle S_{\text{em}} + S_{\text{em-}\pi^0} \rangle \approx S / (1 + \langle \epsilon \rangle). \quad (5.5.9)$$

The factor  $1/(1 + \langle \epsilon \rangle)$  needs to be introduced to preserves the normalization of  $f(S)$  after the approximation.

The correct average value of  $f(S)$  is obtained with this approach

$$\begin{aligned} \langle S \rangle &= \int_0^\infty dS S f(S) = \frac{1}{1 + \langle \epsilon \rangle} \int_0^\infty dS S f_\mu \left( \frac{S}{1 + \langle \epsilon \rangle} \right) \\ &= (1 + \langle \epsilon \rangle) \int_0^\infty dS' S' f_\mu(S') \\ &= (1 + \langle \epsilon \rangle) \langle S_\mu \rangle = \langle S \rangle, \end{aligned} \quad (5.5.10)$$

but fluctuations of  $f(S)$  will be too small. The discussion will return to this issue later.

The ansatz of using  $\langle \epsilon \rangle$  implies a scaling of both  $\langle S_{\text{em}} \rangle$  and  $\langle S_{\text{em-}\pi^0} \rangle$  with the number of muon hits  $k$ :

$$\begin{aligned} \langle \epsilon \rangle &= \frac{\langle S_{\text{em}}^k + S_{\text{em-}\pi^0}^k \rangle}{\langle S_\mu^k \rangle} = \frac{\langle S_{\text{em}}^k \rangle}{\langle S_\mu^k \rangle} + \frac{\langle S_{\text{em-}\pi^0}^k \rangle}{\langle S_\mu^k \rangle} = \frac{k \langle S_{\text{em}}^1 \rangle}{k \langle S_\mu^1 \rangle} + \frac{\langle S_{\text{em-}\pi^0}^k \rangle}{k \langle S_\mu^1 \rangle} \\ &\stackrel{!}{=} \frac{\langle S_{\text{em}}^1 \rangle}{\langle S_\mu^1 \rangle} + \frac{\langle S_{\text{em-}\pi^0}^1 \rangle}{\langle S_\mu^1 \rangle} \quad \Rightarrow \quad \langle S_{\text{em-}\pi^0}^k \rangle \stackrel{!}{=} k \langle S_{\text{em-}\pi^0}^1 \rangle \end{aligned} \quad (5.5.11)$$

This feature was already justified for  $\langle S_{\text{em}} \rangle$ , but it is not apparent for  $\langle S_{\text{em-}\pi^0} \rangle$ . Very roughly, one may argue that the number of muon hits  $k$  in a station is almost proportional the cosmic ray energy  $E$ , while the same is true for  $\langle S_{\text{em}} \rangle$ :

$$k \propto n_\mu \propto E, \quad \langle S_{\text{em-}\pi^0} \rangle \propto E \quad \Rightarrow \quad \langle S_{\text{em-}\pi^0} \rangle \propto k.$$

This approximation is very rough for several reasons. Neither the muon component nor the electromagnetic component of an air shower scale exactly with  $E$ . Furthermore, the depth  $X_{\text{max}}$  of the electromagnetic shower maximum increases with  $\ln E$  as shown in Chapter 3, so that the electromagnetic component is slight less attenuated at higher energies.

The signal ratio  $\langle \epsilon \rangle$  is only a function of the cosmic ray energy  $E$  and mass  $A$  because of the contribution of  $\langle S_{\text{em-}\pi^0} \rangle$ . Neglecting these dependencies is part of the ansatz and only a good approximation because  $\langle S_{\text{em-}\pi^0} \rangle$  has a small impact in the reconstruction of very inclined air showers. The impact of these approximations will be evaluated later.

The remaining dependencies of the signal ratio  $\langle \epsilon \rangle$  can be reduced to the position  $(r, \psi)$  of the SD station in the lateral coordinate system and the direction  $(\theta, \phi)$  of the shower

$$\langle \epsilon \rangle \simeq \langle \epsilon \rangle(r, \psi; \theta, \phi). \quad (5.5.12)$$

The dependencies of  $\langle \epsilon \rangle$  on  $E_\mu$  and  $\theta_\mu$  are implicitly regarded. Both  $E_\mu$  and  $\theta_\mu$  are approximately universal and therefore only functions of the remaining parameters  $(r, \psi; \theta, \phi)$ .

This is the approach developed successively in ref. [11, 13, 15–17]. The approximations may seem rough, but the reconstruction of simulated events in Chapter 6 will show, that they work out surprisingly well in practice.

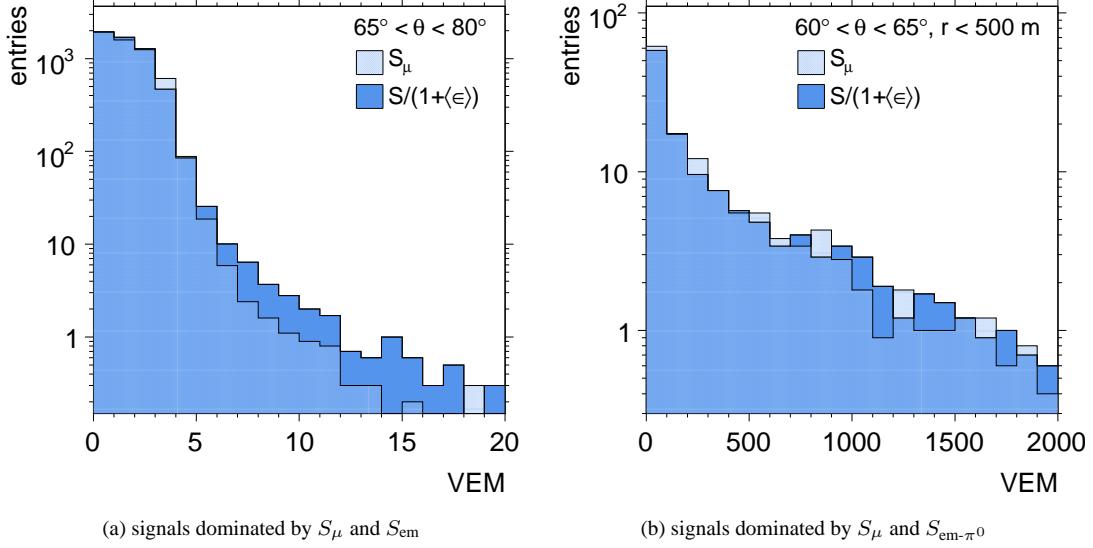


Figure 5.24: a) The histograms compare the simulated distribution of the signal response  $S^1/(1 + \langle\epsilon\rangle)$  to single muons with the distribution of the muon signal  $S_\mu^1$  alone. Cuts on the zenith angle assure, that the contribution of  $S_{\text{em-}\pi^0}$  is negligible. b) The histograms compare the simulated distribution of the signal response  $S/(1 + \langle\epsilon\rangle)$  to any number of muons with the muon signal  $S_\mu$  alone. Cuts on the parameter region assure, that the total signal  $S$  is dominated by  $S_\mu$  and  $S_{\text{em-}\pi^0}$ .

### Fluctuations of the electromagnetic signal components

The discussion shall now return to the neglected fluctuations in Eq. (5.5.8). Approximating  $f_{\text{em}}$  and  $f_{\pi^0\text{-em}}$  by delta-functions neglects fluctuations of the total signal  $S$  caused by the electromagnetic signals  $S_{\text{em}}$  and  $S_{\text{em-}\pi^0}$ . These are good approximations only if  $\sigma[S_\mu] \gg \sigma[S_{\text{em}}]$  and  $\sigma[S_\mu] \gg \sigma[S_{\text{em-}\pi^0}]$ .

The surface detector simulation in the Auger Offline-framework keeps separate records for FADC counts generated by different particle species. The total signal  $S$ , as well as the components made by electrons, muons, and photons may be analyzed separately. This is used in Fig. 5.24 to compare the distribution of the muon signal  $S_\mu$  alone with the distribution of the total signal  $S$  corrected by the average electromagnetic component  $(1 + \langle\epsilon\rangle)$  in simulated events. Both distributions have the same mean, but the distribution of  $S/(1 + \langle\epsilon\rangle)$  also includes the fluctuations caused by the signal contributions from electromagnetic particles. Some cuts on the simulated data allow to compare two regimes. In Fig. 5.24a),  $S_{\text{em-}\pi^0}$  is negligible and the additional fluctuations are caused only by  $S_{\text{em}}$ . In Fig. 5.24b), the contribution from  $S_{\text{em-}\pi^0}$  is relevant or even dominant, as will be shown later.

The analysis shows that the fluctuation of the total signal is indeed dominated by the fluctuation of the muon signal, but it is also apparent that the other contributions are not negligible. The fluctuations of  $S_{\text{em}}$  and  $S_{\text{em-}\pi^0}$  are neglected in the standard approach. It is proposed now to approximately include these fluctuations by re-interpreting Eq. (5.5.8) as the definition of how to construct  $f_\mu(S_\mu)$ . By defining the muon signal as

$$S_\mu \equiv \frac{S}{1 + \langle\epsilon\rangle}, \quad (5.5.13)$$

and building  $f_\mu$  from the distribution of this random variable, the fluctuations of the total signal due to the electromagnetic background are included into the p.d.f.  $f_\mu$ . This ansatz also has the advantage, that a possible bias in the  $\langle\epsilon\rangle$ -model is partly absorbed into  $f_\mu(S_\mu|\theta_\mu, k)$ , so that it effectively cancels in the application of  $f_\mu$  in the reconstruction. It will be shown in Chapter 6, that this step indeed improves the precision event reconstruction.

The technical derivation and implementation of  $f_\mu$  is discussed in Appendix C.

### 5.5.3 Analysis of the electromagnetic signal component

A detailed analysis of the signal ratio  $\langle \epsilon \rangle \equiv \langle S_{\text{em}} + S_{\text{em-}\pi^0} \rangle / \langle S_\mu \rangle$  based on simulations is performed in ref. [16], which also yields a model of  $\langle \epsilon \rangle$ . This model is also used in this study. The predictions of the model are compared with data obtained from the simulated SD events produced for this study in the following.

The analysis exploits again that the simulation separates the contributions of muons, electrons, and photons to the total signal. Lateral profiles of the average muon signal  $\langle S_\mu \rangle$  and the average electromagnetic signal  $\langle S_{\text{em}} + S_{\text{em-}\pi^0} \rangle$  are generated first. Stations with a total signal  $S$  smaller than the lowest trigger threshold  $S_{\text{T1}} = 1.75$  VEM are not used in order to emulate realistic data taking conditions.

Then, the ratio  $\langle \epsilon \rangle$  is calculated from these lateral profiles. The uncertainty of  $\langle \epsilon \rangle$  is approximately

$$\sigma[\epsilon] \simeq \frac{\langle S_{\text{em}} \rangle}{\langle S_\mu \rangle} \left( \frac{\sigma^2[S_{\text{em}}]}{\langle S_{\text{em}} \rangle^2} + \frac{\sigma^2[S_\mu]}{\langle S_\mu \rangle^2} \right)^{1/2}, \quad (5.5.14)$$

under the restrictions  $\sigma[S_{\text{em}}] \ll \langle S_{\text{em}} \rangle$  and  $\sigma[S_\mu] \ll \langle S_\mu \rangle$ . Fig. 5.25 and Fig. 5.26 show the results together with the expectation from the  $\langle \epsilon \rangle$ -model. In order to get a reasonable uncertainty estimate from Eq. (5.5.14) the ratio  $\langle \epsilon \rangle$  is only shown in these figures where  $\sigma[S_{\text{em}}] < 0.5 \langle S_{\text{em}} \rangle$  and  $\sigma[S_\mu] < 0.5 \langle S_\mu \rangle$ .

The signal ratio  $\langle \epsilon \rangle$  shows the previously discussed features. At zenith angles  $\theta > 62^\circ$  and radial distances  $r \gtrsim 1$  km from the shower axis, the signal ratio  $\langle \epsilon \rangle$  is small and almost constant. In this range, the signal contribution  $S_{\text{em-}\pi^0}$  of electromagnetic particles from the hadronic cascade is negligible and  $\langle \epsilon \rangle$  is independent of the cosmic ray energy  $E$  and mass  $A$ .

In the zenith angle range  $60^\circ < \theta < 66^\circ$ , a signal contribution  $S_{\text{em-}\pi^0}$  is present, but it mostly affects stations at  $r < 1$  km. This contribution depends on the cosmic ray energy  $E$  and mass  $A$  and is not universal, as discussed previously. Proton showers yield a ratio which is up to 0.1  $\langle S_\mu \rangle$  larger than iron showers. An effect of the same order is observed as a function of the cosmic ray energy  $E$ .

Both effects are well known [151]. The present  $\langle \epsilon \rangle$ -model neglects them and is derived from proton showers at  $E = 10^{19}$  eV. The reconstructed energy  $E_{\text{SD}}$  of showers close to  $\theta \approx 60^\circ$  will be slightly biased because of this in Chapter 6. The size of this bias is best estimated directly by applying the event reconstruction to simulated SD events. The analysis is performed in Chapter 6 and the bias will turn out to be acceptably small.

Another increase in  $\langle \epsilon \rangle$  is observed close to the shower axis very large zenith angles  $\theta \gtrsim 80^\circ$ , which is not related to a contribution of  $S_{\text{em-}\pi^0}$ . At such large zenith angles the average muon energy on the ground close to the shower axis becomes so high that additional electromagnetic particles are generated by muons through bremsstrahlung and direct  $e^+e^-$ -production in the field of nuclei, as discussed in Chapter 3. This  $S_{\text{em}}$ -contribution is still universal in good approximation.

There are unexpected features as well. The analysis shows an overall discrepancy of up to 0.2  $\langle S_\mu \rangle$  between the model from ref. [16] and this analysis. Generally, the model is in better agreement at a small radial distances  $r \lesssim 2$  km, where the discrepancy is at the level of 0.05  $\langle S_\mu \rangle$  to 0.1  $\langle S_\mu \rangle$ . Far from the shower axis, the simulated data shows a different behavior than the model. The model tends towards  $\langle \epsilon \rangle \approx 0.2$ , while the analysis indicates a drift towards zero. The drift starts closer to the shower axis at lower energies.

Most of the discrepancy between the  $\langle \epsilon \rangle$ -model and this analysis is understood and is related to the rejection of stations with signals below the T1 threshold  $S_{\text{T1}} = 1.75$  VEM in this analysis.

Fig. 5.27 shows the effect of a signal threshold on the signal composition in a station. Without a signal threshold, the muon component is very often zero. A lot of small signals are generated by photons, which are numerous. This fraction is strongly suppressed, if a signal threshold is introduced.

This allows to understand the effect on the signal ratio  $\langle \epsilon \rangle$ . The trigger effectively suppresses signals with no muon contribution if the muon density is very low. The electromagnetic particles make smaller signals on average than muons and are often not able to trigger the station without a simultaneous muon hit. The opposite is not true. Even a single isolated muon is able to trigger the station under optimal conditions at such large inclination angles.

The suppression effect vanishes, if the local muon density is large and the station is triggered in any case. Therefore, the average signal contribution  $\langle S_{\text{em}} \rangle$  from halo particles has a slight non-linear depen-

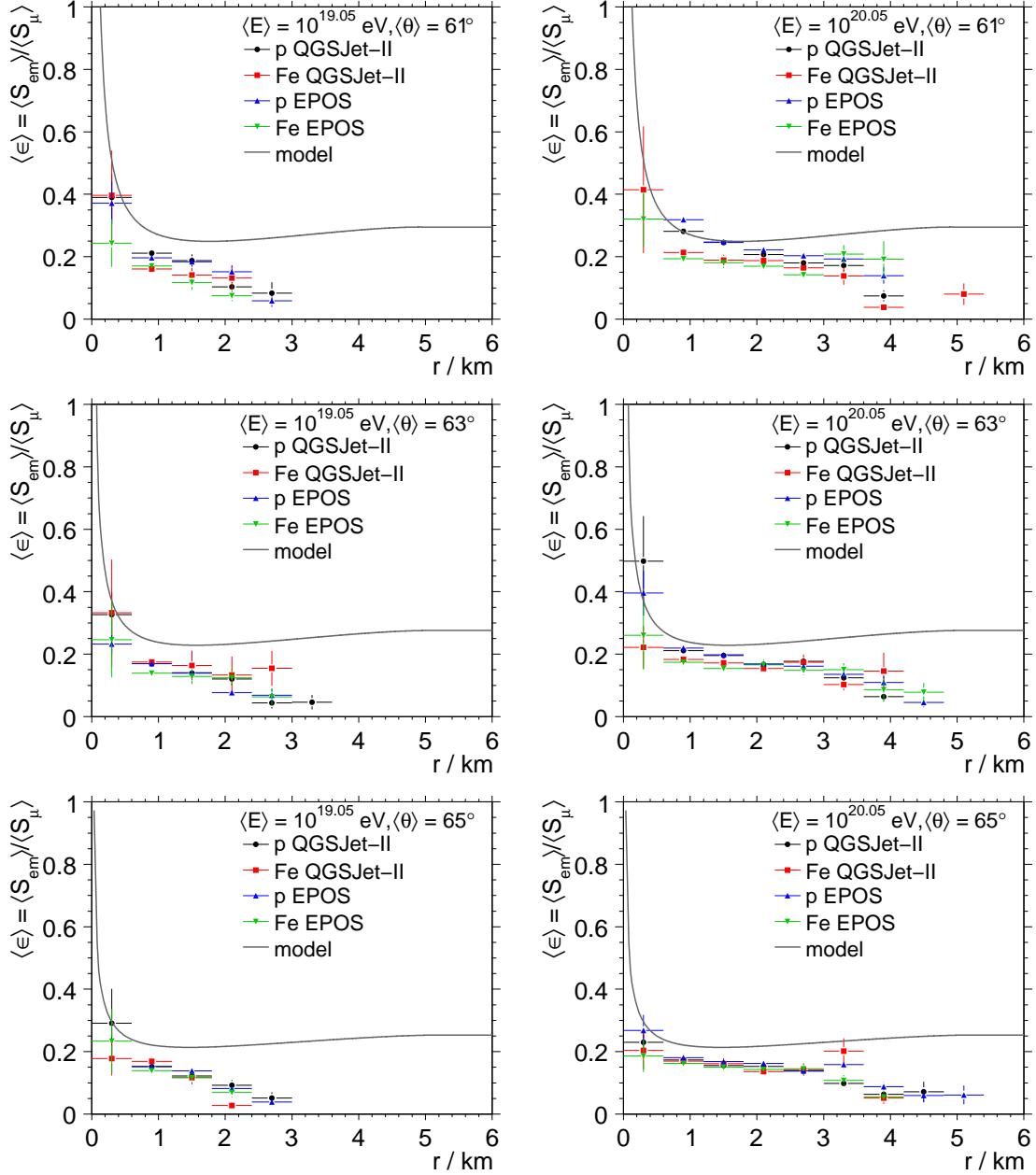


Figure 5.25: The points show the simulated signal ratio  $\langle \epsilon \rangle$  as a function of the radial distance  $r$  to the shower axis, the zenith angle  $\theta$ , and the cosmic ray energy  $E$  and mass, as well as the hadronic interaction model. The curve is the model prediction from ref. [16].

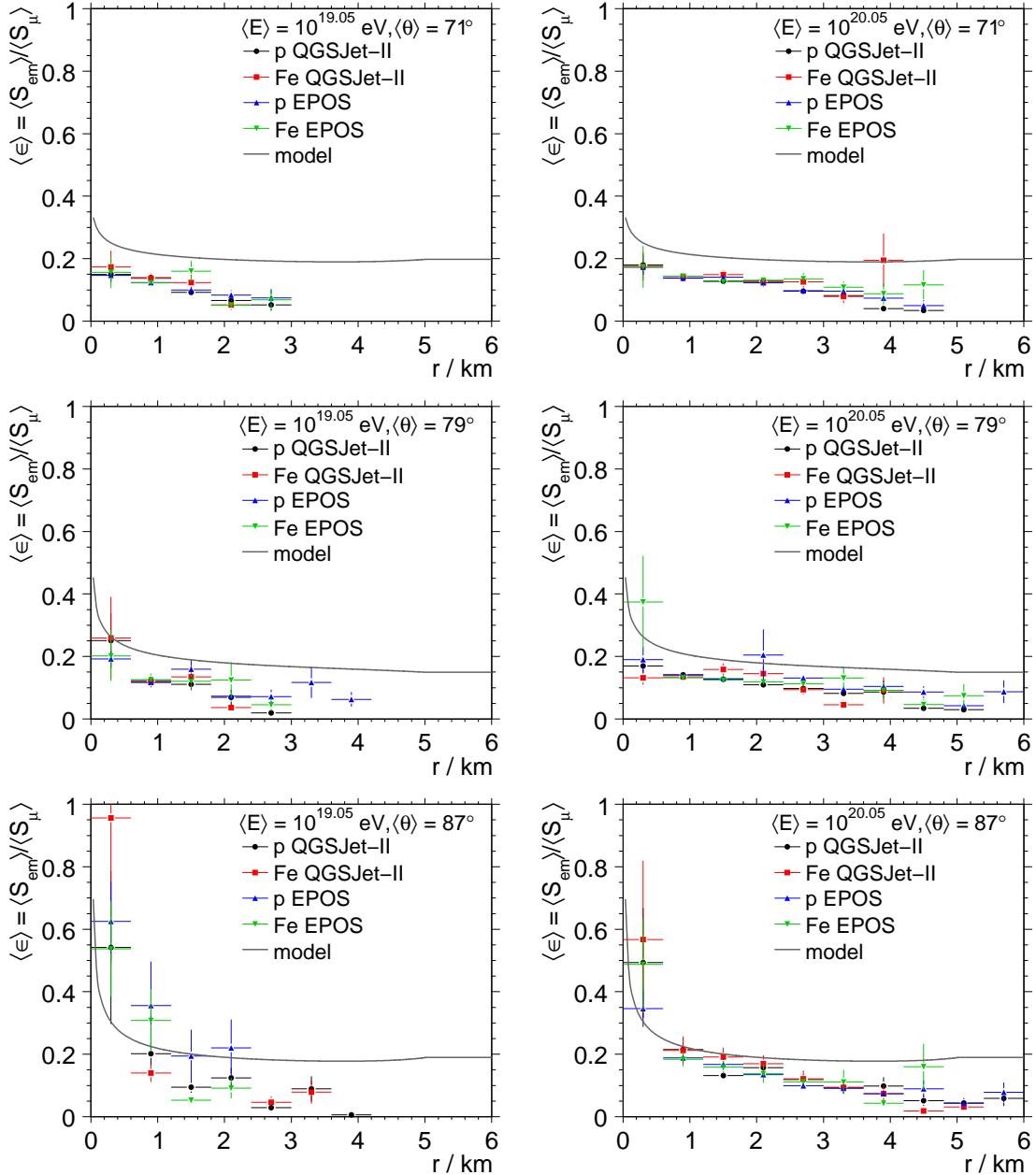


Figure 5.26: The plots are a continuation of Fig. 5.26.

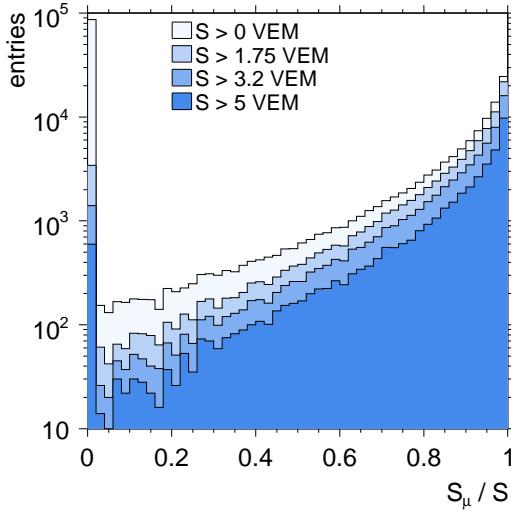


Figure 5.27: The plot shows the simulated ratio  $S_\mu/S$  of the muon generated signal to the total signal in a station. Only stations with  $S < 10$  VEM are included in the analysis to avoid signal contributions from  $\pi^0$ -generated electromagnetic particles. The peak at  $S_\mu/S = 0$  represents signals generated solely by electromagnetic particles. It is actually an isolated peak and only has a finite width due to the resolution of the histogram. The peak is suppressed, if stations with total signals below a certain threshold are rejected.

dency on  $k$ , the number of muon hits in the station. The average signal ratio  $\langle \epsilon \rangle$  effectively depends on the muon density  $n_\mu$  at the position of the station in turn and thus on the energy  $E$  of the cosmic ray.

If stations with signals below the trigger threshold are included into the analysis, a very good agreement with the functional form of the  $\langle \epsilon \rangle$ -model is found. However, a small and almost constant offset remains. The model prediction of  $\langle \epsilon \rangle$  stays about 0.05  $\langle S_\mu \rangle$  above the results of this study.

The available data is not detailed enough to derive a new  $\langle \epsilon \rangle$ -model despite these findings. Building a model of the signal ratio  $\langle \epsilon \rangle$  requires dedicated simulations. The already discussed approach to derive the signal response model  $f_\mu$  absorbs most of the bias found here. A small energy dependent bias present at all zenith angles nevertheless remains, which will be corrected empirically in Chapter 6.

#### 5.5.4 Comparison with another model

The model derived in this study is compared with the model derived in ref. [15]. To ease the discussion, the following naming convention is used:

- The model from ref. [15] is called model A.
- The model derived in this study is called model B.

It is sufficient to compare the signal response to single muon hits Fig. 5.28 shows the result. The shapes of the p.d.f.s are similar. There is an almost constant shift in the average muon response  $\langle S_\mu \rangle$ . The predictions of model B are lower by 10 % to 20 % than those of model A.

This shift is the one generated by the bias of the  $\langle \epsilon \rangle$ -model from ref. [16], which causes an opposite bias in model B. The combination of model B and the  $\langle \epsilon \rangle$ -model will nevertheless be unbiased by construction.

The signal variance  $\sigma[S_\mu]$  in model A and B is comparable. This is a coincidence. Expected is the same bias which is already observed in  $\langle S_\mu \rangle$ . Model B should therefore have a smaller variance than model A. It is still comparable, because model B includes fluctuations from the electromagnetic halo particles that make the variance larger.

The model comparison only shows differences which are expected from the earlier analysis of the  $\langle \epsilon \rangle$ -model or put in by construction. To properly compare the models, their performance in the reconstruction of simulated and real events needs to be assessed. This comparison is done in Chapter 6.

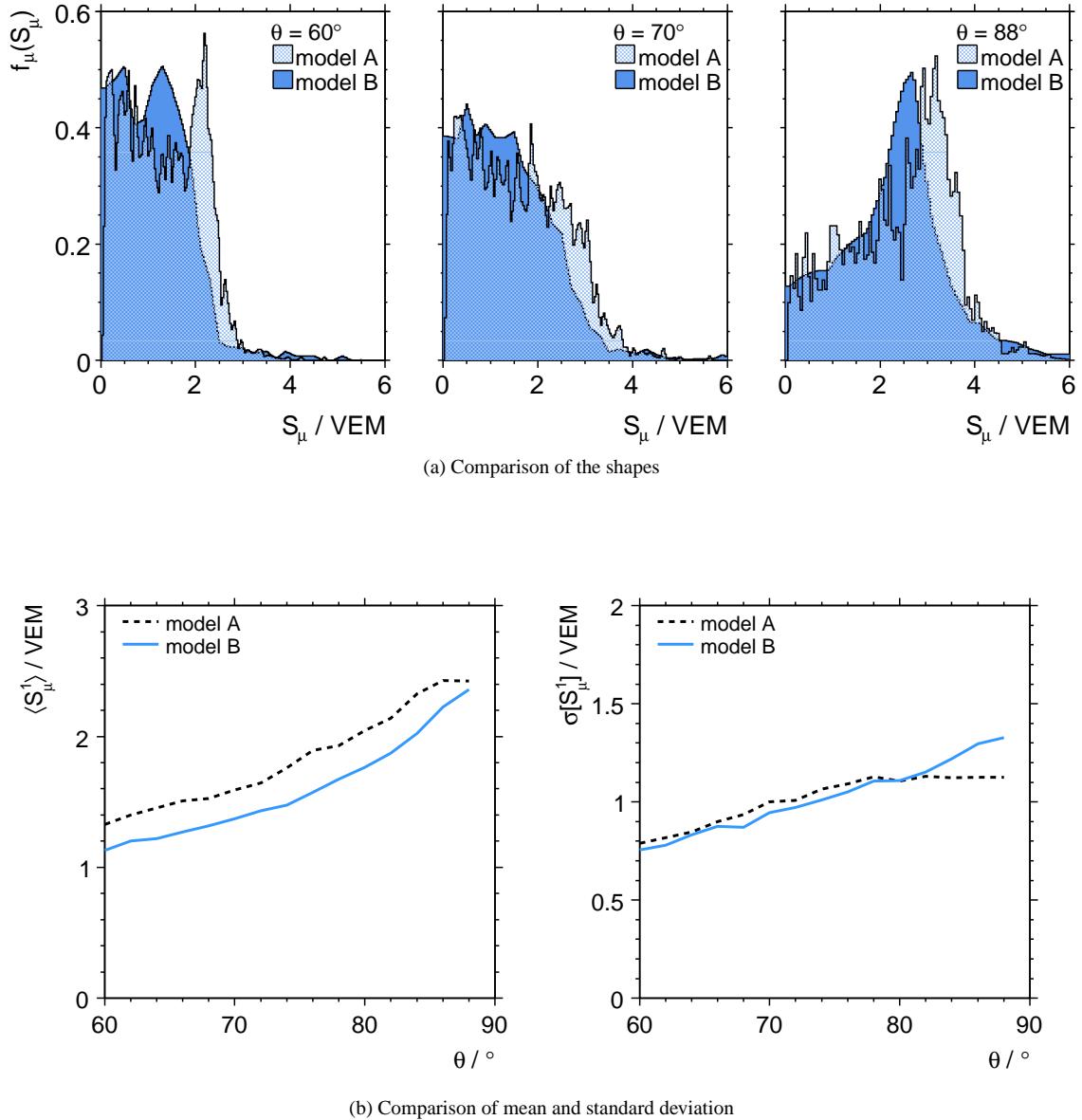


Figure 5.28: The plots compares two models of the signal response of to a single muon hit. Model A is taken from ref. [15], model B from this study.

# Chapter 6

## Reconstruction of cosmic ray properties

The Pierre Auger Observatory reconstructs the properties of cosmic rays by detecting the initiated extensive air showers as shown in Fig. 6.1. In the view of the particle physicist, the atmosphere forms a hadronic and electromagnetic calorimeter, which is instrumented and read out by the observatory.

**The surface detector** (SD) instruments a single slice of this calorimeter with particle detector stations, and samples the particle flux and arrival time structure of the air shower in the slice. It is therefore not able to measure the energy of the cosmic ray in a calorimetric way, because that would require to sample the deposited energy in the whole volume of the “calorimeter”.

Still, the surface detector is able to get the direction of the cosmic ray from the sampled arrival times, and an energy estimator from the sampled profile of the particle flux. This energy estimator is proportional to the particle flux through the slice, and almost proportional to the cosmic ray energy  $E$ .

The energy estimator depends also strongly on the atomic mass  $A$  and arrival direction  $(\theta, \phi)$  of the cosmic ray. It depends also weakly on the current condition of the atmosphere itself, which shows slight seasonal and daily variations. The arrival direction  $(\theta, \phi)$  can be measured unambiguously and the condition of the atmosphere can be monitored, but the dependency of the energy estimator of  $E$  and  $A$  remains entangled.

Thus, if the cosmic ray mass  $A$  is not known, the systematic uncertainty of the cosmic ray energy  $E$  derived from the energy estimator is large. Because the theoretical knowledge about soft hadronic interactions in general and at ultra-high energies in particular is limited, these systematic uncertainties are further increased. More details about this topic can be found in Chapter 5.

**The fluorescence detector** (FD) of the Pierre Auger Observatory does not suffer from these limitations. In some sense, this detector “instruments” the whole volume of the atmospheric calorimeter, and therefore is able to measure the energy of the cosmic ray by integrating over the energy loss along the shower path.

The fluorescence detector does so by collecting fluorescence light generated by collisions of electrons in the shower with nitrogen molecules. The collected light flux is sampled with a pixel camera with a high time resolution. With the orientation of the pixel trace in the camera, the arrival times of the light in the triggered pixels, and the arrival time of the shower front at a single surface tank, it is possible to get an accurate reconstruction of the shower geometry.

The knowledge of the position and orientation of the shower axis relative to the fluorescence detector together with the current state of the atmosphere allows to calculate the total amount of light emitted along the shower axis from the detected light. The total amount of light along the shower axis is a calorimetric measure of the total energy deposited by the electrons in the air shower. The conversion factor is called the *fluorescence yield* and can be measured in laboratory experiments.

The fluorescence detector can only detect the fraction of the cosmic ray energy, which is converted into electromagnetic particles and consequently deposited in the atmosphere. Fortunately, this is about 90 % at ultra-high energies, as discussed in Chapter 3. The exact fraction depends weakly on the cosmic

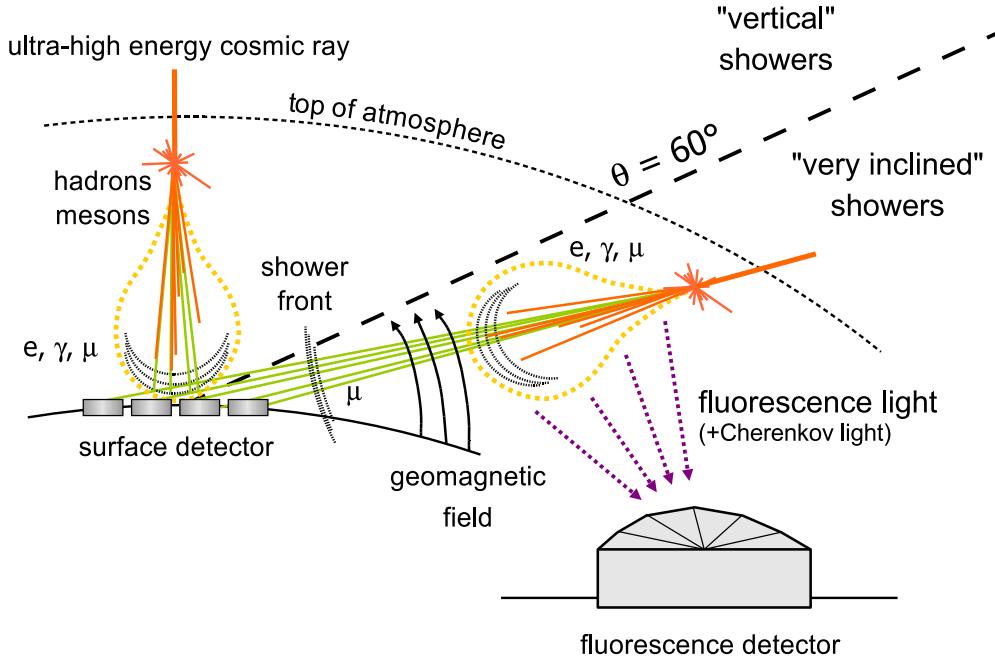


Figure 6.1: The drawing summarises the basic ideas behind the cosmic ray measurement at the Pierre Auger Observatory, and distinguishes between the so called *vertical showers* ( $0^\circ < \theta < 60^\circ$ ) and *very inclined air showers* ( $60^\circ < \theta < 90^\circ$ ). Vertical showers reach the surface detector with photons, electrons (yellow dotted cloud), and muons (green solid lines). In very inclined air showers, the electromagnetic component is (almost) extinct at ground level, and the surface detector observes mainly a muon shower. The fluorescence detector measures the fluorescence light (violet dashed lines) generated by the electromagnetic component of the shower in both cases.

ray energy  $E$ , the mass  $A$ , and on the theoretical modeling of the shower. The systematic uncertainty of the reconstructed energy due to this effect is about 4 % [47], which is even smaller than the precision of the absolute calibration of the FD. It is therefore well justified to call the FD measurement a calorimetric energy measurement. A downside of the fluorescence detector is that it can only operate in clear moonless nights, which limits its duty cycle to about 13 %.

**The combined use of the fluorescence and surface detector** cancels their respective weaknesses. Air showers measured in both detectors can be used to calibrate the energy estimator of the surface detector. If the calibration is done with an unbiased sample of the mass composition of cosmic rays, the FD-calibrated surface detector measures the cosmic ray energy with comparably small systematic uncertainties and a duty cycle of almost 100 %. This approach is followed here.

## 6.1 Vertical and very inclined showers

Important features of the lateral profile of an air shower measured at the Auger South ground altitude depend on the zenith angle, which change the signal patterns recorded in the SD. Two regimes form, which need to be treated by different SD reconstruction methods: the so-called *vertical* showers in the zenith angle range  $0^\circ < \theta < 60^\circ$ , and *very inclined air showers* at  $60^\circ < \theta < 90^\circ$ , see Fig. 6.1 and Chapter 3. These differences are only important for the reconstruction of SD events. The reconstruction of FD events is unaffected.

The SD reconstruction methods for both zenith angle regimes are based on different approximations to

the lateral shower profile, and in consequence yield different energy estimators. In vertical showers, the primary electromagnetic component of the shower is active at the ground level, and the signal measured in the Water-Cherenkov-Detector stations is dominated by the large number of electromagnetic particles. The typical path lengths are small enough, so that the influence of the geomagnetic field on the lateral profile of the shower can be neglected.

The energy estimator  $S_{1000}$  of vertical showers is obtained by fitting a modified Nishimura-Kamata-Greisen function [152–154] to the recorded spatial pattern of signals in the SD

$$S(r) = S_{1000}(E, A, \theta) \left( \frac{r}{r_1} \right)^{\beta(\theta)} \left( \frac{r + r_2}{r_2 + r_1} \right)^{\beta(\theta)+\gamma}, \quad (6.1.1)$$

with  $r_1 = 1000$  m and  $r_2 = 700$  m. The radial symmetric Nishimura-Kamata-Greisen function was originally derived for pure electromagnetic cascades. The  $\theta$ -dependency of  $\beta$  and  $S_{1000}$  can be parameterised from data [155–157]. The constant  $\gamma$  is zero in small showers and a free parameter in events with a high station multiplicity.

In very inclined air showers, the primary electromagnetic component is (almost) extinct, and the signal measured in the Water-Cherenkov-Detector stations is dominated by muons. The flux of low energy electromagnetic particles into the detectors is still larger than the flux of muons, but this is overcompensated by the better signal conversion properties of muons, as shown in Chapter 5.

The energy estimator<sup>1</sup>  $R_\mu$  of very inclined showers is obtained by fitting a reference profile of the muon density  $n_\mu^{\text{ref}}$  to the recorded spatial pattern of SD signals [10–17]

$$S(r, \psi) = \underbrace{\langle S_\mu^1 \rangle(\theta_\mu) (1 + \langle \epsilon \rangle(r, \psi; \theta, \phi))}_{\text{muon signal + em-background per muon}} \times \underbrace{A_{\text{station}}(\theta_\mu) R_\mu(E, A) n_\mu^{\text{ref}}(r, \psi; \theta, \phi)}_{\text{no. of muons per SD station}}. \quad (6.1.2)$$

This approximate signal model is build from several components which were previously discussed along with their systematic uncertainties in Chapter 5:

- $\langle S_\mu^1 \rangle(\theta_\mu)$  is the average signal generated by a single, isolated muon of with the inclination  $\theta_\mu$ .
- $\langle \epsilon \rangle = \langle S_{\text{em}} + S_{\text{em-}\pi^0} \rangle / \langle S_\mu \rangle$  is the average contribution of electromagnetic particles to the total signal, which arrive together with the muon. It is a function of the direction of the air shower  $(\theta, \phi)$  and the position of the station  $(r, \psi)$  in the lateral coordinate system of the shower.
- $A_{\text{station}}$  is the area of a surface detector station projected onto the ground as seen by a muon with the inclination  $\theta_\mu$ . The calculation is discussed in Appendix A.3.
- $R_\mu(E, A) n_\mu^{\text{ref}}(r, \psi; \theta, \phi)$  is the number density of muons on the ground. The first term  $R_\mu(E, A) = N_\mu(E, A, \theta) / N_\mu^{\text{ref}}(E, A, \theta)$  is the ratio between the total number of muons on the ground in the event and a reference shower. The second term is the lateral profile of the reference, which is obtained from air shower simulations. The reference model represents an average air shower at a given fixed mass and fixed energy. The standard choice is a  $10^{19}$  eV proton shower.

The zenith angle dependencies in  $N_\mu(E, A, \theta)$  and  $N_\mu^{\text{ref}}(E, A, \theta)$  cancel in good approximation, so that the energy estimator  $R_\mu$  of very inclined air showers is  $\theta$ -independent.

An in-depth discussion of the reconstruction of vertical showers is given in ref. [158]. The study focuses on the reconstruction of very inclined showers.

## 6.2 SD event reconstruction of very inclined showers

The current reconstruction procedure for very inclined air showers is based on a long line of achievements [10–17, 84, 159–166]. In the past, some developments were done in parallel, which formerly lead to two separate reconstruction programs: **SdHorRec** [14] and **efit** [17, 161–165].

<sup>1</sup>The energy estimator is also called N19 in other references:  $R_\mu \equiv N19$ .

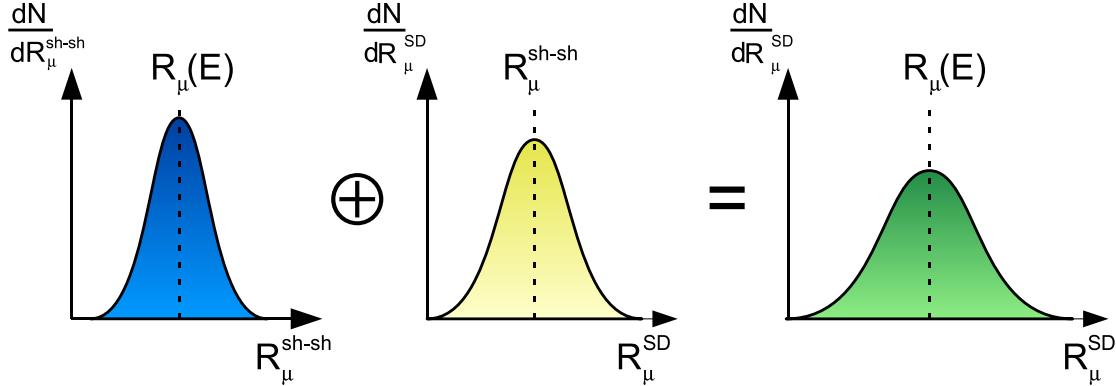


Figure 6.2: The figure illustrates, how the fluctuation of the reconstructed energy estimator  $R_\mu^{\text{SD}}$  (right graph) is composed of natural shower-to-shower fluctuations to the shower development (left graph) and the fluctuations due to the sampling of the detector (centered graph). In this definition,  $R_\mu$  is the ideal average of the physically realised energy estimator  $R_\mu^{\text{sh-sh}}$ . The latter may be measured with a perfect detector, while the former only exists on a statistical basis. Only  $R_\mu$  is directly related to the cosmic ray energy  $E$ .

Both are very sophisticated stand alone programs, but so far they are not implemented in the official analysis framework `Offline` of the Pierre Auger Collaboration. Their approaches and concepts were recently merged into a new reconstruction module called `SdHorizontalReconstruction` [166], which is now part of in the `Offline` analysis framework. Large parts of this module were contributed in the course of this work. The source code of the module can be obtained online [167].

The new module implements two coupled fits, one to reconstruct the energy estimator  $R_\mu$  and another one to reconstructed the shower arrival direction  $(\theta, \phi)$ . The energy estimator reconstruction is strongly based on the structure and the established concepts of the `efit` program, but was rewritten from scratch to make use of the software facilities of the `Offline` framework. The directional reconstruction on the other hand was copied from the standard reconstruction module for vertical showers [158, 168, 169] and is therefore identical in vertical and inclined showers.

The `Offline` framework aims to be a comprehensive modular framework which is easy to understand, extend and modify. It is distributed with a set of standard analysis modules which can be freely combined to execute a certain analysis task. The module concept is build on the idea, that analysis tasks like the reconstruction of very inclined SD events can be divided into a sequence of independent steps. Each step is realised with its own software module.

The standard modules also serve as templates, which others can improve with their own ideas and return these improvements into the standard module. `SdHorizontalReconstruction` was written in this spirit in a collaborative effort [166].

The reconstruction is discussed in detail in the following. As an introduction, a simplified reconstruction procedure for an air shower measured with an idealised SD measurement is regarded first. The simplified reconstruction is then expanded into a full reconstruction for real data.

Each reconstruction yields an energy estimator, among other observables. For the following discussion in this and the next chapter, it will be sometimes necessary to explicitly distinguish between different kinds of energy estimators:

- The *ideal* energy estimator  $R_\mu \propto E^{1/\gamma}$  is proportional to the average number of muons generated by a cosmic ray with the energy  $E$ . The real number of muons fluctuates from shower-to-shower, as discussed in Chapter 3 and Chapter 5.
- The *true* energy estimator  $R_\mu^{\text{sh-sh}}$  is proportional to the real number of muons in a shower. The true energy estimator  $R_\mu^{\text{sh-sh}}$  fluctuates from shower-to-shower with respect to  $R_\mu$ . For identical cosmic

rays that hit the atmosphere:  $R_\mu = \langle R_\mu^{\text{sh-sh}} \rangle$ . In contrast to  $R_\mu$ ,  $R_\mu^{\text{sh-sh}}$  could actually be measured with a perfect detector, that counts every muon that arrives the ground. The true energy estimator  $R_\mu^{\text{sh-sh}}$  is available in simulated air showers.

- The *reconstructed* energy estimator  $R_\mu^{\text{SD}}$  is the result of the measurement of  $R_\mu^{\text{sh-sh}}$  with a non-perfect detector. Sampling fluctuations additionally randomise  $R_\mu^{\text{SD}}$  with respect to  $R_\mu^{\text{sh-sh}}$ , because not every muon is counted.

The variables are further illustrated in Fig. 6.2. The explicit variables  $R_\mu^{\text{sh-sh}}$  and  $R_\mu^{\text{SD}}$  will be used, where the distinction is important. Where the distinction is not important, the energy estimator is simply called  $R_\mu$ , as before.

### 6.2.1 Idealised reconstruction

The surface detector consists of an array of identical Water-Cherenkov detector stations, which are excellent muon counters. In first approximation, the SD shall be regarded as a perfect muon counter and an SD event as a spatial configuration of muon counts. The shower direction  $(\theta, \phi)$  and the local muon arrival direction  $(\theta_{\mu,i}, \phi_{\mu,i})$  at the position of each station shall also be perfectly known.

In this case, the average number of muon counts  $\langle N_{\mu,i} \rangle$  in a station  $i$  can be calculated by multiplying the effective area of the station  $A_{\text{station}}(\theta_{\mu,i})$  with the muon density  $n_\mu$ . The effective area depends on the muon inclination  $\theta_{\mu,i}$  at the position of the station and is discussed in Appendix A.3. With the factorisation approach from Eq. (5.4.1), this can be formulated as the second part of Eq. (6.1.2):

$$\langle N_{\mu,i} \rangle(\tilde{\mathbf{r}}_i; R_\mu, \mathbf{r}_c, \theta, \phi) = A_{\text{station}}(\theta_{\mu,i}) R_\mu n_\mu^{\text{ref}}(\tilde{\mathbf{r}}_i - \mathbf{r}_c; \theta, \phi), \quad (6.2.1)$$

whereas  $\tilde{\mathbf{r}}_i$  is the position of station  $i$  in ground plane coordinates and  $\mathbf{r}_c$  is the point where the shower axis intersects with the ground. The latter position is also called the *shower core*.

Since the shower direction and the  $\theta_{\mu,i}$  are fixed, the equation has only three free parameters: the energy estimator  $R_\mu$  and the two coordinates of the shower core  $\mathbf{r}_c$  on the ground surface. By expanding Eq. (6.2.1) into a statistical model of the sampled muon counts  $N_{\mu,i}$ , the equation can be used to determine the free parameters from the sampled muon counts.

The reference profile  $n_\mu^{\text{ref}}$  is only defined in a tangential ground plane, while the SD stations have vertical offsets with respect to this ideal plane due to Earth's curvature and local altitude changes, as shown in Chapter 4. The true station position  $\mathbf{r}_i$  therefore needs to be projected into the ground plane of the model, which has the shower core point  $\mathbf{r}_c$  as its origin:

$$\mathbf{r}_i \xrightarrow{\text{proj.}} \tilde{\mathbf{r}}_i \quad \text{with} \quad (\tilde{\mathbf{r}}_i - \mathbf{r}_c) \mathbf{e}_z \stackrel{!}{=} 0, \quad (6.2.2)$$

whereas  $\mathbf{e}_z$  is the local vertical direction at the position of the shower core.

The solution with negligible bias at all zenith angles is the projection along the average muon arrival direction  $(\theta_{\mu,i}, \phi_{\mu,i})$  at the position of the station. The muon arrival direction is *a priori* known in the idealised case and can be modeled well in the realistic case. Different projection approaches are compared in Appendix A.4.

A good statistical model for the muon counts is the Poisson distribution. The model holds if the muon positions on the ground are uncorrelated over small scales. This is a good approximation, because muons are generated far from the ground with individual random directions. When they reach the ground, their positions are sufficiently randomised. This leads to the following likelihood function

$$L(\{N_{\mu,i}\}|R_\mu, \mathbf{r}_c) = \prod_i \frac{1}{N_{\mu,i}!} \langle N_{\mu,i} \rangle^{N_{\mu,i}} e^{-\langle N_{\mu,i} \rangle}, \quad (6.2.3)$$

with  $\langle N_{\mu,i} \rangle$  calculated by Eq. (6.2.1). Stations not hit by a muon are called *silent stations*. They have  $N_{\mu,i} = 0$ , but provide valid information to the likelihood function, too. The probability to get a silent station can be calculated from the Poisson distribution and thus silent stations help to constrain the model parameters, in particular the shower core position.

Optimal estimates for the free parameters are obtained from a numerical minimisation of  $-\ln L$  with the MINUIT package [148]. Minimising  $-\ln L$  is equivalent to maximising  $L$ , but can be better handled numerically. The numerical minimisation needs start values, which are as close to the final solution as possible. The barycenter  $\mathbf{r}_b$  of the muon counts yields a good initial estimate for the shower core:

$$\mathbf{r}_c \approx \mathbf{r}_b = \frac{\sum_i \mathbf{r}_i N_{\mu,i} / \sigma[N_{\mu,i}]}{\sum_i 1 / \sigma[N_{\mu,i}]} = \frac{\sum_i \mathbf{r}_i \sqrt{N_{\mu,i}}}{\sum_i 1 / \sqrt{N_{\mu,i}}}. \quad (6.2.4)$$

An initial estimate of  $R_\mu$  is obtained by fixing the shower core to the barycenter and solving

$$\frac{\partial \ln L(\{N_{\mu,i}\} | R_\mu, \mathbf{r}_b)}{\partial R_\mu} \stackrel{!}{=} 0 \quad (6.2.5)$$

analytically, which leads to

$$R_\mu = \frac{\sum_{i=1}^k N_{\mu,i} A_{\text{station}}(\theta_{\mu,i}) n_\mu^{\text{ref}}(\tilde{\mathbf{r}}_i - \mathbf{r}_c; \theta, \phi)}{\sum_{i=1}^k A_{\text{station}}(\theta_{\mu,i}) n_\mu^{\text{ref}}(\tilde{\mathbf{r}}_i - \mathbf{r}_c; \theta, \phi)} \quad (6.2.6)$$

with  $k$  as the total number of regarded stations, including silent stations.

### Analysis of resolution and bias

In the simplified reconstruction, the resolution and bias of the reconstructed energy estimator  $R_\mu^{\text{SD}}$  is only a function of the data statistic and the  $n_\mu$ -model from Chapter 5. It is worthwhile to evaluate the performance of the model by applying the simplified reconstruction to the 6480 simulated SD events from Chapter 5.

The full information of particle hits in each station is available in the simulated SD events, so that the ideal muon count  $N_{\mu,i}$  and the mean muon arrival direction  $(\theta_{\mu,i}, \phi_{\mu,i})$  can be calculated for each station. The reconstruction uses the  $n_\mu$  model derived from proton showers simulated with QGSJet-II. The reconstruction efficiency over the whole library is 99.9 %.

The resolution and bias of the fit are derived by comparing reconstructed parameters with the input values of the simulation. The reconstructed energy estimator  $R_\mu^{\text{SD}}$  has to be compared with the true energy estimator  $R_\mu^{\text{sh-sh}}$  in this case, not the ideal estimator  $R_\mu \propto E^{1/\gamma}$ , because only the resolution of the detector and the reconstruction is of interest.

The true energy estimator  $R_\mu^{\text{sh-sh}}$  can be calculated in simulated showers by counting the total number of muons on the ground  $N_\mu$  and dividing the number by the value obtained from the reference model  $N_\mu^{\text{ref}}$

$$R_\mu^{\text{sh-sh}} = N_\mu / N_\mu^{\text{ref}}. \quad (6.2.7)$$

For the reconstructed shower core position  $\mathbf{r}_c^{\text{SD}}$ , two kinds of residuals are regarded. The shower core resolution is obtained from the deviation distance of the reconstructed and true shower core positions

$$r_c = |\mathbf{r}_c^{\text{SD}} - \mathbf{r}_c|. \quad (6.2.8)$$

The core resolution  $\sigma[r_c]$  is taken as the 68 % quantile of the distribution of  $r_c$ .

The variable  $r_c$  cannot be used to detect a systematic bias in the core offset. A variable better suited for a bias analysis is the component of the core offset along a certain direction. Out of all possible directions, the projected arrival direction of the shower is a particularly interesting, the corresponding shift

$$\tilde{r}_c = (\mathbf{r}_c^{\text{SD}} - \mathbf{r}_c) (\mathbf{e}_x \cos \phi + \mathbf{e}_y \sin \phi), \quad (6.2.9)$$

is called the early-late component of the shower core offset. It is positive (negative), if the shower core is shifted towards the early (late) arriving part of the shower. A bias in this direction has the largest impact on the reconstructed zenith angle  $\theta_{\text{SD}}$  of the shower, as will become apparent in the full reconstruction later.

Fig. 6.3 to Fig. 6.5 show the results of the analysis. The distribution of  $R_\mu^{\text{SD}}$  around the true value is a Gaussian. This is in exact agreement with the expectation.

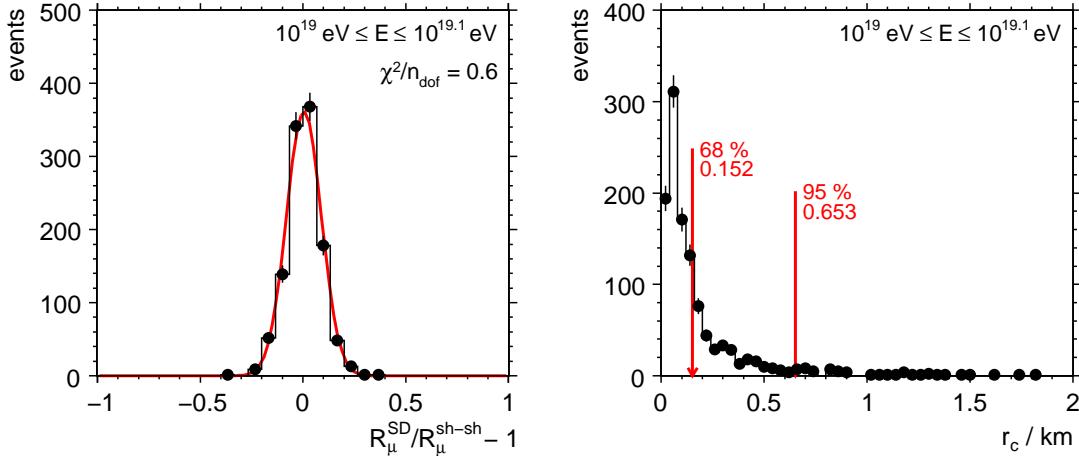


Figure 6.3: The plots show example distributions of the  $R_\mu^{\text{SD}}$ -residual and the shower core offset  $r_c$ ; as obtained from the reconstruction of simulated ideal SD events. Only events in the energy range of  $10^{19}$  eV to  $10^{19.1}$  eV are regarded. Events from proton and iron cosmic rays, simulated with QGSJet-II and EPOS, are combined. The red curve in the left plot is a fit of a Gaussian. The red arrows in the right plot show the 68 % and 95 % quantiles.

The bias analysis of  $R_\mu^{\text{SD}}$  shows a systematic uncertainty smaller than 5 %, which is an excellent result. The systematic uncertainty is small compared to the estimated systematic uncertainty of 10 % of the reference model derived in Chapter 5. Some systematics in the model apparently cancel in its application to the reconstruction of SD events.

The resolution  $\sigma[R_\mu^{\text{SD}}]/R_\mu^{\text{SD}}$  of the energy estimator depends strongly on the cosmic ray energy and to a lesser degree on the zenith angle. The average resolution at  $10^{19}$  eV is about 10 %.

The core resolution  $\sigma[r_c^{\text{SD}}]$  shows a dramatic decrease at  $\theta > 84^\circ$  with a corresponding increase in the core bias  $r_c$ . The increase is expected to some degree, since the lateral profile of the muon density becomes flatter with rising zenith angle, so that the center is less pronounced. The magnitude, however, is surprising. Events at these large zenith angles are excluded in some resolution and bias analyses where they would otherwise totally dominate the result, see for example the left plot in Fig. 6.5b).

Overall, the core shows a small shift of about 50 m towards the early arriving shower part. This shift is a function of the azimuth. It is larger if the deflections through the geomagnetic field are small. The shift is small enough to be acceptable, but further improvement seems to be possible here in the future.

The average estimated resolutions of  $R_\mu^{\text{SD}}$  and  $r_c^{\text{SD}}$  from the reconstruction are in good agreement with the respective true resolutions, especially in the case of  $R_\mu^{\text{SD}}$ . This indicates that the statistical model of the measurement covers the dominating effects and that the reconstruction bias is small compared to the reconstruction resolution.

Proton induced showers are measured with a slightly lower resolution than iron showers. The resolution in proton showers simulated with QGSJet-II is particularly low. The effect is explained by the varying number of muons  $N_\mu$  produced in the showers at the same cosmic ray energy  $E$ , as shown in Table 5.1 in Chapter 5. Proton showers simulated with QGSJet-II have the lowest number of muons.

The idealised reconstruction avoids some additional uncertainties, which are present in the reconstruction of realistic SD events. In this sense, the resolutions obtained here are an upper limit for the attainable resolution in a realistic reconstruction. The steps of a realistic reconstruction are discussed in the following.

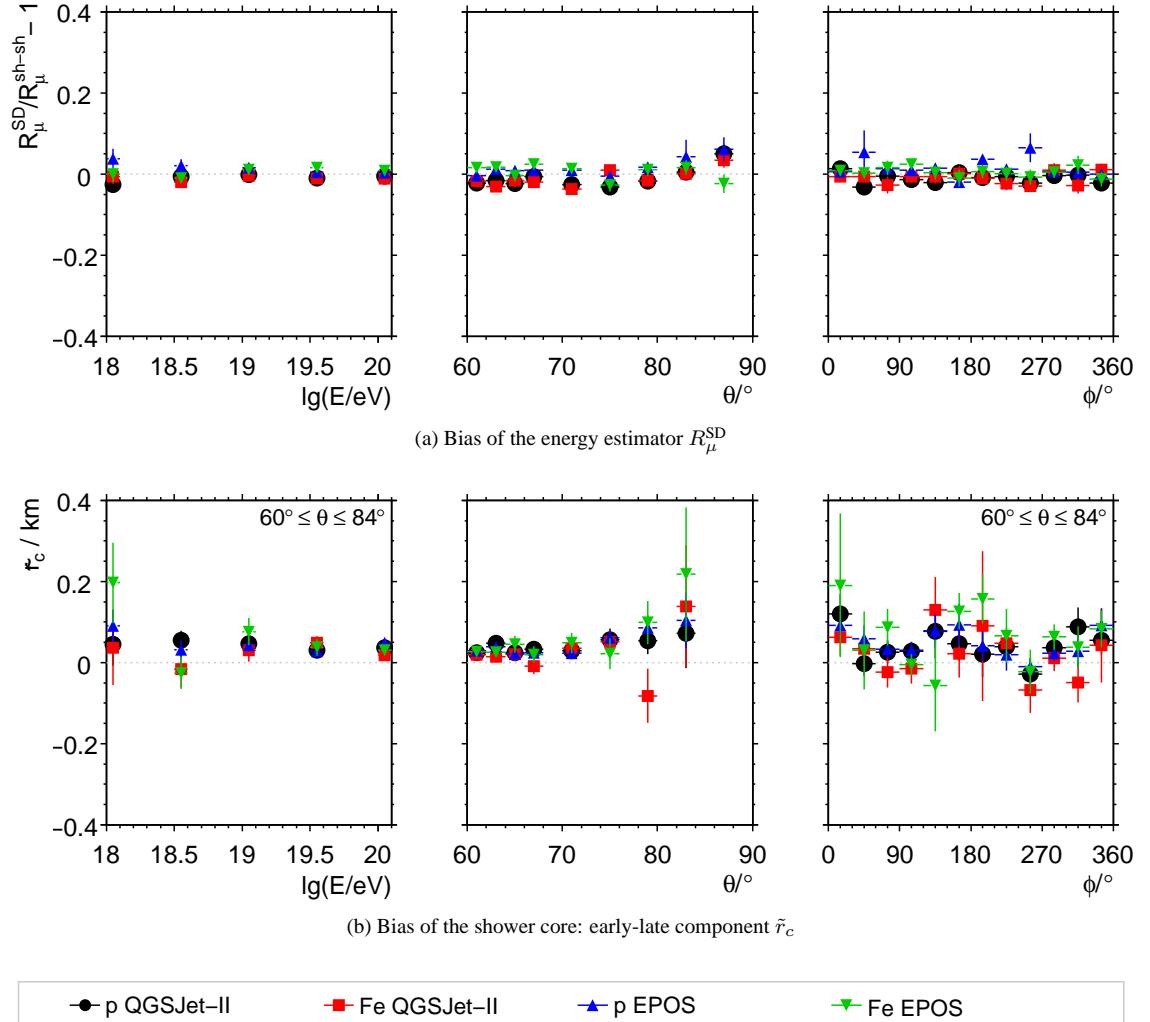


Figure 6.4: The points show biases of the energy estimator  $R_\mu^{\text{SD}}$  and the shower core  $r_c$  as a function of the cosmic ray energy  $E$  and direction  $(\theta, \phi)$ . The biases are derived from the simplified reconstruction of simulated ideal SD events. The core bias  $\tilde{r}_c$  at  $87^\circ$  is outside of the scale, it is around 0.7 km. The cosmic ray nuclei and hadronic interaction models used in the simulation are distinguished with different markers and colors.

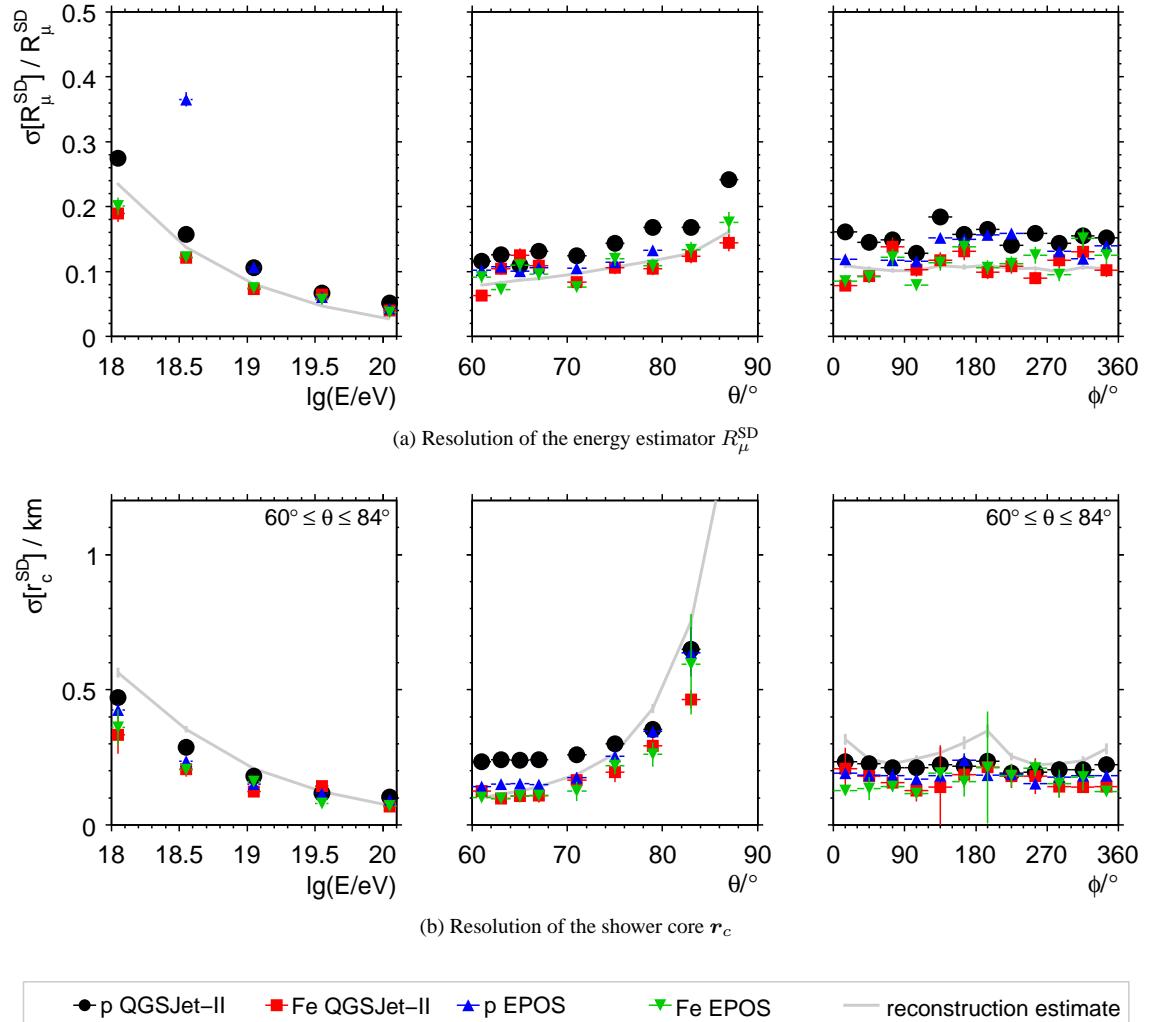


Figure 6.5: The points show the resolution of the reconstructed energy estimator  $R_\mu^{\text{SD}}$  and shower core  $r_c$  as a function of the cosmic ray energy  $E$  and direction ( $\theta, \phi$ ). The resolution is obtained from the simplified reconstruction of simulated ideal SD events. The true core resolution  $\sigma[r_c^{\text{SD}}]$  at  $87^\circ$  is poor and outside of the scale, it is around 2.4 km. The cosmic ray nuclei and hadronic interaction models used in the simulation are distinguished with different markers and colors. The reconstruction estimates of the respective resolutions are shown with a solid gray line.

### 6.2.2 Realistic SD reconstruction

The treatment of real SD events adds more complexity to the reconstruction. The full SD reconstruction has to deal with signal backgrounds, an estimation of the shower direction from data, and the ambiguous relation between signals measured in SD stations and their corresponding muon counts.

The full reconstruction procedure follows a sequence of several steps:

- (1) **Background rejection.** Real SD events contain various background signals, which need to be subtracted.
- (2) **Preliminary reconstruction of the shower direction.** A preliminary estimate of the shower direction can be obtained by fitting a flat shower front model to the signal start times of at least three SD stations.
- (3) **Reconstruction of the energy estimator and shower core.** With an established shower direction, the reconstruction of the shower core  $r_c$  and the energy estimator  $R_\mu$  may be started.
- (4) **Improved reconstruction of the shower direction.** The reconstruction of the shower direction is improved by a fit of a more realistic curved shower front model to the signal start times of at least four stations, which depends weakly on the reconstructed shower core position. Step (3) and (4) are iterated at least once. The iteration is stopped if the result has converged, which usually takes only one cycle.
- (5) **Acceptance selection.** A selection is applied to the reconstructed event to ensure minimal systematic uncertainties in the reconstructed parameters. Essentially, the selection assures that the shower is well contained in the SD array. The selection also assures a well defined exposure of the surface detector.

The individual steps are discussed in the next sections.

Technical details about the reconstruction in the Auger Offline-framework can be found in Appendix D.

### Background rejection

The event trigger of the surface detector of the Pierre Auger Observatory is permissive. Only about 10 % of all triggered very inclined air showers are real air showers [68]. The main background is generated by the background flux of atmospheric muons, which is constantly generated by low energy cosmic rays. Another source of background are lightning strikes in the SD array.

The latter are a rare but spectacular background. Lightning generates an electromagnetic pulse, which causes oscillating signals in the SD stations. The pulse probably enters the station DAQ via the grounding of the electronics. Signals generated in this way can fulfill the SD trigger condition and produce events with strange shapes. An example as shown in Fig. 6.6. The oscillations in the signal trace are used to rejected lightning events. If a single station in an event shows these oscillations, the whole event is rejected. Since this is a rare phenomenon, the influence on the SD acceptance is negligible.

Atmospheric muons are the main background, which hit each SD station with a rate  $\dot{N}_{\text{acc}} \approx 2.5 \text{ kHz}$ . The rate  $\dot{N}_{\text{T1}} = 100 \text{ Hz}$  of first level triggers in each SD station is dominated by these *accidental* muons. Accidental muons can both generate fake events and contaminate real air showers.

The average number of accidentally triggered stations per event can be estimated by multiplying the T3 time window  $\Delta t_{\text{T3}} = 60 \mu\text{s}$  with the T1 rate  $\dot{N}_{\text{T1}}$  and the number of read out stations  $N_{\text{tot}} \approx 300$

$$\Delta t_{\text{T3}} \dot{N}_{\text{T1}} N_{\text{tot}} \approx 2. \quad (6.2.10)$$

The actual number of stations that are read out as part of an event varies, as described in Chapter 4. Only a rough estimate is used here.

The probability for a station to be contaminated with an accidental muon is obtained by multiplying the time window  $t_{\text{trace}} = 768 \times 25 \text{ ns} = 19.2 \mu\text{s}$  of a signal trace with the rate of accidental muons  $\dot{N}_{\text{acc}}$

$$t_{\text{trace}} \dot{N}_{\text{acc}} \approx 5 \%. \quad (6.2.11)$$

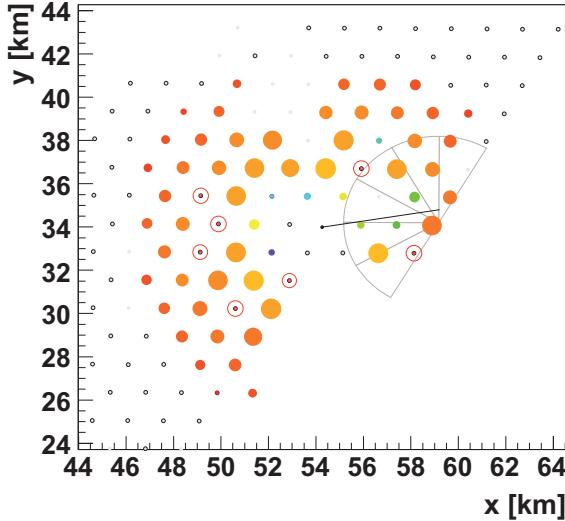


Figure 6.6: This top view on the SD array shows a spectacular lightning event (event 1332966). Colored circles represent triggered stations. The color indicates the arrival time of the signal, blue are early signals, red late signals. Full and open circles represent stations with recorded signals. The radii of the circles are proportional to the logarithm of the recorded signal. Dark small gray circles represent active un-triggered stations, light small gray circles represent stations that are inactive at the time of the event or not yet deployed in the field. The event formed like a doughnut of triggered stations impossible to get from a real air shower. Other signatures of lightning events are described in the text (plot adapted from [170]).

A contamination is therefore likely. Nevertheless, the impact of these additional signals on the reconstructed energy estimator  $R_\mu$  is usually negligible. The contamination probability is reduced to about 1 % by finding the start of the trace and regarding only signals in time window of  $5 \mu\text{s}$  from there.

The situation is more severe for the SD reconstruction of the shower direction, which is based on the start time of the signal trace. If the accidental muon hit occurs shortly before the actual real signal, the shower direction may be wrongly reconstructed. A time window of  $t_{\text{pre}} = 2.5 \mu\text{s}$  is read out before the time of the station trigger. The probability for an accidental muon hit is

$$t_{\text{pre}} \dot{N}_{\text{acc}} \approx 0.6 \%. \quad (6.2.12)$$

Two strategies are applied to reject the accidental muons. The *trace cleaning* algorithm is applied to every station signal. It is designed to remove accidental signals before the actual start of the real signal and reduce the signal contamination.

The *T4 algorithm*<sup>2</sup> works on the level of a whole event. It rejects the accidentally triggered stations from the event based on a number of heuristics. If the event is not reconstructable afterwards, it is rejected altogether.

### Trace cleaning

The trace cleaning algorithm is applied to the calibrated VEM trace of each station independently. It does not take the global event structure into account. The currently used algorithm is based on ref. [171]. The algorithm separates the signal trace into a set of separated time segments which contain a part of the total signal. A main segment is reconstructed, which was most likely generated by the passing shower front. Small separated peaks before the main segment are rejected.

Fig. 6.7 shows an example result of the algorithm. The technical implementation of the algorithm can be found in Appendix E.

<sup>2</sup>The name T4 implies an online event trigger, like the T1 - T3 triggers from Chapter 4. Instead, it is an algorithm for *offline* event selection, applied after the data taking. Nevertheless, it is often called “physics trigger” and regarded as a continuation of the online trigger hierarchy, because it distinguishes between background and signal events.

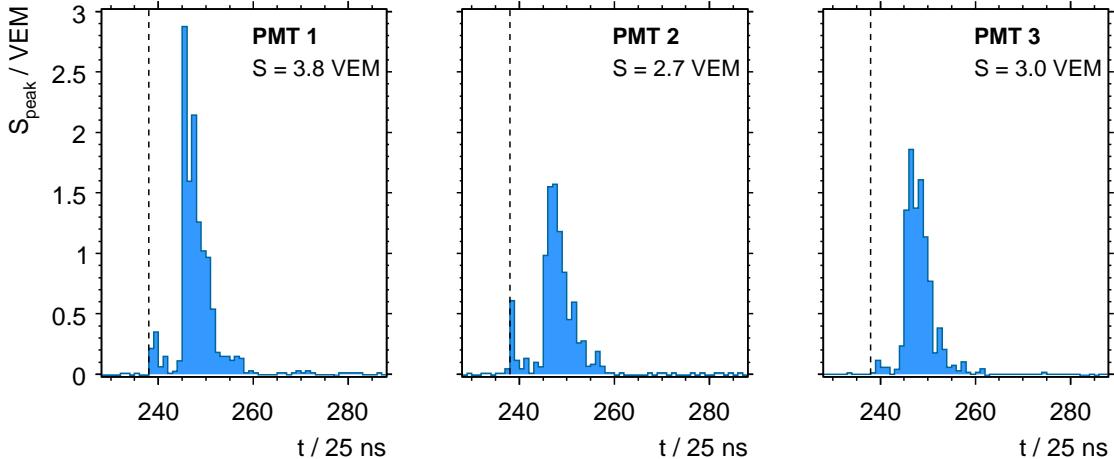


Figure 6.7: An example of a station with a good start time. The plots show the peak-calibrated signal in three PMTs of a single station, the start time found by the trace cleaning algorithm is the dashed line. The small peak before the main signal is correctly included as being a part of the shower front in this case.

#### T4 selection

The T4 selection algorithm finds and marks accidentally triggered stations in the event. These stations are not used in subsequent reconstruction steps. The remaining triggered stations are required to fulfill the T3 condition of the SD event trigger. Otherwise, the whole event is rejected.

Several heuristics are applied to separate accidentally triggered stations from real stations. The main criteria are that a station triggered by an air shower should have a compact spatial configuration and their signal start times should be compatible with a shower front moving with the speed of light. Thus, stations without close neighbors are rejected and those with large time offsets from a preliminary reconstructed shower front.

Currently, separate T4 selections are for vertical and for very inclined air showers. The T4 selection for vertical showers [172] uses a bottom-up approach, which starts with a minimal seed of good stations which define a preliminary shower front. The other stations are then accepted or rejected based on the compatibility with this seed.

The T4 selection for very inclined showers [173] is based on a top-down approach. The top-down approach first regards all stations as part of the event. It uses these stations to perform a preliminary reconstruction of the shower front. If this fit does not pass certain quality criteria, it is repeated after subsequently rejecting stations, until an acceptable configuration is found. The algorithm first tries all possible combinations where one station is rejected, then all combinations where two stations are rejected, and so on.

It was shown in ref. [174] that the top-down approach has an overall higher efficiency at zenith angles larger than  $60^\circ$ . Fig. 6.8 shows an example result of the T4 selection. The technical implementation of the algorithm can be found in Appendix E.

#### Preliminary reconstruction of the shower direction

An early reconstruction of the shower front is already done in the T4 algorithm, but the reconstructed axis is only used internally. A better but still preliminary reconstruction of the shower direction is done in the current reconstruction step.

The preliminary shower direction is obtained by fitting a flat shower front to the signal start times in the triggered stations of the event. The full procedure is documented in ref. [168]. The model of a flat shower front is already a good approximation, if the event has only a few stations.

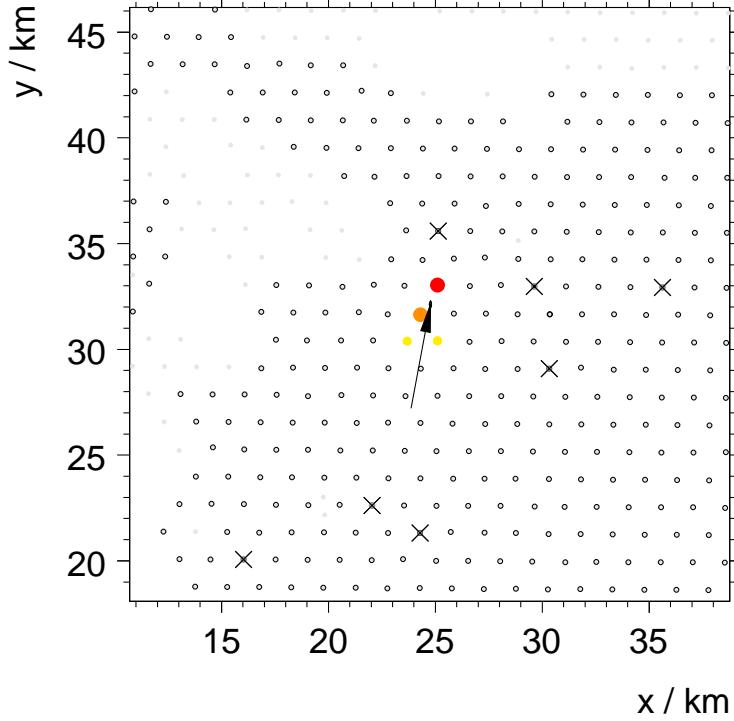


Figure 6.8: The top view on the SD array shows the result of the T4 selection on event 1904422. The coding is the same as in Fig. 6.6. Stations rejected by the T4 algorithm are additionally marked with a cross. The rejected station directly above the shower foot print has an incompatible timing, the others lack close neighbors.

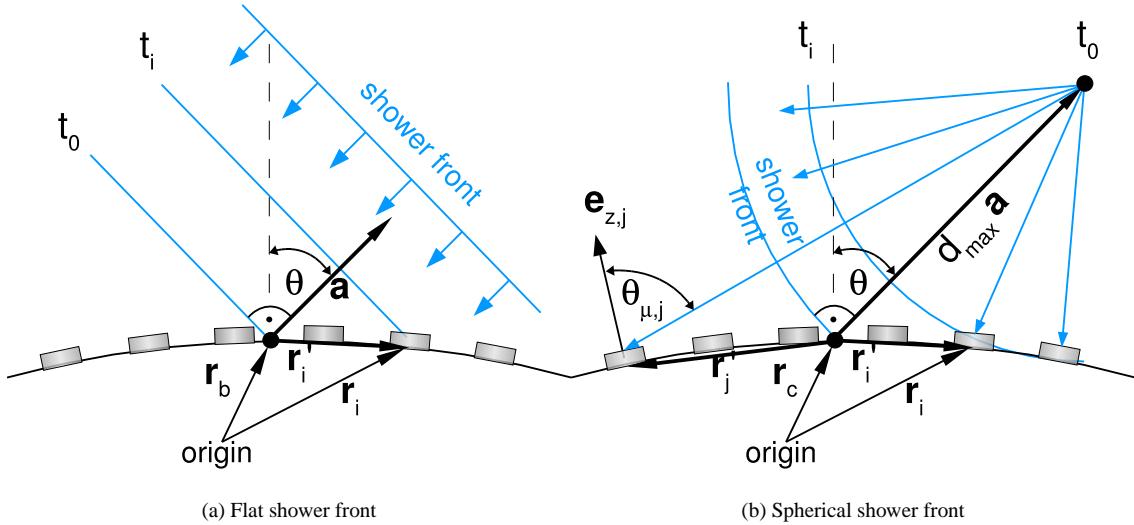


Figure 6.9: The drawings shows an arriving shower front corresponding to a normalised incoming direction vector  $\mathbf{a}$ , in the picture of the flat and the spherical shower front approximation. The vector  $\mathbf{r}_i$  points to the station  $i$ ,  $\mathbf{r}_c$  points to the shower core, while  $\mathbf{r}'_i$  is the relative vector. The distance  $d_{\max}$  is the radius of the spherical shower front at the position of the shower core. The right drawing also illustrates the local muon inclination  $\theta_{\mu,j}$  at the position of station  $j$ , with the local vertical direction  $\mathbf{e}_{z,j}$ . The drawings are not to scale.

The algorithm starts by calculating the barycenter  $\mathbf{r}_b$  of the event as a reference point

$$\mathbf{r}_b = \frac{\sum_{i=1}^{n-k} \mathbf{r}_i S_i^{1/2}}{\sum_{i=1}^{n-k} S_i^{1/2}}, \quad (6.2.13)$$

with  $\mathbf{r}_i$  and  $S_i$  as defined before. The power  $S_i^{1/2}$  is motivated by assuming Poisson statistics for  $S_i$  and considering a weight  $w_i = S_i/\sigma[S_i] \approx S_i/\sqrt{S_i}$ , analogue to Eq. (6.2.4). The reference point in the flat shower front approximation is arbitrary, but the barycenter is a convenient choice.

The model of a flat shower front moving with the speed of light predicts the average signal start time in a station as:

$$c(\langle t_i \rangle - t_0) = -\frac{1}{c} \mathbf{a}(\mathbf{r}_i - \mathbf{r}_b) \Leftrightarrow \langle t_i \rangle = t_0 - \frac{1}{c}(ux'_i + vy'_i + \sqrt{1-u^2-v^2}z'_i), \quad (6.2.14)$$

with  $\mathbf{r}'_i = \mathbf{r}_i - \mathbf{r}_b = (x'_i, y'_i, z'_i)^T$ , and  $t_0$ ,  $u$ , and  $v$  as free parameters. The full model is not linear in  $u$  and  $v$ , so that the general least-squares method is applied to obtain the best parameter estimates.

The MINUIT package [148] is used to minimise the  $\chi^2$ -function

$$\chi^2(\{t_i\}|t_0, u, v) = \sum_i \frac{(t_i - \langle t_i \rangle)^2}{\sigma^2[t_i]}, \quad (6.2.15)$$

whereas  $\sigma^2[t_i]$  is a model of the time uncertainty of the measured signal start time.

The standard model [175] of the time variance  $\sigma^2[t_i]$  is

$$\sigma^2[t](S, \theta_\mu, t_{50}) = 212 \text{ ns}^2 + 0.36 \left( \frac{2t_{50}}{n(S, \theta_\mu)} \right)^2 \frac{n(S, \theta_\mu) - 1}{n(S, \theta_\mu) + 1}, \quad (6.2.16)$$

whereas  $t_{50}$  is the time interval between the start of the VEM trace and the point, where 50 % of the total signal is accumulated and  $n(S, \theta_\mu)$  is an equivalent number of muon hits with an inclination  $\theta_\mu$ , which would generate the same signal  $S$  in the station.

The first term in this equation is given by the jitter of the GPS clock and the FADC bin resolution of 25 ns. The second term models the variance due to the statistical sampling of the true start time of the shower front. The term depends on the thickness of the shower front which is proportional to  $t_{50}$  and the number of trials  $n$ .

The number of trials  $n$  is very roughly approximated by the average equivalent number of muon hits that would generate the measured signal  $S$

$$n(S, \theta_\mu) \simeq S / \langle S_\mu^1 \rangle(\theta_\mu), \quad (6.2.17)$$

whereas  $h$  and  $r$  are height and radius of the water volume of the SD station. The formula for  $\langle S_\mu^1 \rangle$  is derived in Appendix A.3.

The numerical parameters of this time variance model are obtained from a fit to real data. An overview and comparison with other time variance models is given in ref. [169].

The muon inclination in Eq. (6.2.16) is approximated by the shower inclination  $\theta_\mu \approx \theta$ . Through  $n(S, \theta_\mu)$ , the time variance model  $\sigma^2[t]$  becomes a function of the shower inclination  $\theta$  itself, which introduces another kind of non-linear dependency on the parameters  $u$  and  $v$  in Eq. (6.2.15).

The numerical minimisation of Eq. (6.2.15) needs good initial values for  $u$ ,  $v$ , and  $t_0$ . With the approximations  $z'_i \simeq 0$  and  $\sigma^2[t_i] \simeq \text{const.}$ , Eq. (6.2.16) becomes a linear function of the parameters and its minimisation can be done analytically.

## Reconstruction of the energy estimator and shower core

With a preliminary shower axis, it is possible to run the main reconstruction of the energy estimator  $R_\mu$  and the shower core  $\mathbf{r}_c$ . The general concept at this point is the same as in the ideal reconstruction, but this time the signal response of the SD station to a muon hit has to be taken into account. The electromagnetic particles which arrive together with the muons also need to be regarded.

In analogy to the discussion of the reconstruction of ideal SD events, a model of the average values is regarded first. It is then expanded into a full statistical model, that also describes the signal fluctuations.

The following model is used to describe the average signal in a SD station

$$\langle S \rangle(\tilde{\mathbf{r}}_i; R_\mu, \mathbf{r}_c, \theta, \phi) = \langle S_\mu^1 \rangle(\theta_{\mu,i}) (1 + \langle \epsilon \rangle(\tilde{\mathbf{r}}_i - \mathbf{r}_c; \theta, \phi)) \times A_{\text{station}}(\theta_{\mu,i}) R_\mu n_\mu^{\text{ref}}(\tilde{\mathbf{r}}_i - \mathbf{r}_c; \theta, \phi), \quad (6.2.18)$$

whereas  $\langle S_\mu^1 \rangle$  is the signal generated by an isolated single muon,  $\langle \epsilon \rangle = \langle S_{\text{em}} \rangle / \langle S_\mu \rangle$  is the average ratio of signals generated by electromagnetic particles and muons, and  $\theta_{\mu,i}$  is the average direction of the incident muons. This model of the signal includes several approximations, which are described and quantised in Chapter 5.

The second part of the equation is the same as in Eq. (6.2.1), it describes the expected number of muons in the station. The position  $\tilde{\mathbf{r}}_i$  is the *projected* station position into the ground plane defined at the shower core position  $\mathbf{r}_c$ , with the same projection applied as described in Section 6.2.1.

The model has three free parameters:  $R_\mu$  and the two coordinates of  $\mathbf{r}_c$ . The shower direction  $(\theta, \phi)$  can be obtained independently from the measured arrival times of the shower front and is considered fixed.

The muon inclination  $\theta_{\mu,i}$  at the station  $i$  can be derived from the spherical shower front model, which assumes a common origin for all muons in a distance  $d_{\max}$  to the shower core along the shower axis, as indicated in Fig. 6.9b). The model predicts the muon inclination as

$$\cos \theta_{\mu,i} = (d_{\max} \mathbf{a} - \mathbf{r}'_i) \mathbf{e}_{z,i} = (d_{\max} \mathbf{a} - (\mathbf{r}_i - \mathbf{r}_c)) \mathbf{e}_{z,i}. \quad (6.2.19)$$

Over the collection area of a SD station, the muons arrive in parallel in excellent approximation, so that  $\theta_\mu$  essential has no spread.

The sphere radius  $d_{\max}$  depends only on the zenith angle  $\theta$  in good approximation for  $\theta > 60^\circ$ , as shown in Chapter 3. An initial value is derived from an approximate analytical calculation in Appendix A.1, a final value is obtained from the fit of the spherical shower front to the signal start times in the next reconstruction step.

In order to estimate the free parameters from the signal model with a maximum likelihood method, also the fluctuations of the signal need to be considered. There are three fluctuating quantities: the signal response  $S_\mu^1$  to an isolated muon, the electromagnetic signal background  $\epsilon$ , and the number of muon hits per SD station  $A_{\text{station}} R_\mu n_\mu^{\text{ref}}$ . The Poisson distribution is a reasonable probability density function (p.d.f.) for the latter, as already discussed in Section 6.2.1.

The p.d.f.  $f_\mu$  of the signal response  $S_\mu^1$  of the SD station to a single muon can be obtained from simulations, as shown in Chapter 5. A p.d.f. of the electromagnetic background  $\epsilon$  could be derived from simulations as well, but is not considered in the standard approach. In this study, the fluctuations of  $\epsilon$  are approximately included into the p.d.f.  $f_\mu$ , as described in Chapter 5.

The probability to observe a signal  $S_i$  is the convolution of the p.d.f. to observe  $S_i$  with a given number of muon hits  $k$  and the p.d.f. of getting  $k$  muon hits in case of a given local muon density  $n_\mu$  at the position of station  $i$ . *Saturated* stations only measure a lower limit for the true signal, while *silent* stations only provide an upper limit due to the T2 threshold  $S_{\text{th}} = 3.2$  VEM for the true signal. The probability to get a saturated or silent station can be calculated in this approach by integrating over all possibilities for the signal, if these limits are given.

The likelihood function  $L$  is the product of all these probabilities:

$$L(\{S_i\}|R_\mu, \mathbf{r}_c; d_{\max}, \theta, \phi) = L_{\text{saturated}} \times L_{\text{triggered}} \times L_{\text{silent}} \quad (6.2.20)$$

$$L_{\text{saturated}} = \prod_i \sum_{k=0}^{\infty} \int_{S_i/(1+\langle\epsilon\rangle)}^{\infty} ds f_\mu(s|\theta_{\mu,i}, k) \frac{1}{\langle N_{\mu,i} \rangle} e^{-k/\langle N_{\mu,i} \rangle} \quad (6.2.21)$$

$$L_{\text{triggered}} = \prod_j \sum_{k=0}^{\infty} f_\mu\left(\frac{S_j}{1+\langle\epsilon\rangle} \middle| \theta_{\mu,j}, k\right) \frac{1}{\langle N_{\mu,j} \rangle} e^{-k/\langle N_{\mu,j} \rangle} \quad (6.2.22)$$

$$L_{\text{silent}} = \prod_\ell \sum_{k=0}^{\infty} \int_0^{S_{\text{th}}/(1+\langle\epsilon\rangle)} ds f_\mu(s|\theta_{\mu,\ell}, k) \frac{1}{\langle N_{\mu,\ell} \rangle} e^{-k/\langle N_{\mu,\ell} \rangle}, \quad (6.2.23)$$

with

$$\begin{aligned} \langle\epsilon\rangle &= \langle\epsilon\rangle(\tilde{\mathbf{r}}_i - \mathbf{r}_c; \theta, \phi) \\ \langle N_{\mu,i} \rangle &= A_{\text{station}}(\theta_{\mu,i}) R_\mu n_\mu^{\text{ref}}(\tilde{\mathbf{r}}_i - \mathbf{r}_c; \theta, \phi) \\ \theta_{\mu,i} &= \arccos\left(\left(d_{\max} \mathbf{a} - (\tilde{\mathbf{r}}_i - \mathbf{r}_c)\right) \mathbf{e}_{z,i}\right). \end{aligned}$$

A maximisation of  $L$  yields best estimates for the free parameters  $R_\mu$  and  $\mathbf{r}_c$ . A minimisation of  $-\ln L$  is equivalent and done numerically with the MINUIT package [148]. To speed up the calculation of  $L$ , only silent stations within a fixed radius of 5 km around the shower axis are included.

The sum over the number of muon hits  $k$  has to be carried out numerically. Up to 100 terms are calculated, starting at the integer  $k_0$  closest to  $\langle N_{\mu,i} \rangle$  and stepping upward and downward by one subsequently. The summation is stopped earlier if the so far accumulated result is not zero and the next term only contributes a fraction smaller than  $10^{-3}$ .

The numerical minimisation of  $-\ln L$  needs initial values for the free parameters, which are already close to the final result. The barycenter  $\mathbf{r}_b$  from Eq. (6.2.13) serves again as a starting point for the shower core  $\mathbf{r}_c$ .

An initial value for  $R_\mu$  is obtained by fixing the shower core in Eq. (6.2.20) to  $\mathbf{r}_b$  and solving

$$\frac{\partial \chi^2(\{S_j\}|R_\mu)}{\partial R_\mu} \stackrel{!}{=} 0 \quad (6.2.24)$$

analytically, with

$$\begin{aligned} \chi^2(\{S_j\}|R_\mu) &= \sum_j \frac{(S_j - \langle S_{\mu,j}^1 \rangle (1 + \langle \epsilon_j \rangle) \langle N_{\mu,j} \rangle)^2}{\sigma^2[S_j]} \quad (6.2.25) \\ \langle \epsilon_j \rangle &= \langle \epsilon \rangle(\tilde{\mathbf{r}}_j - \mathbf{r}_c; \theta, \phi) \\ \langle N_{\mu,j} \rangle &= A_{\text{station}}(\theta_{\mu,j}) R_\mu n_\mu^{\text{ref}}(\tilde{\mathbf{r}}_j - \mathbf{r}_c; \theta, \phi) \\ \langle S_{\mu,j}^1 \rangle &= \int ds s f(s|\theta_{\mu,j}, 1) \\ \sigma^2[S_j] &= \langle N_{\mu,j} \rangle \sigma^2[S_{\mu,j}^1] = \langle N_{\mu,j} \rangle \left( \int ds s^2 f(s|\theta_{\mu,j}, 1) - \left( \int ds s f(s|\theta_{\mu,j}, 1) \right)^2 \right). \end{aligned}$$

This leads to

$$R_\mu = \left( \frac{\sum_i S_i^2 / (N_{\mu,i}^{\text{ref}} \sigma^2[S_{\mu,i}^1])}{\sum_j (\langle S_{\mu,j}^1 \rangle^2 N_{\mu,j}^{\text{ref}}) / \sigma^2[S_{\mu,j}^1]} \right)^{1/2} \quad (6.2.26)$$

with  $N_{\mu,i}^{\text{ref}} = A_{\text{station}}(\theta_{\mu,i}) n_\mu^{\text{ref}}(\tilde{\mathbf{r}}_i - \mathbf{r}_c; \theta, \phi)$ . Only normal stations can be included into this calculation, saturated and silent stations cannot be treated.

The reconstruction of an example event is shown in Fig. 6.10.

## Reconstruction of the shower direction

With an estimated shower core position  $\mathbf{r}_c$  from the previous step, it is possible to get a final estimate of the shower axis. The shower front is now treated as the surface of a sphere, expanding with the speed of light, as indicated in Fig. 6.9b). The approach used here is identical to the standard one used in vertical showers, as described in ref. [168]. An alternative Monte-Carlo based reconstruction of the shower direction is developed in ref. [57].

In principle, the predefined shower core position  $\mathbf{r}_c$  is not needed for this reconstruction step. The arrival time of the shower front at station  $i$  with the position  $\mathbf{r}_i$  in this model is already fully determined by the origin  $\mathbf{r}_0$  of the shower front sphere and the common start time  $t_0$  of all particles at the origin:

$$c(t_0 - \langle t_i \rangle) = |\mathbf{r}_i - \mathbf{r}_0|. \quad (6.2.27)$$

The four free parameters  $t_0$  and  $\mathbf{r}_0$  of this model can be fitted in a completely independent way from the previous reconstruction step. The reconstruction of the shower front sphere is decoupled.

However, the reconstructed shower origin  $\mathbf{r}_0$  alone does not define the shower axis. The shower core position  $\mathbf{r}_c$  is needed as a second reference point. In order to directly obtain the shower axis from the fit, the model is expressed in a different but equivalent way, see Fig. 6.9b):

$$c(t_0 - \langle t_i \rangle) = |\mathbf{r}'_i - d_{\max} \mathbf{a}| \Leftrightarrow \langle t_i \rangle = t_0 + \frac{1}{c} |\mathbf{r}_i - (\mathbf{r}_c + d_{\max} \mathbf{a})|, \quad (6.2.28)$$

whereas  $d_{\max} = |\mathbf{r}_0 - \mathbf{r}_c|$  is the length of the normalised arrival vector  $\mathbf{a} = (u, v, \sqrt{1 - u^2 - v^2})^T$ .

This introduces a weak coupling between the last two steps of the reconstruction, since the reconstruction of the shower core  $\mathbf{r}_c$  relies on the shower axis  $\mathbf{a}$  and vice versa. Both reconstruction steps are therefore iterated until the results converge. One iteration is usually sufficient.

The iteration neglects correlations between the energy estimator  $R_\mu$  and the shower core  $\mathbf{r}_c$  on the one side and the parameters  $t_0$ ,  $d_{\max}$ ,  $\theta$ , and  $\phi$  on the other side. The correlations are expected to be small, since the uncertainty of the shower core position is small compared to the length of the shower axis

$$\sigma[r_c] \ll d_{\max},$$

and therefore shifts in the core position affect the reconstruction of the shower direction only weakly.

The dependency of the spherical shower front model on its free parameters is non-linear. They are obtained with the least-squares method, by minimising

$$\chi^2(\{t_i\} | t_0, d_{\max}, u, v) = \frac{(t_i - \langle t_i \rangle)^2}{\sigma^2[t_i]}, \quad (6.2.29)$$

numerically with the MINUIT package [148]. The time variance  $\sigma^2[t_i]$  is again given by Eq. (6.2.16). The time variance model produces a minor dependency on the total signals  $\{S_i\}$  in the event, but otherwise the input data for the fit of the energy estimator  $R_\mu$  and this fit are independent. Initial values for the numerical minimisation are adopted from the preliminary reconstruction of the shower axis.

The reconstruction of an example event up to this point is shown in Fig. 6.10.

## Acceptance and quality selection

Very inclined air showers can have very elongated patterns in the SD array. Events like in those in Fig. 6.11 are recorded and reconstructed, which have core position that is not well constrained by actual measurements around it. The shower core uncertainty and the uncertainty of the energy estimator  $R_\mu$  are correlated. The uncertainty of  $R_\mu$  may become large in such cases.

The reconstruction may also be biased in such a scenario. If, for example, the profile of the muon density  $n_\mu$  was steeper in the model than in reality, then this systematic would be counter-balanced to some degree by measurements close and far to the shower axis under normal circumstances. But if only measurements far from the shower axis are available, the assumed model bias has a direct impact on the energy estimator  $R_\mu$ .

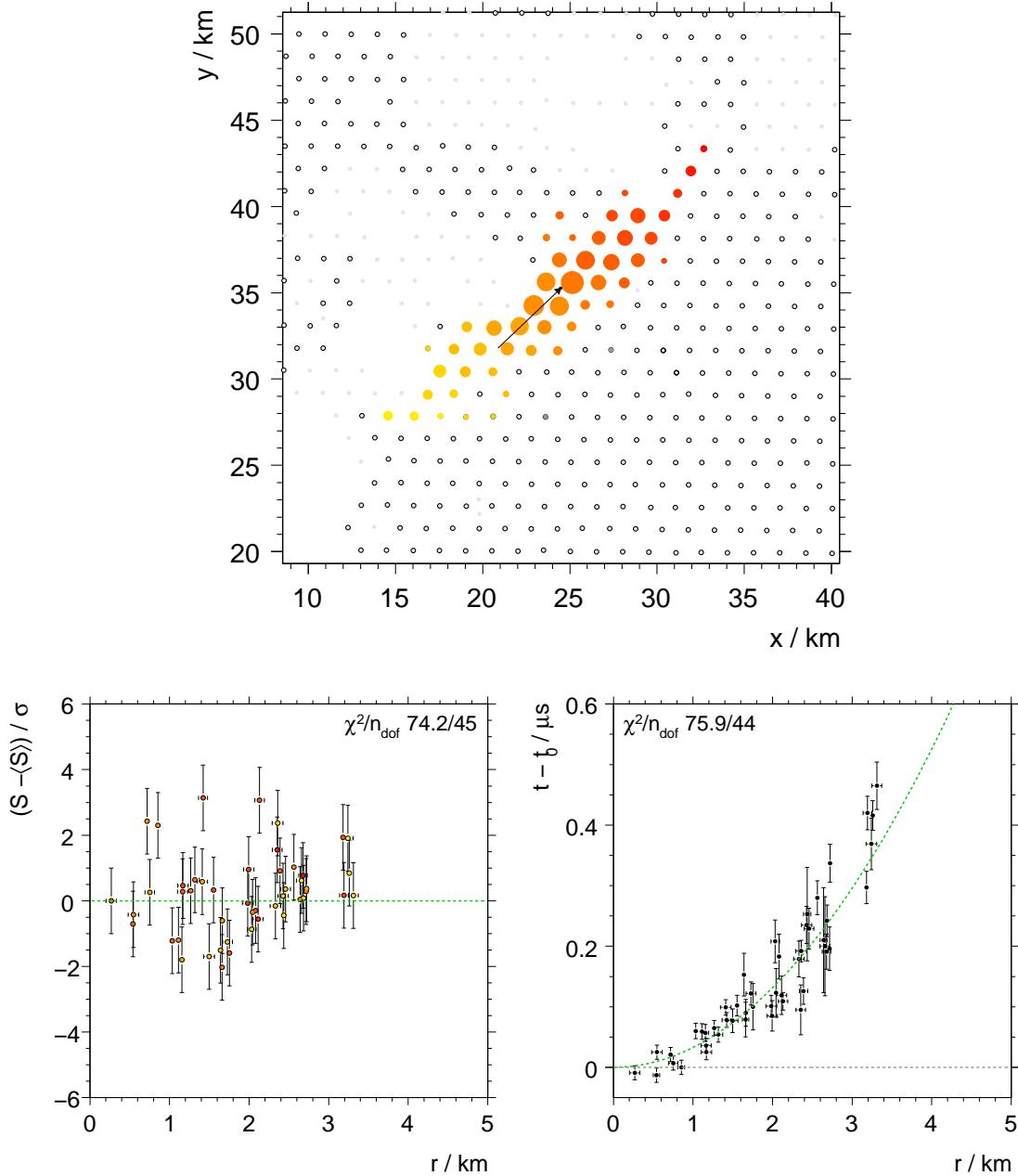


Figure 6.10: The plots show the surface detector event 2344346, which is reconstructed with  $E \approx 7 \times 10^{19}$  eV,  $\theta \approx 76^\circ$ , and  $\phi \approx 224^\circ$ . This spectacular event triggered 48 stations. Top: Shown is a top view on the surface detector, the symbols have the same meaning as in Fig. 6.6. Bottom left: The points show the residuals of the signals  $S$  with respect to the model over the radial distance  $r$  to the shower axis. Bottom right: The points shows the time residual with respect to a plane shower front moving with the speed of light (blue dashed line) over the radial distance. The prediction of the spherical shower front model is represented with a green dashed line.

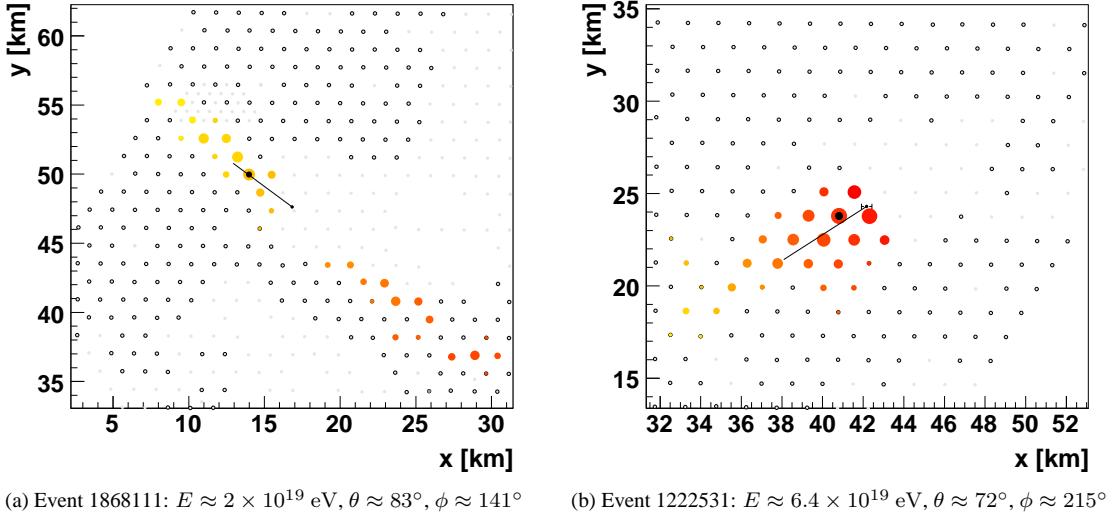


Figure 6.11: The top views on the SD array show example events whose core positions are not well constrained by measurements. The symbols have the same meaning as in Fig. 6.6. A big black dot marks the station with the largest signal in the event.

Because of these reasons, the acceptance of the SD array for air showers is artificially reduced by rejecting events whose shower core is well confined in the array. This artificial reduction can be performed in such a way that the remaining acceptance of the SD array can be calculated very precisely.

The event rejection is based on a class of conditions [17, 68, 168, 176–178], the *T5 criteria*<sup>3</sup>. They require that the station closest to the shower core is surrounded by other active stations. Active means that the station is ready to measure signals, but not necessarily has a triggered signal.

- **T5-Prior.** The station with the largest signal in the event has to be surrounded by six active stations.
- **T5-Posterior.** The T5-Posterior is a relaxed T5-Prior. The station with the largest signal in the event has to be surrounded by at least five active stations. In addition, the reconstructed shower core has to be located in a triangle of active stations.
- **Strict T5-Posterior.** The strict T5-Posterior is a strict version of the T5-Posterior. The station with the largest signal in the event has to be surrounded by six active stations. In addition, the reconstructed shower core has to be located in a triangle of active stations. It can also be regarded as a stricter version of the T5-Prior, in the sense, that every event accepted by the strict T5-Posterior is also accepted by the T5-Prior but not vice versa.
- **T5-Core.** The station closest to the reconstructed shower core is surrounded by six active stations.

The names “prior” and “posterior” refer to whether the criterion can be applied before or after the reconstruction. The working of the T4-Prior, the T5-Core, and the strict T5-Posterior are illustrated in Fig. 6.12.

The T5 criteria are devised as such that it is possible determine every second, whether a particular station in the array would be able to fulfill the criterion, if the shower would fall next to it. Only if this is the case, the elementary cell area attributed to the station contribute to the acceptance of the SD array. For a vertical shower, the elementary cell is a hexagon<sup>4</sup> as indicated in Fig. 6.12a). The details of this calculation are covered in Chapter 8.

<sup>3</sup>Like the T4, the T5 is an offline event selection criterion, not an online trigger. Nevertheless, it is often called “acceptance trigger” or “quality trigger”, in continuation of the online trigger hierarchy T1 - T3.

<sup>4</sup>In general, the elementary cell of the SD array is the first Brillouin zone of the array projected into the lateral coordinate system, which has a varying shape with the direction of the shower [179]. Using the instructive picture of the hexagon cell nevertheless is sufficient. Only the shape of the cell changes, but not its area.

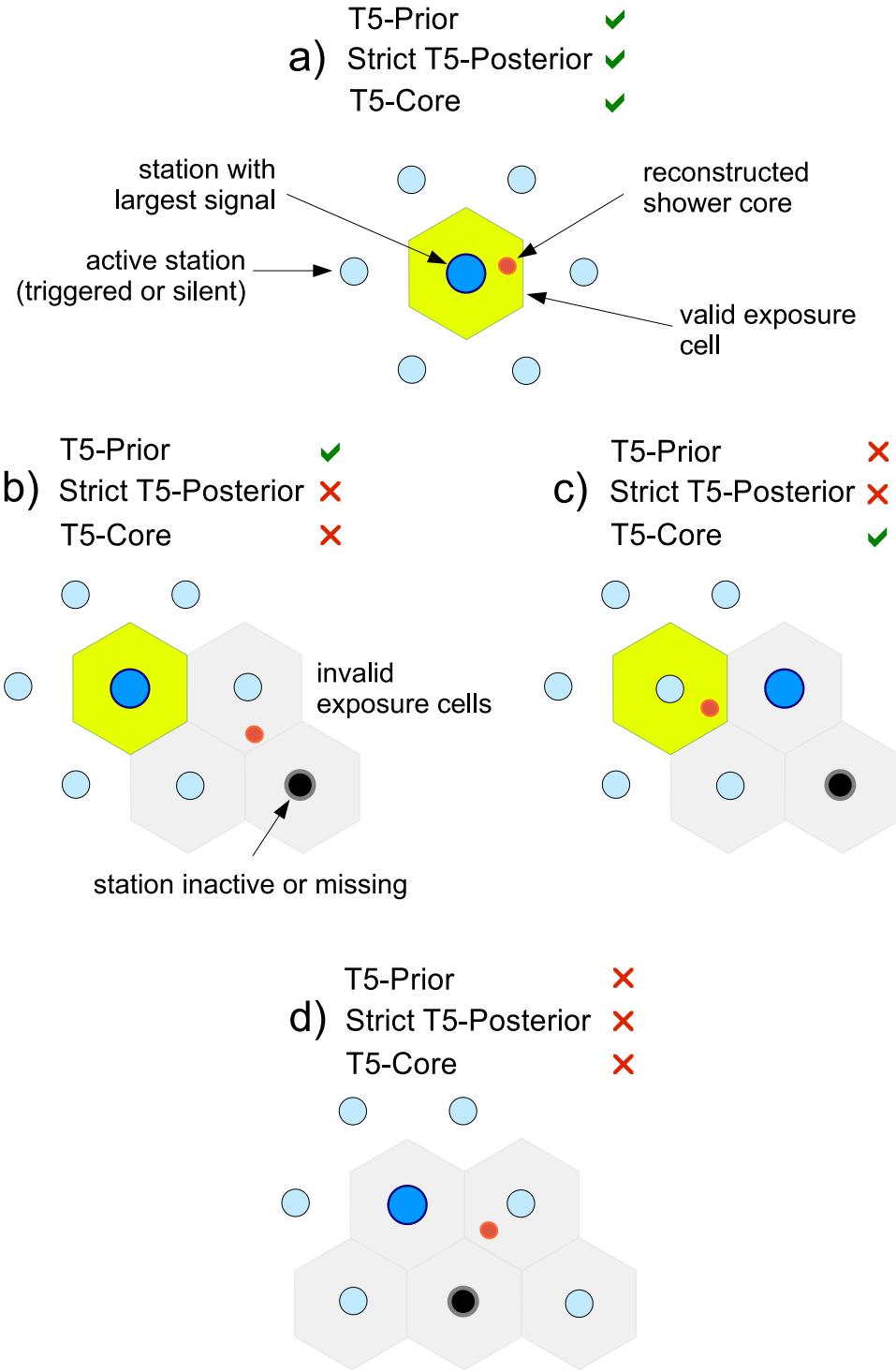


Figure 6.12: The drawings illustrate different SD event configurations, which pass different T5 variants. All configurations are meant to be embedded in an otherwise regular SD array. The valid exposure cells are only drawn around the station that is relevant for the respective T5 variant. Invalid exposure cells are only drawn around the stations in the plane.

The T5-Prior and T5-Posterior were developed for vertical showers. They exploit that the lateral density profile of an air shower has a monotonic and steep decline as a function of the radial distance  $r$  to the shower axis. The station with the largest signal is most certainly the station closest to the true impact point of the shower. The T5-Prior has the advantage that its outcome is completely independent of the reconstruction and hence unaffected by possible biases in the reconstruction of the shower core position.

The T5-Prior cannot be used in very inclined air showers. As the zenith angle  $\theta$  increases, the lateral shower profile comes flatter, distorted by geomagnetic deflections, and weaker. The core position is not as well pronounced and the station with the largest signal in an event is not necessarily the one closest to the true shower core. The examples in Fig. 6.11 illustrate this. The station with the largest signal fulfills the T5-Prior condition, but the shower core is still not well constrained.

The reconstructed core position needs to be included into the criterion, even though this requires to make the criterion more susceptible to possible reconstruction biases. The T5-Core completely relies on the reconstructed core for the decision. The strict T5-Posterior tries to extend the T5-Prior to rejected the cases in Fig. 6.11. It should be less dependent on possible reconstruction biases of the shower core. This study is based on the strict T5-Posterior.

In fact, Fig. 6.13 shows that both T5-Core and the strict T5-Posterior yield the same acceptance, which is encouraging. The impact of the strict T5-Posterior compared to the T5-Prior is significant at cosmic ray energies larger than  $10^{18.5}$  eV, where the strict T5-Posterior selects up to 5 % less events than the T5-Prior.

### 6.2.3 Reconstruction resolution and bias

Possible biases of the reconstruction and its resolution are assessed in the following with an analysis of simulated SD events. The analysis is performed on the library of 6480 SD events from Chapter 5. The reconstruction uses the parameters summarised in Table 6.1. Each event of the library is a full end-to-end simulation of an air shower. The only differences between simulated and real events are:

- the simulation uses a perfectly regular and fully developed SD array without holes or gaps,
- events always fall approximately in the center of the SD array, and
- accidental muons are not simulated.

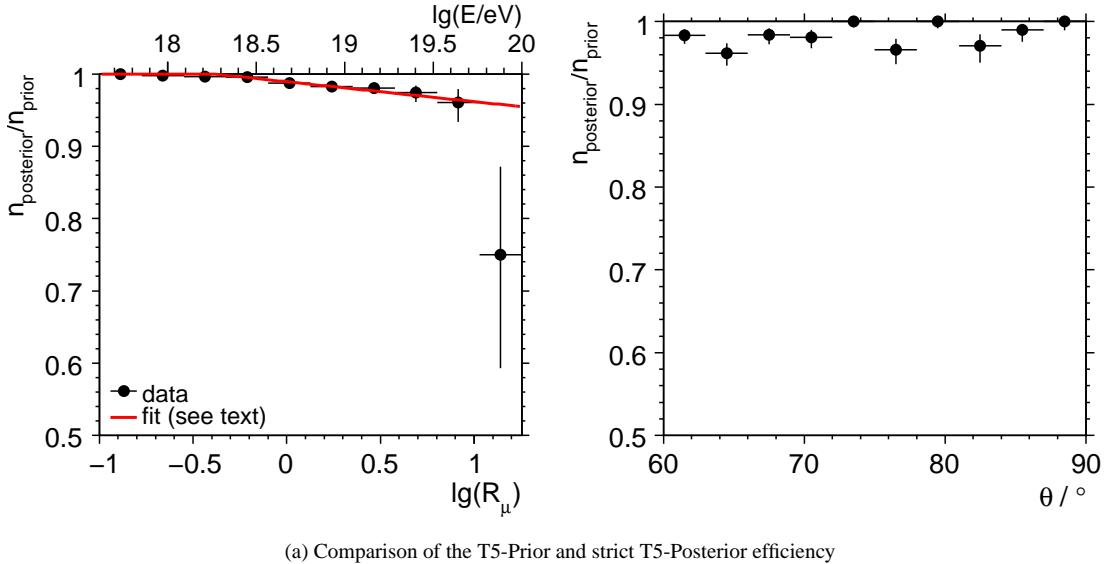
Fig. 6.14 shows an example event of the library. The reconstruction efficiency of simulated events that pass the T4 selection is 99.8 %.

The signal model of the SD station is built from the same data, the reconstruction of the simulation input therefore is as an end-to-end cross-check of the involved models. The comparisons of the reconstructed energy estimator  $R_\mu^{\text{SD}}$  and the shower core position  $\mathbf{r}_c$  are done in the same way as in Section 6.2.1. New is the analysis of the space angle  $\alpha$  between the true and the reconstructed shower direction. It is defined as

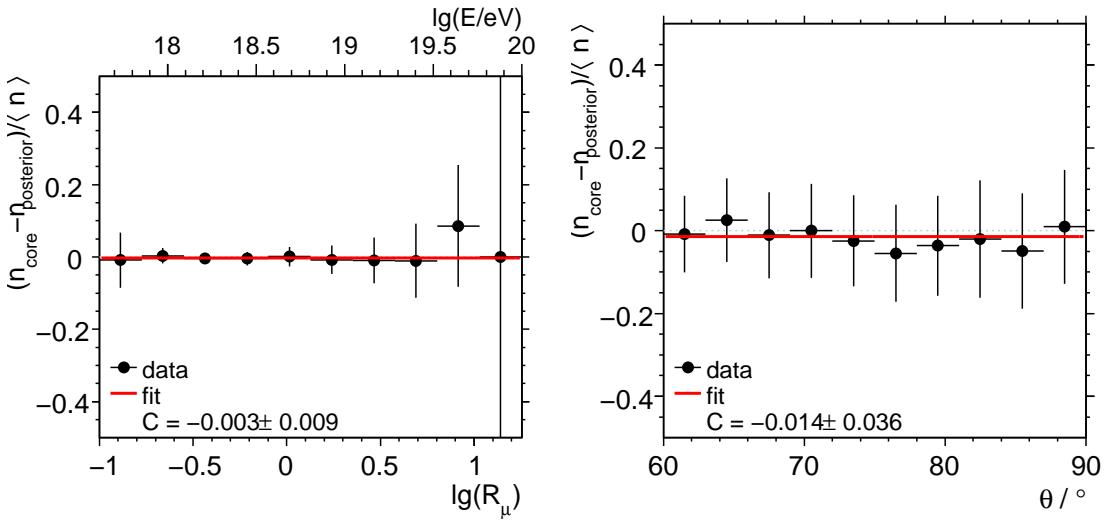
$$\cos \alpha = \mathbf{a}_{\text{sim}} \cdot \mathbf{a}, \quad (6.2.30)$$

Table 6.1: Settings for the `SdHorizontalReconstruction` module, which was used to produce the events in this note.

Parameter	Value
Model of the muon density at ground	this study
Model of the tank response	this study
Model of the electromagnetic signal component	ref. [16]
Signal threshold of silent stations	3.2 VEM
Maximum distance of silent stations included in the fit	5 km
Minimum number of triggered stations per event	4
T4 selection	ref. [173]
T5 criterion	strict T5-Posterior



(a) Comparison of the T5-Prior and strict T5-Posterior efficiency



(b) Comparison of the strict T5-Posterior and the T5-Core efficiency

Figure 6.13: The points show the relative efficiency of the T5 variants described in the text, if they are applied to real SD events. The red lines are fits. The analysis uses all events from 01/2004 to 12/2008, which pass the T4 selection, have at least a reconstructed energy and fall in the zenith angle range  $60^\circ < \theta < 90^\circ$ . The shower energy scale corresponding to the  $\lg R_\mu$ -scale is an anticipation of the energy calibration in Chapter 7. The last data point in the top left plot is a statistical fluctuation.

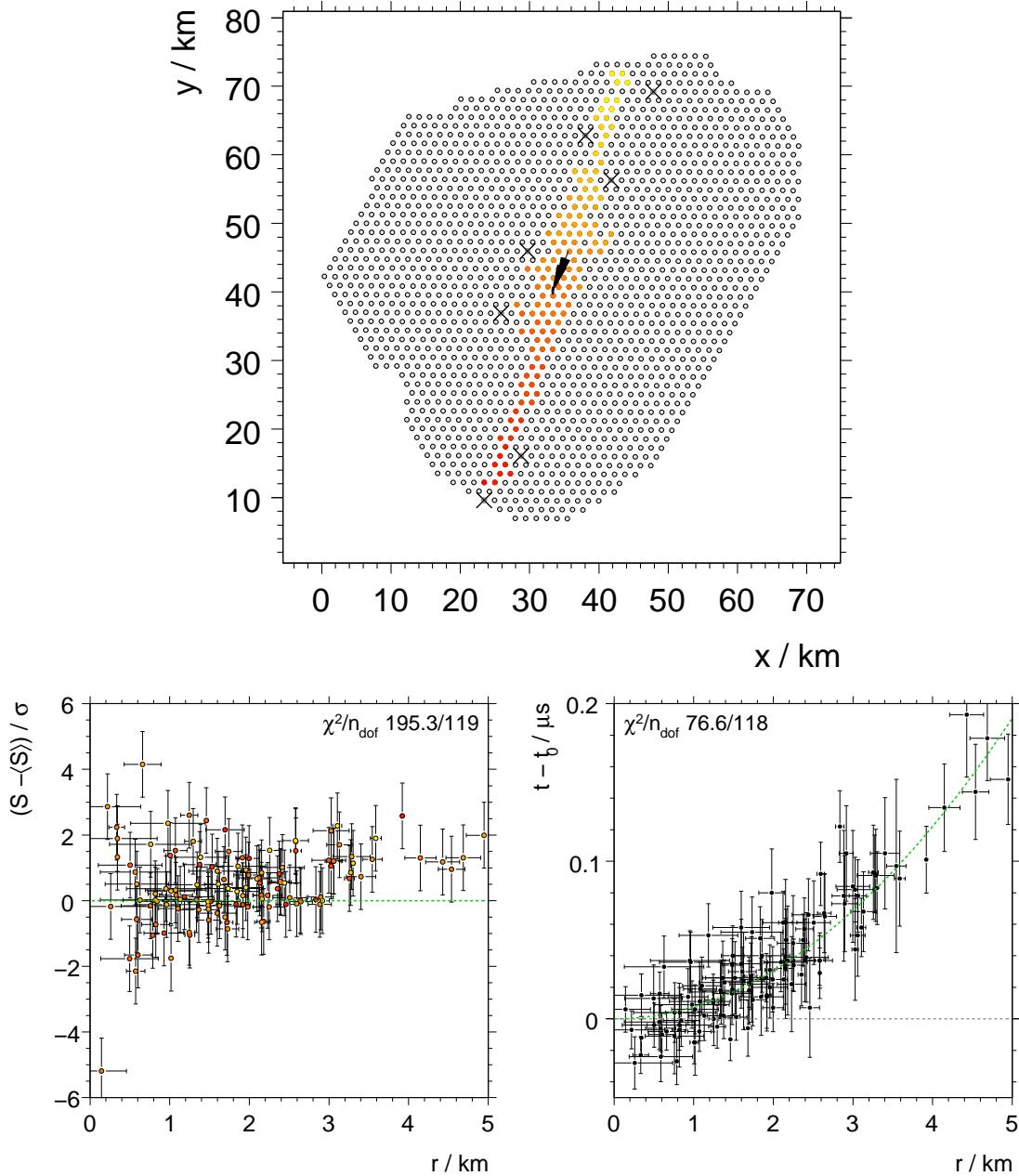


Figure 6.14: The plots show an example of a simulated surface detector event. The event is generated by a proton with  $E \approx 1.25 \times 10^{20} \text{ eV}$ ,  $\theta \approx 87^\circ$ , and  $\phi \approx 70^\circ$ . The high energy interaction model in the air shower simulation is EPOS. This particular cosmic ray triggered more than 120 stations. The upper plot shows the a top view of the surface detector, the symbols have the same meaning as in Fig. 6.6. The bottom left plot shows the residual of station signal and reconstruction model over the radial distance to the shower axis. The bottom right plot shows the time residual to a plane shower front moving with the speed of light (blue dashed line) over the radial distance. The prediction of the spherical shower front model is represented with a green dashed line.

whereas  $\mathbf{a}_{\text{sim}}$  is the true arrival direction of the shower in the simulation.

Fig. 6.15 shows examples of the residual distribution of the energy estimator  $R_{\mu}^{\text{SD}}$ , and the distributions of  $r_c$  and  $\alpha$ . The distribution of the energy estimator  $R_{\mu}^{\text{SD}}$  around its true value around  $10^{19}$  eV is a Gaussian in very good approximation. The resolution of the core deviation  $r_c$  and the space angle  $\alpha$  are taken as the 68 % quantiles of their respective distributions.

Fig. 6.16 and Fig. 6.17 show the reconstruction biases and resolution as a function of the input parameters of the shower. In some cases, the analyses shown in these figures are restricted to a certain energy or zenith angle range, because they would otherwise be completely dominated by large biases or low resolutions at energies below  $10^{18.5}$  eV or zenith angles larger than  $84^\circ$ . Why this happens is discussed in more detail below. Possible restrictions are shown in the legend of each figure.

### Analysis of bias

Fig. 6.16a) shows the bias of the reconstructed energy estimator  $R_{\mu}^{\text{SD}}$ . The most apparent features is a large overestimation around  $10^{18}$  eV and a beginning underestimation above  $\theta > 80^\circ$ .

The bias at the lowest energies is mainly caused by a threshold effect of the SD. If an air shower is barely able to trigger the minimum of four SD stations, it is more likely for events to be above the reconstruction threshold, if the sampling of the muon density in one or more SD stations shows random upward fluctuations.

Close to the threshold, where  $R_{\mu}^{\text{SD}}$  is reconstructed from very few sampled muons, these upward fluctuations are propagated to a large degree into the reconstructed energy estimator  $R_{\mu}^{\text{SD}}$ . This is a selection bias but has nothing to do with the precision of the reconstruction.

The other bias of  $R_{\mu}^{\text{SD}}$  at large zenith angles is a real reconstruction bias. Below  $80^\circ$ , the bias is smaller than 5 %. This is only slightly worse than the bias observed in the simplified reconstruction of ideal SD events from Section 6.2.1. This is a very good result, concerning that much more approximate model input is necessary to reconstruct real events.

Another encouraging result is that the reconstruction shows no significant bias of  $R_{\mu}^{\text{SD}}$  around  $\theta \approx 60^\circ$  as a function of the cosmic ray mass  $A$  or the hadronic interaction model used in the simulation. Some bias is expected at these zenith angles due to the contribution of electromagnetic particles from the hadronic cascade to the signal, as pointed out in Chapter 5. Apparently, the reconstruction is dominated by stations farther away from the shower axis, where the contribution of these electromagnetic particles is negligible.

Fig. 6.16b) shows the core shift along the arrival direction of the shower. A positive bias is a shift towards the early arriving shower part, a negative bias a shift to the late arriving shower part. The shift is negligible in the range  $60^\circ < \theta < 70^\circ$  and smaller than 200 m up to  $\theta = 80^\circ$ . A dramatic increase in the shift towards the late arriving shower part is observed at even larger zenith angles.

A core shift  $\tilde{r}_c$  has an influence on the reconstructed zenith angle  $\theta$ . The impact on  $\theta$  can be approximated as

$$\Delta\theta_{\text{core}} \approx -\frac{\tilde{r}_c}{d_{\max}} \cos \theta, \quad (6.2.31)$$

for small core shifts. With the approximate calculation of  $d_{\max}$  from Chapter 3 and the observed core shift, it is possible to estimate a bias  $\Delta\theta_{\text{core}} < 0.04^\circ$  in the zenith angle range  $60^\circ < \theta < 80^\circ$ . Below  $80^\circ$ , the effect on the zenith angle  $\theta$  is negligible.

Still, an energy-dependent bias in the zenith angle  $\theta$  of about  $\pm 0.1^\circ$  is observed in Fig. 6.16c). The bias is too large to be explained with the core bias  $\tilde{r}_c$  and therefore appears to be generated by the fit of the spherical shower front itself.

While anisotropy studies need to be aware of a bias of this order, the effect has only a mild effect on the reconstructed energy estimator  $R_{\mu}^{\text{SD}}$ . The effect of the zenith angle bias on  $R_{\mu}^{\text{SD}}$  can be estimated with the model reference of the total number of muons on the ground  $N_{\mu}^{\text{ref}} \propto R_{\mu}^{\text{SD}}$

$$\frac{\Delta R_{\mu}^{\text{SD}}}{R_{\mu}^{\text{SD}}} \approx \frac{1}{N_{\mu}^{\text{ref}}(\theta)} \frac{\partial N_{\mu}^{\text{ref}}(\theta)}{\partial \theta} \Delta\theta. \quad (6.2.32)$$

The calculation shows, that the zenith angle bias of  $0.1^\circ$  changes  $R_{\mu}^{\text{SD}}$  by about 0.5 % (1%, 2%) at  $60^\circ$  ( $80^\circ$ ,  $88^\circ$ ). The zenith angle bias is not able to explain the observed bias in the reconstructed energy estimator

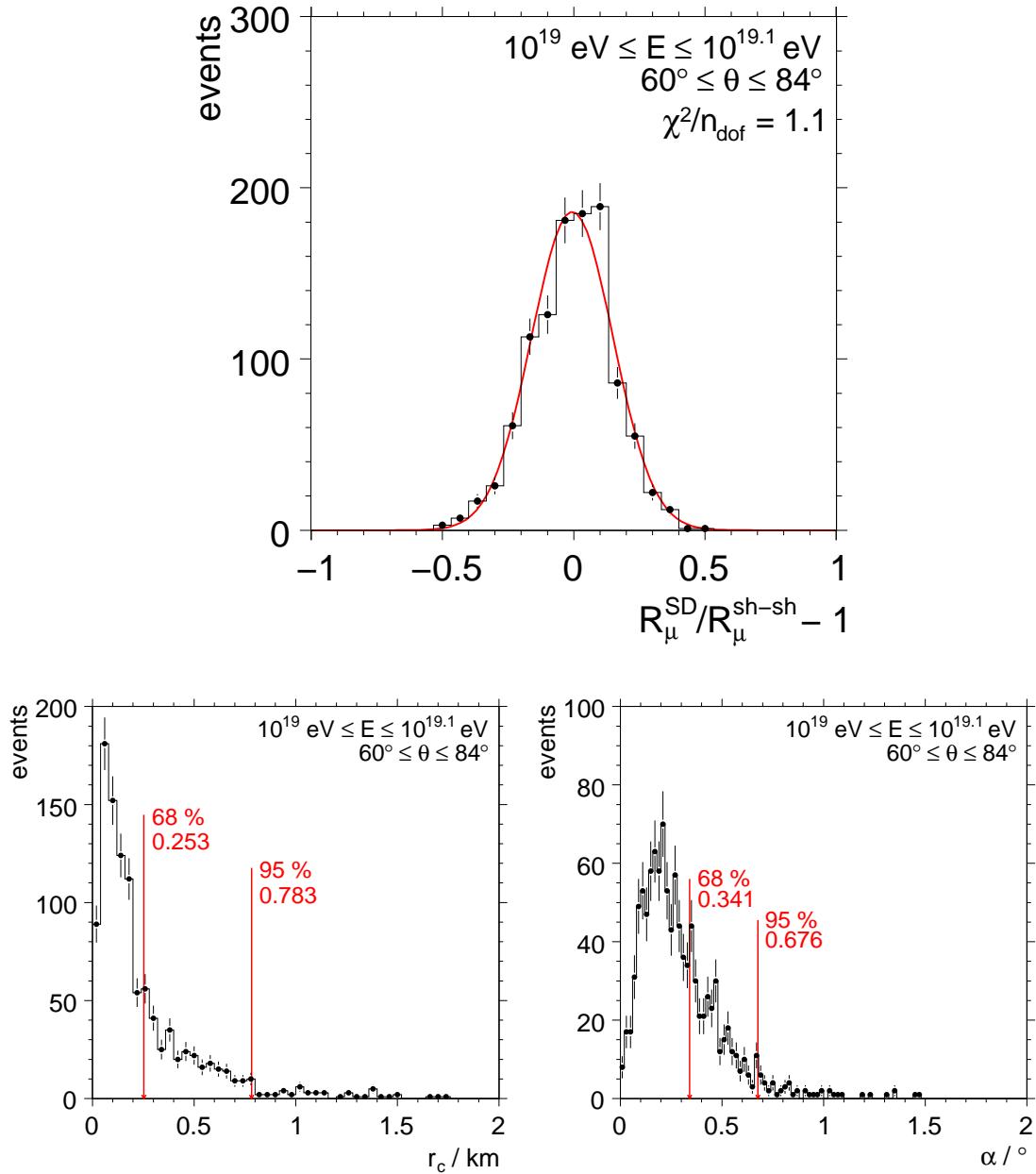


Figure 6.15: The plots show example distributions of the reconstructed energy estimator  $R_\mu^{\text{SD}}$ , shower core deviation  $r_c$ , and the angular deviation  $\alpha$  of the shower axis; as obtained from the reconstruction of simulated SD events. Only events in the energy range of  $10^{19} \text{ eV}$  to  $10^{19.1} \text{ eV}$  are included in the analysis. Events from proton and iron cosmic rays, simulated with QGSJet-II and EPOS, are combined. The red curve in the upper plot is the fit of a Gaussian. The red arrows in the lower plots indicate 68 % and 95 % quantiles.

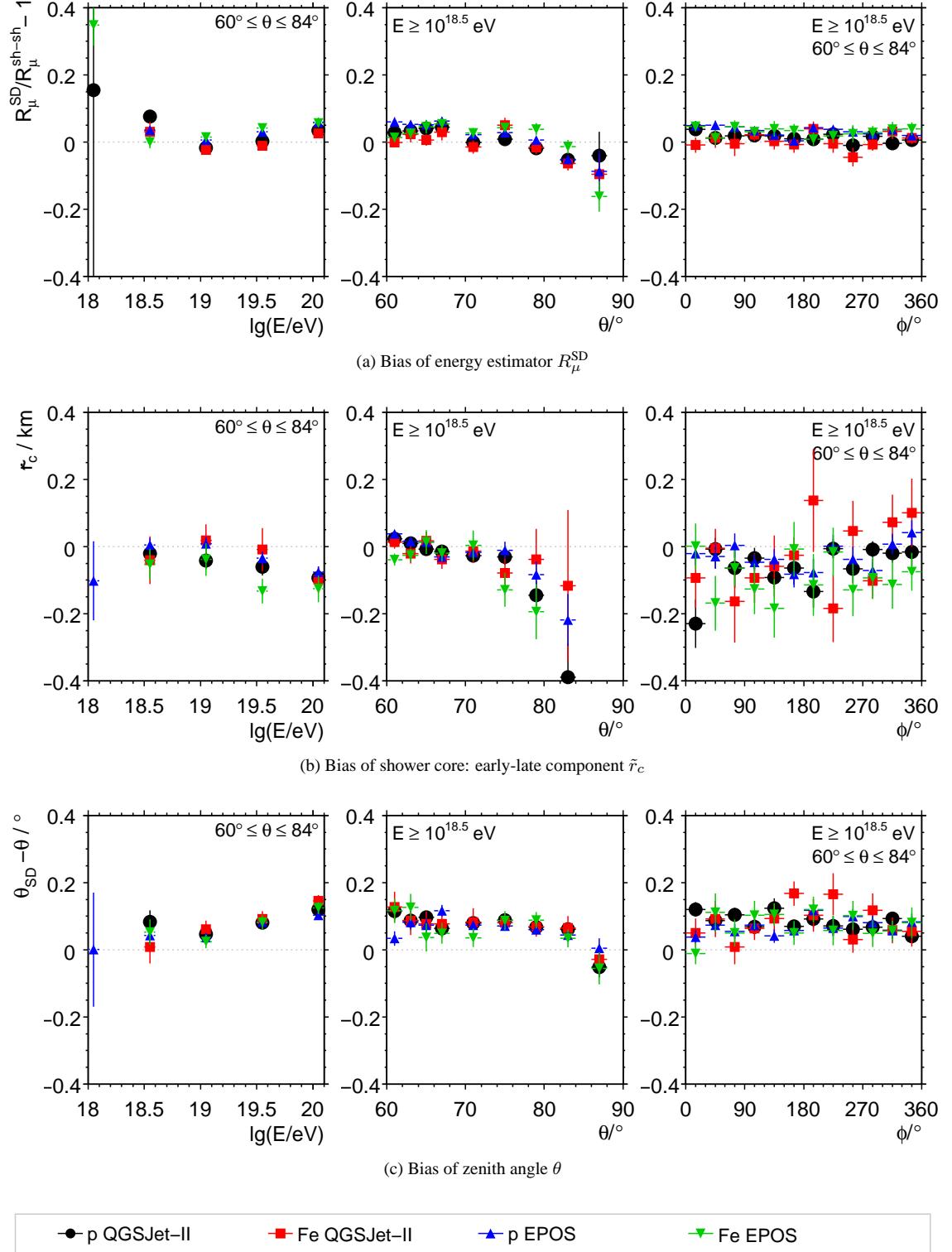


Figure 6.16: The plots show biases of some reconstructed parameters over the cosmic ray energy, zenith, and azimuth angle; as obtained from the reconstruction of simulated SD events. The core-bias at  $87^\circ$  is outside of the scale, it is around 3.5 km. The cosmic ray nuclei and hadronic interaction models used in the simulation are distinguished with different markers and colors.

$R_\mu^{\text{SD}}$ , which is apparently produced by yet another effect.

### Analysis of resolution

Fig. 6.17 shows the obtained parameter resolution. There is a general trend towards better resolutions at high energies, as the sampling of the SD becomes more precise. Another general observation is, that the reconstruction overestimates the true resolution.

The resolution of the energy estimator  $R_\mu^{\text{SD}}$  is shown in Fig. 6.17a). It increases almost with the logarithm of the primary energy, while it stays rather constant in the range  $60^\circ < \theta < 80^\circ$ . A typical value at  $10^{19}$  eV is 16 %. This is nearly a factor of two worse than the typical resolution obtained in the reconstruction of vertical showers [158]. The lower resolution in the reconstruction of very inclined showers is mostly an effect of the worse sampling of very inclined showers, since the primary electromagnetic component is missing as an additional signal source. As already noted in Section 6.2.1, the  $R_\mu^{\text{SD}}$ -resolution is lower for proton than for iron showers, because proton showers produce less muons.

The  $R_\mu^{\text{SD}}$ -resolution estimate from the maximum likelihood method is not far from the true resolution. This is a good indicator, as the true resolution includes statistical uncertainties and systematic biases. If the resolution estimate nearly agrees with the true resolution, then the systematic biases are small compared to the statistical resolution of the measurement.

The resolution of the shower core position  $r_c$  is shown in Fig. 6.17b). It is again weakly affected by the cosmic ray energy, but strongly so by the zenith angle. A typical resolution at  $10^{19}$  eV is 250 m, which is also by about a factor of two worse than the resolution obtained in analyses of vertical showers [158].

The estimated  $r_c$ -resolution does not agree well with the true resolution of the shower core above  $10^{19}$  eV. This is an indicator for remaining systematic effects, which reduce the core resolution.

At zenith angles larger than  $80^\circ$ , the core resolution becomes very bad. A proper calculation of the exposure of the SD array for showers with very large core uncertainties cannot be guaranteed by the usual T5 criteria anymore, and therefore they need to be excluded in analyses that are sensitive to the exposure.

The angular resolution is shown in Fig. 6.17c). It increases both with the energy and the zenith angle. The increase with the cosmic ray energy is particularly fast between  $10^{18}$  eV and  $10^{19}$  eV: the resolution improves from about  $1.2^\circ$  to  $0.4^\circ$ . The typical resolution at  $10^{19}$  eV of  $0.4^\circ$  is by about a factor of two better than the typical resolution found in vertical showers [158]. The angular resolution benefits from the compact time structure of the muon dominated shower front and the larger station multiplicity in very inclined showers.

The resolution of the energy estimator  $R_\mu^{\text{SD}}$  is affected by the angular resolution. The correlations between the reconstructed shower direction and the energy estimator are currently neglected, as explained in Section 6.2.2. The impact of the observed angular resolution is approximated by assuming a zenith angle resolution equal to angular resolution and using Eq. (6.2.32). The propagation yields a neglected contribution to the  $R_\mu^{\text{SD}}$ -uncertainty estimate of 4 % (2 %, 1 %) at  $10^{18.5}$  eV ( $10^{19}$  eV,  $10^{20}$  eV). The contribution may therefore be neglected.

In a similar way, the angular resolution is affected by the core resolution. The correlations are neglected here as well. The impact on the observed angular resolution is approximated by assuming a early-late shift equal to the core resolution and using Eq. (6.2.31). The propagation yields a neglected contribution to the  $\alpha$ -uncertainty estimate of  $0.4^\circ$  ( $0.07^\circ$ ,  $0.03^\circ$ ) at  $60^\circ$  ( $80^\circ$ ,  $88^\circ$ ). This shows, that the core resolution is an important factor for the angular resolution, it is even the dominant effect at  $60^\circ$ . A future revision of the reconstruction should therefore respect all correlations between the fits of the lateral signal profile and the shower axis.

The resolutions obtained from this analysis may be higher than those obtained from real events, where the array is not perfectly regular and fake signals generated by accidental muon hits are present.

### 6.2.4 Comparison with references

The software module `SdHorizontalReconstruction`, which implements the reconstruction of very inclined air showers in the `Offline` framework, was designed to allow an easy exchange of the models of

- the lateral profile of the muon density  $n_\mu^{\text{ref}}$ ,

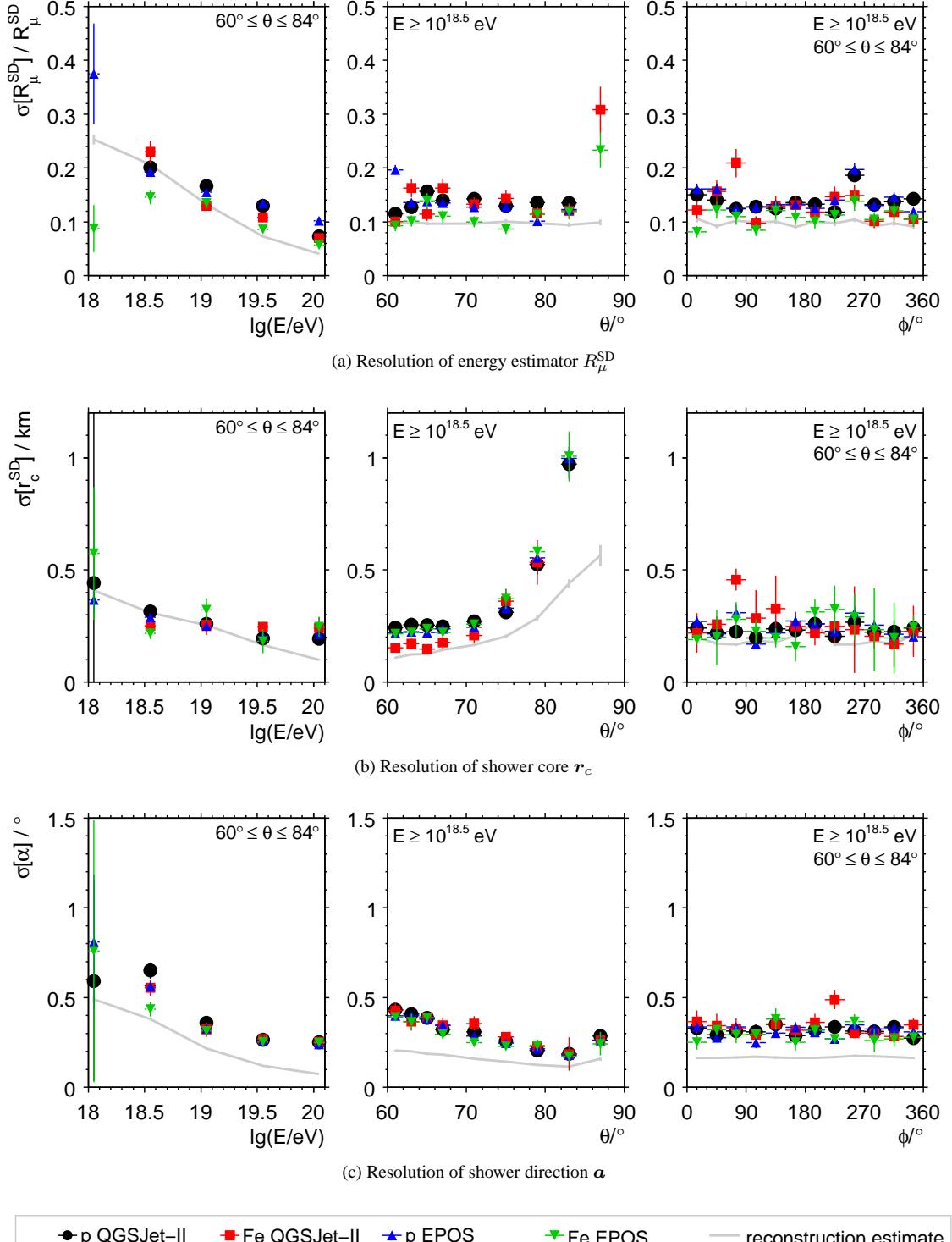


Figure 6.17: The plots show the resolution of the main reconstructed parameters over the cosmic ray energy, zenith, and azimuth angle; as obtained from the reconstruction of simulated ideal SD events. The core-resolution at  $87^\circ$  is outside of the scale, it is around 6 km. The cosmic ray nuclei and hadronic interaction models used in the simulation are distinguished with different markers and colors. The reconstruction estimate of the resolution of these parameters is represented with a solid gray line.

- the signal response  $f(S_\mu)$  of a station to muon hits, and
- the electromagnetic signal background  $\langle \epsilon \rangle$

in the reconstruction analysis. Models can be exchanged at run time with simple change in the setup of the module. This allows an easy comparison of models in a consistent analysis framework.

This feature is used to compare the models of  $n_\mu^{\text{ref}}$  and  $f(S_\mu)$  derived in this study with the standard models [10, 11, 15, 17] in the reconstruction software `efit` that were included into the `SdHorizontalReconstruction` module. The model of the electromagnetic background  $\langle \epsilon \rangle$  is the same in both reconstruction programs.

To ease the discussion, the following naming convention for the comparison is used:

- The standard model is called A.
- The corresponding model from this study is called B.
- There are four possible combinations of the lateral profile of the muon density  $n_\mu^{\text{ref}}$  and the signal response  $f(S_\mu)$  with two models for each type. All combinations are tested, they are distinguished with this key:
  - AA = both models are taken from the references,
  - AB =  $n_\mu^{\text{ref}}$  is taken from the references,  $f(S_\mu)$  from this study,
  - BA =  $n_\mu^{\text{ref}}$  is taken from this study,  $f(S_\mu)$  from the references,
  - BB = both models are taken from this study.

There are two remarks concerning the comparisons. The signal response model  $f(S_\mu)$  used in `efit` is a function of the muon energy  $E_\mu$  at the position of each station. This dependency is not implemented in `SdHorizontalReconstruction` and thus neglected in  $f(S_\mu)$ -A and  $f(S_\mu)$ -B. An extrapolation approach for the model of the muon density  $n_\mu^{\text{ref}}$  towards arbitrary large radial distances  $r$  from the shower axis is discussed in Chapter 5. This extrapolation is used for model  $n_\mu^{\text{ref}}$ -A and  $n_\mu^{\text{ref}}$ -B.

The models are compared according to the bias and resolution that they show in the reconstruction of simulated SD events. Another comparison is performed on real SD events, which uses a statistical test to rank the models according their agreement with the data.

### Simulated events

The model combinations are compared with the set of 6480 simulated SD events from Chapter 5. The standard models are not able to reproduce the true energy estimator  $R_\mu^{\text{sh-sh}}$  in these events very well. An overestimation of up to 20 % is observed, depending on the model combination. This bias is expected and its sources were already discussed in Chapter 5.

An overall bias is not of interest for the comparison. Such a bias would be absorbed in the energy calibration with FD events, which is described in the next chapter. Correction factors are introduced to make the models comparable. They are summarised in Table 6.2. The factors are chosen so that the average bias in the zenith angle range  $60^\circ < \theta < 70^\circ$  vanishes for cosmic ray energies above  $10^{18.5}$  eV.

The results of the comparison are shown in Fig. 6.18 and Fig. 6.19. The overall performance of all model combinations is similar. The  $f(S_\mu)$ -models have little effect on the reconstruction. Most of the differences are caused by the  $n_\mu^{\text{ref}}$ -models.

It is interesting to view these results in the light of the direct comparison of the  $n_\mu^{\text{ref}}$ -models in Chapter 5. The standard model  $n_\mu^{\text{ref}}$ -A neglects early-late asymmetries, which are present in  $n_\mu^{\text{ref}}$ -B. It appears that this explains a core shift towards the early arriving part of the shower caused by  $n_\mu^{\text{ref}}$ -A in the range  $60^\circ < \theta \lesssim 75^\circ$ .

This core bias seems to counter-balance an intrinsic bias in the reconstruction of the zenith angle  $\theta$  by coincidence. In consequence,  $n_\mu^{\text{ref}}$ -A shows a slightly smaller zenith angle bias than  $n_\mu^{\text{ref}}$ -B.

The bias of the energy estimator  $R_\mu^{\text{SD}}$  is the same for  $n_\mu^{\text{ref}}\text{-A}$  and  $n_\mu^{\text{ref}}\text{-B}$  in the zenith angle range  $60^\circ < \theta < 80^\circ$ . Differences in the resolution of  $R_\mu^{\text{SD}}$  are caused by the  $n_\mu^{\text{ref}}$ -models at  $\theta > 70^\circ$ , whereas  $n_\mu^{\text{ref}}\text{-B}$  yields a better resolution.

Considering the large differences of up to 100 % found in the direct comparison of  $n_\mu^{\text{ref}}\text{-A}$  and  $n_\mu^{\text{ref}}\text{-B}$  in Chapter 5, this result is quite surprising and shows that the reconstruction is rather insensitive to the shape of  $n_\mu^{\text{ref}}$ . The situation might be different, if only a part of the lateral profile is sampled, for example, when the shower falls close to the border of the SD array.

Differences between the  $n_\mu^{\text{ref}}$ -models also show up in the reconstruction of the shower core position. In the zenith angle range  $60^\circ < \theta < 80^\circ$ ,  $n_\mu^{\text{ref}}\text{-B}$  yields a smaller bias and a better shower core resolution than  $n_\mu^{\text{ref}}\text{-A}$ . However, at larger zenith angles,  $n_\mu^{\text{ref}}\text{-A}$  shows a smaller core bias than  $n_\mu^{\text{ref}}\text{-B}$ . This should be investigated further in the future.

In conclusion, the models of the lateral profile of the muon density  $n_\mu^{\text{ref}}$  and the signal response  $f(S_\mu)$  roughly show the same performance. The models derived in this study show some improvements with respect to the standard ones. Improvements in the  $n_\mu^{\text{ref}}$ -model are responsible for the gains.

### Real events

Simulated SD events have the advantage that the true values of the energy estimator  $R_\mu^{\text{SD}}$ , the shower core position  $r_c$  and the shower axis  $a$  are known. With this information, the models used in the reconstruction can be directly compared with respect to the bias and the resolution that they achieve.

The same analysis is not possible with real SD events, because the true values are not known. It is nevertheless important to compare models with real events, since simulated SD events may not describe all aspects of real events adequately.

There are other possibilities to compare reconstruction models on the basis of real events. One way is to compare the goodness-of-fit (GOF) of the model to the data or use the framework of hypothesis testing to decide between models, see *e.g.* ref. [147].

A GOF test well known to physicists is based on the  $\chi^2$ -statistic

$$\chi^2 = \sum_i \left( \frac{x_i - \langle x_i \rangle}{\sigma[x_i]} \right)^2,$$

defined as the sum of the quadratic residuals  $(x_i - \langle x_i \rangle)/\sigma[x_i]$  over many events. Asymptotically, the ratio  $\chi^2/n_{\text{dof}}$  approaches one for a model that describes the data correctly, whereas  $n_{\text{dof}}$  is the sum of the degrees of freedom per event.

Unfortunately, the  $\chi^2$  statistic cannot include the information obtained from silent or saturated stations as both yield only a limit for  $x_i$ . The likelihood  $L$  is another statistic which does not have this limitation. The *likelihood ratio test* (LR test) allows to decide between different models (hypotheses) based on  $L$ . It is not necessarily the most powerful test, but it is very simple to calculate.

The LR test shall be summarised in the following. Let  $\mathbf{X}$  be a vector of data, and let

$$f(\mathbf{X}|\xi), \quad g(\mathbf{X}|\zeta)$$

be the probability density functions (p.d.f.s) of two models of the data. The vectors  $\xi$  and  $\zeta$  shall be the free

Table 6.2: The table summarises the correction factors, which correct the overall bias in the reconstructed energy estimator  $R_\mu^{\text{SD}}$  in the compared model combinations.

model combination	AA	AB	BA	BB
$R_\mu^{\text{SD}}$ -factor	1.21	1.00	1.17	0.97

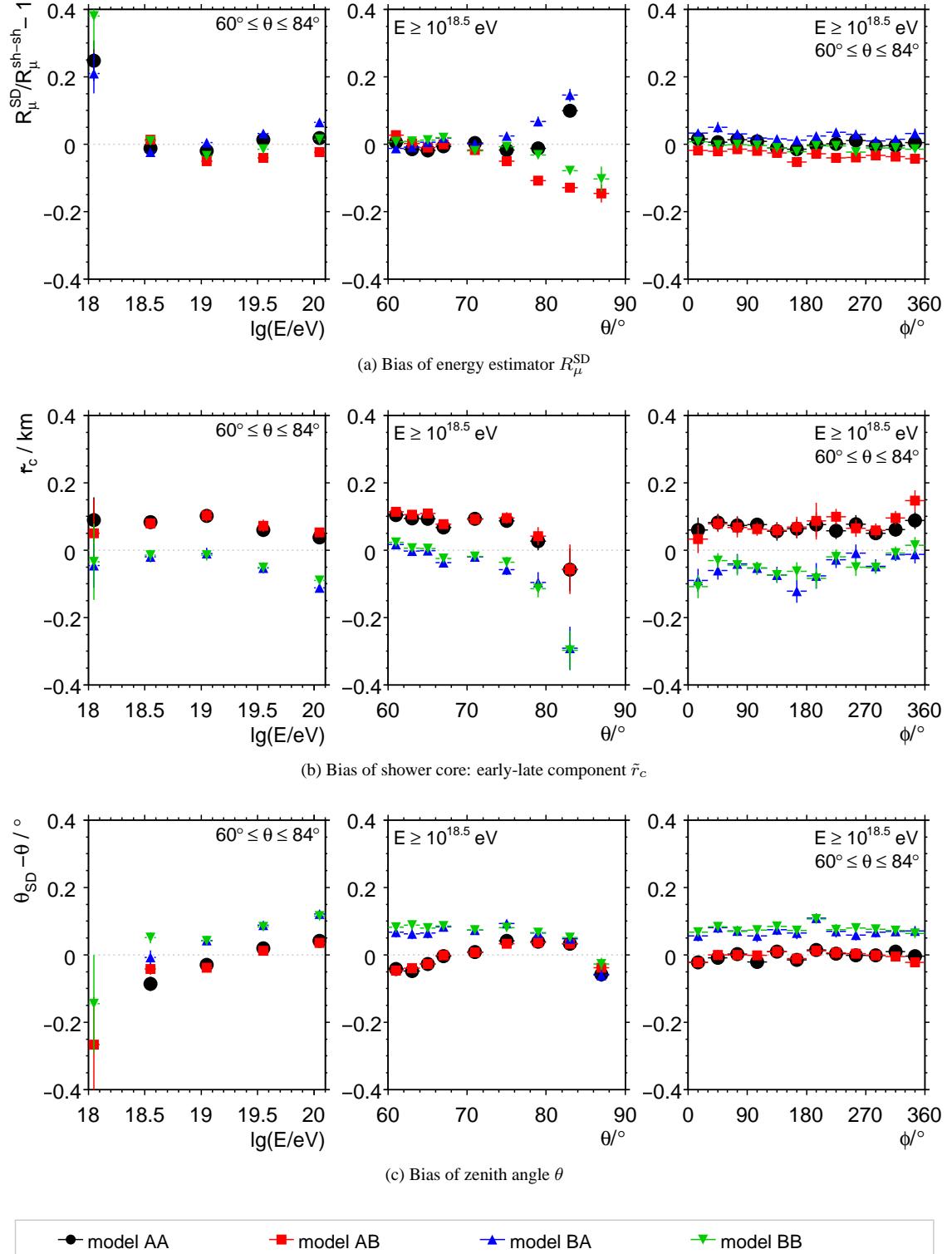


Figure 6.18: The plots compare the biases obtained from different reconstruction models over the cosmic ray energy, zenith, and azimuth angle; as obtained from the reconstruction of simulated SD events. The reconstruction models are distinguished by different markers and colors. The abbreviations “model AA”, “model AB”, and so on are explained in the text.

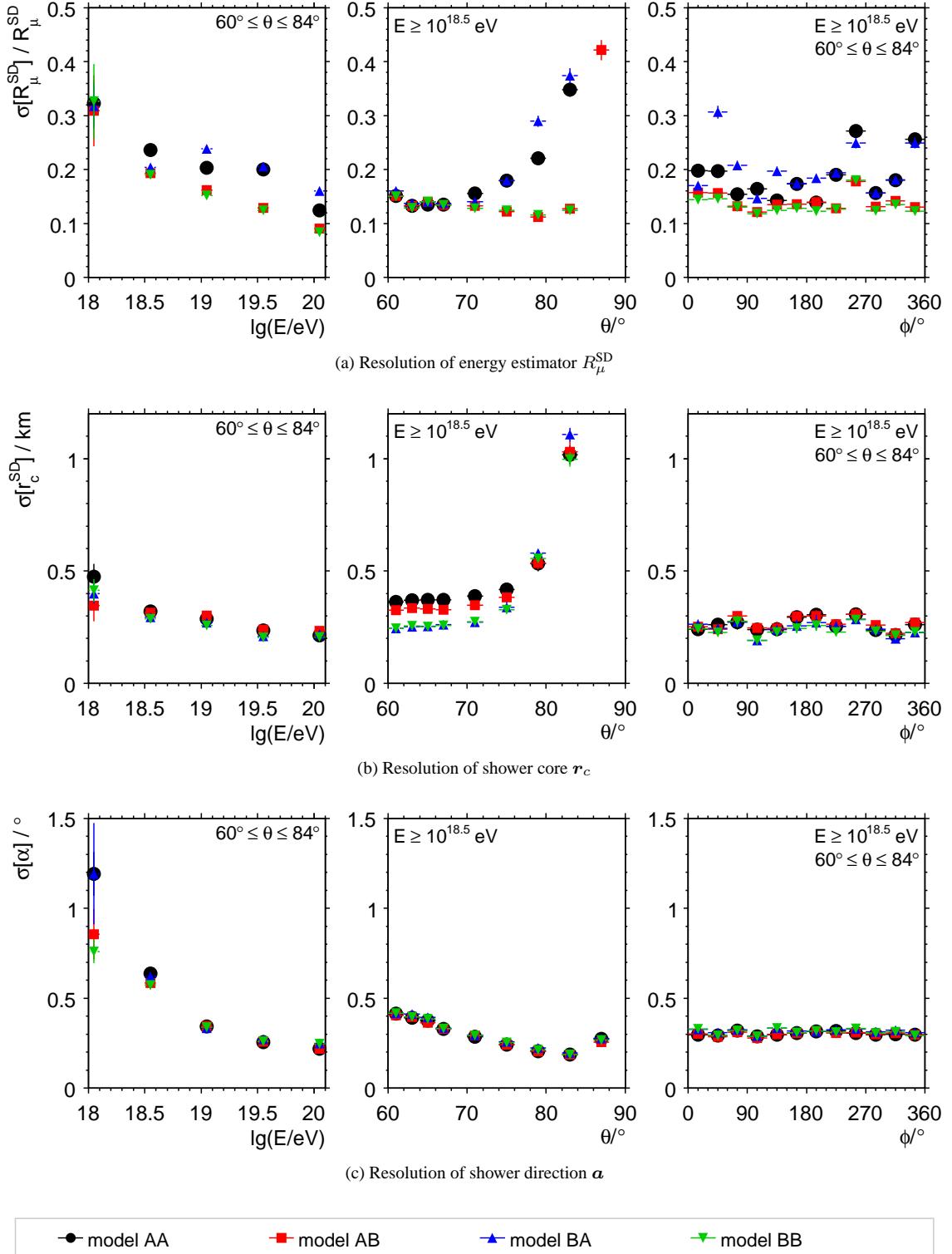


Figure 6.19: The plots compare the resolution obtained from different reconstructed models over the cosmic ray energy, zenith, and azimuth angle; as obtained from the reconstruction of simulated ideal SD events. The reconstruction models are distinguished by different markers and colors. The abbreviations “model AA”, “model AB”, and so on are explained in the text.

parameters of the models. The parameters may be fitted independently to the data  $\{X_i\}$ , by maximising

$$L_f(\boldsymbol{\xi}) = \prod_i f(X_i|\boldsymbol{\xi}), \text{ and}$$

$$L_g(\boldsymbol{\zeta}) = \prod_i g(X_i|\boldsymbol{\zeta})$$

respectively. The solutions of this maximisation are called  $\boldsymbol{\xi}'$  and  $\boldsymbol{\zeta}'$ .

Now, it is always possible to define a composite p.d.f.  $h$  as

$$h(\mathbf{X}|\eta, \boldsymbol{\xi}, \boldsymbol{\zeta}) = (1 - \eta)f(\mathbf{X}|\boldsymbol{\xi}) + \eta g(\mathbf{X}|\boldsymbol{\zeta}), \quad (6.2.33)$$

with  $0 \leq \eta \leq 1$ . The question of choosing  $g$  over  $f$ , given the data  $\mathbf{X}$ , can be reformulated as testing the hypothesis

$$H_0 : \eta = 0, \quad \boldsymbol{\xi}, \boldsymbol{\zeta} \text{ unspecified}$$

against the hypothesis

$$H_1 : \eta \neq 0, \quad \boldsymbol{\xi}, \boldsymbol{\zeta} \text{ unspecified.}$$

If  $H_0$  can be rejected with a high confidence, then  $f$  is disfavored compared to  $g$ .

The parameters of model  $h$  shall be fitted with a maximum likelihood method as well:

$$L_h(\eta, \boldsymbol{\xi}, \boldsymbol{\zeta}) = \prod_i h(X_i|\eta, \boldsymbol{\xi}, \boldsymbol{\zeta}),$$

yielding the solutions  $\eta'', \boldsymbol{\xi}''$ , and  $\boldsymbol{\zeta}''$ . The statistic of the LR test is

$$-2 \ln \lambda = -2 \ln \left( \frac{L_f(\boldsymbol{\xi}')}{L_h(\eta'', \boldsymbol{\xi}'', \boldsymbol{\zeta}'')} \right) = -2 \ln L_f(\boldsymbol{\xi}') + 2 \ln L_h(\eta'', \boldsymbol{\xi}'', \boldsymbol{\zeta}''). \quad (6.2.34)$$

It is possible to show, that the test statistic  $-2 \ln \lambda$  under the hypothesis  $H_0$  is asymptotically distributed like the  $\chi^2$ -distribution with one degree of freedom. This allows to calculate the probability  $P$  for observing  $-2 \ln \lambda$  under the hypothesis  $H_0$ . The value  $(1 - P)$  is then the confidence of disfavoring  $f$  over  $g$ , given the data.

This is the full LR test. With a loss of some precision it is possible to avoid the implementation of  $h$  and its numerical parameter optimisation, which makes the test very simple to apply. The simplification is based on the inequality

$$L_h(\eta'', \boldsymbol{\xi}'', \boldsymbol{\zeta}'') \geq L_g(\boldsymbol{\zeta}'), \quad (6.2.35)$$

which says that  $h$  can only fit equally well or better to the data than  $g$  alone. This is apparently so, because  $h$  includes  $g$ , but has more degrees of freedom to adapt to the data.

Finally, if the likelihood ratio  $\ell$  of the independently optimised models  $g$  and  $f$  is considered

$$\ell = \frac{L_f(\boldsymbol{\xi}')}{L_g(\boldsymbol{\zeta}')}, \quad (6.2.36)$$

and combined with Eq. (6.2.35), the following inequality is obtained

$$\begin{aligned} -2 \ln \ell &= 2 \ln L_g(\boldsymbol{\zeta}') - 2 \ln L_f(\boldsymbol{\xi}') \\ &\leq 2 \ln L_h(\eta'', \boldsymbol{\xi}'', \boldsymbol{\zeta}'') - 2 \ln L_f(\boldsymbol{\xi}') = -2 \ln \lambda. \end{aligned} \quad (6.2.37)$$

A sufficiently large value of  $-2 \ln \ell$  guarantees that  $H_0$  is rejected, because  $-2 \ln \ell \leq -2 \ln \lambda$ . If for example  $-2 \ln \ell \geq 9$ , the model  $f$  is disfavored compared to  $g$  with a confidence greater than 99.8 %. These numbers are obtained from the  $\chi^2$ -distribution with one degree of freedom, as discussed before. The simplified version of the LR test is very handy, because it only relies on information, which is calculated during the reconstruction anyway: the logarithms of the maximum likelihood values.

The simple LR test is performed over the subset of all real SD events from 01/2004 to 01/2009 that pass the following requirements with each model combination:

- the event is accepted by the T4 selection,
- the reconstruction of the energy estimator was successful,
- the event is accepted by the strict T5-Posterior, and
- the zenith angle is in the range  $60^\circ < \theta < 82^\circ$ .

Fig. 6.20a) shows the  $-2 \ln \ell$  value of the test. All other model combinations are compared to the combination  $n_\mu^{\text{ref}}\text{-B}$  and  $f_\mu\text{-B}$ . The model of the signal response  $f(S_\mu)\text{-B}$  derived in this study fits better to the data than the standard model  $f(S_\mu)\text{-A}$ . On first sight, the muon density model of this study  $n_\mu^{\text{ref}}\text{-B}$  only yields better results if the events have more than seven stations, but appears to be worse than  $n_\mu^{\text{ref}}\text{-A}$  otherwise.

This lack of performance is traced back to a class of 29 out of 45000 events. Fig. 6.21 shows an example out of this class. The events have a saturated station, whose signal is greatly underestimated by  $n_\mu^{\text{ref}}\text{-B}$ , while the same station fits well with model  $n_\mu^{\text{ref}}\text{-A}$  in most cases. This produces a huge drop in the overall likelihood of model  $n_\mu^{\text{ref}}\text{-B}$ .

The model  $n_\mu^{\text{ref}}\text{-A}$  fits better in this case, because it predicts a much higher muon density close to the shower core than  $n_\mu^{\text{ref}}\text{-B}$ , as shown in Chapter 5. Both models yield about the reconstructed values despite the bad fits in this class of events. The difference in the reconstructed energy estimators  $R_\mu^{\text{SD}}$  is larger than  $1\sigma$  in only three events. Model  $n_\mu^{\text{ref}}\text{-B}$  outperforms  $n_\mu^{\text{ref}}\text{-A}$  if these 29 events are excluded, as shown in Fig. 6.20b).

In conclusion, the signal response model  $f_\mu\text{-B}$  derived in this study agrees better with the data than the standard model  $f_\mu\text{-A}$ . The gain is achieved by including some signal fluctuations in  $f_\mu\text{-B}$  which are neglected in  $f_\mu\text{-A}$ . The model of the muon density  $n_\mu^{\text{ref}}\text{-B}$  derived in this study agrees better with the data in most cases, but very badly in a rare class of 29 out of 45000 events. These events are fitted well with  $n_\mu^{\text{ref}}\text{-A}$ , which points to an underestimation of the muon density  $n_\mu$  very close to the shower core in model  $n_\mu^{\text{ref}}\text{-B}$ . The bad fit has no significant impact on the reconstruction and thus model  $n_\mu^{\text{ref}}\text{-B}$  is still slightly in favor due to its better overall performance.

### 6.2.5 Constant intensity analysis

The reconstruction of simulated SD events shows, that the energy estimator  $R_\mu$  does not depend on the zenith angle  $\theta$  within certain limits. This result can be cross-checked with real SD events under certain conditions, based on the principle that the arrival directions of cosmic rays are isotropic in very good approximation. An analysis of this kind is often called method of the *constant intensity cut*, see e.g. ref. [155, 156, 158].

The basic idea is that the expected  $\theta$ -distribution in a flat detector that measures an isotropic flux is a well known distribution. It is obtained by projecting the zenith angle distribution of an isotropic source onto a plane:

$$f(\theta) \propto \sin \theta \cos \theta \Leftrightarrow \tilde{f}(\sin^2 \theta) \propto 1. \quad (6.2.38)$$

The measured  $\sin^2 \theta$ -distribution therefore should be flat, if the SD reconstruction probability<sup>5</sup> for cosmic rays is about 100 % over the regarded zenith angle range. This is not true at low energies, but it can be assured by regarding only events above an energy estimator threshold  $R_\mu > R_{\mu,\text{thr}}$ .

The cut introduces a dependency of the distribution on  $R_\mu$ , which is the key point of the analysis. By design,  $R_\mu$  should be independent of  $\theta$ . If this not the case, then the  $\sin^2 \theta$ -distribution will not be flat for any threshold  $R_{\mu,\text{thr}}$ . The implicit assumption in this argument is, that the reconstructed zenith angle  $\theta$  is not significantly biased, too. A bias of about  $0.1^\circ$  is expected, its influence can be checked by varying  $\theta$  at the level of the bias.

The analysis is very sensitive to a possible bias in  $R_\mu$ , which is a consequence of the steeply falling energy spectrum of cosmic rays. If a power law spectrum  $AE^{-\alpha}$  is assumed, then the number of events

<sup>5</sup>Including the trigger probability.

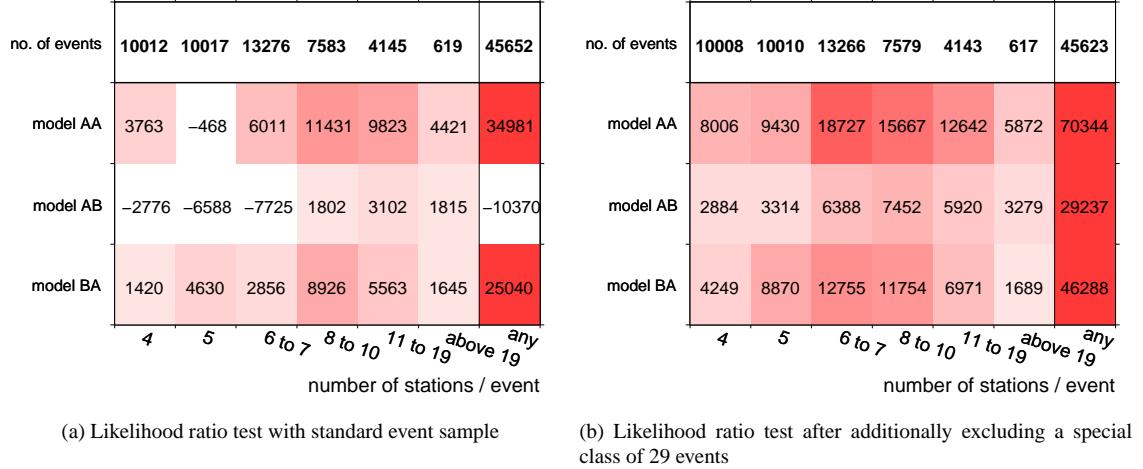


Figure 6.20: The plots shows the criterion  $-2 \ln \ell$  of the likelihood ratio test described in the text. The test statistic  $-2 \ln \ell$  is calculated against the model combination BB. The total value of the test statistic  $-2 \ln \ell$  is shown as well as the contributions from events with certain station multiplicities. If  $-2 \ln \ell > 9$ , the test favors the model combination BB. If  $-2 \ln \ell < -9$ , the test favors compared model combination.

above a threshold energy  $E_{\text{thr}}$  is

$$N(E > E_{\text{thr}}) = \int_{E_{\text{thr}}}^{\infty} dE A E^{-\alpha} = \frac{A}{\alpha - 1} E_{\text{thr}}^{-\alpha+1} \quad (6.2.39)$$

An energy-bias  $\Delta E$  is equivalent to a change  $-\Delta E_{\text{thr}}$  in the threshold, which can be propagated into the event number  $N$ :

$$\begin{aligned} \frac{\Delta N}{N} &\simeq (1 - \alpha) \frac{\Delta E_{\text{thr}}}{E_{\text{thr}}} = (\alpha - 1) \frac{\Delta E}{E} \\ \Rightarrow \quad \frac{\Delta R_{\mu}^{\text{SD}}}{R_{\mu}^{\text{SD}}} &\approx \frac{\Delta E}{E} \approx \frac{1}{2} N^{-1/2}, \end{aligned} \quad (6.2.40)$$

whereas a power law index  $\alpha \approx 3$  and Poisson statistics for  $N$  are assumed in the last step. Thus, 100 events per  $\sin^2 \theta$ -bin are already sufficient to detect a 5 %-bias in  $R_{\mu}^{\text{SD}}$ .

The analysis is performed with the settings in Table 6.1 on SD events from 01/2004 to 01/2009. Fig. 6.22 shows the resulting distributions above several energy estimator thresholds.

A flat  $\sin^2 \theta$ -distribution is indeed observed above a threshold of about  $R_{\mu, \text{thr}} \approx 1$ . The value fits well to the point of 100 % reconstruction efficiency obtained in the trigger threshold analysis in Chapter 7. A flat plateau is observed in the zenith angle range  $60^\circ < \theta < 82^\circ$ . The deviation above  $82^\circ$  is expected partly because of the reconstruction bias at these inclinations and partly because of the trigger saturation threshold which increases rapidly at these large inclinations. Below  $60^\circ$ , the event number drops down due to a zenith angle cut which is implemented in the T4 selection of very inclined air showers.

There may be a small lack of events between  $60^\circ$  and  $62^\circ$ . Statistically, the lack is not significant, but it is apparent by eye. It turns out that the flatness can be artificially restored by adding 2 % to the reconstructed energy estimator in this zenith angle region. A possible bias of at level is acceptable and smaller than the expected systematic uncertainty of  $R_{\mu}$  of the order of 5 %. As a final cross-check, the analysis is repeated with all zenith angles shifted by  $-0.1^\circ$  according to the expected bias in the reconstructed zenith angle. The result remains unchanged.

In conclusion, the analysis indicates a safe range of zenith angles  $60^\circ < \theta < 82^\circ$ , in which the energy estimator  $R_{\mu}$  is independent of  $\theta$  to a level of 2 %. The trigger saturation of the SD is reached at about  $R_{\mu} \approx 1$ . By rejecting the largest zenith angles  $82^\circ < \theta < 90^\circ$ , 8 % of the theoretical acceptance above  $60^\circ$  is lost. In practice it is much less due to the increased SD trigger threshold at the largest zenith angles.

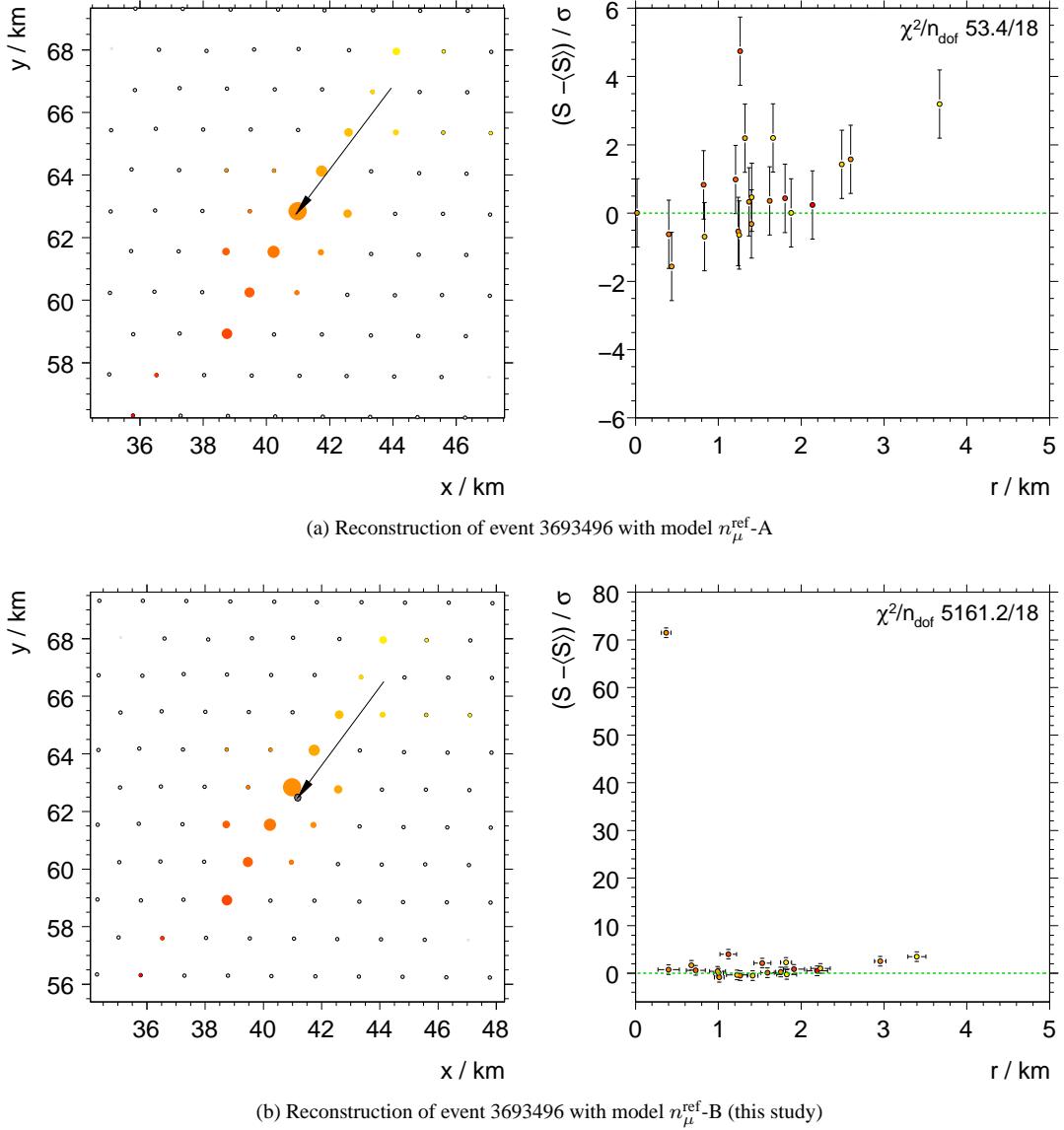


Figure 6.21: The plots show one of 29 events with saturated stations, which are significantly better fitted with model  $n_{\mu}^{\text{ref-A}}$  than with  $n_{\mu}^{\text{ref-B}}$ . The event has the reconstructed parameters  $E \approx 9 \times 10^{18}$  eV,  $\theta \approx 75^\circ$ , and  $\phi \approx 54^\circ$ . The left side shows a top view on the SD array, the symbols have the same meaning as in Fig. 6.6. The right side shows the residuals of the station signals with respect to the fitted reconstruction model. The station closest to the shower core is a saturated station.

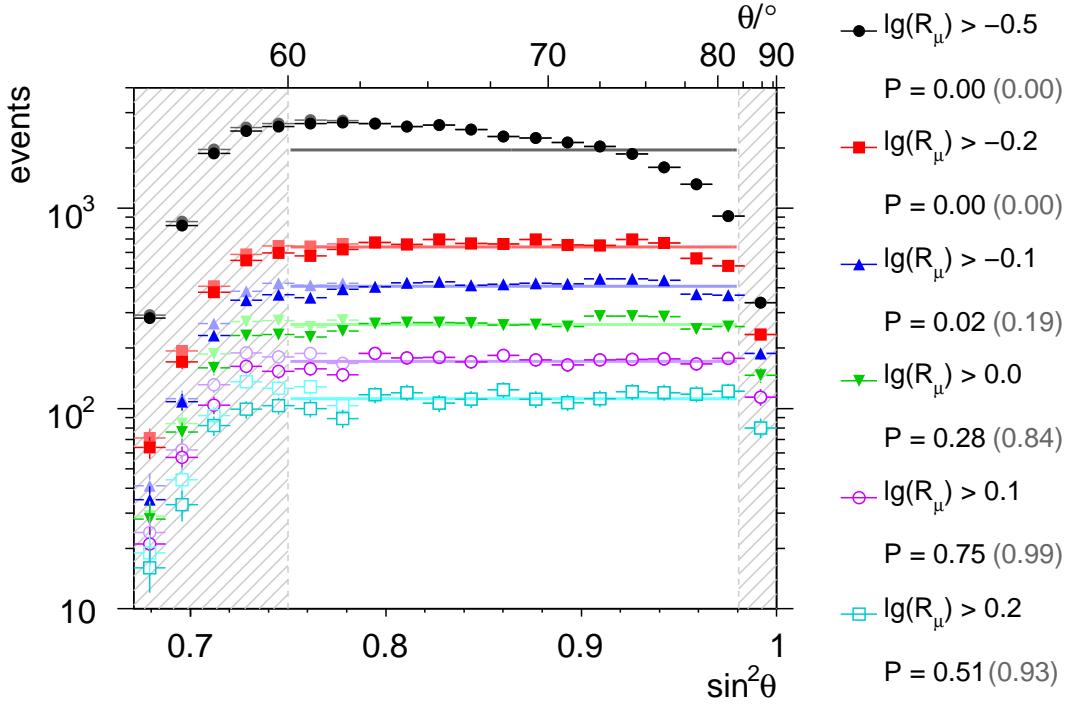


Figure 6.22: The plot shows the  $\sin^2 \theta$ -distribution of real SD events above several energy estimator thresholds. The hatched areas represent excluded zenith angle regions. The horizontal lines are fits to the distributions. The hypothesis that the distributions are flat is tested with the  $\chi^2$ -method. The probability  $P$ , that the observed fluctuations are by chance, is given in the legend. The lighter colored symbols show the distributions after applying the correction explained in the text. The probability in braces is the result of the  $\chi^2$ -test after the correction.

The constant intensity analysis does not exclude the use of events with  $R_\mu \lesssim 1$ , but the limited detector efficiency at these energies needs to be properly modeled. This will be done in Chapter 7.

### 6.2.6 Bias correction of the energy estimator

The previous analysis showed no indication of a significant bias of the reconstructed energy estimator  $R_\mu^{\text{SD}}$  as a function of the zenith angle  $\theta$  in the regarded range, but the energy-dependent bias observed in simulated events remains. Fig. 6.23a) shows the bias as a function of the true energy estimator  $R_\mu^{\text{sh-sh}}$ .

As discussed before, the bias at  $\lg R_\mu < 0$  can be attributed to a selection effect generated by the SD trigger, the shower-to-shower fluctuations and the SD sampling fluctuations. This apparent bias is a selection effect and cannot be corrected through the energy estimator. It will be corrected by properly regarding the limited detection efficiency in the following chapters.

The smaller bias at  $\lg R_\mu > 0$  is a reconstruction bias and therefore correctable. If this bias is not corrected, it will be absorbed into the energy calibration function in Chapter 7

$$E = E_{\text{cal}} \times R_\mu^\gamma. \quad (6.2.41)$$

The calibrated SD-energy  $E_{\text{SD}}$  would be almost bias free in any case. However, the fitted calibration constants would be biased in turn and not directly comparable with predictions from simulations. Therefore, the bias is corrected directly in the reconstructed energy estimator  $R_\mu^{\text{SD}}$ .

The bias grows almost linearly with  $\lg R_\mu^{\text{sh-sh}}$ . It is fitted in the range  $-0.2 < \lg R_\mu^{\text{sh-sh}} < 1.2$  with a

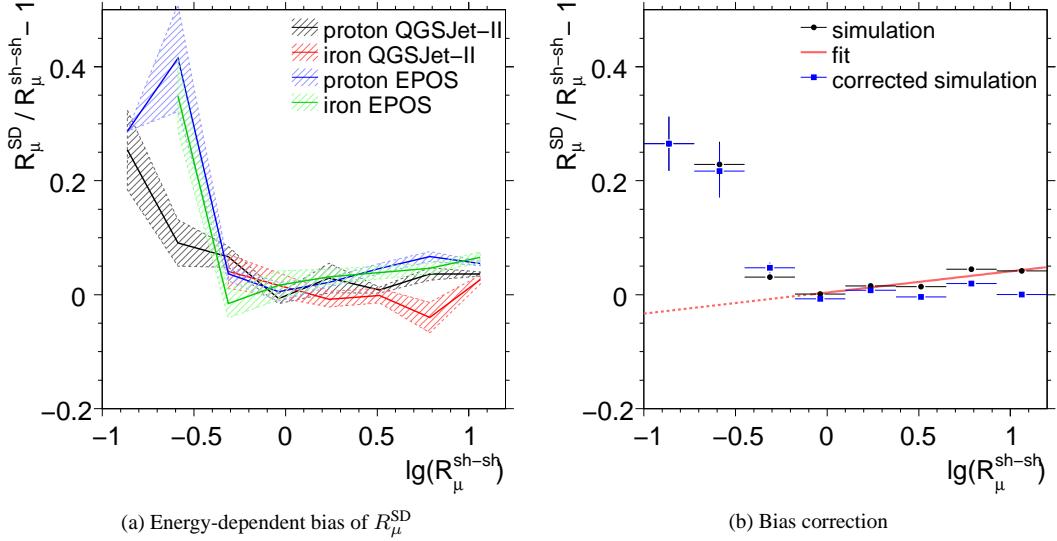


Figure 6.23: The plots show the energy-dependent bias of the reconstructed energy estimator  $R_\mu^{\text{SD}}$  in simulated events. The bias is shown as a function of the true energy estimator  $R_\mu^{\text{sh-sh}}$  in a) for different cosmic ray particles and hadronic interaction models. The hatched bands indicate the statistical uncertainty in the simulation. All simulated data is combined in b), which also shows the fitted correction function, and the event-by-event corrected data for comparison.

polynomial of first order

$$\frac{R_\mu^{\text{SD}} - R_\mu^{\text{sh-sh}}}{R_\mu^{\text{sh-sh}}} = a + b \lg R_\mu^{\text{sh-sh}} \quad (6.2.42)$$

$$\text{with } a = 0.0038 \pm 0.0035, \quad b = 0.0372 \pm 0.0041.$$

The fitted correction can be applied event-by-event by approximately replacing  $R_\mu^{\text{sh-sh}}$  with  $R_\mu^{\text{SD}}$ . This is possible, because the correction is at the level of  $10^{-2}$  and depends only on the logarithm of the energy estimator.

Fig. 6.23b) shows the uncorrected data, the fit, and the event-by-event corrected data, which is now unbiased above  $R_\mu \approx 1$ . This bias correction will be applied consistently in all analyses in the following chapters to real SD events and simulated SD events.

The constant intensity analysis showed that the reconstructed energy estimator  $R_\mu^{\text{SD}}$  is not biased as a function of the zenith angle  $\theta$  in the range  $60^\circ < \theta < 82^\circ$ . The bias of  $R_\mu^{\text{SD}}$  as a function of the cosmic ray energy  $E$  is corrected here. The remaining systematic uncertainty of  $R_\mu^{\text{SD}}$  is estimated to be smaller than 3 %, which is negligible. The reconstructed energy estimator  $R_\mu^{\text{SD}}$  will therefore be regarded as un-biased in the following chapters in the zenith angle range  $60^\circ < \theta < 82^\circ$ .

## 6.2.7 Summary

The section discussed the complete reconstruction of the energy estimator  $R_\mu$  and the shower direction  $(\theta, \phi)$  of very inclined air showers from events recorded with the surface detector. The energy estimator  $R_\mu$  is almost proportional to the shower energy  $E$  and will be calibrated to  $E$  in Chapter 7. The performance of the reconstruction was evaluated with simulated SD events. The reconstruction efficiency for events which pass the T4 selection is 99.8 %.

The resolution of the energy estimator  $R_\mu$  and the shower direction  $(\theta, \phi)$  depend on the cosmic ray energy  $E$ , its mass  $A$ , and the zenith angle  $\theta$ . For a  $10^{19}$  eV shower with an inclination  $\theta = 70^\circ$ , the resolution of the energy estimator is 16 % and the angular resolution  $0.4^\circ$ .

Based on bias-analyses of reconstructed cosmic ray parameters, the reconstruction is restricted to the zenith angle range  $60^\circ < \theta < 82^\circ$ . The restriction is mainly based on a rapidly increase in the bias of the shower core position  $\mathbf{r}_c$  at  $\theta > 80^\circ$ . The shower core resolution also degrades fast above  $80^\circ$ . Resolution and bias eventually become larger than the station-to-station distance of 1.5 km.

The adopted T5 selection depends on the reconstructed shower core position. The T5 selection rejects poorly reconstructed air showers and defines the effectively exposed area of the SD. An error in calculation of the exposed area directly propagates into reconstructed cosmic ray flux, as will be shown in Chapter 8. In order to avoid this the shower core bias should be small compared to the grid size of the SD array.

In the restricted zenith angle range, the reconstruction shows small or negligible systematic biases. The shower core position is biased by less than 200 m and the zenith angle by less than  $0.1^\circ$ . The energy estimator  $R_\mu$  shows a relative bias as a function of the cosmic ray energy of up to 3 % between  $10^{19}$  eV and  $10^{20}$  eV. This bias was corrected event-by-event with an empirical parameterisation. Any relative bias on  $R_\mu$  as a function of the zenith angle  $\theta$  is smaller than 2 %. The analyses of simulated events indicates that the energy estimator  $R_\mu$  reconstructs the true number of muons  $N_\mu$  on the ground with a systematic uncertainty of 3 %.

The reconstruction of the energy estimator  $R_\mu$  is based on a model of the muon density  $n_\mu$  on the ground and a model of the signal response  $f_\mu(S_\mu)$  of a SD station to muon hits. The models developed and used in this study are compared with other models [10, 11, 15, 17]. The compared models show an overall bias of 20 % in the reconstructed energy estimator  $R_\mu$ . The overall bias is generated by the model of electromagnetic signal component from ref. [16], as discussed in Chapter 5. The signal response model  $f_\mu(S_\mu)$  derived in this study absorbs this bias by construction is thus not affected.

If the overall bias is corrected, the performance of the compared models similar to those derived in this study in the restricted zenith range  $60^\circ < \theta < 82^\circ$ . Still, the models of this study achieve slightly better resolutions and fit better to real events in most cases.

Finally, an apparent bias in the reconstructed energy estimator  $R_\mu$  is observed below  $10^{18.5}$  eV which is not caused by the event reconstruction. The apparent bias is the consequence of random fluctuations of  $R_\mu$  and a selection effect of the detector which causes only the upward fluctuations to trigger. The effect will be analysed further and modeled in Chapter 7.

## 6.3 FD event reconstruction

The reconstruction of FD events is documented in ref. [180, 181] and the references therein. An overview is presented in the following. Technical details about the FD reconstruction can be found in Appendix D.

The reconstruction follows three steps. The signal pulses are extracted from the raw signal traces in every pixel and noise pixels are rejected. Then, the shower axis is reconstructed, which uses the signal arrival times in the pixels and a single SD station. Finally, the total shower energy is reconstructed from the detected light at the aperture based on the shower axis geometry and the current Mie scattering properties.

### 6.3.1 Pulse finding and rejection of random pixels

The reconstruction starts with a set of triggered camera pixels. Each pixel  $i$  covers a small solid angle of the sky with a pointing direction  $\mathbf{p}_i$ . The pixel has an ADC-trace of counts above a baseline with in intervals of 100 ns, as discussed in Chapter 4.

The first task is to locate the signal pulse generated by the air shower in each trace, which is done with a window search. The pulse is located by optimising the signal to noise ratio  $S/N$  within an adjustable time window  $\Delta t$

$$S/N = \frac{S(\Delta t)}{\sqrt{\Delta t} \times \sigma_b}, \quad (6.3.1)$$

whereas  $S(\Delta t)$  is the integrated signal in the time interval  $\Delta t$  above the baseline and  $\sigma_b$  is the fluctuation of baseline. Only pulses with a signal to noise ratio  $S/N > 5$  after the optimisation are accepted. After the pulse finding, each accepted pixel has a defined signal size  $w_i = S_i(\Delta t)$  and a signal arrival time  $t_i$ , which is the barycenter of the pulse in the time domain.

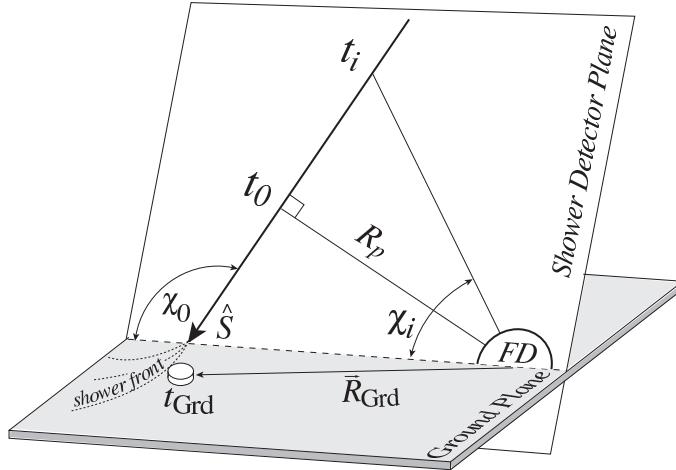


Figure 6.24: The drawing illustrates the shower axis reconstruction. The shower axis and the telescope position define the shower detector plane. The position and orientation of the axis inside the plane is defined by the shortest distance  $R_p$  to the telescope and its angle  $\chi_0$  to the ground plane. Light emitted at a certain elevation  $\chi_i$  above the horizon at the shower axis arrives in the telescope at the time  $t_i$ . The arrival time  $t_0$  corresponds to the point with the shortest distance to the telescope. The shower front arrives at the time  $t_{\text{Grd}}$  at position  $R_{\text{Grd}}$ .

The so far accepted pixels still contain a lot of accidental pixels, which are not part of the air shower. These are rejected in subsequent steps. Isolated pixels are rejected first. A pixel is isolated, if the angle between its own direction and that of every other signal pixel is larger than  $5^\circ$ .

To reject more accidentals, the shower detector plane (SDP) of the event is estimated. The SDP contains the shower axis and the fluorescence telescope, as shown in Fig. 6.24. It is fully defined by the position of the fluorescence telescope and the normal vector  $\mathbf{n}$  of the plane.

A preliminary SDP is obtained by calculating all possible normal vectors  $\tilde{\mathbf{n}}_{ij}$  from all pairs of pixel directions  $(\mathbf{p}_i, \mathbf{p}_j)$

$$\tilde{\mathbf{n}}_{ij} = \mathbf{p}_i \times \mathbf{p}_j. \quad (6.3.2)$$

The normal vectors  $\tilde{\mathbf{n}}_{ij}$  together with the telescope position define candidate SDPs. The best estimate is the candidate SDP, which has the largest number of compatible pixels. A pixel is compatible, if its direction  $\mathbf{p}_k$  diverges less than  $2^\circ$  from the candidate SDP. Finally, all pixels farther away than  $2^\circ$  from the estimated SDP are rejected.

This is the basic noise rejection. More heuristics are applied during the rest of the reconstruction to either further reject or re-include pixels, which are not covered here. The rejection up to this point is quite strict and optimal for showers which are distant so that they appear as a one-dimensional line in the telescope camera. If the shower is very close, the pixel track has a lateral extension and pixels may be rejected which are actually part of the event. Once the basic geometry of the shower axis is established, the reconstruction tries to recover such pixels based on the compatibility with the time structure of a developing shower.

### 6.3.2 Reconstruction of the shower axis

A precise reconstruction of the shower detector plane is performed with the cleaned set of pixels from the previous step. The SDP has only two degrees of freedom, the two components  $u$  and  $v$  of its normal vector  $\mathbf{n} = (u, v, \sqrt{1 - u^2 - v^2})^T$ . Best estimates of  $u$  and  $v$  are obtained by minimising the sum  $Q^2$  of the quadratic deviations from the SDP weighted with the signal size  $w_i$

$$Q^2 = \frac{1}{\sum_i w_i} \sum_i w_i \left( \frac{\arccos(\mathbf{p}_i \cdot \mathbf{n}) - \pi/2}{\sigma_p} \right)^2, \quad (6.3.3)$$

whereas  $\sigma_p = 0.35^\circ$  is a pointing uncertainty derived from an analysis of laser shots [182].

The position and orientation of the shower axis inside the shower detector plane is reconstructed next by fitting a model of the time-development of the light spot in the camera to the observed signal arrival times. The shower is modeled as a light source moving with the speed of light along the shower axis. A geometrical analysis yields

$$\langle t_i \rangle = t_0 + \frac{R_p}{c} \tan \left( \frac{1}{2} (\chi_0 - \chi_i) \right), \quad (6.3.4)$$

see Fig. 6.24, which also explains the meaning of the variables  $t_0$ ,  $R_p$ ,  $\chi_0$ , and  $\chi_i$ . In principle, the arrival times  $t_i$  measured in the pixels are enough to constrain the free parameters of the model  $t_0$ ,  $R_p$ , and  $\chi_0$ , but the precision is limited. In the worst case, the track is short and the measured elevations are close to  $\chi_0$ , so that the model degenerates to a line fit:

$$\chi_i - \chi_0 \approx 0 \Rightarrow \langle t_i \rangle \approx t_0 + \frac{R_p}{c} \frac{1}{2} (\chi_0 - \chi_i). \quad (6.3.5)$$

The parameters  $t_0$ ,  $R_p$  and  $\chi_0$  become strongly correlated and the shower axis is not well defined anymore.

In order to avoid this situation, the signal arrival time at a single SD station is included into the reconstruction. If the shower front curvature is neglected, the arrival time  $t_{\text{Grd}}$  in the SD station can be modeled as

$$\langle t_{\text{Grd}} \rangle = t_0 + \frac{1}{c} \mathbf{R}_{\text{Grd}} \cdot \hat{\mathbf{S}}, \quad (6.3.6)$$

the variables are again explained in Fig. 6.24.

The least-squares method is used to derive the free parameters of the model. The following sum is minimised:

$$\chi^2 = \sum_i \left( \frac{t_i - \langle t_i \rangle}{\sigma[t_i]} \right)^2 + \left( \frac{t_{\text{Grd}} - \langle t_{\text{Grd}} \rangle}{\sigma[t_{\text{Grd}}]} \right)^2, \quad (6.3.7)$$

whereas  $\sigma[t_i]$  and  $\sigma[t_{\text{Grd}}]$  are the respective uncertainties of the arrival time measurements in the pixels and the SD station.

The SD station used in the fit is selected out of those, which have a maximum distance from the shower detector plane of 2 km. The station with the largest signal is tried first. If this station does not lead to an acceptable fit, the station with the next to largest signal is used and so on.

This is the basic reconstruction concept. Some additional heuristics are applied to get a good SD station for the fit, which are not covered here. The effect of the shower front curvature on the arrival time  $t_{\text{Grd}}$  is also treated in an approximate way, which is not described here.

The typical angular resolution of the axis reconstruction is about  $0.5^\circ$  [183].

### 6.3.3 Reconstruction of the shower energy

The energy reconstruction of the Pierre Auger Observatory uses fluorescence and Cherenkov light emitted by an air shower to reconstruct the energy of the cosmic ray. The scattering and attenuation of the light in the atmosphere needs to be modeled. Both Rayleigh and Mie scattering processes are considered. The former describes scattering at air molecules, the latter scattering at aerosols and water vapor.

The reconstruction neglects the lateral extension of air showers. They are treated as one-dimensional objects. Fig. 6.25 illustrates the contributions of fluorescence light as well as direct and scattered Cherenkov light to the observed number of photons at the aperture of the telescope.

Electrons are the second most abundant particle species in an air shower after photons. They loose energy in the atmosphere rapidly, partly by colliding with nitrogen molecules. The molecules are excited in the process and partly de-excite by emitting fluorescence photons. The number of fluorescence photons  $N_\gamma^f$  emitted in a slant depth interval  $\Delta X_i$  is proportional to the average energy loss in that interval  $(dE/dX)_i$  and the fluorescence yield  $Y_i^f$  at that point in the atmosphere

$$N_\gamma^f(X_i) = Y_i^f (dE/dX)_i \Delta X_i, \quad (6.3.8)$$

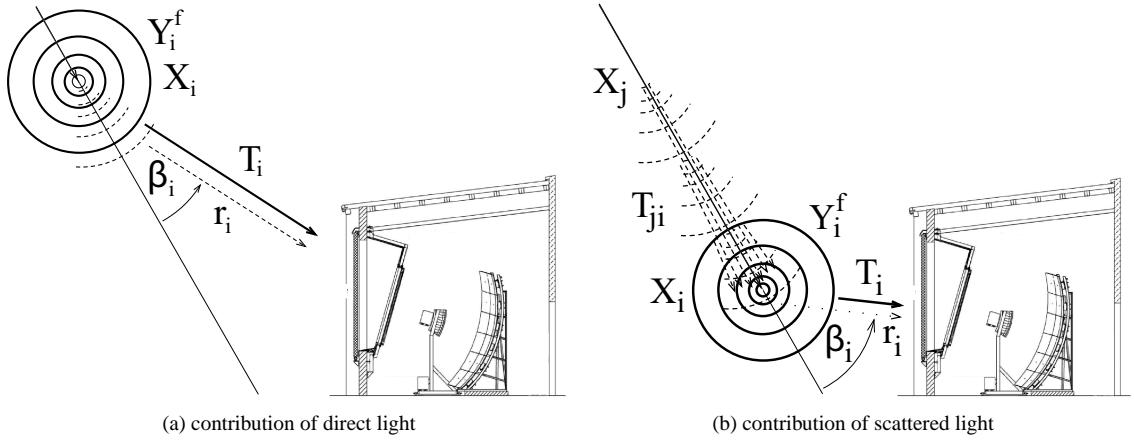


Figure 6.25: Illustrated are the contributions of direct and scattered light to the amount of photons detected at the aperture. The thick solid lines represent fluorescence light, which is emitted isotropically. The dashed lines represent direct Cherenkov light. The dotted line represents scattered Cherenkov light. In a), fluorescence light and Cherenkov light are emitted at  $X_i$ . The fluorescence yield  $Y_i^f$  at the emission point has to be regarded for the former, the emission angle  $\beta_i$  to the shower axis for the latter. The mixed light travels the distance  $r_i$  to the telescope and is attenuated by the factor  $T_i$  on its way. In b), the shower development has progressed further. A concentrated beam of Cherenkov light has formed along the shower axis. Direct fluorescence light from point  $X_i$  mixes with scattered Cherenkov light, which was initially emitted at  $X_j$  and attenuated between  $X_j$  and  $X_i$  by the factor  $T_{ji}$ .

The fluorescence yield  $Y_i^f$  is not constant during the shower development. It depends on the air density, temperature, and humidity, as discussed in Chapter 4.

Fluorescence light is emitted isotropically. The number of fluorescence photons  $y_i^f$  registered in a telescope with aperture  $A$  and light detection efficiency  $\epsilon$  is

$$y_i^f = \frac{A \epsilon T_i}{4\pi r_i^2} Y_i^f (\mathrm{d}E/\mathrm{d}X)_i \Delta X_i, \quad (6.3.9)$$

whereas  $T_i$  is the Rayleigh and Mie attenuation factor along the light path with the length  $r_i$ .

Most electrons in the shower are highly relativistic and therefore emit Cherenkov light in air. The Cherenkov light forms a concentrated beam along the shower axis. The number of Cherenkov photons in the frequency range of the FD telescopes is comparable to the number of fluorescence photons. Still in most cases the shower is observed from the side and only a small amount of scattered Cherenkov light reaches the telescope. If the shower faces the telescope, the detected signal may be completely dominated by Cherenkov light.

The number of Cherenkov photons  $N_\gamma^C$  emitted in the frequency interval of the telescopes is proportional to the number of electrons  $N_e(X_i)$  above the Cherenkov production threshold at a given slant depth  $X_i$ . The conversion factor can be modeled and is called Cherenkov yield  $Y_i^C$  in analogy to the previous case

$$N_\gamma^C = Y_i^C N_e(X_i). \quad (6.3.10)$$

The contributions of other fast charged particles, most notably the muons, can be neglected.

The number of direct Cherenkov photons  $y_i^{Cd}$  registered in the telescope is

$$y_i^{Cd} = \frac{A \epsilon T_i}{4\pi r_i^2} f_C(\beta_i) Y_i^C N_e(X_i), \quad (6.3.11)$$

whereas  $f_C(\beta_i)$  is the fraction of Cherenkov photons emitted at an angle  $\beta_i$  from the shower axis. The Cherenkov light cone of each individual electron is narrow. The total divergence of the beam is considerably larger due to multiple scattering of the electrons.

Finally, there is an indirect contribution from scattered Cherenkov light. The direct Cherenkov light emitted at earlier points of the shower development forms a Cherenkov beam along the shower axis, as indicated in Fig. 6.25b). The number of beam photons  $N_\gamma^{beam}(X_i)$  at a certain slant depth  $X_i$  of the shower development is the sum of all these earlier contributions

$$N_\gamma^{beam}(X_i) = \sum_{j=0}^i T_{ji} Y_j^C N_e(X_j), \quad (6.3.12)$$

whereas  $T_{ji}$  is the attenuation factor for light traveling from  $X_j$  to  $X_i$ .

The number of scattered Cherenkov photons  $y_i^{Cs}$  registered in the telescope is

$$y_i^{Cd} = \frac{A \epsilon T_i}{4\pi r_i^2} f_s(\beta_i) \sum_{j=0}^i T_{ji} Y_j^C N_e(X_j), \quad (6.3.13)$$

whereas  $f_s(\beta_i)$  is the fraction of Cherenkov photons emitted at an angle  $\beta_i$  from the shower axis.

The total light received at the detector in a given pixel  $i$ , which has the interval  $\Delta X_i$  of the air shower development in its field of view, is the sum of the three light contributions

$$y_i = y_i^f + y_i^{Cd} + y_i^{Cs}. \quad (6.3.14)$$

The fluorescence light contribution is directly related to the energy loss profile  $(dE/dX)_i$  of the shower. The Cherenkov light contributions depend on the electron number profile  $N_e(X_i)$ , but can be related to the energy loss, too:

$$(dE/dX)_i = \alpha_i N_e(X_i), \quad (6.3.15)$$

whereas  $\alpha_i$  is a universal conversion factor that depends only on the shower age  $s_i = 3/(1 + 2X_{\max}/X_i)$  in good approximation, whereas  $X_{\max}$  is the depth of the electromagnetic shower maximum. Thus, the energy loss profile  $(dE/dX)_i$  can be reconstructed from the detected light  $y_i$ . The contribution  $y_i^{Cs}$  from scattered Cherenkov light couples different shower intervals  $\Delta X_i$  with a single point  $(dE/dX)_j$  in the energy loss profile. The relation between  $y_i$  and  $(dE/dX)_i$  therefore is a matrix equation.

In general, the telescope is not able to observe the full profile because of its limited field of view. The extrapolation to depths outside the field of view is done with the Gaisser-Hillas function [184]

$$f_{GH}(X) = (dE/dX)_{\max} \left( \frac{X - X_0}{X_{\max} - X_0} \right)^{(X_{\max} - X_0)/\lambda} \exp((X_{\max} - X)/\lambda), \quad (6.3.16)$$

whereas  $(dE/dX)_{\max}$ ,  $X_{\max}$ ,  $\lambda$ , and  $X_0$  are free parameters. The full profile is also necessary to calculate the correct amount of scattered Cherenkov light in the observed fraction of the profile.

The Gaisser-Hillas function completes the data model. It is fitted to the available data with the least-squares method. The energy deposit  $E_{\text{em}}$  of the cosmic ray in the atmosphere is the integral of the fitted Gaisser-Hillas function  $f_{GH}(X)$

$$E_{\text{em}} = \int_0^\infty dX f_{GH}(X). \quad (6.3.17)$$

It was already discussed in Chapter 3, that  $E_{\text{em}}$  is smaller than the total energy  $E$  of the cosmic ray. About 10 % to 15 % is dumped into the ground by muons or carried away by neutrinos [47]. It is possible to model the average amount of this invisible energy and define a correction factor  $f_{\text{inv}}$ , which accounts for it. The final correction

$$E = f_{\text{inv}} E_{\text{em}} \quad (6.3.18)$$

completes the energy reconstruction.

The statistical uncertainty  $\sigma[E]$  of the reconstructed energy is derived from a full propagation of all event-by-event uncertainties. The uncertainty in the light flux at the aperture  $\sigma_{\text{flux}}[E_{\text{em}}]$  can be calculated from the uncertainty of the fitted Gaisser-Hillas function. Another contribution  $\sigma_{\text{geom}}[E_{\text{em}}]$  is generated by

the uncertainty in the reconstructed shower axis which affects the reconstructed energy. The last contribution is the event-by-event uncertainty  $\sigma[f_{\text{inv}}]$  of the correction factor  $f_{\text{inv}}$ , which shows shower-to-shower fluctuations. The total statistical uncertainty of the reconstructed energy is

$$\sigma_{\text{stat}}^2[E] = E^2 \sigma^2(f_{\text{inv}}) + \left( \frac{df_{\text{inv}}}{dE_{\text{em}}} E_{\text{em}} + f_{\text{inv}} \right)^2 (\sigma_{\text{geom}}^2[E_{\text{em}}] + \sigma_{\text{flux}}^2[E_{\text{em}}]). \quad (6.3.19)$$

The energy resolution is better than 10 % [181]. A detailed analysis of the FD energy resolution of very inclined air showers is shown in Chapter 7. Fig. 6.26 shows an example of a reconstructed fluorescence detector event.

### 6.3.4 Systematic uncertainty of the energy

In Chapter 7, the energy estimator of the surface detector is calibrated to the energy measurement of the fluorescence detector. The SD energy  $E_{\text{SD}}$  obtained in this way inherits the systematic uncertainty of the FD energy  $E_{\text{FD}}$ . The contributions to this systematic uncertainty were carefully estimated in ref. [185]. A listing of the uncertainties is shown in Table 6.3. The contributions can be divided into four large groups that concern the fluorescence yield, the signal calibration of the telescopes, the atmospheric model, and the reconstruction procedure.

The fluorescence yield is measured in the laboratory. The Pierre Auger Observatory uses the absolute value from Nagano et al. [73]. Relative measurements of the dependency on pressure, humidity, and temperature of the air are taken from the AIRFLY experiment [71, 72]. The absolute measurement of the fluorescence yield is technically difficult. Relative measurements are much more precise. The aim of the ongoing AIRFLY experiment is to reduce the absolute uncertainty of 14 % by at least a factor of two.

The absolute calibration of the FD telescopes at the Pierre Auger Observatory is performed with a light drum, as discussed in Chapter 4. The output of the light drum is measured in the laboratory. The light drum is then carried into the field to calibrate the telescopes. The absolute calibration of the light drum is one source of uncertainty. Another one is the drift of the calibration over time, which followed to some degree by the relative calibration procedure.

The atmospheric models have only very small systematic uncertainties as long as the parameters of the Mie scattering are well measured during the nights of data acquisition.

The uncertainty of the reconstruction method is assessed by comparing the standard method discussed here with the alternative from ref. [186]. The alternative reconstruction also regards the lateral extension of the shower, but does not model the scattered Cherenkov light. It is currently not possible to decide, which reconstruction is less biased. The fluorescence detector simulation also neglects the lateral extension of the shower and cannot be used to distinguish between the reconstructions. Reconstructed laser shots from the central laser facility cannot be used, because they are also one-dimensional.

The correction for invisible energy is derived from air shower simulations. The correction depends on the mass of the cosmic ray and the hadronic interaction model used in the simulation [47]. A variation of the models and the cosmic ray mass yields the uncertainty estimate given in the table.

The contributions to the systematic uncertainty  $\sigma_{\text{sys}}[E]$  of the reconstructed energy amount to 22 %. The systematic uncertainty will become smaller by successive improvements in the future. A reduction to 15 % could already be achieved by improving the accuracy of the fluorescence yield by factor of two and by eliminating the systematic uncertainty of the reconstruction method.

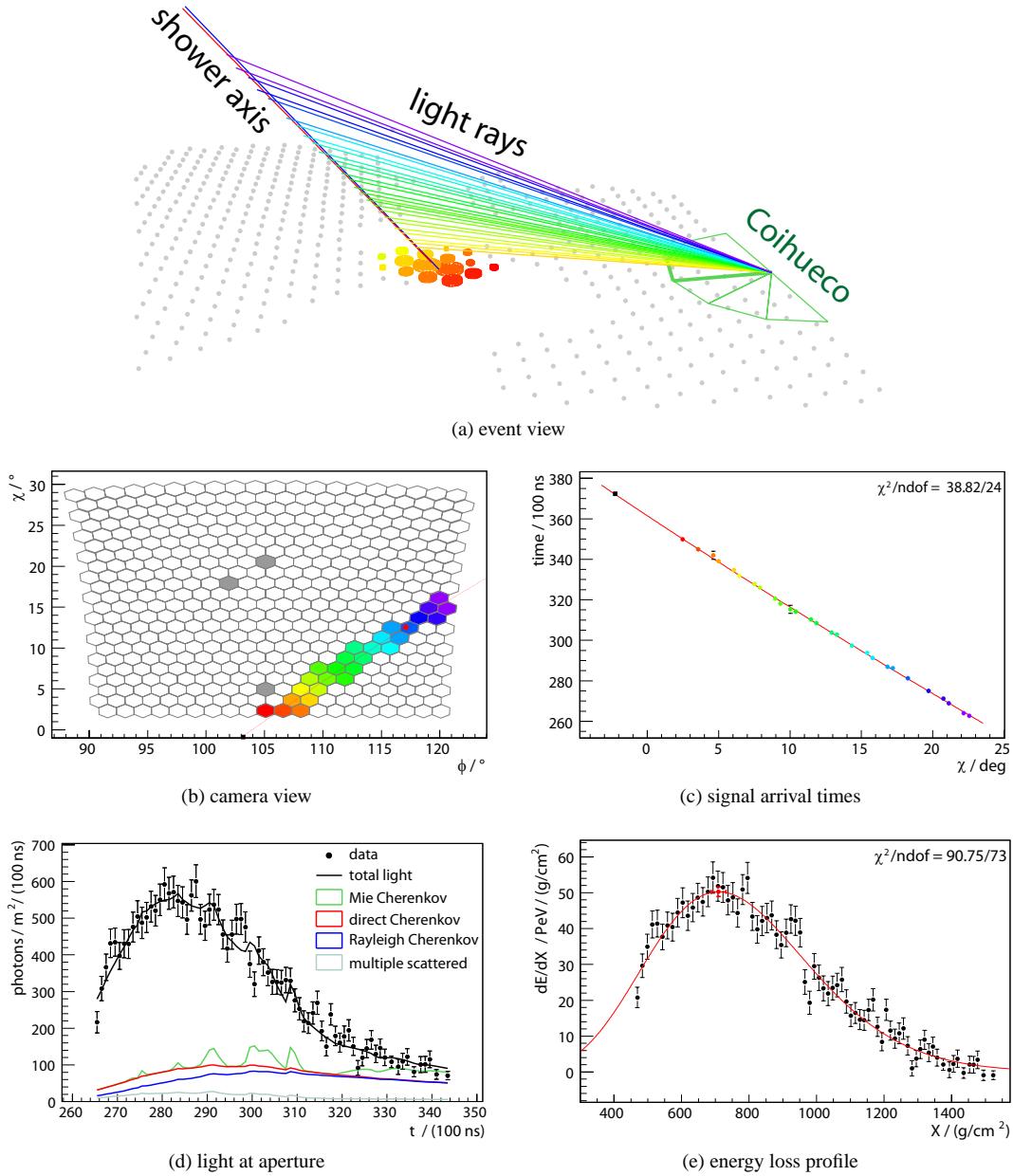


Figure 6.26: Shown in an example of a reconstructed FD event. Event 2151605 has a zenith angle  $\theta \approx 61^\circ$ , an azimuth angle  $\phi \approx 280^\circ$  and an energy of about  $3 \times 10^{19} \text{ eV}$ . a) The event view in shows the development of the shower, the signal arrival time is color coded. The shower also triggered several SD stations. b) The camera view shows the triggered pixels. The shower detector plane is indicated by a red line. c) The points show the signal arrival times and the solid line the fitted model. The black square represents the arrival time in the SD station with the largest signal. d) The points refer to the detected light at the aperture. The stacked profiles are the reconstructed contributions from fluorescence light as well as direct and scattered Cherenkov light. The contribution from Mie scattering is irregular, because the aerosol density is irregular. e) The points show the reconstructed energy loss profile, the solid line the fitted Gaisser-Hillas function.

Table 6.3: Listed are the sources, which contribute to the systematic uncertainty of the reconstructed FD energy [185]. The total uncertainty is obtained by adding the contributions in quadrature.

Source		$\Delta E/E [\%]$
Fluorescence yield	absolute	14
	pressure dependence	1
	humidity dependence	1
	temperature dependence	5
Telescope calibration	absolute	11
	wavelength dependence	3
Atmosphere	Rayleigh scattering	1
	wavelength dependence of aerosol scattering	1
	aerosol phase function	1
Reconstruction	method	10
	invisible energy	4
Total		22

# Energy calibration of the Surface Detector

This chapter discusses the energy calibration of the reconstructed energy estimator  $R_\mu^{\text{SD}}$  of the surface detector (SD) from Chapter 6. It is based on a correlation between the energy  $E_{\text{FD}}$  measured by the fluorescence detector (FD) and the energy estimator  $R_\mu^{\text{SD}}$ , which can be observed in events that triggered the fluorescence detector and the surface detector simultaneously.

The basic relation between the SD energy estimator  $R_\mu^{\text{SD}}$  based on the total number of muons which arrive the ground and the cosmic ray energy  $E$  was already suggested in Chapter 3 and confirmed through simulations in Chapter 5. It is a power law in good approximation, if the composition of the cosmic rays does not change too rapidly and the regarded energy range is not too large:

$$E = E_{\text{cal}} \times R_\mu^\gamma, \quad (7.0.1)$$

whereas  $R_\mu$  is the ideal energy estimator. The distinction between the ideal energy estimator  $R_\mu$ , the true energy estimator  $R_\mu^{\text{sh-sh}}$ , and the reconstructed energy estimator  $R_\mu^{\text{SD}}$  was explained in Chapter 6 and is again important in this chapter.

A theoretical prediction of the calibration constants  $E_{\text{cal}}$  and  $\gamma$  is a major challenge, as shown in Chapter 5. This is partly so, because the average cosmic ray mass  $\langle A \rangle$  at a given cosmic ray energy  $E$  is not well known and partly because of theoretical uncertainties concerning the hadronic interactions. It is more precise to derive the constants  $E_{\text{cal}}$  and  $\gamma$  from measurements. The FD provides an almost calorimetric measurement of the total energy  $E$  in an air shower, as discussed in Chapter 6, which can be used to calibrate  $R_\mu$ .

The standard approach [17, 161–163, 165, 187] expresses the average measured energy estimator  $\langle R_\mu^{\text{SD}} \rangle$  as a function of the measured FD energy  $E_{\text{FD}}$

$$\langle R_\mu^{\text{SD}} \rangle = \left( \frac{E_{\text{FD}}}{E_{\text{cal}}} \right)^{1/\gamma}. \quad (7.0.2)$$

This equation is fitted to the data with a least squares method that takes the estimated uncertainties from the reconstructions into account. An analogue approach is used to calibrate the energy estimator of vertical air showers, see *e.g.* ref. [156, 157, 188].

The method is simple and straight-forward, but has some issues, partly discussed already *e.g.* in ref. [17, 156, 157]:

- The standard least-squares fit offers no natural way to take the uncertainty on the FD energy  $E_{\text{FD}}$  into account, although it is only possible to approximately include it. The least-squares fit cannot handle fluctuations around the model that are not Gaussian.
- The event sample below the threshold energy of the SD trigger saturation is biased due to a selection effect. The least-squares method cannot model this bias and therefore events in affected energy range need to be excluded. Most showers are observed at low energies and thus a major amount of data cannot be used to constrain  $E_{\text{cal}}$  and  $\gamma$ .

- The low energy cut itself introduces another bias on the fit which is complicated to avoid.

This study proposes an improved approach for the energy calibration. The idea is to develop a complete statistical model of the FD and SD measurement processes, formulated as a probability density function (p.d.f.) in the  $(E_{\text{FD}}, R_{\mu}^{\text{SD}})$  parameter space. Deriving the *p.d.f.* is a complex task, but once it is formulated, the calibration constants and a lot more information can be derived from the data by fitting this model with a standard likelihood method to the observed event distribution.

The new method has many advantages, the two most important ones are:

- Events below the point of SD trigger saturation can be included naturally. This allows to use about three times as much data for the fit as the standard approach.
- The method allows to fit the unknown shower-to-shower fluctuations of the energy estimator  $R_{\mu}^{\text{SD}}$  from the data. These fluctuations cannot be derived from the data otherwise and are sensitive to the cosmic ray mass  $A$ . Thus, a comparison of the fitted shower-to-shower fluctuations with predictions from simulations allows to draw conclusions about the composition of cosmic rays.

The new method introduces a lot more modeling of the data, which could bias the result if it is not done correctly. The method and its model components are therefore checked thoroughly in this chapter.

First part of this chapter deals with an important prerequisite of the energy calibration: a well understood data set. The T5 criterion assures a good reconstruction with controlled systematics for SD events. An event selection with the same purpose needs to be applied to FD events, but in the FD case it is much more complex. The FD selection is based on cuts on many event parameters, which may introduce possible biases themselves. A self-consistent algorithm [158] is used to derive unbiased cuts.

The new energy calibration method is derived and applied to the available data in the second part of this chapter. Several cross-checks are applied, which show the general consistency of the approach and allow to estimate the systematic uncertainty of the result.

## 7.1 Event selection

A prerequisite to a fit of the energy calibration function from the data is a well understood and unbiased data set. It is particularly important, that the selected event sample does reflect the true composition of cosmic rays at every energy interval. Heavy nuclei with the same energy  $E$  produce more muons and therefore a larger energy estimator  $R_{\mu}$ . If the cosmic ray composition was biased in the event sample, then also the energy calibration would be biased.

It was shown in Chapter 6, that the SD energy estimator  $R_{\mu}^{\text{SD}}$  is unbiased if the SD is fully efficient. If it is not fully efficient, the average energy estimator  $\langle R_{\mu}^{\text{SD}} \rangle$  is biased at a given cosmic ray energy  $E$ . In particular, the limited efficiency would bias a mixed composition of cosmic rays. Iron nuclei with the same energy as protons have a higher chance to trigger the SD, because they generate more muons.

A similar situation can arise for the fluorescence detector. Fig. 7.1 shows an example. One of the basic requirements for a reliable FD reconstruction is the observation of the shower maximum in the field of view of the telescope. Proton showers develop deeper in the atmosphere than iron showers and thus have a larger probability to be accepted by this requirement if they develop close to the telescope. Because the FD exposure increases with the cosmic ray energy  $E$ , showers close to the telescope usually have low energies. At larger energies, the corresponding average distance to the telescope is also larger and no bias occurs.

There are two ways to deal with this kind of situation. If the bias in the events is well understood, it can be included into the data model, thus canceling its effect. This approach will be used for the SD energy estimator  $R_{\mu}^{\text{SD}}$  in the region where the SD is not fully efficient. If the bias is not well understood, cuts have to be applied which avoid the bias. This approach will be used for the FD energy  $E_{\text{FD}}$ .

The SD events are well understood and need no further event selection. This section will focus on the selection of FD events. The FD data set contains a fair amount of showers of poor quality since its event trigger is very permissive. A lot of these showers are systematically biased. There is not a single criterion like the T5, which allows to reject these showers. Only a complex combination of several cuts allows to avoid the biased events.

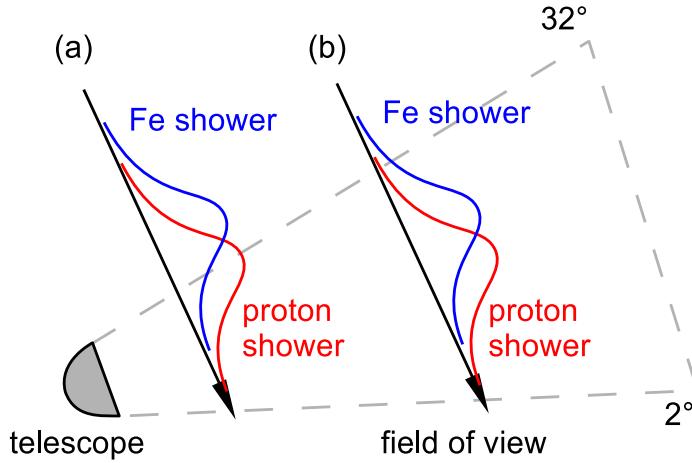


Figure 7.1: The drawing shows the field of view of a fluorescence telescope, and in a symbolic way, the energy loss profile of a proton and an iron shower along its shower axis. In case a), both showers arrive very close to the telescope, in case b), they are farther away.

The method to derive the FD cuts is adopted from ref. [157], where it was applied to select FD events for the energy calibration of vertical showers. The approach is based on a self-consistency argument. It exploits that the SD energy estimator  $R_{\mu}^{\text{SD}}$  is unbiased above a certain threshold and therefore may be used as a reference to obtain unbiased cuts for the FD. The method has the advantage that relies only on the data itself and not on air shower simulations.

### 7.1.1 SD event selection

Compared to the complex selection of FD events, the SD event selection is simple and robust. The applied cuts and selections are discussed in detail in Chapter 6. The only new cut regards certain time periods, where the central data acquisition was unstable or buggy. More information about these so called bad periods is given in Chapter 8.

To simplify the discussion, the following convention is used to name the cuts.

- **SdBadPeriod:** The event is outside a time period of normal data acquisition.
- **SdT4Level:** The event passes the T4 selection.
- **SdT5Level:** The event passes the T5 criterion.
- **SdEnergyReconstructionLevel:** The event passes the reconstruction of the energy estimator. The reconstruction of the shower front curvature is optional and not required.
- **SdThetaMin:** The event has a zenith angle  $\theta$  larger than  $60^\circ$ .
- **SdThetaMax:** The event has a zenith angle  $\theta$  smaller than  $82^\circ$ .

Table 7.1 summarises the individual cut efficiencies in the data from 2004/01 to 2009/01, Fig. 7.2 shows the cut correlations. Appendix F shows event examples, which are rejected by these cuts.

The strongest correlations are found on the one hand between the cuts **SdT4Level** and **SdReconstructionLevel**, and on the other between **SdT4Level** and **SdThetaMax**. The former is expected, since the T4 selection is designed to rejects events which are not reconstructable. The latter correlation may indicate that the T4 selection is too strict for almost horizontal showers.

Table 7.1: The table summarises the cuts to select high quality SD events in the time span from 2004/01 to 2009/01. The number of dropped events and the efficiency of each cut are calculated under exclusion of all other cuts. The cuts are correlated, which is why the efficiencies do not multiply up to the total efficiency.

	All SD events	144053	
	Accepted SD events	45063	31.3 %
Cut		Events dropped (excl.)	Efficiency (excl.) / %
SdBadPeriod		5929	96
SdT4Level		53643	63
SdT5Level		40961	72
SdEnergyReconstructionLevel	has $R_\mu$	869	99
SdThetaMin/ $^\circ$	$\geq 60$	27607	81
SdThetaMax/ $^\circ$	$\leq 82$	4057	97

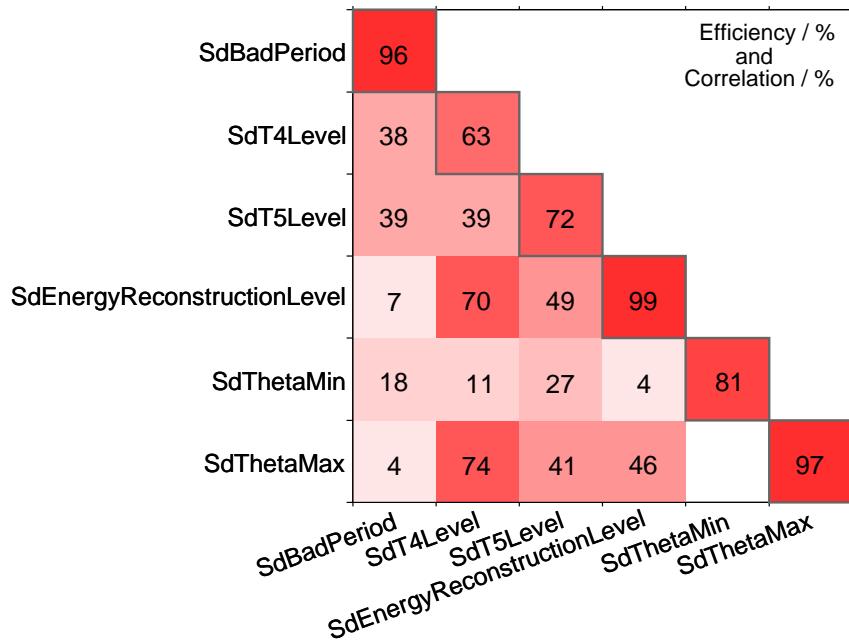


Figure 7.2: The matrix shows the efficiency and correlation of the cuts in Table 7.1. The diagonal entries represent the efficiencies of each cut under exclusion of all other cuts. The entries in the lower triangle show the correlation of the cuts in percent. Correlations below 1 % are not shown.

### 7.1.2 FD event selection

The FD selection combines many cuts on reconstructed observables and their estimated uncertainties to obtain a high quality data set. The choice of cuts and the self-consistency analysis to obtain the cut thresholds is based on ref. [157]. The cut variables can be motivated *a priori*, but not the cut thresholds. The self-consistency analysis allows to place the thresholds as such that the selection efficiency is maximised.

The analysis is based on FD events collected between 2004/01/01 and 2009/01/01. However, there are early FD measurements in the data set, which cannot be used in this analysis. The events in question were recorded at a time when no absolute FD calibration was available. The date of the first reliable measurement for each FD building is given in Table 7.2.

The FD cuts are listed and motivated in the following. The general idea is to rather place cuts on many variables. The amount of trust which can be placed on the unbiasedness of the result increases with the number of cuts. A naming convention is introduced again to ease the discussion. The final cut values will also be given in advance, the optimisation procedure is described separately in the next section.

- **FdDistanceXmaxFoV:** This cut is placed on the distance of the shower maximum  $X_{\text{max}}$  from the borders of the field of view of the telescope in units of slant depth. The cut ensures that the shower maximum is observed well within in the field of view of the telescope.
- **FdRelativeEnergyUncertainty:** This cut is placed on the relative uncertainty of the energy measurement. Since the FD reconstruction propagates geometrical uncertainties in the reconstruction of the shower axis and uncertainties due to the atmospheric conditions into the uncertainty of the shower energy, showers of poor quality can be rejected with this variable.
- **FdXmaxUncertainty:** This is a cut on the reconstructed uncertainty of the depth  $X_{\text{max}}$  of the shower maximum. A cut may be placed on this variable for the same reasons as for the previous cut. In addition, a poor resolution of the shower maximum may be an indicator for a weak shower, which is observed barely above the noise of the PMTs.
- **FdPixelNumber:** A cut on the number of the pixels with signal in the telescope camera. The number of pixels is relevant for the reconstruction of the shower axis. A low number of pixels may also indicate a weak shower, barely observed above the noise of the PMTs.
- **FdCherenkovFraction:** This cut is placed on the fraction of Cherenkov light in the reconstructed shower light profile. Cherenkov light is treated as part of the signal, as described in Chapter 6. Very inclined events often have a fair amount of Cherenkov light in the reconstructed photon flux at the telescope. Ideally, there should be no bias depending on the fraction of Cherenkov light, if it is correctly modeled in the reconstruction. Its influence has to be checked in the self-consistency analysis.
- **FdReducedChi2GaisserHillas:** This is a cut on the reduced  $\chi^2$ -value of the Gaisser-Hillas fit to the longitudinal profile of the energy loss of the shower. This is a very powerful cut to reject general anomalies in the recorded energy loss profile, which are mostly due to clouds in the field of the view. The absorption or reflection of fluorescence or Cherenkov light on clouds can severely distort the measurement.
- **FdReducedChi2Line:** This cut sets a lower limit on the reduced  $\chi^2$ -value of a simple line fit to the energy loss profile. If the data is compatible with a straight line, neither the shower maximum nor the shower energy are well defined. This is another cut to remove weak or partly recorded showers.

Table 7.2: The table gives the earliest date for each fluorescence building, when an absolute calibration of all its telescopes is available [189].

Fluorescence building	Los Leones	Los Morados	Loma Amarilla	Coihueco
First trusted date	2004/12/01	2005/06/02	-	2004/12/01

- **FdDistanceTankCore**: This cut is placed on the distance between the estimated shower core position of the FD reconstruction and the SD station which is used in the reconstruction of the shower axis. The station with the largest signal is a good first guess for the position of the shower core. If the core of the FD reconstruction is far away from this station, this may be an indicator for a bad axis reconstruction. Also, as the FD reconstruction only uses a flat shower front model, the use of the timing information from a far away station can introduce a bias.
- **FdTrackLength**: This is a cut on the observed length of the shower profile in units of slant depth. It is another cut to reject faint showers and those only partly recorded.

The last three cuts are not part of the optimisation.

- **FdMieMeasurement**: The last cut is a selection criterion. It rejects events, which do not have a corresponding measurement of the Mie-scattering and -attenuation length and the vertical optical atmospheric depth available for the night of the data taking. Events without such measurements are reconstructed with average values, which can be arbitrarily wrong and have a strong impact on the reconstructed energy. Events of this kind are generally rejected in this analysis.
- **FdThetaMin**: This is another cut on the reconstructed zenith angle. It selects only the very inclined showers with  $\theta \geq 60^\circ$  from the data. A corresponding upper limit is not necessary, because geometrical constraints make it almost impossible to observe a near horizontal air shower in the FD and the SD simultaneously.
- **FdActiveCrown**: This cut criterion requires six active SD stations in a hexagon around the SD station with the largest signal in the FD event. The purpose of this criterion is explained below.

Two sets of FD events are selected with these cuts. The first set only contains events which triggered the FD and the SD simultaneously. They are called *golden hybrids* by convention. These events are later used in the fit of the energy calibration constants  $E_{\text{cal}}$  and  $\gamma$ .

The second set is a superset of the first. It contains more FD events, because an independent SD trigger is not required anymore. These events are called *potential golden hybrids* in this chapter. The additional **FdActiveCrown** cut selects only showers that fell into an active part of the SD array. Therefore, this event set only contains showers which could have triggered the SD, if the shower front would have been intense enough. The potential golden hybrids will be used to constrain some aspects of the data model, which is used to describe the distribution of golden hybrids. It will also be used to derive the SD efficiency as a function of the cosmic ray energy  $E$  at the end of this chapter.

The efficiencies of the optimised cuts for the golden hybrids are shown in Table 7.3, the cut correlations in Fig. 7.3. The same is done for the potential golden hybrids in Table 7.4 and Fig. 7.4. Examples of rejected events are shown in Appendix F.

Most of the cuts are highly correlated, since they are sensitive to similar event traits. The strongest correlation is found between the cuts **FdTrackLength** and **FdDistanceXmaxFoV**, which reject almost the same events. Most of these events are nearby showers with a shower maximum outside the field of the view of the telescope. Another strong correlation is found between the cuts **FdRelativeEnergyUncertainty** and **FdXmaxUncertainty** which can be expected since both are obtained from the fit of the Gaisser-Hillas function, as shown in Chapter 6.

Only 411 golden hybrid events pass the event selection. Since the data statistic is so small, there was a strong motivation to develop a method for the energy calibration which uses all these events without imposing any more cuts.

Table 7.3: The table shows the cuts which are used to select golden hybrids, as described in the text. Only events are regarded here, which triggered both the SD and the FD and passed the SD cuts from Table 7.1 already. The number of dropped events and the efficiency of each cut are given exclusive of all other cuts. The cuts are correlated, which is why the efficiencies do not multiply up to the total efficiency.

Golden hybrid events that pass the SD cuts		2453	
Accepted golden hybrid events		411	16.8 %
Cut		Events dropped (excl.)	Efficiency (excl.) / %
FdDistanceXmaxFoV/g cm <sup>-2</sup>	$\geq$ 200	901	63
FdRelativeEnergyUncertainty	$\leq$ 0.2	746	70
FdXmaxUncertainty/g cm <sup>-2</sup>	$\leq$ 150	508	79
FdPixelNumber/pixels	$\geq$ 6	148	94
FdCherenkovFraction	$\leq$ 0.5	676	72
FdReducedChi2GaisserHillas	$\leq$ 2	550	78
FdReducedChi2Line	$\geq$ 0.5	53	98
FdDistanceTankCore/km	$\leq$ 1	151	94
FdTrackLength/g cm <sup>-2</sup>	$\geq$ 600	489	80
FdMieMeasurement		321	87
FdThetaMin/ $^{\circ}$	$\geq$ 60	810	67

Table 7.4: The table shows the cuts which are used to select potential golden hybrids, as described in the text. More FD events are regarded as input here, because the independent SD trigger is optional now. However, the events are still required to pass the `SdBadPeriod` cut. The number of dropped events and the efficiency of each cut are given exclusive of all other cuts. The cuts are correlated, which is why the efficiencies do not multiply up to the total efficiency.

FD events with optional SD trigger		342271	
Accepted FD events		1299	0.4 %
Cut		Events dropped (excl.)	Efficiency (excl.) / %
FdDistanceXmaxFoV/g cm <sup>-2</sup>	$\geq$ 200	260486	24
FdRelativeEnergyUncertainty	$\leq$ 0.2	154809	55
FdXmaxUncertainty/g cm <sup>-2</sup>	$\leq$ 150	73989	78
FdPixelNumber/pixels	$\geq$ 6	17694	95
FdCherenkovFraction	$\leq$ 0.5	55694	84
FdReducedChi2GaisserHillas	$\leq$ 2	43852	87
FdReducedChi2Line	$\geq$ 0.5	23603	93
FdDistanceTankCore/km	$\leq$ 1	12295	96
FdTrackLength/g cm <sup>-2</sup>	$\geq$ 600	246736	28
FdMieMeasurement		51555	85
FdThetaMin/ $^{\circ}$	$\geq$ 60	313445	8
FdActiveCrown		132174	61

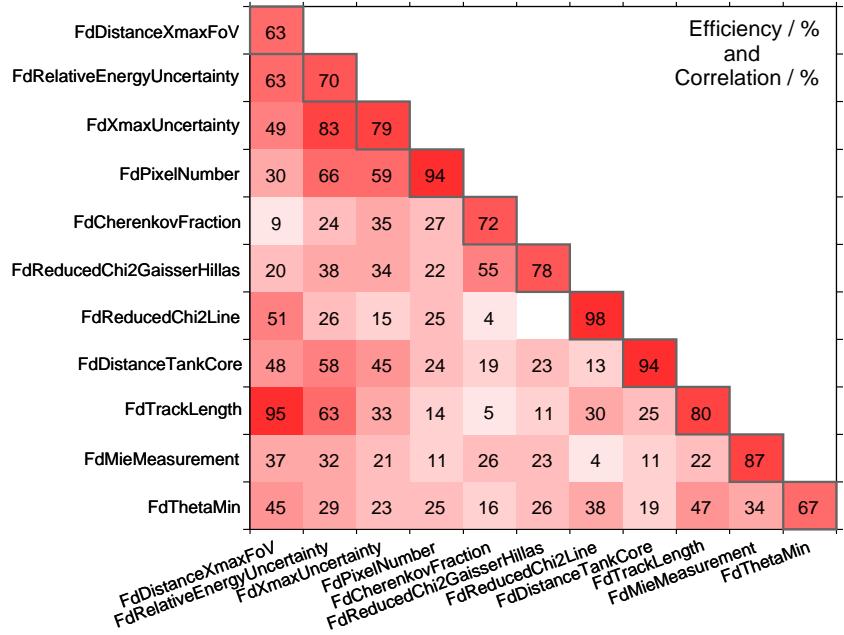


Figure 7.3: The plot shows the efficiency and correlation matrix corresponding to Table 7.3. The diagonal entries show the efficiencies of each cut alone, without applying the other cuts. The lower triangle entries show the correlations of the cuts. Large values mean that the cuts are very correlated and mostly reject the same events. Correlations below 1 % are not shown.

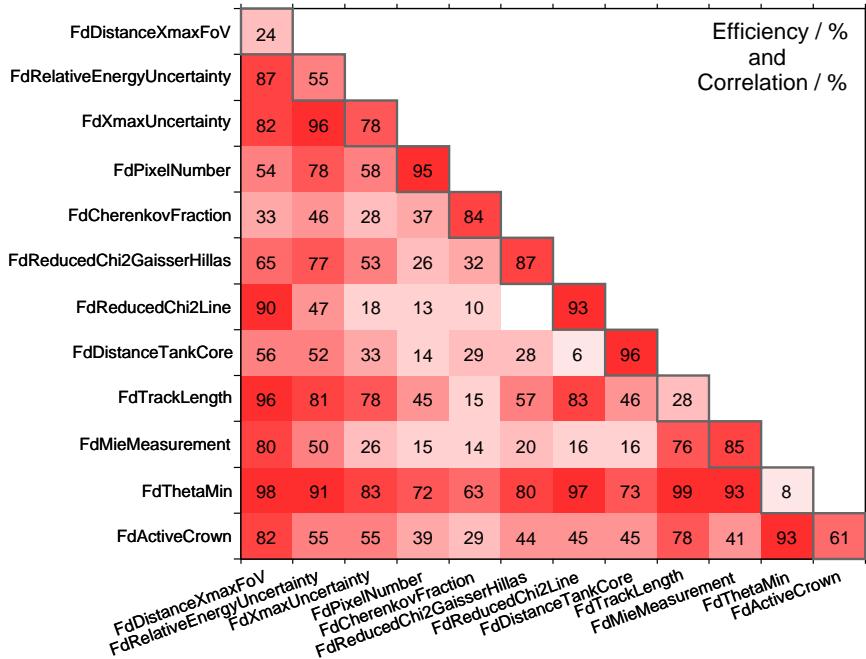


Figure 7.4: The plot shows the efficiency and correlation matrix corresponding to Table 7.4. The diagonal entries show the efficiencies of each cut alone, without applying the other cuts. The lower triangle entries show the correlations of the cuts. Large values mean that the cuts are very correlated and mostly reject the same events. Correlations below 1 % are not shown.

### 7.1.3 Cut optimisation

In order to optimise the FD cuts shown in Table 7.3, an iterative self-consistency analysis is used. The analysis is based on two points: the SD works independently of the FD and the reconstructed SD energy  $E_{SD}$  is unbiased if the SD is fully efficient. Thus, the SD energy  $E_{SD}$  can be used as a reference to detect biases in the reconstructed FD energy  $E_{FD}$ . The optimal cut threshold is obtained by plotting the bias of the variable  $(E_{FD} - E_{SD})/\langle E \rangle$  as a function of the threshold, whereas  $\langle E \rangle$  is the average of  $E_{FD}$  and  $E_{SD}$ . The optimal cut threshold avoids the bias but is as permissive as possible. In order to avoid biased SD events in this analysis, only events with an energy above  $10^{18.5}$  eV are used. At this energy, the SD is almost fully efficient.

This kind of analysis either needs a preliminary energy calibration to convert the SD energy estimator  $R_\mu^{SD}$  into the SD energy  $E_{SD}$  or an initial guess of the cut thresholds so that a preliminary calibration analysis may be performed. Optimal cut thresholds are obtained by iterating the cut optimisation and the energy calibration several times. The method converges quickly. A wrong energy calibration usually causes only an overall bias in this analysis, which can be distinguished by eye from the bias that should be avoided with the cut.

Fig. 7.5 to Fig. 7.13 show the control plots of the cut optimisations with the final energy calibration. Black circles show the bias  $\langle (E_{FD} - E_{SD})/\langle E \rangle \rangle$  as a function of the cut threshold which has to be zero. Gray circles show the individual values of  $(E_{FD} - E_{SD})/\langle E \rangle$ . The unselected data is shown on the left of each figure, the selected data as a cross-check on the right. The hatched region in the left plot of each figure is rejected by the cut. The red solid line shows the cut efficiency as a function of the cut value. The analysis shows that the FD energy  $E_{FD}$  is indeed unbiased after the cuts.

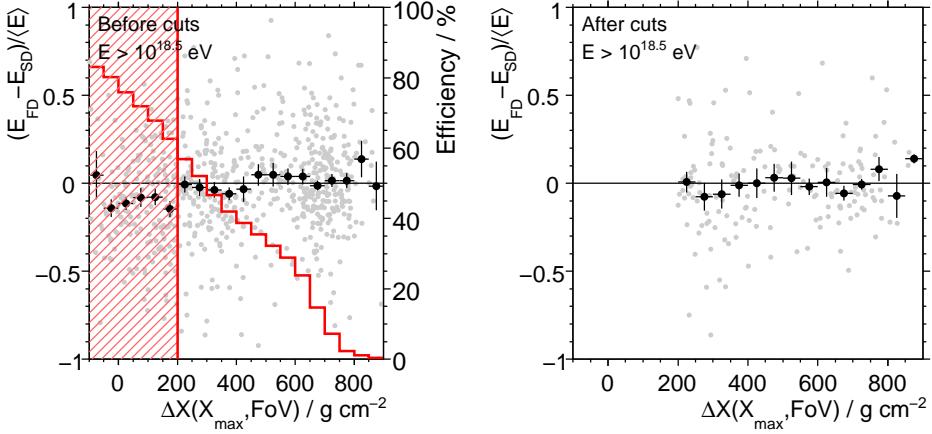


Figure 7.5: Optimisation of the cut on the distance between the reconstructed  $X_{\max}$  and the borders of the field of view of the fluorescence telescope (see text).

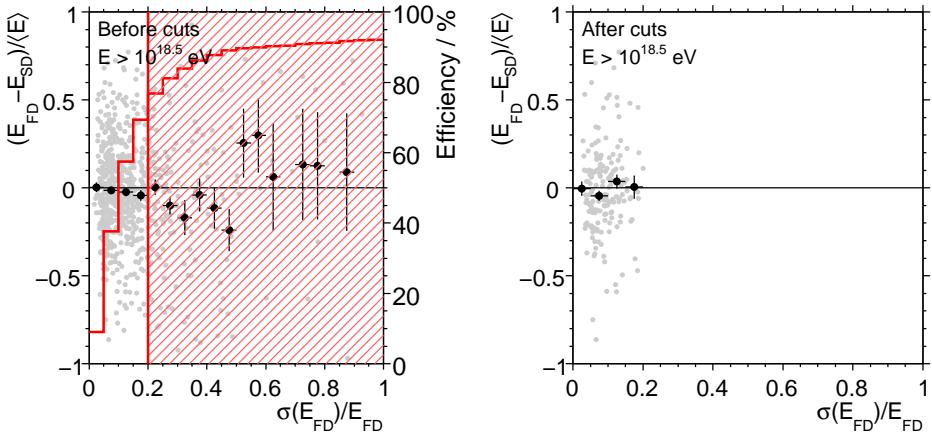


Figure 7.6: Optimisation of the cut on the relative uncertainty of the reconstructed energy (see text).

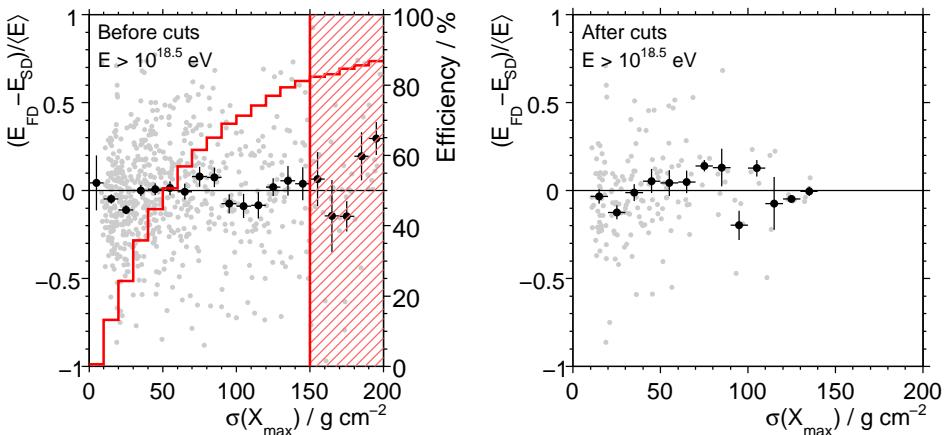


Figure 7.7: Optimisation of the cut on the uncertainty of the reconstructed  $X_{\max}$ .

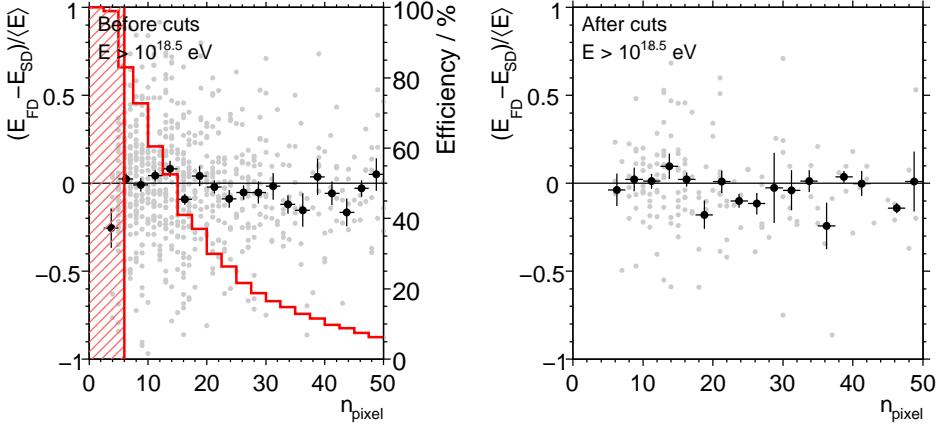


Figure 7.8: Optimisation of the cut on the number of triggered pixels in the telescope camera (see text).

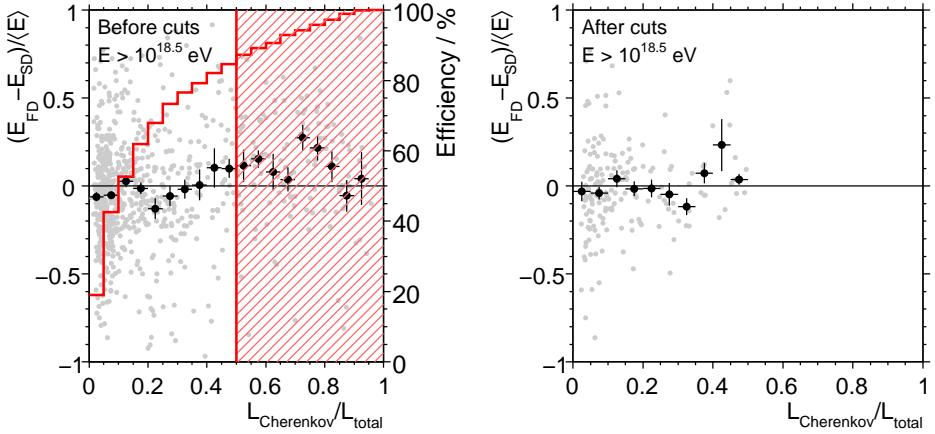


Figure 7.9: Optimisation of the cut on the fraction of Cherenkov light in the event (see text).

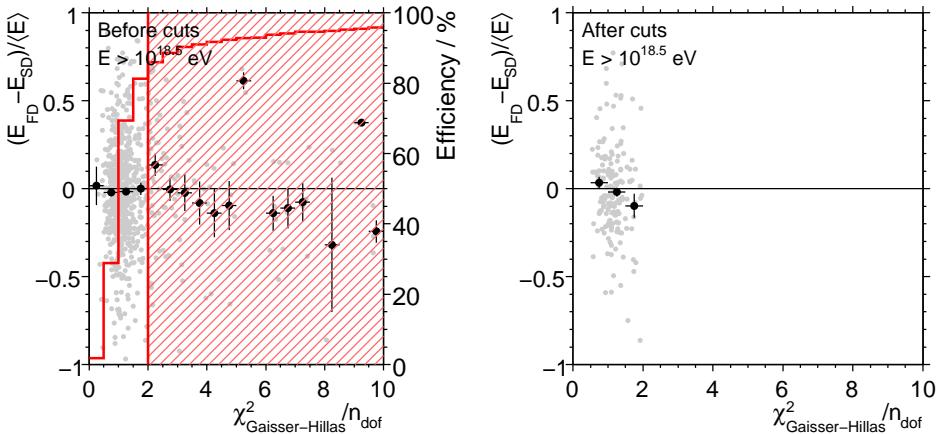


Figure 7.10: Optimisation of the cut on the value  $\chi^2/n_{\text{dof}}$ , which measures the agreement of the observed energy loss profile with the theoretical Gaisser-Hillas function (see text).

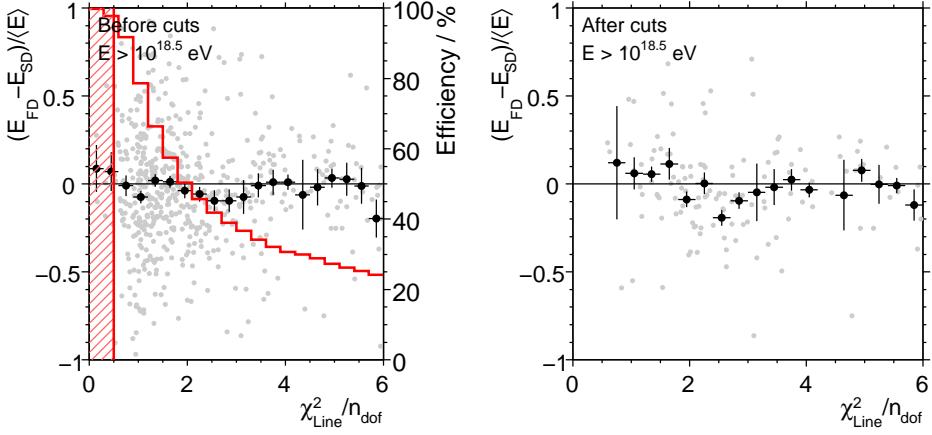


Figure 7.11: Optimisation of the cut on the value  $\chi^2/n_{\text{dof}}$ , which measures the agreement of the observed energy loss profile with a line fit (see text).

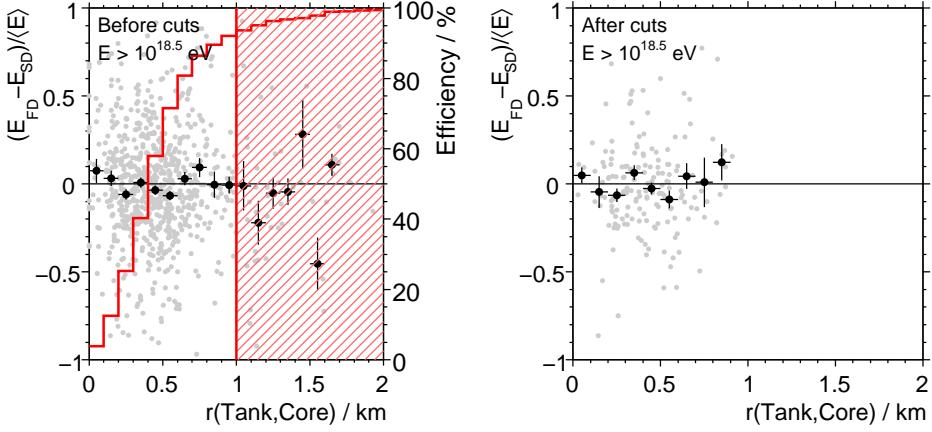


Figure 7.12: Optimisation of the cut on the distance between the reconstructed shower core and the SD station with the largest signal, which was used in the reconstruction of the shower axis (see text).

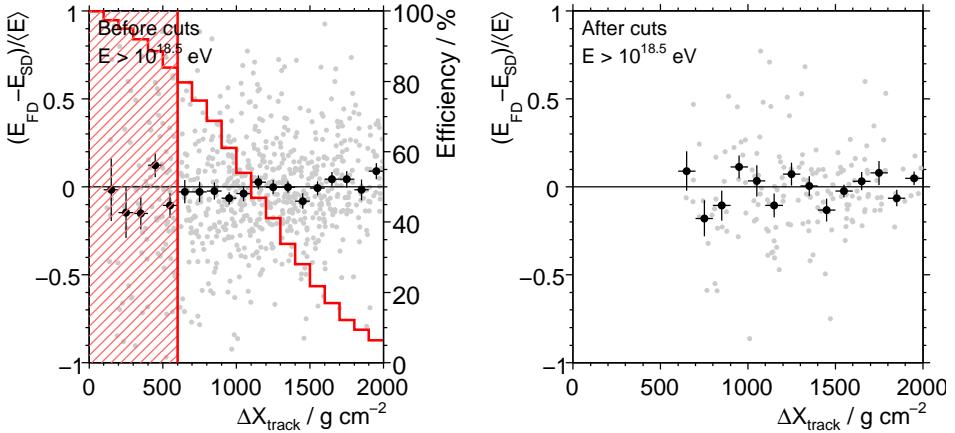


Figure 7.13: Optimisation of the cut on the observed track length in the telescope (see text).

## 7.2 Calibration method

The event selection assures that the input for the following statistical analysis is unbiased and well understood. Fig. 7.14 shows the event distribution of the data set which will be analysed in the following. A clear correlation between the energy estimator  $R_\mu^{\text{SD}}$  and the FD energy  $E_{\text{FD}}$  is visible. It is the purpose of the calibration method to extract the calibration constants  $E_{\text{cal}}$  and  $\gamma$  from this distribution.

Eq. (7.0.1) models the basic relation between the average FD energy  $\langle E_{\text{FD}} \rangle$  and the average SD energy estimator  $\langle R_\mu^{\text{SD}} \rangle$ . So far, no clear deviation from this simple power law was found, see e.g. [157, 188]. It is therefore assumed in the following that Eq. (7.0.1) is valid between  $10^{18}$  eV and a few  $10^{20}$  eV.

In some sense, all data points originate from the curve defined by Eq. (7.0.1), but they are scattered around it by various effects. The effects are illustrated in Fig. 7.15.

- **Shower-to-shower fluctuations.** The energy estimator  $R_\mu^{\text{SD}}$  of the SD is by design proportional to the total number of muons  $N_\mu$  on the ground. This muon number fluctuates from shower to shower, even if the cosmic ray energy  $E$  is fixed. More about this effect can be found in Chapter 3 and Chapter 5.

The observable energy  $E_{\text{FD}}$  in the fluorescence detector is weakly anti-correlated with these fluctuations because of energy conservation. However, since the muon component carries only about 10 % of the total energy in the shower, the correlated effect on  $E_{\text{FD}}$  is one magnitude smaller than the effect on  $R_\mu^{\text{SD}}$ , and can therefore be neglected.

- **Trigger effects.** The SD reaches the point of trigger saturation at about  $5 \times 10^{18}$  eV in very inclined air showers. The trigger basically depends on the number of muons on the ground, which is modified by shower-to-shower fluctuations. Below the trigger saturation point, an air shower with less muons than average has a lower probability to trigger than a shower with more muons than average. This introduces a principal bias in that energy region, because more events will be found on the right side of the calibration line in Fig. 7.15 than on the left side.
- **Measurement uncertainties.** The measurements of both the fluorescence and the surface detector show fluctuations, which are caused by the statistical sampling in the measurement. The sampling fluctuations are independent of each other. The fluctuations can be assumed to be Gaussian in first approximation.

The observed event distribution is further shaped by the fact that most golden hybrids are observed at lower energies due to the steep energy spectrum. The energy spectrum in the golden hybrid sample is less steep than the true energy spectrum, because the effective exposure of the FD grows with the cosmic ray energy.

All these effects can be included into a detailed statistical model of the data in form of a p.d.f. of the event distribution. The maximum likelihood method then allows to extract the parameters of interest from this model. The basic idea of this approach was presented in ref. [190, 191], which inspired this study.

The maximum likelihood method maximises the total probability of all data points  $\{(R_\mu^{\text{SD}}, E_{\text{FD}})_i\}$  in the event distribution by modifying free parameters  $\mathbf{p}$  of the data model  $f_{\text{tot}}$ :

$$L(\mathbf{p}) = \prod_i f_{\text{tot}}((R_\mu^{\text{SD}})_i, (E_{\text{FD}})_i | \mathbf{p}). \quad (7.2.1)$$

In practice,  $-\ln L$  is minimised, which is equivalent. The minimisation is done with the MINUIT package [148].

If the data model  $f_{\text{tot}}$  is accurate, it is possible to show that the likelihood function is an *optimal* parameter estimator [147], which means that no statistical information in the data is lost and the parameter estimates have the smallest possible variance.

The un-normalised data model  $f'_{\text{tot}}$  can be written as a weighted convolution of several component p.d.f.s which describe the individual statistical effects:

$$\begin{aligned} f'_{\text{tot}}(R_\mu^{\text{SD}}, E_{\text{FD}} | \mathbf{p}) &= \int d\theta \int dE g_{\text{FD-rec}}(E_{\text{FD}} | E, \theta, \mathbf{p}) h_{\text{FD-hyb}}(E, \theta | \mathbf{p}) \times \\ &\quad \int dR_\mu^{\text{sh-sh}} P_{\text{SD}}(R_\mu^{\text{SD}} | R_\mu^{\text{sh-sh}}, \theta, \mathbf{p}) g_{\text{SD-rec}}(R_\mu^{\text{SD}} | R_\mu^{\text{sh-sh}}, \theta, \mathbf{p}) h_{\text{SD-sh-sh}}(R_\mu^{\text{sh-sh}} | R_\mu(E), \mathbf{p}) \end{aligned} \quad (7.2.2)$$

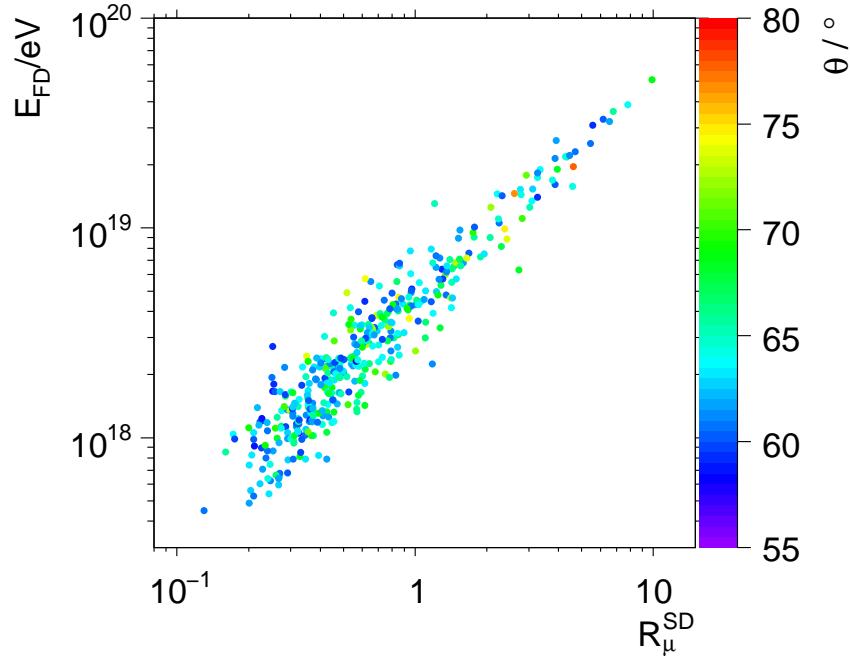


Figure 7.14: The points represent the golden hybrid sample selected by the cuts in Table 7.1 and Table 7.3. The color of a point indicates the zenith angle  $\theta$  of the event. Most events are found below  $70^\circ$ .

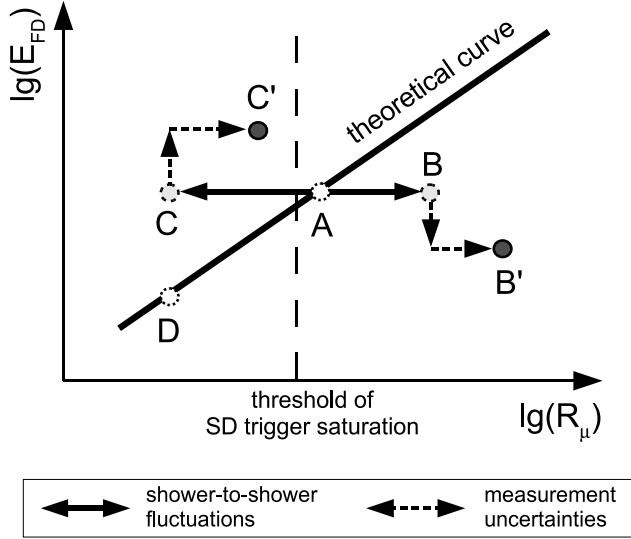


Figure 7.15: The drawing illustrates the effects which lead to the observed data points. Point A is the ideal average on top of the calibration function. The physical points B and C are upward and downward fluctuations of this ideal average, further randomised by measurement uncertainties into B' and C'. Because Point C fluctuated below the saturation threshold of the SD, it has a chance to be lost. In the analysis of the resulting data, the experimenter cannot distinguish anymore whether point C' originated from point A or D, or any other ideal point of the calibration function. All possibilities have to be regarded, weighted with the underlying energy spectrum of the FD.

whereas  $g_{\text{FD-rec}}$  and  $g_{\text{SD-rec}}$  model the measurement uncertainty of the FD and the SD respectively,  $h_{\text{FD-hyb}}$  is the distribution of the true energies observed in the FD,  $P_{\text{SD}}$  is the SD reconstruction probability (which is not a p.d.f.), and  $h_{\text{SD-sh-sh}}$  is the model of the shower-to-shower fluctuations. The energy  $E$  and the zenith angle  $\theta$  are the true values of the cosmic ray,  $R_\mu$  is the ideal energy estimator corresponding to  $E$ , and  $R_\mu^{\text{sh-sh}}$  is the true energy estimator realised in the particular event. The calibration function Eq. (7.0.1) predicts the relation between  $R_\mu$  and  $E$ , which removes the remaining degree of freedom in the integral.

To obtain a valid p.d.f.,  $f'_{\text{tot}}$  has to be normalised by integrating over the input range of  $E_{\text{FD}}$  and  $R_\mu^{\text{SD}}$ , respectively:

$$f_{\text{tot}}(R_\mu^{\text{SD}}, E_{\text{FD}}) = \frac{f'_{\text{tot}}(R_\mu^{\text{SD}}, E_{\text{FD}})}{\int_{R_\mu^{\text{SD,min}}}^{R_\mu^{\text{SD,max}}} dR_\mu^{\text{SD}} \int_{E_{\text{FD,min}}}^{E_{\text{FD,max}}} dE_{\text{FD}} f'_{\text{tot}}(R_\mu^{\text{SD}}, E_{\text{FD}})}. \quad (7.2.3)$$

The function  $f'_{\text{tot}}$  is not automatically normalised because of the SD reconstruction probability  $P_{\text{SD}}$ . The limits of these integrals may be chosen freely, but only data points within these limits may be used in the calculation of the likelihood function  $L(p)$ .

The numerical calculation of  $f_{\text{tot}}$  is described in Appendix A.5. Computational speed is an issue, because  $f_{\text{tot}}$  is recalculated several hundred times in a single maximum likelihood fit. Some approximations are made to speed up the calculation. The approximations introduce a bias to the fitted parameters of the data model  $f_{\text{tot}}$ . The size of the bias is evaluated at the end of this section.

### 7.2.1 Model components

Fig. 7.16 gives an overview over the model components of the data model  $f_{\text{tot}}(R_\mu^{\text{SD}}, E_{\text{FD}})$ . The heart of the p.d.f. is the calibration function, which has two free parameters. The other components introduce many additional parameters, too many in fact to fit them all from the sample of golden hybrids. Fortunately, several components can be fitted to larger data sets or derived from event simulations.

The model components are presented and discussed in the following. As this is a pioneering work, most components are modeled in the simplest possible way and some approximations are made. The overall framework however is general enough so that these simple models may be replaced by much more complex and sophisticated ones in the future. The systematic uncertainty which is introduced through the simplifications is carefully estimated in several cross-checks at the end of this chapter.

A summary of the numerical values of the model parameters is given by Table 7.5. All simulated events in the following analyses are taken from the library developed in Chapter 5.

#### SD-Reconstruction probability $P_{\text{SD}}$

A model of the SD reconstruction probability  $P_{\text{SD}}$ , which is in good approximation given by the SD trigger probability, can be developed from an analysis of simulated events.

The reconstruction probability  $P_{\text{SD}}$  depends on the intensity and lateral distribution of the shower, and its particle content. These properties depend on the cosmic ray energy  $E$  and the zenith angle  $\theta$  of the shower. The reconstruction probability  $P_{\text{SD}}$  depends to a much lesser degree also on the azimuth  $\phi$ , partly because of the grid structure and partly because of geomagnetic field effects. The azimuthal dependency is neglected.

Very inclined showers are muon dominated and treating them as pure muon showers is a good first approximation. In that case, the SD trigger probability can only depend on the total number of muons on the ground  $N_\mu \propto R_\mu^{\text{sh-sh}}$  at a fixed zenith angle  $\theta$ , because the normalised lateral muon density profile  $p_\mu$  is approximately universal with respect to the cosmic ray energy and mass, as shown in Chapter 5. At  $\theta \approx 60^\circ$ , a small but significant electromagnetic component is still active, and some additional influence of the energy and mass of the cosmic ray is expected.

Fig. 7.17 shows the reconstruction probability as obtained from the simulation as a function of the total number of muons  $N_\mu$  on the ground, which is readily available in simulated showers. An error function fits the data well:

$$P_{\text{SD}}(\lg N_\mu) = \frac{1}{2} \left( 1 - \text{erf} \left( -\frac{\lg N_\mu - q_0}{\sqrt{2} q_1} \right) \right), \quad (7.2.4)$$

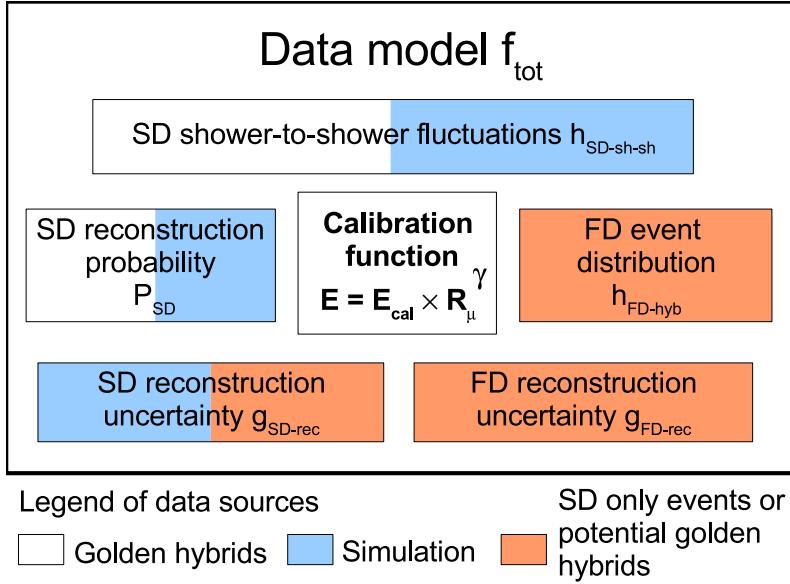


Figure 7.16: The graphic gives an overview over the model components that are needed to calculate the data model  $f_{\text{tot}}$ , as described in the text. The colors indicate possible data sources to constrain the parameters of the model components.

Table 7.5: A table of the parameters introduced in each model component of the total probability density function  $f_{\text{tot}}$  of the golden hybrid event distribution. The parameter uncertainties are of statistical nature, if not specified otherwise.

Model	Data source	Parameters	
$P_{\text{SD}}$	Surface detector simulation	$q_0$	$6.82^{+0.12}_{-0.12}$ (sys.)
		$q_1$	$0.15^{+0.07}_{-0.10}$ (sys.)
$g_{\text{SD-rec}}$ and $g_{\text{FD-rec}}$	Surface detector simulation Potential golden hybrids	$p_0$	$0.0390 \pm 0.0024$
		$p_1$	$0.1128 \pm 0.0043$
		$p_2$	$0.091 \pm 0.021$
		$p_3$	$0.03044 \pm 0.00033$
$h_{\text{FD-hyb}}$	Potential golden hybrids	$s_0$ $s_1$ $s_2$ $s_3$ $s_4$	$17.89 \pm 0.44$ $0.23 \pm 0.14$ $-1.06 \pm 0.77$ $-6.45 \pm 4.93$ $-44.88 \pm 24.47$

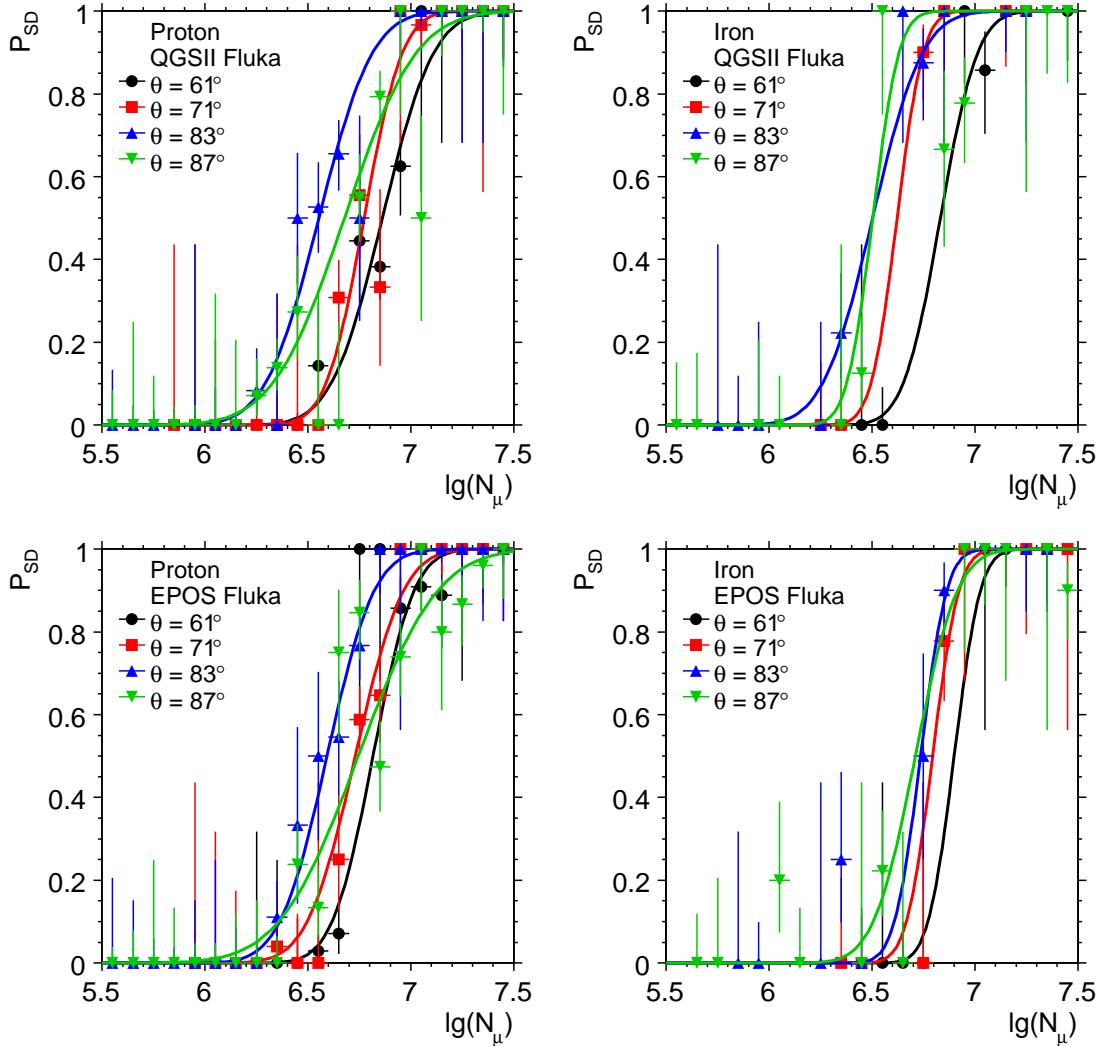


Figure 7.17: The plots show the efficiency of the SD reconstruction including the SD trigger as a function of the total number of muons  $N_\mu(\theta)$  on the ground at a given zenith angle  $\theta$ . The simulation is organised in small zenith angle bins of  $2^\circ$  width, the plots show a selection of these bins. The four plots separate the simulations by the cosmic ray mass (proton, iron) and the hadronic interaction model (QGSJet-II, EPOS). The continuous lines are fits to the simulated data.

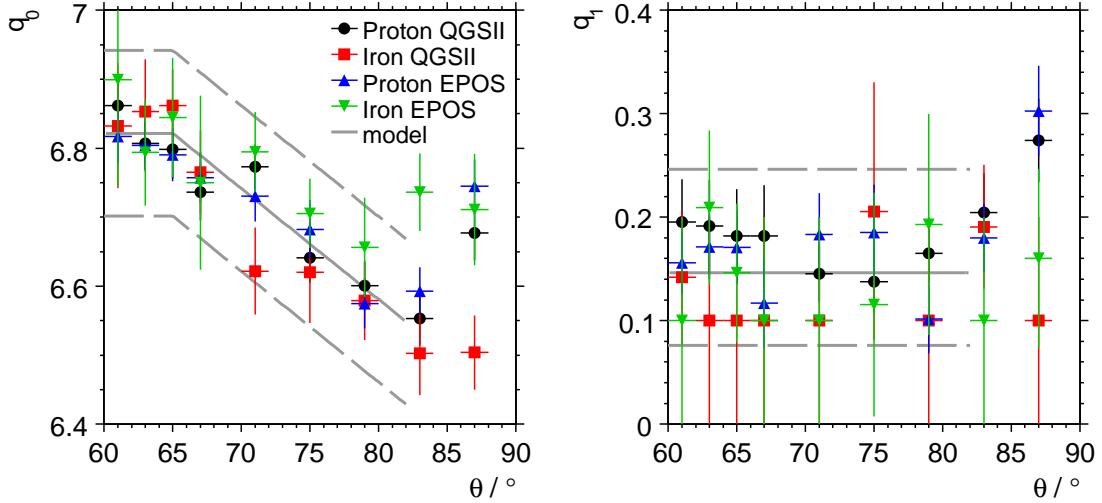


Figure 7.18: The plots show the parameters of the fits from Fig. 7.17 as a function of the zenith angle  $\theta$ . The continuous lines are fits of a model. The dashed lines estimate the systematic uncertainty of the model.

with  $q_0$  and  $q_1$  as free parameters. At a given zenith angle  $\theta$ , the number of muons is proportional to the true energy estimator  $R_\mu^{\text{sh-sh}}$ , and one can be calculated from the other with the reference model of the lateral muon profile  $n_\mu^{\text{ref}}$ .

The available amount of data for these fits is not optimal. In particular, the fits generated from iron showers are not well constrained. Fig. 7.18 shows the fitted parameters of  $q_0$  and  $q_1$  as a function of the zenith angle  $\theta$ . The width  $q_1$  of the threshold function is not defined in many cases due to a lack of data points in the threshold region, and has to be constrained to a lower value of  $q_1 = 0.1$ .

This lack of data is caused by the fact, that the simulation does not cover a continuous spectrum of cosmic ray energies. The showers are distributed in comparably distant and narrow energy bins. This produces corresponding gaps in the distribution of the total number of muons  $N_\mu \propto E^{1/\gamma}$ . Shower-to-shower fluctuations of  $N_\mu$  fill the gaps somewhat, but only for proton showers, where these fluctuations are large enough. More simulations at intermediate energies are necessary for this kind of analysis, and should be performed in the future.

Fig. 7.18 shows, that the dependency of the threshold function on the zenith angle  $\theta$  is moderate. It is much stronger, if the threshold would be parameterised in  $R_\mu^{\text{sh-sh}}$  instead of  $N_\mu$ , which confirms  $N_\mu$  is the right variable for the parameterisation. Within the limited resolution of the analysis, no significant dependency on the cosmic ray mass  $A$  or the hadronic interaction model is observed.

The parameter  $q_1$  is assumed to be independent of the zenith angle. The  $\theta$ -dependency of  $q_0$  is modeled very roughly by eye as

$$q_0(\theta) = q_0 \times \begin{cases} 1 & 60^\circ \leq \theta < 65^\circ \\ 1 - 0.04(\theta - 65^\circ)/17^\circ & 65^\circ \leq \theta < 82^\circ. \end{cases} \quad (7.2.5)$$

A parameterisation up to  $82^\circ$  is sufficient, as it covers the complete zenith angle range used in this study. The final average values of  $q_0$  and  $q_1$  are obtained in a fit to the available data in Fig. 7.18 and shown in Table 7.5. To account for the crudeness of the model, rather large systematic uncertainties are assigned to  $q_0$  and  $q_1$ , which approximate the observed spread in Fig. 7.18. Fig. 7.19 shows the individual fitted threshold functions and the average model in comparison. The individual curves are within the average model for zenith angles up to  $82^\circ$ .

This model with its large uncertainties is not of much use by itself. The parameters  $q_0$  and  $q_1$  will therefore be left free in  $f_{\text{tot}}$  and fitted from the golden hybrid sample. The fitted result will be only be compared with the simulation result. Furthermore, it will be cross-checked with an analysis of the reconstruction probability based on the potential golden hybrids at the end of this chapter.

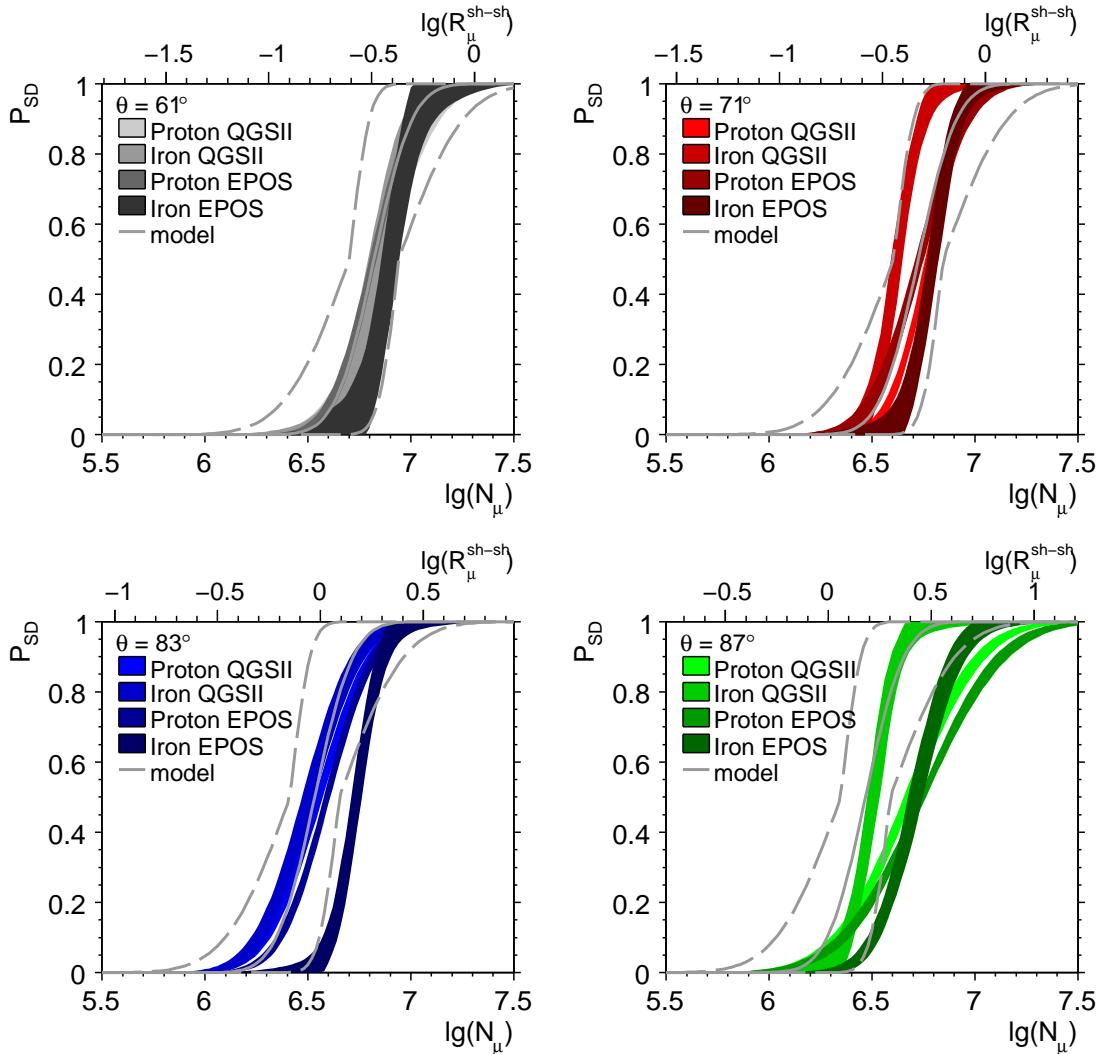


Figure 7.19: The plots compare the fitted threshold functions from Fig. 7.17 with the average model and its upper and lower systematic limit. The limits are obtained by varying the model parameters within their systematic uncertainties and taking the maximum and minimum of the variation. The second axis shows the true energy estimator  $R_\mu^{\text{sh-sh}}$  corresponding to the total number of muons on the ground  $N_\mu$ .

Finally, there is an important point to make. It seems reasonable at first to parameterise the reconstruction probability  $P_{\text{SD}}$  as a function of the true energy estimator  $R_{\mu}^{\text{sh-sh}} \propto N_{\mu}$ . The trigger decision should depend on the true number of muons that arrive at the ground. In this particular analysis it is also unavoidable, because the reconstructed energy estimator  $R_{\mu}^{\text{SD}}$  is simply not defined in events that did not trigger the SD.

However, it turns out over course of all analyses and cross-checks in this chapter that a consistent description of the trigger threshold behaviour can only be achieved by making  $P_{\text{SD}}$  a function of the reconstructed energy estimator  $R_{\mu}^{\text{SD}}$  and not the true energy estimator  $R_{\mu}^{\text{sh-sh}}$  in the data model  $f_{\text{tot}}$ . The fluctuations of  $R_{\mu}^{\text{SD}}$  with respect to  $R_{\mu}^{\text{sh-sh}}$  are not random. They are generated by sampling fluctuations in the detector. If the SD samples by chance more of the available muons,  $R_{\mu}^{\text{SD}}$  has an upward fluctuation and vice versa. This affect the trigger decision, so that  $P_{\text{SD}}$  is actually a function of  $R_{\mu}^{\text{SD}}$  and not  $R_{\mu}^{\text{sh-sh}}$ . Evidence for this hypothesis will be shown below.

### Measurement uncertainties $g_{\text{SD-rec}}$ and $g_{\text{FD-rec}}$

The FD and SD measurements are independent and their resolution functions can be modeled by normal distributions:

$$g_{\text{SD-rec}}(R_{\mu}^{\text{SD}}, R_{\mu}^{\text{sh-sh}}, \sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]) = \frac{1}{\sqrt{2\pi}\sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]} \exp\left(-\frac{(R_{\mu}^{\text{SD}} - R_{\mu}^{\text{sh-sh}})^2}{2\sigma_{\text{rec}}^2[R_{\mu}^{\text{SD}}]}\right) \quad (7.2.6)$$

$$g_{\text{FD-rec}}(E_{\text{FD}}, E, \sigma_{\text{rec}}[E_{\text{FD}}]) = \frac{1}{\sqrt{2\pi}\sigma_{\text{rec}}[E_{\text{FD}}]} \exp\left(-\frac{(E_{\text{FD}} - E)^2}{2\sigma_{\text{rec}}^2[E_{\text{FD}}]}\right), \quad (7.2.7)$$

whereas  $\sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]$  and  $\sigma_{\text{rec}}[E_{\text{FD}}]$  are the respective experimental uncertainties, which in general are not constant. Simple models can be found, which describe  $\sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]$  and  $\sigma_{\text{rec}}[E_{\text{FD}}]$  well.

The SD-resolution of the energy estimator was already analysed in detail in Chapter 6. Based on simulations, it was found that the resolution of the reconstructed energy estimator  $R_{\mu}^{\text{SD}}$  is normal in good approximation. In the zenith angle range  $60^\circ \leq \theta \leq 82^\circ$ , the resolution depends approximately only on the cosmic ray energy, or more precisely, the true energy estimator  $R_{\mu}^{\text{sh-sh}}$ .

Fig. 7.20 compares the true resolution of the energy estimator  $R_{\mu}^{\text{SD}}$  in simulations with the reconstruction estimate of the resolution in real events. Some unexpected features are visible in the unbinned resolution estimates, like the accumulation of points at  $\sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]/R_{\mu}^{\text{SD}} \approx 0.1$ . The affected events did not show any special common feature in an eye scan. This point need further study in the future.

The reconstruction estimate of  $\sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]$  differs from the simulated resolution, because the statistical models used in the reconstruction are not yet perfect. The simulated resolution is biased at  $\lg R_{\mu} \lesssim -0.4$ . It underestimates the true resolution due to the influence of the SD trigger in this region. The variance of  $R_{\mu}^{\text{SD}}/R_{\mu}^{\text{sh-sh}} - 1$  in the simulation gets smaller, because more and more events in the lower tail of the Gaussian are rejected by the SD event trigger.

The simulation provides the more reliable estimate of  $\sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]$  and is used for the parameterisation. It will be shown in a moment that it is possible to properly model the detector fluctuations at  $\lg R_{\mu} < -0.4$  by convoluting  $g_{\text{SD-rec}}$  with  $P_{\text{SD}}(R_{\mu}^{\text{SD}})$ . However, right now it is necessary to parameterise the resolution  $\sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]$  without this effect and therefore only data in the range  $-0.4 \leq \lg R_{\mu} \leq 1.2$  is used. The relative resolution  $\sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]/R_{\mu}^{\text{SD}}$  is well described by an empirical expansion in  $R_{\mu}^{\text{SD}-\frac{1}{2}}$  up to first order:

$$\sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]/R_{\mu}^{\text{SD}} = p_0 + p_1 R_{\mu}^{\text{SD}-\frac{1}{2}}, \quad (7.2.8)$$

with  $p_0$  and  $p_1$  as free parameters. The values are given in Table 7.5. Since the simulation result is not confirmed with a real measurement, a systematic uncertainty will be assigned to the resolution obtained from this parameterisation later.

Fig. 7.21 shows that the detector fluctuations at  $\lg R_{\mu} < -0.4$  are not Gaussian but well described by weighting  $g_{\text{SD-rec}}(R_{\mu}^{\text{SD}}, R_{\mu}^{\text{sh-sh}}, \sigma_{\text{rec}}[R_{\mu}^{\text{SD}}])$  with  $P_{\text{SD}}(R_{\mu}^{\text{SD}}, \theta)$ . This combination models the trigger effect on the distribution of the detector fluctuations. The model predictions agree well with the distribution found in simulated events.

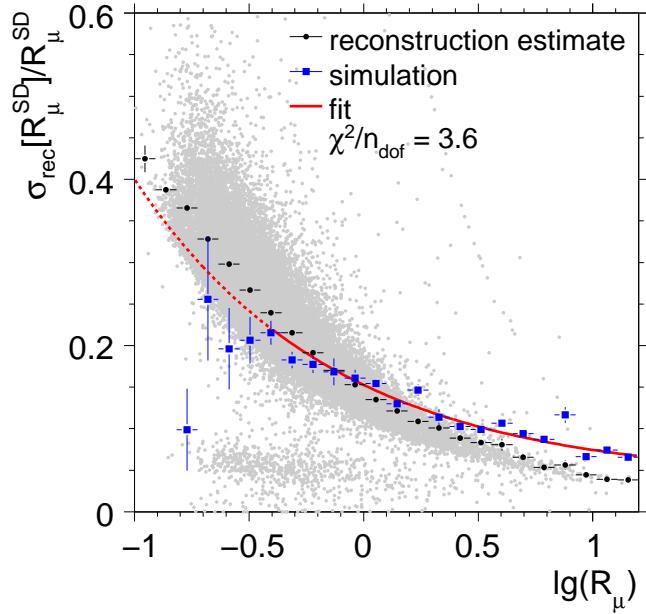


Figure 7.20: The resolution of the reconstructed energy estimator  $R_\mu^{\text{SD}}$  is shown as a function of the energy estimator. Light gray points represent the estimated uncertainty of the energy estimator as obtained event-by-event from the reconstruction of real events, the black points are binned averages. The blue points represent the true resolution observed in simulated events, which is the variance of the variable  $R_\mu^{\text{SD}} / R_\mu^{\text{sh-sh}} - 1$ . Horizontal error bars indicate bin widths. The solid line is a fit to the simulated resolution, the dashed line its extrapolation beyond the data range which was used in the fit.

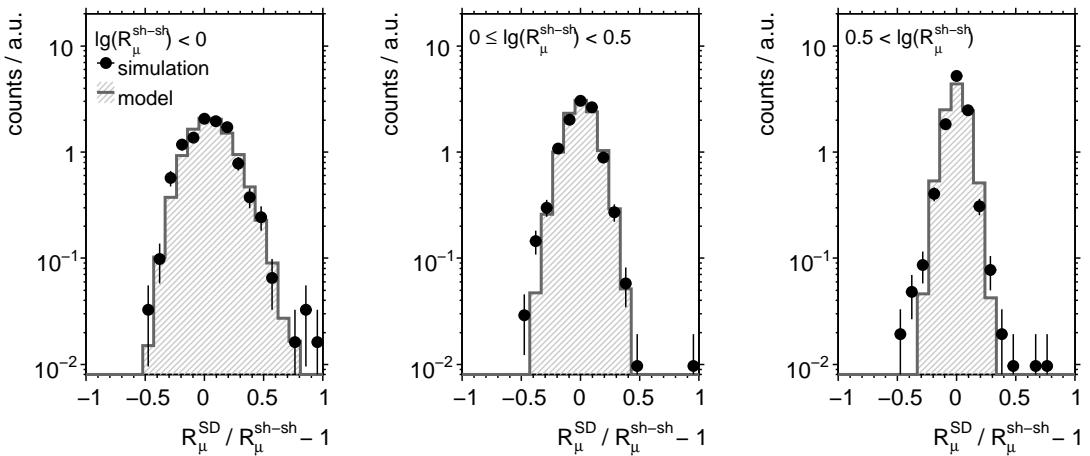


Figure 7.21: The plots show the distribution of the variable  $R_\mu^{\text{SD}} / R_\mu^{\text{sh-sh}} - 1$  in simulated events for three different  $R_\mu$ -ranges. The hatched histograms show the corresponding prediction of the resolution model. A correct description at  $\lg R_\mu^{\text{sh-sh}} < 0$  is only obtained, if the resolution model is weighted with the reconstruction probability  $P_{\text{SD}}$  (see text).

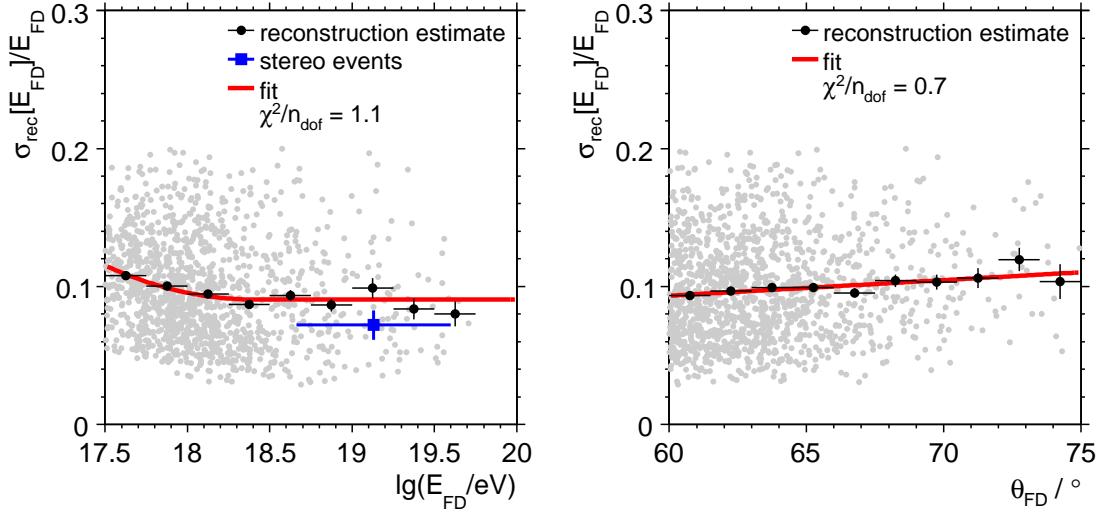


Figure 7.22: The resolution of the FD energy  $E_{\text{FD}}$  is shown as a function the energy and the zenith angle. Light gray points represent the estimated uncertainty of the energy as obtained event-by-event from the reconstruction of real events, the black points are binned averages. The blue point in the left plot shows the resolution obtained from the analysis of the stereo events for comparison. Horizontal error bars indicate bin widths. The solid lines are fits of a parameterisation.

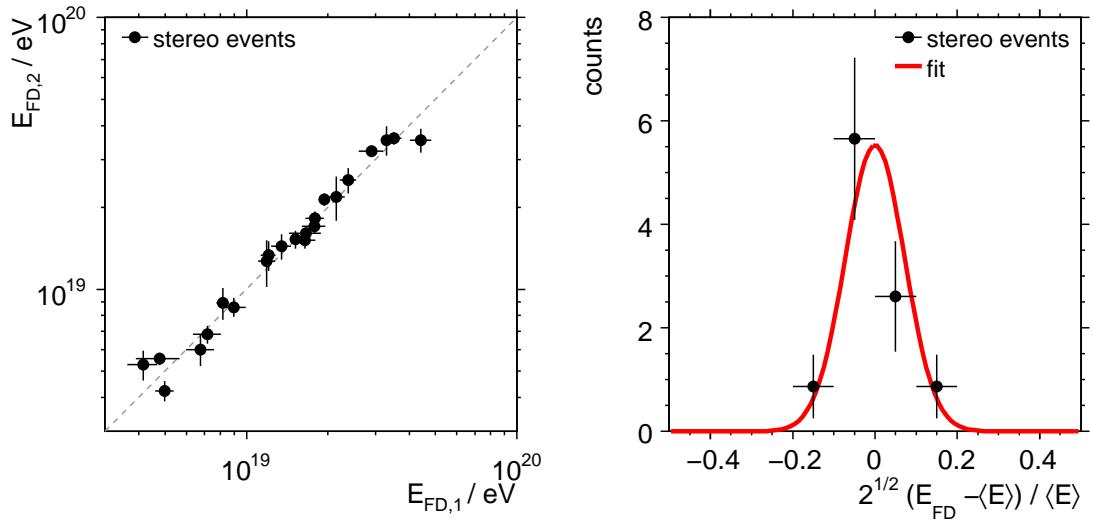


Figure 7.23: The plots shows an analysis of 21 stereo hybrid events. These are events, which are observed in two FD buildings simultaneously. The left plot shows the correlation of the individual measurements, the right plot shows the distribution of the resolution variable (see text) and a fitted Gaussian.

The Gaussian shape of  $g_{\text{FD-rec}}$  is demonstrated *e.g.* in ref. [192] with reconstructed laser shots from the central laser facility of the Pierre Auger Observatory. The FD resolution can be estimated with laser shots, but using real showers is more reliable. Fig. 7.22 shows the reconstruction estimate of the FD resolution as a function of the reconstructed energy  $E_{\text{FD}}$  and the reconstructed zenith angle  $\theta_{\text{FD}}$ . The resolution depends only weakly on both observables. The  $\theta$ -dependency is neglected in the following, while the former is parameterised empirically as

$$\sigma_{\text{rec}}[E_{\text{FD}}]/E_{\text{FD}} = \begin{cases} p_3 + p_4 [\lg(E_{\text{FD}}/\text{eV}) - 18.4]^2 & \text{if } E_{\text{FD}} \leq 10^{18.4} \text{ eV}, \\ p_3 & \text{if } E_{\text{FD}} > 10^{18.4} \text{ eV}, \end{cases} \quad (7.2.9)$$

with  $p_3$  and  $p_4$  as free parameters. The values are given in Table 7.5. There are possible other dependencies of the resolution on the details of the shower axis orientation and distance to the telescope, but they average out if the resolution is only regarded as a function of the cosmic ray energy.

The weak dependency of the relative FD resolution  $\sigma_{\text{rec}}[E_{\text{FD}}]/E_{\text{FD}}$  on the cosmic ray energy  $E$  is generated by an exposure effect. On the one hand, the measurement of the FD energy is more precise if more light falls into the telescopes. The collected light for a shower at a fixed distance to a telescope is proportional to the cosmic ray energy. On the other hand, the effective exposure of the FD increases with the energy, so that also more distant and faint showers are observed at higher energy. The exposure growth cancels the first effect so that the average FD resolution remains roughly constant with the cosmic ray energy.

The Pierre Auger Observatory is equipped with four FD telescope buildings, which allows to check the reconstruction estimate of the energy resolution with a real measurement. In this study, this is done with stereo events: showers which are observed by two telescopes independently and simultaneously.

There are 21 stereo events in the data set of potential hybrids. The experimental energy resolution from  $k$  redundant measurements is given by the variance of the variable

$$\left( \frac{E_{\text{FD}i} - \langle E_{\text{FD}} \rangle}{\langle E_{\text{FD}} \rangle} \right) \times \sqrt{\frac{k}{k-1}}, \quad (7.2.10)$$

whereas  $i < k$  is counting the  $k$  independent measurements of the same true energy  $E$ , and  $\langle E_{\text{FD}} \rangle$  is the average energy of the individual measurements. This formula is taken from ref. [193], where it is first used in an analogue analysis in a different context.

The stereo measurements are shown in Fig. 7.23. The obtained resolution is slightly lower than the reconstruction estimate, as shown in Fig. 7.22, but still within one standard deviation of the uncertainty of the parameterisation.

### Energy and zenith angle distribution of FD events $h_{\text{FD-hyb}}$

The distribution  $h_{\text{FD-hyb}}$  of cosmic ray energies and zenith angles in the golden hybrid data set can be extracted from the event sample of potential golden hybrids. Fig. 7.24 shows the distributions of the measured energies and zenith angles in this event set. Empirical parameterisations are fitted to these distributions, under the assumption that the dependencies approximately factorise:

$$h_{\text{FD-hyb}}(E, \theta) \approx \tilde{h}_{\text{FD-hyb}}(E) \times \tilde{h}_{\text{FD-hyb}}(\theta). \quad (7.2.11)$$

The p.d.f.  $\tilde{h}_{\text{FD-hyb}}(E)$  of the energy  $E$  is proportional to the cosmic ray flux multiplied with the effective aperture of the FD. The distribution is parameterised with a piece-wise power law in  $E$ . In practice,  $\tilde{h}_{\text{FD-hyb}}(E)$  is formulated in the variable  $x = \lg E/\text{eV}$  so that the power law becomes and an exponential function

$$E^\alpha = 10^{\alpha \lg E} = 10^{\alpha x}, \quad (7.2.12)$$

which has better numerical properties. The aperture effect below  $10^{18}$  eV is modeled with an error function in the variable  $x$ . The full parameterisation is

$$\tilde{h}_{\text{FD-hyb}}(x) = C^{-1} \left[ 1 - \text{erf} \left( -\frac{x - s_0}{\sqrt{2} s_1} \right) \right] \times \begin{cases} 10^{(s_2 - 0.3)x} & \text{if } 17.0 < x \leq 18.3 \\ 10^{s_2 x - 5.49} & \text{if } 18.3 < x \leq 19.6 \\ 10^{(s_2 - 1.2)x + 18.03} & \text{if } x > 19.6, \end{cases} \quad (7.2.13)$$

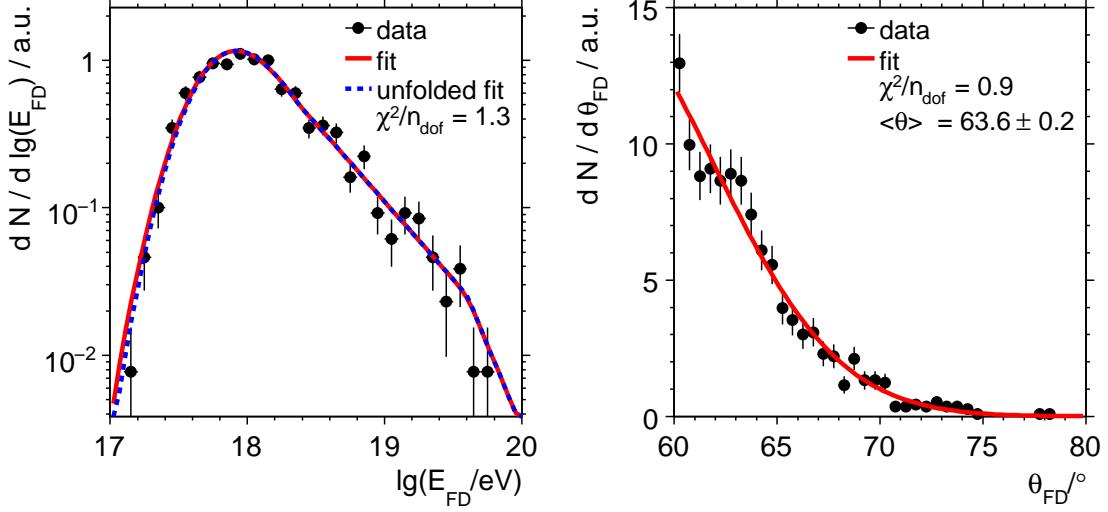


Figure 7.24: The figure shows the distribution of energy  $E$  (left) and zenith angle  $\theta$  (right) of potential golden hybrid events. The points represent histograms of the event distributions, the lines are empirical fits to these histograms. The continuous lines show fits to the unprocessed distributions, which are distorted to some degree by the effect of limited detector resolution. The dashed line in the left plot shows a normalised fit, where this effect was unfolded (see text).

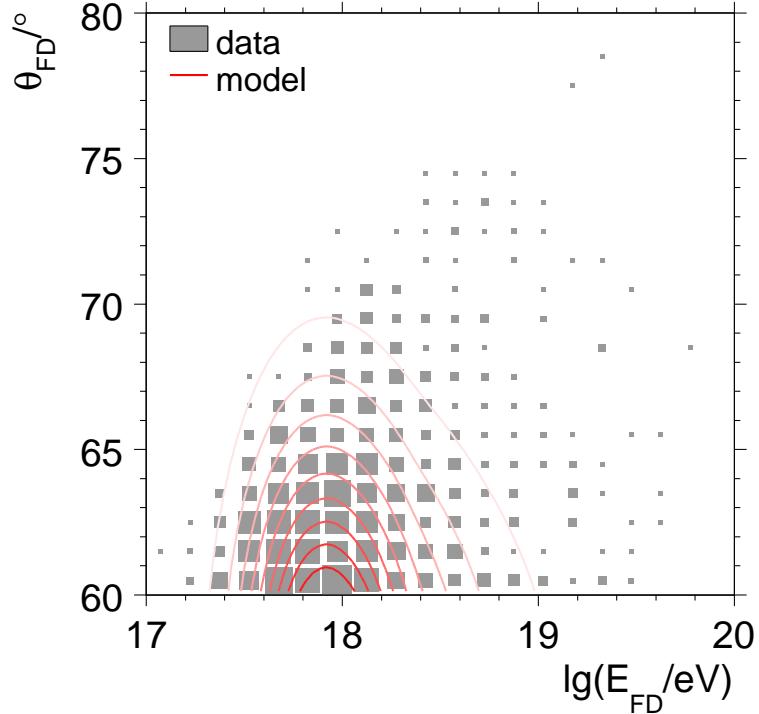


Figure 7.25: The plot shows the distribution of the observed energies  $E_{FD}$  and zenith angles  $\theta_{FD}$  in the potential hybrid events. The corresponding model  $h_{\text{FD-hyb}}$  is overlaid, the solid lines represent contours of constant probability.

whereas  $C$  is a normalisation constant and  $s_0, s_1, s_2$  are free parameters. The break points and the relative changes in spectral indices are taken from the analysis in ref. [188] to reduce the number of degrees of freedom in this fit, but the overall spectral index  $s_2$  is left free. The spectral index  $s_2$  is larger than the spectral index of the true energy distribution, because the FD exposure growths with the cosmic ray energy  $E$ .

It is necessary to regard in the fit of Eq. (7.2.13), that the observed distribution in Fig. 7.24 is modified by the limited detector resolution of the fluorescence detector. The p.d.f.  $\tilde{h}_{\text{FD-hyb}}(E)$  is supposed to parameterise the energy distribution without this effect. To obtain the correct result, the model  $\tilde{h}_{\text{FD-hyb}}$  is not fitted directly to the data distribution, but first convoluted with the detector resolution  $g_{\text{FD-rec}}$ :

$$\tilde{h}_{\text{FD-hyb}}^{\text{folded}}(E_{\text{FD}}) = \int dE g_{\text{FD-rec}}(E_{\text{FD}}, E) \tilde{h}_{\text{FD-hyb}}(E). \quad (7.2.14)$$

The convoluted result  $h_{\text{FD-hyb}}^{\text{folded}}(E_{\text{FD}})$  is then fitted to the data. Fig. 7.24 shows the folded and unfolded distribution together with the data. Above  $10^{18}$  eV, both distributions are almost equal. In the threshold region below a difference of up to 50 % is observed.

The p.d.f.  $\tilde{h}_{\text{FD-hyb}}(\theta)$  of the zenith angle distribution is parameterised as a function of  $y = (\theta - 60^\circ)/\text{rad}$  with the following empirical formula:

$$\tilde{h}_{\text{FD-hyb}}(y) = C^{-1} \exp(s_3 y + s_4 y^2), \quad (7.2.15)$$

whereas  $C$  is a normalisation constants and  $s_3, s_4$  are the remaining free parameters. This model is not used in detail in the correct approach and therefore the effect of the limited detector resolution is not considered, although the same argument applies as in the case of  $\tilde{h}_{\text{FD-hyb}}(E)$ . The final parameter values are shown in Table 7.5.

The factorised model is compared with the two-dimensional distribution of the events in Fig. 7.25. The plot shows a moderate positive correlation between  $E$  and  $\theta$ : the largest zenith angles only occur at the highest energies. This correlation is neglected in the current approach.

### Shower-to-shower fluctuations $h_{\text{SD-sh-sh}}$

The shower-to-shower fluctuations of the true energy estimator are derived from the fluctuations of the total number of muons  $R_\mu^{\text{sh-sh}} \propto N_\mu$ , which were analysed in detail in Chapter 5 with air shower simulations. The simulations show that the relative size of the fluctuations  $\sigma_{\text{sh-sh}}[R_\mu^{\text{sh-sh}}]/R_\mu^{\text{sh-sh}}$  is only a function of the cosmic ray mass  $A$ . No significant dependency on the energy  $E$  and direction  $(\theta, \phi)$  of the cosmic ray was found. Systematically, the fluctuations further depend on the hadronic interaction model used in the simulation. The relative size of the fluctuations ranges from about 3 % for iron showers to 13 % – 20 % for protons. The p.d.f. of the fluctuations is approximately normal, with a slight asymmetry in the tails. Downward fluctuations are slightly larger than upward fluctuations.

Fig. 7.26 shows the fluctuations of protons and iron nuclei, simulated with the hadronic interaction models QGSJet-II and EPOS. The true energy estimator  $R_\mu^{\text{sh-sh}}$  is by about 30 % – 40 % larger for iron showers than for proton showers at the same energy  $E$ . The offset is constant in good approximation. Therefore, a mixed cosmic ray flux consisting of proton and iron nuclei can have larger shower-to-shower fluctuations than proton or iron showers alone. Fig. 7.27 show the size of the fluctuations in such a mixed scenario. Fluctuations up to 27 % are possible. An iron fraction of 40 % yields the largest fluctuations.

Since the cosmic ray composition is not well known and the systematic uncertainties in the hadronic interaction models are considerable, the shower-to-shower fluctuations need to be fitted from the golden hybrid events. There are indications for a change in the composition between  $10^{18}$  eV and  $10^{20}$  eV [194] and thus the fluctuation model should allow a variation with the cosmic ray energy.

This analysis uses the simplest possible model for the fluctuations under these circumstances to keep the additional number of free parameters reasonably small. The fluctuations are modeled with a Gaussian distribution. A linear transition as a function of the logarithm of the energy between  $10^{18.3}$  eV and  $10^{19.5}$  eV

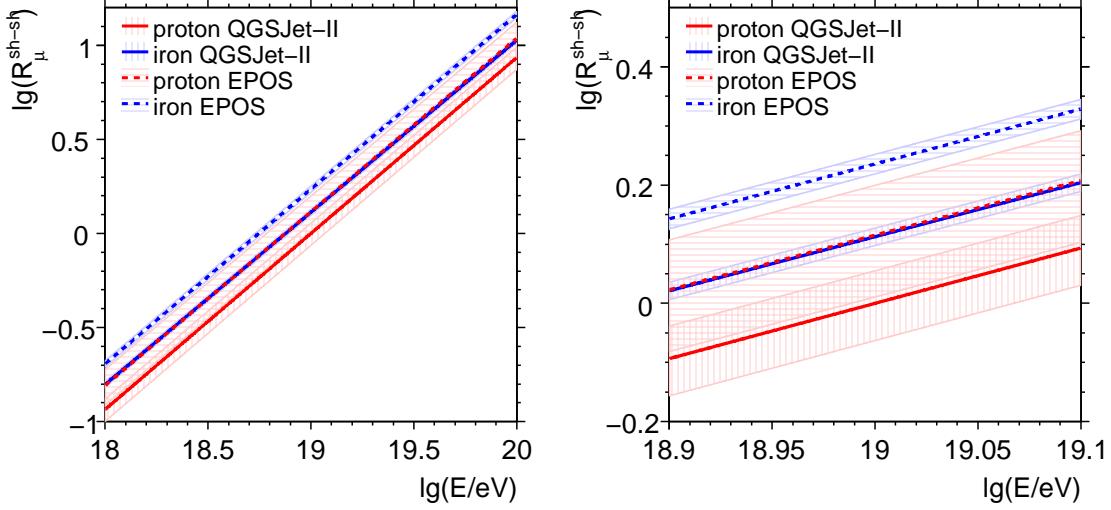


Figure 7.26: Both plots shows the energy estimator  $R_\mu^{\text{sh-sh}}$  as a function of the cosmic ray energy for proton and iron showers and the two hadronic interaction models QGSJet-II and EPOS. The hatched bands represent the shower-to-shower fluctuations. The right plot is a zoom of the left plot.

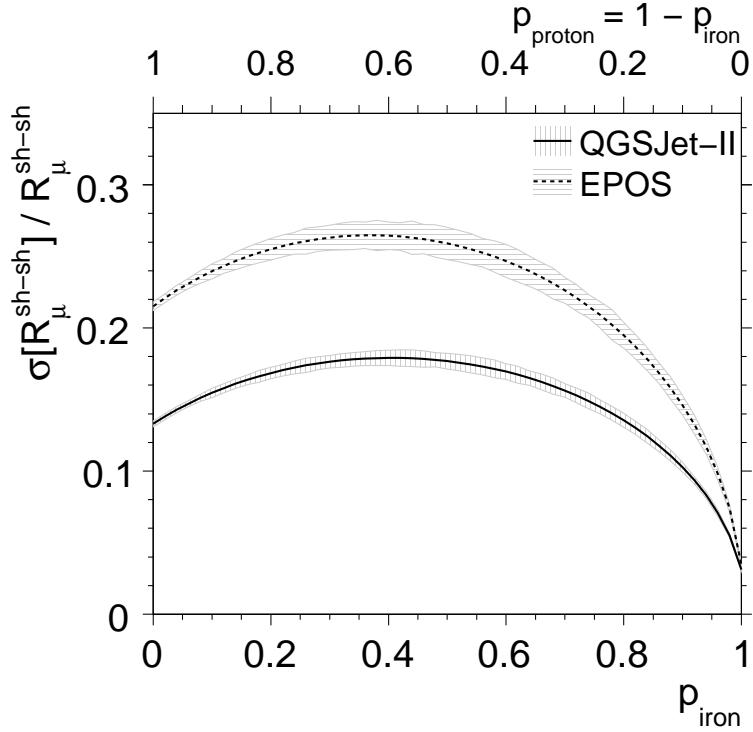


Figure 7.27: The curves shows the shower-to-shower fluctuations which would be observed in case of a mixed flux of cosmic rays of proton and iron nuclei at  $10^{19}$  eV as a function of the iron fraction  $p_{\text{iron}}$ . The bands represent the statistical uncertainty of the simulation. The number of muons generated by proton and iron showers of the same energy differs by 30 % to 40 %, thus a mixed flux can produce larger shower-to-shower fluctuations than a single component. The mixed fluctuations are in not Gaussian.

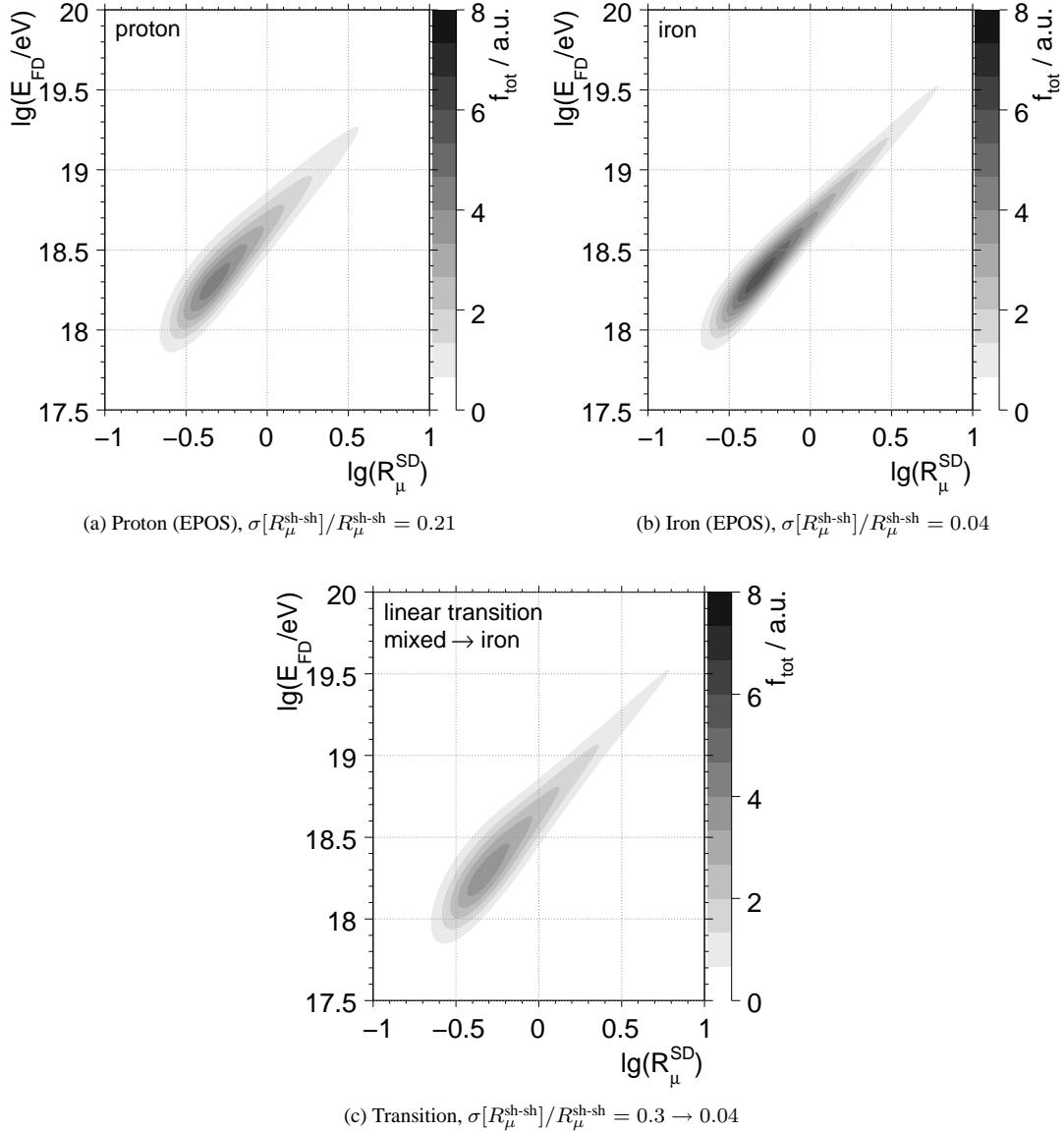


Figure 7.28: The plots show the probability density function  $f_{\text{tot}}(\lg R_\mu^{\text{SD}}, \lg E_{\text{FD}})$  as modeled in this study. The calibration function in these examples uses the constants  $E_{\text{cal}} = 5 \times 10^{18} \text{ eV}$  and  $\gamma = 1.07$ . Constant shower-to-shower fluctuations are assumed in a) and b), corresponding to pure proton or iron scenarios. In c), the shower-to-shower fluctuations evolve from a mixed composition to a pure iron scenario.

is allowed for the relative size of the fluctuations:

$$\sigma_{\text{sh-sh}}[R_\mu^{\text{sh-sh}}]/R_\mu^{\text{sh-sh}} = \begin{cases} u_0 & E < 10^{18.3} \text{ eV} \\ u_0(1-z) + u_1 z & 10^{18.3} \text{ eV} \leq E < 10^{19.5} \text{ eV} \\ u_1 & E > 10^{19.5} \text{ eV} \end{cases} \quad (7.2.16)$$

with  $z = \frac{\lg(E/\text{eV}) - 18.3}{19.5 - 18.3}$ ,

whereas  $u_0$  is the size of  $\sigma_{\text{sh-sh}}[R_\mu^{\text{sh-sh}}]/R_\mu^{\text{sh-sh}}$  at  $10^{18.3}$  eV and  $u_1$  the size of  $\sigma_{\text{sh-sh}}[R_\mu^{\text{sh-sh}}]/R_\mu^{\text{sh-sh}}$  at  $10^{19.5}$  eV.

The choice of the energy range for the transition is based on the apparent change of the elongation rate of the shower maximum in ref. [194] (see also Fig. 3.2 in Chapter 3), but still somewhat arbitrary. The quantitative result of the fit will depend on the choice and on the type of transition and therefore has to be interpreted carefully. However, the qualitative result alone is already interesting, whether the data imply a transition or not.

It is also important to point out, that approximating the fluctuations by a Gaussian is only appropriate in the case of pure proton or iron showers or as a first approximation. A mixed composition may have a completely different shape, depending on the relative fractions of different cosmic ray nuclei. With enough data however, an extended approach would even allow to fit these relative fractions as a function of the energy, similar to ref. [195].

Fig. 7.28 shows the predictions of the data model  $f_{\text{tot}}$  for different composition scenarios. The shape of  $f_{\text{tot}}$  is apparently quite sensitive to the size of the shower-to-shower fluctuations.

### 7.2.2 Statistical and numerical bias

The previously developed data model  $f_{\text{tot}}$  allows to obtain several parameters from a fit to the sample of golden hybrid events:

- the calibration constants  $E_{\text{cal}}$  and  $\gamma$ ,
- the position  $q_0$  and shape  $q_1$  of the SD trigger threshold, and
- the size of the shower-to-shower fluctuations  $\sigma[R_\mu^{\text{sh-sh}}]/R_\mu^{\text{sh-sh}}$ .

These parameters may be biased with respect to the true values for several reasons:

- (1) the corresponding model or a correlated model is wrong,
- (2) the approximations in  $f_{\text{tot}}$  are too strong, or
- (3) the amount of data is too small.

Cross-checks can be used in case of the first point, which allow to estimate the systematic uncertainty introduced by some models. This will be done in the next section. The third point is often negligible, but it is possible to show theoretically, that both the maximum likelihood and the least-squares method introduce a bias to the fit result under general conditions in the case of a low number of events, see *e.g.* ref. [147].

The present analysis regards the second and third point. The data distribution of golden hybrids is completely modeled by the components of  $f_{\text{tot}}$ , which allows to build up a simple Monte-Carlo simulation of golden hybrid events. It is not necessary to solve the integrals in Eq. (7.2.2) to perform the simulation, instead each model component p.d.f. can be sampled individually. The Monte-Carlo simulation of a golden hybrid event is done with the following steps:

- (1) Pick a random energy  $E$  and zenith angle  $\theta$  from  $h_{\text{FD-hyb}}$ . Calculate the corresponding ideal energy estimator  $R_\mu = (E/E_{\text{cal}})^{1/\gamma}$ .
- (2) Fluctuate  $R_\mu$  according to  $h_{\text{SD-sh-sh}}$  to obtain the true energy estimator  $R_\mu^{\text{sh-sh}}$ .
- (3) Fluctuate  $R_\mu^{\text{sh-sh}}$  according to  $g_{\text{SD-rec}}$  to obtain the reconstructed energy estimator  $R_\mu^{\text{SD}}$ .

- (4) Calculate the reconstruction probability  $P_{\text{SD}}$  as a function of  $R_{\mu}^{\text{SD}}$  and the zenith angle  $\theta$ . Make a random decision whether the event is accepted or rejected based on  $P_{\text{SD}}$ . If the event is rejected, jump back to (1).
- (5) Fluctuate the true energy  $E$  according to  $g_{\text{FD-rec}}$  to obtain the reconstructed FD energy  $E_{\text{FD}}$ .

Such a fast simulation of golden hybrid events is not a replacement for a full air shower and detector simulation, but it is valuable for this study. This fast simulation is based on a different mathematical concept, which allows to cross-check the use of the approximative lattice calculation of  $f_{\text{tot}}$ . If the bias in the fitted parameters is negligible, the quality of the approximative lattice calculation of  $f_{\text{tot}}$  is sufficient.

To access the bias of the result of fitting the lattice calculated  $f_{\text{tot}}$ , 50 Monte-Carlo experiments with 400 events each are fast-simulated and reconstructed for a particular set of input parameters. Nine sets of input parameters are tried, based on the possible combinations of

- $E_{\text{cal}}/\text{EeV} \in \{4.5, 5.0, 5.5\}$ ,
- $\gamma \in \{1.00, 1.05, 1.10\}$ , and
- $\sigma[R_{\mu}^{\text{sh-sh}}]/R_{\mu}^{\text{sh-sh}}(10^{18.3} \text{ eV}) = \sigma[R_{\mu}^{\text{sh-sh}}]/R_{\mu}^{\text{sh-sh}}(10^{19.5} \text{ eV}) = 0.15$ .

All other models use the parameters from Table 7.5.

The fit is run with the same setting as in the final application to real events, which means that the following six parameters are left free in these analyses:

$$E_{\text{cal}}, \quad \gamma, \quad \frac{\sigma[R_{\mu}^{\text{sh-sh}}]}{R_{\mu}^{\text{sh-sh}}}(10^{18.3} \text{ eV}), \quad \frac{\sigma[R_{\mu}^{\text{sh-sh}}]}{R_{\mu}^{\text{sh-sh}}}(10^{19.5} \text{ eV}), \quad q_0, \quad q_1.$$

The last two parameters describe the threshold of the SD reconstruction probability  $P_{\text{SD}}$ , as shown before.

Fig. 7.29 shows the result of the statistical analysis. Some bias is observed, but it is small compared to the statistical resolution in most cases. The reconstruction tends to be slightly biased with respect to the true parameters. The estimated systematic uncertainty derived from this bias is summarised in Table 7.6. The parameters of the shower-to-shower fluctuation model show the largest bias and tend to be underestimated. The bias appears to be caused by the calculation of  $f_{\text{tot}}$  on a lattice. The bias gets smaller if the lattice is made more dense, but then the computation times increase dramatically.

The parameter uncertainties reported by the fit seem consistent with the expectation, if the bias is not too large, as in the case of the shower-to-shower fluctuations. The uncertainty ellipses of the calibration constants contain the input values in  $37 \pm 2\%$  of the cases, which agrees very well with the statistical expectation.

In conclusion, the statistical analysis of the output of the energy calibration method shows that the numerical approximations involved in the grid-computation of  $f_{\text{tot}}$  introduce some bias to the result. However, the magnitude of the observed bias is acceptable compared to the statistical resolution of the result. The statistical uncertainties reported by the fit agree with the expectation.

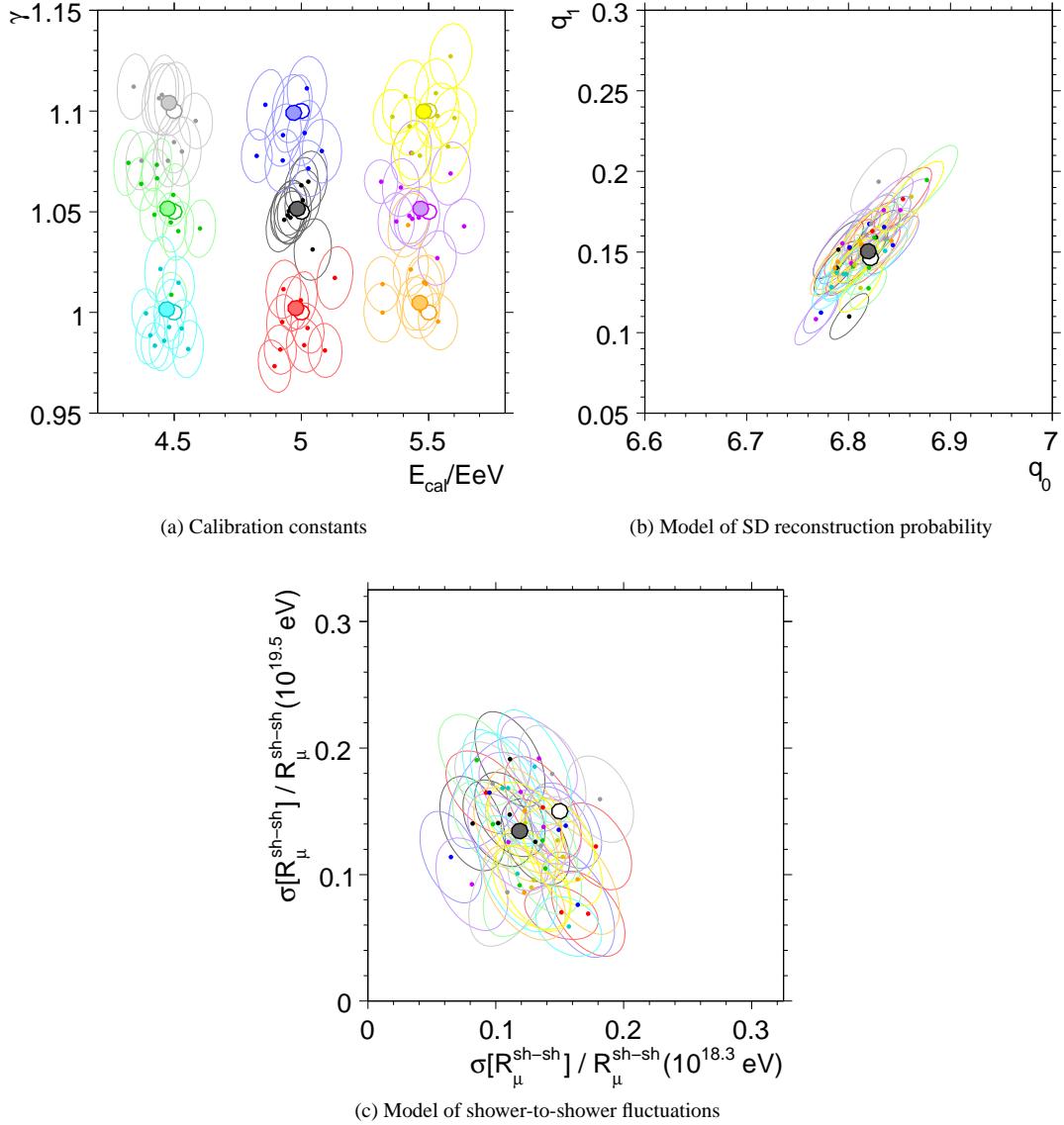


Figure 7.29: The plots show the fitted parameters of the statistical analysis described in the text. The ellipses around each point represent a 38 % contour, which corresponds to one standard deviation in two dimensions. The nine sets of input parameters described in the text are distinguished by different colors. The large circles represent the input parameters (white) and the average of the output of the fit (light colored). The relative shift estimates the bias of the fit. Only a subset of all fits are shown for the sake of clarity.

Table 7.6: The table summarises the estimated systematic uncertainty on fitted parameters in the calibration method, based on Fig. 7.29.

Calibration function	SD shower-to-shower fluctuations			SD reconstruction probability		
	$E_{\text{cal}}$	$\gamma$	$q_0$	$q_1$	$\frac{\sigma[R_\mu^{\text{sh-sh}}]}{R_\mu^{\text{sh-sh}}}(10^{18.3} \text{ eV})$	$\frac{\sigma[R_\mu^{\text{sh-sh}}]}{R_\mu^{\text{sh-sh}}}(10^{19.5} \text{ eV})$
0.03	-0.005	0.03		0.015	0.03	0.05

### 7.3 Application to data

The calibration analysis is now applied to the 411 real golden hybrid events selected earlier. In order to estimate the systematic uncertainty of the fit and as a consistency test, the fit is applied several times with the following variations.

- Model of the SD shower-to-shower fluctuations: The linear transition of the SD shower-to-shower fluctuations in  $\lg E$  is tested against a constant model.
- Model of the SD reconstruction probability: The model impact is partly or completely removed by rejecting events from the fit, which are in a region where the SD reconstruction probability is not 100 %.
- Models of the SD and FD resolution: The resolution of the SD is varied by 3 %, the resolution of the FD by 2 %. These values represent the estimated uncertainty of the respective parameterisations.

The fit variations are listed in Table 7.7. The results of the fits of type A to D are shown in Fig. 7.30. Fig. 7.31 shows the good agreement of the data model  $f_{\text{tot}}$  with the data distribution. The variations of the fit conditions yield consistent results. Slight changes in the fitted parameters are observed but nothing unexpected. The variations of the fitted parameters are used as another estimate of the systematic uncertainty of the fit.

The data model describes the event distribution even in the region where the SD reconstruction probability  $P_{\text{SD}}$  is not 100 % and affects the fluctuations of the SD energy estimator  $R_{\mu}^{\text{SD}}$  around the true value  $R_{\mu}^{\text{sh-sh}}$ . As mentioned before, the standard analysis cannot regard this effect and thus has to exclude events in this energy range. The fit of type D approximates this situation. The comparison of type A and type D shows that the new method triples the amount of usable data compared to the standard analysis and roughly halves the statistical uncertainty of the fitted calibration constants  $E_{\text{cal}}$  and  $\gamma$ .

The fits of type E to H show that variations of the SD and FD resolutions lead to opposite changes in the fitted SD shower-to-shower fluctuations, while the other fit parameters essentially remain unaffected. For example, if the SD resolution is increased by 2 %, the reconstructed shower-to-shower fluctuations become smaller by 2 %. In order to understand this, the FD resolution  $\sigma_{\text{rec}}[E_{\text{FD}}]/E_{\text{FD}}$  is effectively regarded as a contribution to the total resolution  $\sigma[R_{\mu}^{\text{SD}}]/R_{\mu}^{\text{SD}}$  of the SD energy estimator:

$$\left( \frac{\sigma[R_{\mu}^{\text{SD}}]}{R_{\mu}^{\text{SD}}} \right)^2 \simeq \left( \frac{\sigma_{\text{sh-sh}}[R_{\mu}^{\text{sh-sh}}]}{R_{\mu}^{\text{sh-sh}}} \right)^2 + \left( \frac{\sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]}{R_{\mu}^{\text{SD}}} \right)^2 + \left( \frac{\sigma_{\text{FD}}[R_{\mu}^{\text{SD}}]}{R_{\mu}^{\text{SD}}} \right)^2, \quad (7.3.1)$$

whereas  $\sigma_{\text{sh-sh}}[R_{\mu}^{\text{sh-sh}}]$  is contribution of the shower-to-shower fluctuations,  $\sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]$  is the contribution of the detector sampling fluctuations, and  $\sigma_{\text{FD}}[R_{\mu}^{\text{SD}}]$  is the propagated effect of the fluctuations of the FD energy  $E_{\text{FD}}$ :

$$\langle R_{\mu}^{\text{SD}} \rangle = \left( \frac{\langle E_{\text{FD}} \rangle}{E_{\text{cal}}} \right)^{1/\gamma} \Rightarrow \frac{\sigma_{\text{FD}}[R_{\mu}^{\text{SD}}]}{R_{\mu}^{\text{SD}}} \approx \frac{\sigma_{\text{rec}}[E_{\text{FD}}]}{E_{\text{FD}}} \text{ with } \gamma \approx 1. \quad (7.3.2)$$

The total fluctuations  $\sigma[R_{\mu}^{\text{SD}}]/R_{\mu}^{\text{SD}}$  are fixed by the data. It is possible to show with differential calculus that a small increase in  $\sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]/R_{\mu}^{\text{SD}}$  or  $\sigma_{\text{FD}}[R_{\mu}^{\text{SD}}]/R_{\mu}^{\text{SD}}$  leads to an equal decrease in  $\sigma_{\text{sh-sh}}[R_{\mu}^{\text{sh-sh}}]/R_{\mu}^{\text{sh-sh}}$  and vice versa. Essentially, the knowledge about the SD shower-to-shower fluctuations can only be as good as the knowledge of the SD and FD resolutions.

The final result with all correlations and systematic uncertainties is shown in Table 7.8. In case of the calibration constants  $E_{\text{cal}}$  and  $\gamma$  and the shower-to-shower fluctuation parameters  $\sigma[R_{\mu}^{\text{sh-sh}}]/R_{\mu}^{\text{sh-sh}}$ , the systematic uncertainties are comparable or larger than the statistical uncertainties. The situation is better for the position  $q_0$  and width  $q_1$  of the threshold in the SD reconstruction probability function  $P_{\text{SD}}$ . Here, the statistical uncertainty is at least by a factor of two larger than the corresponding systematic uncertainty.

The analysis is more limited by systematic uncertainties than statistical uncertainties, despite the small amount of data. Future improvements of the data model  $f_{\text{tot}}$  are possible, which will reduce these systematic uncertainties. In order to draw quantitative conclusions about the cosmic ray composition, it is important to make the measurements of the SD and FD resolutions more precise.

Table 7.7: The table summarises the fit variations, which are applied to the data set of real golden hybrid events.

Type	Free parameters	Data cut	Events	Comment
A	$E_{\text{cal}}$ , $\gamma$ , $q_0$ , $q_1$ , $\frac{\sigma[R_{\mu}^{\text{sh-sh}}]}{R_{\mu}^{\text{sh-sh}}}$ ( $10^{18.3}$ eV), $\frac{\sigma[R_{\mu}^{\text{sh-sh}}]}{R_{\mu}^{\text{sh-sh}}}$ ( $10^{19.5}$ eV)	–	411	reference result, linear transition in $\lg E$ for shower-to-shower fluctuations
B	$E_{\text{cal}}$ , $\gamma$ , $q_0$ , $q_1$ , $\frac{\sigma[R_{\mu}^{\text{sh-sh}}]}{R_{\mu}^{\text{sh-sh}}}$	–	411	constant shower-to-shower fluctuations
C	$E_{\text{cal}}$ , $\gamma$ , $\frac{\sigma[R_{\mu}^{\text{sh-sh}}]}{R_{\mu}^{\text{sh-sh}}}$	$\lg R_{\mu}^{\text{SD}} > -0.4$	282	constant shower-to-shower fluctuations, reduced impact of threshold events
D	$E_{\text{cal}}$ , $\gamma$ , $\frac{\sigma[R_{\mu}^{\text{sh-sh}}]}{R_{\mu}^{\text{sh-sh}}}$	$\lg R_{\mu}^{\text{SD}} > -0.1$	124	constant shower-to-shower fluctuations, negligible impact of threshold events
E - H	$E_{\text{cal}}$ , $\gamma$ , $q_0$ , $q_1$ , $\frac{\sigma[R_{\mu}^{\text{sh-sh}}]}{R_{\mu}^{\text{sh-sh}}}$ ( $10^{18.3}$ eV), $\frac{\sigma[R_{\mu}^{\text{sh-sh}}]}{R_{\mu}^{\text{sh-sh}}}$ ( $10^{19.5}$ eV)	–	411	fit of type A but with $\sigma_{\text{rec}}[E_{\text{FD}}]/E_{\text{FD}} \pm 0.02$ and $\sigma_{\text{rec}}[R_{\mu}^{\text{SD}}]/R_{\mu}^{\text{SD}} \pm 0.03$ (not shown in Fig. 7.30, Fig. 7.31)

Table 7.8: The table shows the final result of the energy calibration analysis. The systematic uncertainty of each variable is calculated from the systematic uncertainty obtained from the intrinsic bias of  $f_{\text{tot}}$  shown in Table 7.6 and the estimated uncertainty from the fit variations summarised in Table 7.7. Systematic uncertainties with corresponding signs are added ( $x_{-\sigma_{1d}}^{+\sigma_{1u}}$ ,  $x_{-\sigma_{2d}}^{+\sigma_{2u}} \rightarrow x_{-\sigma_{1d}-\sigma_{2d}}^{+\sigma_{1u}+\sigma_{2u}}$ ).

Parameter	Value	Correlation coefficients				
$E_{\text{cal}}/\text{EeV}$	$4.717 \pm 0.071 {}^{+0.080}_{-0.050}$ (sys.)					
$\gamma$	$1.053 \pm 0.016 {}^{+0.012}_{-0.015}$ (sys.)	0.037				
$\frac{\sigma[R_{\mu}^{\text{sh-sh}}]}{R_{\mu}^{\text{sh-sh}}}$ ( $10^{18.3}$ eV)	$0.191 \pm 0.028 {}^{+0.070}_{-0.040}$ (sys.)	0.330	-0.478			
$\frac{\sigma[R_{\mu}^{\text{sh-sh}}]}{R_{\mu}^{\text{sh-sh}}}$ ( $10^{19.5}$ eV)	$0.079 \pm 0.039 {}^{+0.030}_{-0.050}$ (sys.)	-0.096	0.338	-0.497		
$q_0$	$6.724 \pm 0.024 {}^{+0.008}_{-0.005}$ (sys.)	0.013	0.205	-0.166	0.122	
$q_1$	$0.148 \pm 0.019 {}^{+0.002}_{-0.010}$ (sys.)	-0.026	0.181	-0.133	0.098	0.748

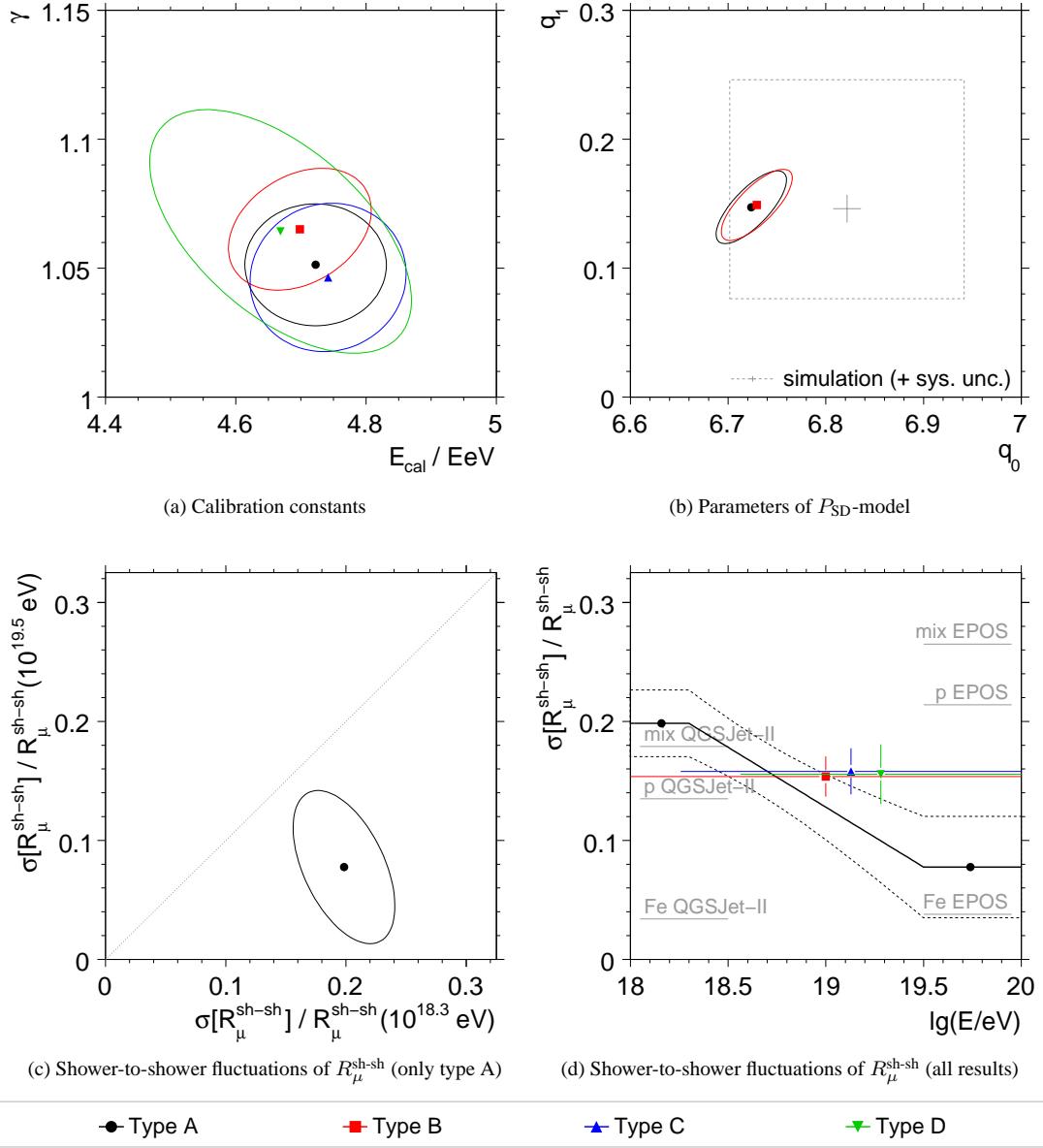


Figure 7.30: The plots show the result of the calibration fit applied to the data set of 411 selected golden hybrid events. The fit is done several times under varying conditions which are listed in Table 7.7. Not all model parameters are free in every type of fit. Parameters which are fixed are not shown. All uncertainty contours are 68 % estimates. a) The calibration constants show no systematic trend as the  $R_\mu^{\text{SD}}$ -threshold increases from type B to D. b) The + -sign and the dashed gray box shows the result and the systematic uncertainty of the  $P_{\text{SD}}$ -model obtained from air shower simulations. c) Only type A allows a linear transition for the shower-to-shower fluctuations and has two fluctuation parameters, which are shown together with their correlation. d) The fits of type B to D enforce constant shower-to-shower fluctuations. Their results are represented by points, the horizontal bars indicate the energy range used to constrain the fit. The fit of type A is shown with a solid black line, the dashed band indicates the uncertainty of the fit. Solid gray lines in the background represent the expected size of the shower-to-shower fluctuations of proton or iron showers and in case of a mixed composition with 40 % iron and 60 % proton. The predictions depend on the hadronic interaction model used in the simulation of the air showers (QGSJet-II or EPOS), but not on the cosmic ray energy  $E$ .

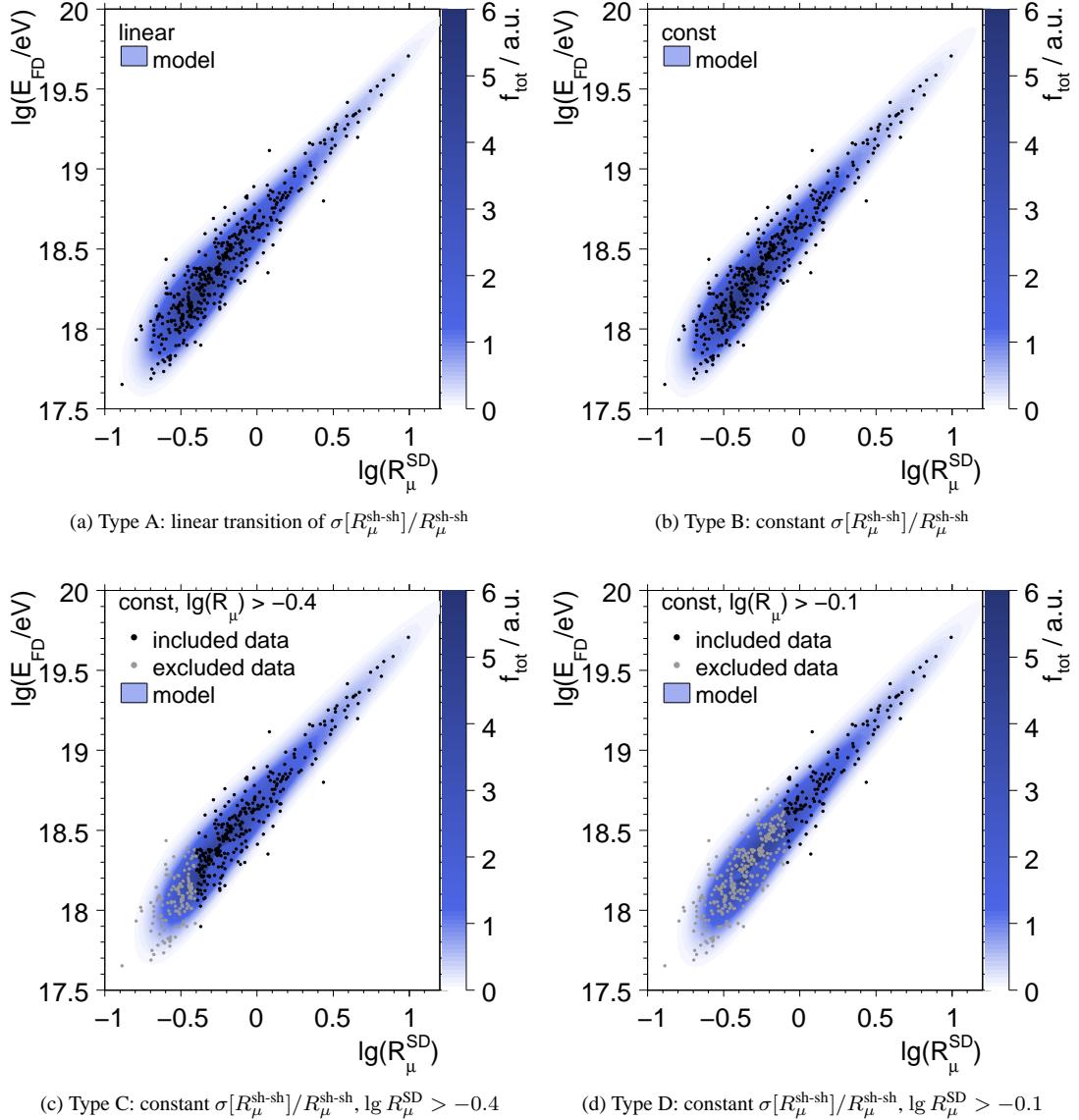


Figure 7.31: The fitted data model  $f_{\text{tot}}$  is compared with the point distribution for the fit variations in Fig. 7.30.

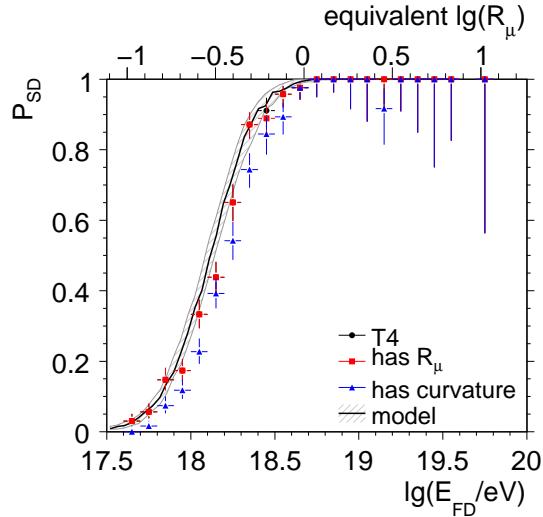


Figure 7.32: The plot shows the SD reconstruction probability  $P_{\text{SD}}$  as a function of the FD energy  $E_{\text{FD}}$ . The legend distinguishes between the probability to pass the T4 selection (black circles), the probability  $P_{\text{SD}}$  to reconstruct the energy estimator (red squares), and the probability to reconstruct the shower front curvature (blue triangles). The latter is optional in this study, but shown nevertheless for comparison. The solid black line shows the prediction of the  $P_{\text{SD}}$ -model with the fitted parameters from Table 7.8. The gray hatched area represents the statistical uncertainty of the model.

As a last cross-check, the potential golden hybrids are used to derive the SD reconstruction probability  $P_{\text{SD}}$  in order to compare it with the fitted  $P_{\text{SD}}$ -model. Since the SD and FD operate independently, it is possible to derive the SD reconstruction probability  $P_{\text{SD}}$  from the conditional probability  $P(\text{SD}|\text{FD})$  to detect an event with the SD if it is detected with the FD

$$P(\text{SD}|\text{FD}) = \frac{P(\text{SD} \cap \text{FD})}{P(\text{FD})} = \frac{P(\text{SD}) P(\text{FD})}{P(\text{FD})} = P_{\text{SD}}. \quad (7.3.3)$$

Fig. 7.32 shows the result of the cross-check. Relevant for this study is the probability  $P_{\text{SD}}$  to reconstruct the SD energy estimator  $R_{\mu}^{\text{SD}}$ . The result obtained from the potential golden hybrids is in good agreement with the  $P_{\text{SD}}$ -model fitted to golden hybrid events. Since the golden hybrid events are a subset of the potential golden hybrids, this analysis is not an independent. But it is based on different concept and thus a good cross-check.

The results found here have several implications, which are discussed below. They also finally allows to convert the SD energy estimator  $R_{\mu}^{\text{SD}}$  event-by-event into a measurement  $E_{\text{SD}}$  of the cosmic ray energy  $E$ . The  $E_{\text{SD}}$ -distribution measured with the SD is used in the next chapter to reconstruct the energy-dependent cosmic ray flux  $J(E)$ .

### Shower-to-shower fluctuations

The shower-to-shower fluctuations  $\sigma_{\text{sh-sh}}[R_{\mu}^{\text{sh-sh}}]/R_{\mu}^{\text{sh-sh}}$  of the SD energy estimator are allowed to make a transition in the data model  $f_{\text{tot}}$ , one obtains fluctuations of about 20 % around  $10^{18}$  eV and about 8 % above  $10^{19.5}$  eV, as shown in Fig. 7.30d).

This indicates a transition from a light or mixed composition to a heavier one between  $10^{18}$  eV and  $10^{20}$  eV, since small shower-to-shower fluctuations are a sign of iron nuclei, as illustrated in Fig. 7.27 and discussed in Chapter 5. The observation of such a transition is in qualitative agreement with other analyses, *e.g.* ref. [29].

The interpretation still has to be tentative, because the statistical and systematic uncertainties are rather large. The statistical significance of the deviation from a constant composition is only slightly larger

than one standard deviation. It reduces to less than one standard deviation, if systematic uncertainties are considered.

The hypothesis of constant shower-to-shower fluctuations can be compared with the hypothesis of a transition with the simplified likelihood ratio test described in Chapter 6. The difference of the log-likelihood values yields

$$-2 \ln \ell = 2 \ln L_{\text{transition}} - 2 \ln L_{\text{constant}} \approx 3.45 \quad (7.3.4)$$

which is not significant for either model.

In order to derive significant results from this kind of analysis in the future, it will be necessary to at least double the amount of events, to improve the models of the FD and SD resolutions, and to clear up the intrinsic systematic bias of the fit. With enough data, it should be possible to fit the contributions of the most dominant cosmic ray nuclei as a function of the cosmic ray energy  $E$  with this approach, similar to the analysis shown in ref. [195].

If the cosmic rays at the highest energies are iron nuclei, a composition analysis based on the SD shower-to-shower fluctuations can be very powerful. Iron showers have the smallest possible shower-to-shower fluctuations, which are consistently predicted by all hadronic interaction models. Therefore, iron showers have a very clean experimental signature.

### **Increase of the muon number on the ground with the cosmic ray energy**

The calibration constant  $\gamma$  measures the rate of increase of the number of muons  $N_\mu$  on the ground with cosmic ray energy  $E$

$$N_\mu \propto E^{1/\gamma}. \quad (7.3.5)$$

Fig. 7.33 compares the fitted calibration constant  $\gamma$  with the values found in air shower simulations of Chapter 5. Because the reconstructed energy estimator  $R_\mu^{\text{SD}}$  was made bias-free with respect to the cosmic ray energy  $E$  in Chapter 6, the fitted value of  $\gamma$  is directly comparable with the simulations. The data value is slightly lower than the range of the simulation predictions, but still agrees within one standard deviation with the lowest simulated value.

This observation can be interpreted. It was shown with a simplified shower model in Chapter 3, that  $\gamma$  is a universal constant in good approximation. It was also shown, that this no longer true if the average cosmic ray mass  $A$  changes with the cosmic ray energy  $E$ . Cosmic rays with a larger mass  $A$  produce more muons. If the mass  $A$  changes with  $E$ , the rate of increase  $1/\gamma$  of number of muons  $N_\mu$  on the ground with  $E$  changes, too.

The following relation was derived between the observed constant  $\gamma_A$  in case of a changing composition and the universal constant  $\gamma$  in case of a constant composition in Chapter 3:

$$\gamma_A = \frac{\gamma}{1 + \beta(\gamma - 1)},$$

whereas  $\beta$  is the rate of change of the average cosmic ray mass  $A$ , described by the formula

$$A(E) = A_0 \times E^\beta.$$

If the composition gets heavier ( $\beta > 0$ ), the observed constant  $\gamma_A$  is smaller than the universal constant  $\gamma$ . This agrees qualitatively with the observation. The simple model for  $\gamma_A$  is not reliable enough to make quantitative predictions, but it serves to illustrate that  $\gamma_A$  carries information about the change of the cosmic ray composition, which may be used in future analyses.

### **Ratio of the muon number in real events and simulations**

Because the reconstructed energy estimator  $R_\mu^{\text{SD}}$  was made bias-free with respect to the cosmic ray energy  $E$  in Chapter 6, the average value  $R_\mu(E) = \langle R_\mu^{\text{SD}} \rangle(E)$  at a particular energy is an estimator of the true total number of muons on the ground:  $N_\mu \propto R_\mu$ . By design, the energy estimator  $R_\mu$  equals one, if the number of muons  $N_\mu$  is equal to the average number of muons in a proton shower simulated with the hadronic interaction model QGSJet-II at  $10^{19}$  eV.

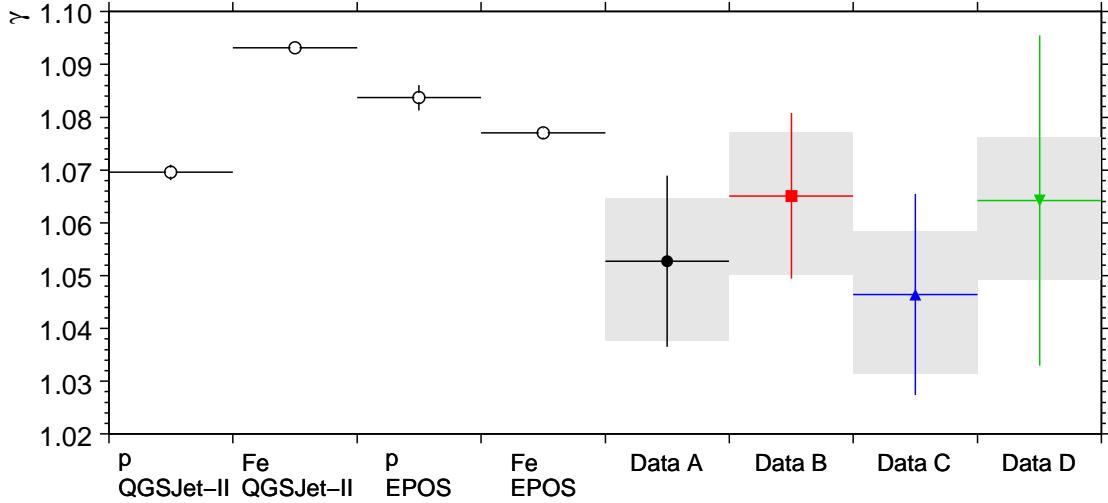


Figure 7.33: The plot shows the exponent  $\gamma$  in the power law relation of the number of muons  $N_\mu$  on the ground and the cosmic ray energy  $E$ :  $N_\mu \propto E^{1/\gamma}$ . The simulation results from Chapter 5 are compared with the fitted values from this chapter. The data entries correspond to the fit setups in Table 7.7. The gray boxes represent the systematic uncertainties.

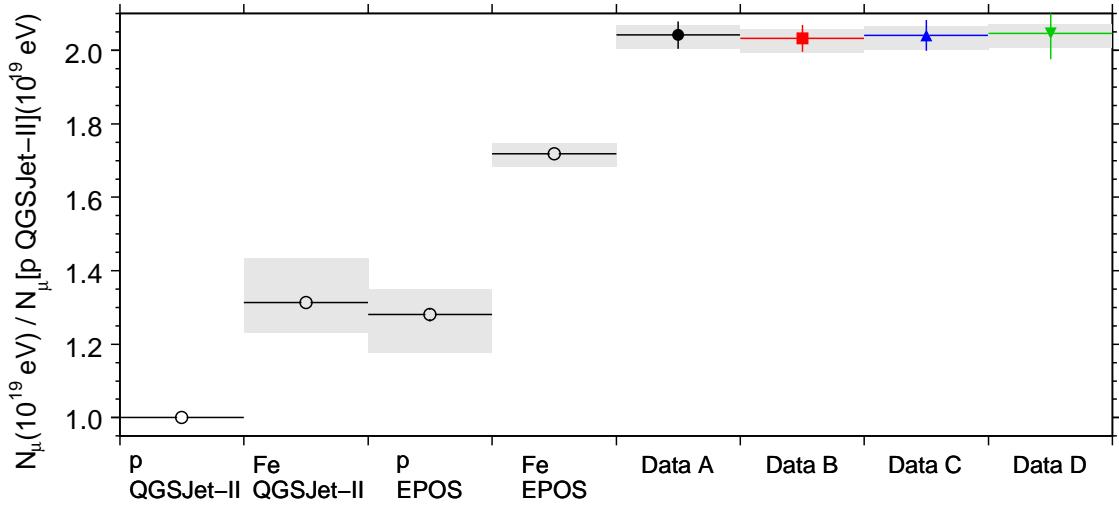


Figure 7.34: The plot shows the number of muons  $N_\mu$  on the ground at a zenith angle  $\theta = 60^\circ$  and a cosmic ray energy  $E = 10^{19}$  eV, normalised to an average proton shower simulated with QGSJet-II. The simulation results from Chapter 5 are compared with the fitted values from this chapter. The data entries correspond to the fit setups in Table 7.7. The gray boxes represent the systematic uncertainties.

The fitted energy calibration function allows to calculate the ratio of the number of muons at  $10^{19}$  eV in the data and in the simulation:

$$\frac{N_\mu[\text{data}](10^{19} \text{ eV})}{N_\mu[\text{p,QGSJet-II}](10^{19} \text{ eV})} = R_\mu(10^{19} \text{ eV}) = \left( \frac{10^{19} \text{ eV}}{E_{\text{cal}}} \right)^{1/\gamma} = 2.042 \pm 0.037^{+0.030}_{-0.040} (\text{sys.}) \quad (7.3.6)$$

Fig. 7.34 compares this result with simulation predictions from Chapter 5. Real events appear even more muon rich than iron showers simulated with the EPOS model. The exact ratio depends slightly on the cosmic ray energy  $E$ . The predictions of air shower simulations are apparently inconsistent with the data in this case. This is a well-known discrepancy, see *e.g.* ref. [196–198], and still under study. It does not affect the other results presented here.

## 7.4 Summary

A new method was applied to derive the calibration constants  $E_{\text{cal}}$  and  $\gamma$  from a special class of 411 events which are detected simultaneously in the FD and SD. The method is based on a complete statistical model  $f_{\text{tot}}$  of the distribution of such events. Thorough tests were applied to the new method which all produced consistent results.

The new method is able to use events with energies between  $10^{18}$  eV and  $10^{18.7}$  eV for the energy calibration where the SD is not fully efficient. This was not possible before. The effect of the limited SD efficiency on the distribution of the observed SD energy estimator  $R_\mu^{\text{SD}}$  is well understood.

The statistical model  $f_{\text{tot}}$  has a component which describes the SD reconstruction probability  $P_{\text{SD}}$  and another component which describes the shower-to-shower fluctuations  $\sigma_{\text{sh-sh}}[R_\mu^{\text{SD}}]$  of the SD energy estimator. Both have free parameters which are fitted to the data. The fitted  $P_{\text{SD}}$ -model will be used to derive the cosmic ray flux  $J(E)$  from the detected number of events in the SD in Chapter 8.

The fitted shower-to-shower fluctuations  $\sigma_{\text{sh-sh}}[R_\mu^{\text{SD}}]$  qualitatively indicate a transition from a light to a heavy composition of cosmic rays between  $10^{18}$  eV and  $10^{20}$  eV, although not with a strong statistical significance. The fitted calibration constant  $\gamma$  is compatible with a changing composition that is getting heavier as the cosmic ray energy  $E$  increases.

A large excess of muons are found in the data compared to predictions from air shower simulations. Even in the most optimistic case, the number of muons in the data is about 20 % larger than in simulated air showers with the same energy.

Chapter **8**

## Flux of ultra-high energy cosmic rays

This chapter covers the determination of the cosmic ray flux at ultra-high energies, which arrives at the Earth. A flux measurement is basically a counting experiment: an *integral* flux is obtained by counting events above a threshold, a *differential* flux by counting events in intervals of a cosmic ray observable  $x$ . The flux is then calculated by dividing the counts through the *exposure* of the counter.

The focus of this study is the differential flux  $J$  as a function of the cosmic ray energy with the surface detector (SD). In this particular case, the flux is calculated as

$$J = \frac{\Delta N}{\Delta E} \frac{1}{P_{\text{SD}}(E, \theta)} \frac{1}{\Lambda}, \quad (8.0.1)$$

whereas  $\Delta N$  are the counts in an energy interval  $\Delta E$ ,  $P_{\text{SD}}$  is the detection efficiency<sup>1</sup>, and the exposure  $\Lambda$  is the product of the solid angle  $\Omega$  of the monitored sky, the collection area  $A_{\text{SD}}$  of the SD, and the run time  $t$  of the observatory:

$$\Lambda(\theta_1, \theta_2, t) = \Omega(\theta_1, \theta_2) A_{\text{SD}}(t) t. \quad (8.0.2)$$

The collection area  $A_{\text{SD}}$  is a function of the run time  $t$ , because the measurement started already when the SD of the Pierre Auger Observatory was not complete. At low energies, where the detection efficiency is not 100 %, the true number of counts is derived by dividing the collected counts through the average detection efficiency  $P_{\text{SD}}$ , which is a function of the distribution of the cosmic ray energies  $E$  and zenith angles  $\theta$  in each bin.

The principle of the flux measurement is simple, but three complications arise in practice.

- **Collection area.** Determining the collection area  $A_{\text{SD}}$  of the flat surface detector is straight forward, but not trivial. Most events were collected during a time when the array had many holes. Air showers that fall into a hole or close to the border of the array still have a finite probability to trigger the SD, especially at large energies, where the showers are intense. A proper criterion has to be found to reject such events, so that the remaining ones correspond to a well defined collection area.
- **Detection efficiency.** Currently, there is no straight forward way to derive the detection probability  $P_{\text{SD}}$  from the data. Several methods exist, three are already applied in Chapter 7. Nevertheless, every method has some disadvantage. With the infill array, a straight forward way will be available in the future.
- **Bin migration.** The Pierre Auger Observatory cannot determine the *true energy*  $E$  of a cosmic ray, it can only measure a *reconstructed energy*  $E_{\text{rec}}$ , which fluctuates around the true value. The differential flux  $J$  is calculated from event counts in small intervals  $\Delta E$ , therefore there is a finite probability for some events to be counted in the wrong bin, as illustrated in Fig. 8.1.

<sup>1</sup>This is often called trigger efficiency, but not every triggered event is actually reconstructable, see Chapter 7.

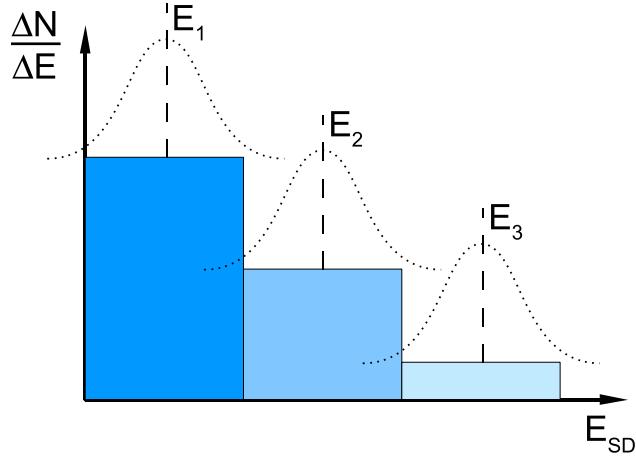


Figure 8.1: The drawing illustrates the effect of bin migration in the measurement of a steeply falling flux spectrum. A true energy  $E_i$  corresponds to each energy bin, equal to the average energy in the bin. If the bin width is close to the experimental resolution of  $E$ , some events fluctuate into adjacent bins. In a steeply falling flux spectrum, the number of events that fluctuate into the lower energy bin make almost no contribution to that energy bin, but events that fluctuate upwards may contribute significantly to the upper energy bin.

The following sections will deal with the first and the third point, while the SD reconstruction probability and its uncertainty are derived in Chapter 7. The bin migration effects are treated by a statistical method, that unfolds the energy resolution effects from the measured flux spectrum. The unfolded flux may then be compared directly with theoretical predictions.

The final result with its statistical and systematic uncertainties will then be compared with the result of the standard analysis, which uses vertical air showers, and a cut-off of the cosmic ray flux above  $4 \times 10^{19}$  eV will be established.

## 8.1 Exposure of the surface detector

The Pierre Auger Observatory uses a powerful yet simple method to assure a well defined exposure of the surface detector [68, 176–178, 199, 200]. On the one hand, the T5 selection is applied to the recorded events which was described in Chapter 6. The T5 selection assures that the shower impact point in the SD array. It does so by rejecting events if the station next to the shower core is surrounded by less than six active stations. This reduces the amount of usable data, but is necessary in any case in order to avoid systematic biases in the SD reconstruction.

From the point of view of each station, the T5 is a geometrical condition which can be tested at any given time. The SD array is a regular grid. Each station is surrounded by an elementary cell of this grid. The area of the elementary cell only contributes to the instantaneous acceptance of the SD, if the station is able to fulfill the T5 criterion at a given point in time. The operational status of every station in the array is monitored by the CDAS every second. The instantaneous configuration of the SD array is therefore available with the time resolution of a second to calculate the SD exposure offline.

This approach has many advantages. The collection area defined in this way

- does not depend on air shower and detector simulations,
- automatically adapts to the growth of the SD array during the construction phase,
- and is able to cope with sudden temporary holes and gaps in the otherwise regular array due to blackouts of individual stations.

The exposure  $\Lambda$  is the product of the solid angle  $\Omega(\theta_1, \theta_2)$  of the sky as seen by the SD, the area  $A_{\text{cell}}$  of the elementary cell and the sum of all instantaneous configurations  $i$ , which lasted for a time  $t_i$  with  $N_{\text{T5},i}$  stations able to fulfill the T5 criterion

$$\Lambda = \Omega(\theta_1, \theta_2) A_{\text{cell}} \sum_i t_i N_{\text{T5},i}. \quad (8.1.1)$$

The cell area in a hexagonal grid with the grid length  $D = 1.5$  km is

$$A_{\text{cell}} = D^2 \frac{\sqrt{3}}{2} \approx 1.949 \text{ km}^2, \quad (8.1.2)$$

and the solid angle of the sky seen by the flat surface detector is

$$\Omega(\theta_1, \theta_2) = \left| \int_0^{2\pi} d\phi \int_{\cos \theta_1}^{\cos \theta_2} d\cos \theta \cos \theta \right| = \pi |\cos^2 \theta_2 - \cos^2 \theta_1|. \quad (8.1.3)$$

The solid angle is derived by regarding the flux of an isotropic source through a flat area element. The SD is a flat detector and therefore its acceptance vanishes at horizontal incidence  $\theta = 90^\circ$ .

Fig. 8.2 shows the collected total exposure for the full sky ( $\Omega = \pi$ ) and the accepted exposure. Certain time periods need to be excluded from the exposure calculation which reduce the total exposure to the accepted exposure. These periods are excluded because the central data acquisition was either off, for example due to maintenance, or not fully operational.

For example, three time periods are excluded due to bugs introduced in software updates of the T3 trigger and communication systems of the CDAS and local trigger logic of the SD stations. The most notable one spans from September 2004 to the beginning of December 2004 and is also the longest excluded period in the regarded time frame.

The majority of the excluded periods lasts less than a day, as shown in Fig. 8.3a). The short periods are rejected because of temporary instabilities in the data acquisition, for example due to communication problems of the CDAS with the SD array during a thunder storm. The signature of such instabilities are drops in the rate of recorded T5 events.

The temporary instabilities are found *a posteriori*. The normal rate of T5 events is about 1.2 event/hexagon/day and quite constant. The arrival of cosmic rays in the surface detector is a Poisson process and thus the time interval between two consecutive T5 events per hexagon follow an exponential distribution. The probability  $P$  to observe an interval  $\Delta t$  larger than the time span  $T$  is

$$P(\Delta t > T) = e^{-\lambda T}, \quad (8.1.4)$$

whereas  $\lambda$  is the expected event rate.

Periods of unstable data acquisition are detected by searching for time intervals  $\Delta t$  with a probability smaller than a threshold  $\alpha$ . Due to the statistical nature of this test, the method will raise some false alarms. But by choosing a very small value of  $\alpha$ , the false alarms appear with a negligible frequency. The currently used value is  $\alpha = 0.5 \times 10^{-6}$ .

The rejected time periods effectively reduce the duty cycle of the surface detector which is shown in Fig. 8.3b). The duty cycle in the first two years of data taking was reduced mostly by technical issues like the software bugs mentioned earlier. These issues were resolved in successive updates of the CDAS soft- and hardware. Since the end of 2005, the data acquisition runs with an average duty cycle of about 98 %. The first two years make up only 10 % of the SD exposure up to the beginning of 2009 and their contribution will become negligible soon.

The final exposure of the surface detector for showers in the zenith angle range  $60^\circ < \theta < 82^\circ$  from 2004/01/01 to 2009/01/01 is

$$\Lambda = 3897 \pm 117 \text{ km}^2 \text{ yr sr}. \quad (8.1.5)$$

The systematic uncertainty of the result is estimated to be at the level of 3 % [68]. It is a conservative combination of uncertainties from the exclusion of time periods, irregularities in the hexagon cells, dead times due to communication problems, and others.

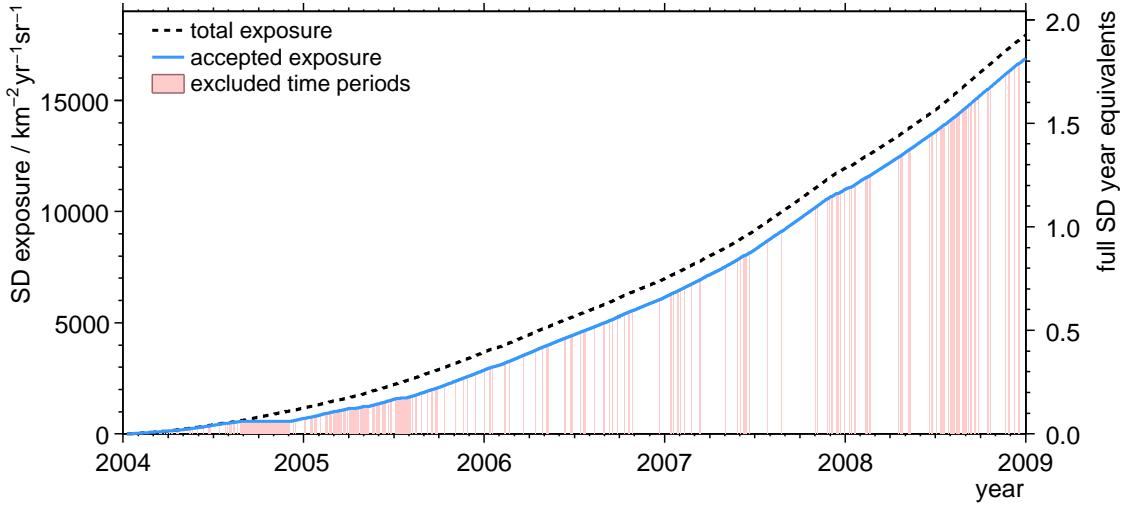


Figure 8.2: The graphs shows the SD exposure for the full sky as a function of the run time of the observatory. The axis on the right hand side shows the exposure in units of the exposure equivalent collected by a full SD array over a year. Certain time periods need to be rejected in which the CDAS was either off or not working reliably. The long excluded period at the end of 2004 is the result of a temporary bug introduced in an update of the T3 trigger software. The excluded time periods reduce the total exposure to the accepted exposure. Only the latter can be used for data analyses.

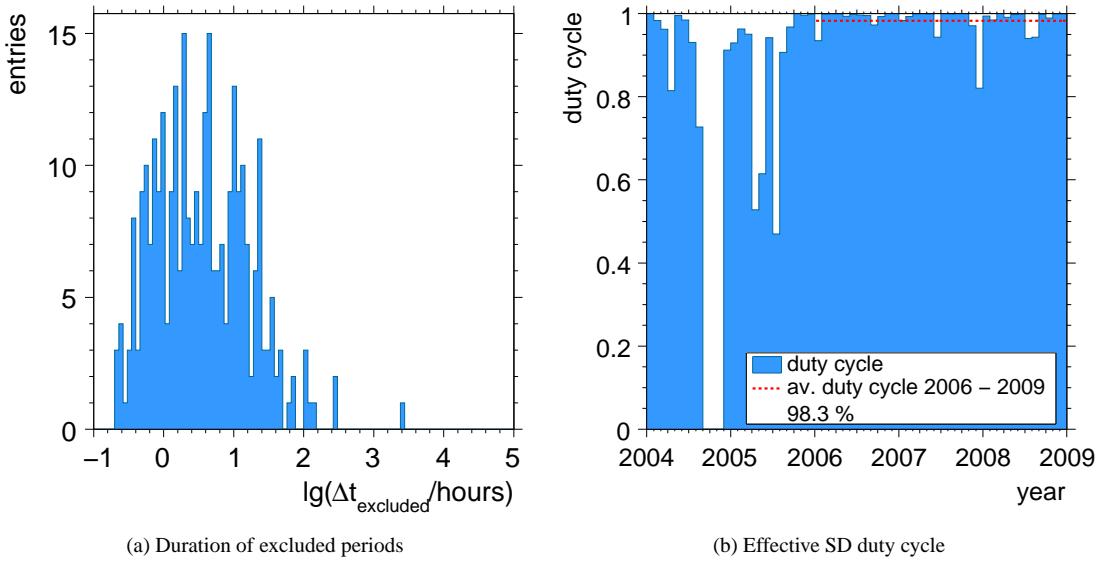


Figure 8.3: a) The histogram shows the distribution of the length  $\Delta t_{\text{excluded}}$  of the excluded time periods from Fig. 8.2. A typical period lasts less than a day. b) The profile shows the effective duty cycle of the SD due to the excluded time periods. Since 2006, the data acquisition runs very stable with an effective duty cycle of 98 %.

## 8.2 Unfolding of the cosmic ray flux

With the exposure  $\Lambda$ , the energy calibration  $E_{\text{SD}} = E_{\text{cal}} R_{\mu}^{\text{SD}\gamma}$  and the model of the reconstruction efficiency  $P_{\text{SD}}$  from Chapter 7, it is possible to derive a first flux estimate with Eq. (8.0.1). However, this flux cannot be compared with other different experiments or with theoretical models, because it is a specific function of the energy resolution of the surface detector in very inclined air showers. The *observed* flux  $g(E_{\text{SD}})$  is a function of the *true* flux  $f(E)$

$$g(E_{\text{SD}}) = \int dE K(E_{\text{SD}}, E) f(E) + \varepsilon(E_{\text{SD}}), \quad (8.2.1)$$

whereas  $K(E_{\text{SD}}, E)$  is the response kernel of the experiment, which describes the finite detector resolution and the limited detection efficiency  $P_{\text{SD}}$ . The function  $\varepsilon(E_{\text{SD}})$  represents the statistical fluctuations in the observed distribution  $g(E_{\text{SD}})$  [201]. It is defined in such a way, that

$$\langle g(E_{\text{SD}}) \rangle = g(E_{\text{SD}}) - \varepsilon(E_{\text{SD}}) \quad (8.2.2)$$

is the average expected distribution. Naturally,  $\varepsilon(E_{\text{SD}})$  is unknown.

Eq. (8.2.1) has to be solved for  $f(E)$  in order to report a comparable result. Ideally, the detector is built in such a way, that the response kernel is close to a delta function  $K(E_{\text{SD}}, E) \approx \delta(E_{\text{SD}}, E)$ , so that  $f(E) \approx g(E_{\text{SD}})$ . In reality however, this is not always possible. For example, the intrinsic shower-to-shower fluctuation of the energy estimator  $R_{\mu}^{\text{SD}}$  in the very inclined showers is unavoidable. In this case, the kernel  $K(E_{\text{SD}}, E)$  has to be simulated or modeled.

Solving Eq. (8.2.1) is still not trivial. It has an exact solution, but this solution is entirely dominated by the fluctuation term  $\varepsilon(E_{\text{SD}})$ . The art of unfolding is to find an approximate solution, which is as close as possible to the true solution but not dominated by the statistical fluctuations.

In the following, the general problem of the unfolding will be illustrated and a short overview of unfolding approaches is given. This introduction is based on ref. [201–207]. Then, the unfolding approach used in this study is discussed, the RUN-method [201]. The systematic uncertainty introduced by the unfolding is derived from a series of Monte-Carlo experiments. Finally, the RUN-method is applied to the SD data.

### 8.2.1 Unfolding problem and solutions

Eq. (8.2.1) is a so called ill-posed problem, because the solution  $f(E)$  depends very sensitively on small fluctuations in the input  $g(E_{\text{SD}})$ . In order to demonstrate this [201, 202], a basic example kernel is considered

$$K(E_{\text{SD}}, E) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(E_{\text{SD}} - E)^2}{2\sigma^2}\right), \quad (8.2.3)$$

which represents the Gaussian fluctuations of the observed energy around the true energy.

A way to solve Eq. (8.2.1) is to expand  $g(E_{\text{SD}})$  and  $f(E)$  into a complete system of orthonormal functions. The Fourier expansion is one possibility:

$$\begin{aligned} f(E) &= \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos\left(k \frac{2\pi}{\Delta E_{\text{SD}}} E\right) + b_k \sin\left(k \frac{2\pi}{\Delta E_{\text{SD}}} E\right) \\ g(E_{\text{SD}}) &= \frac{\tilde{a}_0}{2} + \sum_{k=1}^{\infty} \tilde{a}_k \cos\left(k \frac{2\pi}{\Delta E_{\text{SD}}} E_{\text{SD}}\right) + \tilde{b}_k \sin\left(k \frac{2\pi}{\Delta E_{\text{SD}}} E_{\text{SD}}\right), \end{aligned}$$

whereas  $\Delta E_{\text{SD}}$  is the total range of all observed energies  $E_{\text{SD}}$ . Such an expansion is always possible in real applications, where  $\Delta E_{\text{SD}}$  is finite. The coefficients  $(\tilde{a}_0, \tilde{a}_k, \tilde{b}_k)$  of  $g(E_{\text{SD}})$  can be easily and independently obtained directly from the set of data points  $\{E_{\text{SD},i}\}$ , see e.g. ref. [147]. Note, that the fluctuations  $\varepsilon(E_{\text{SD}})$  are included into the coefficients  $\tilde{a}_0, \tilde{a}_k$ , and  $\tilde{b}_k$ .

The remainder of the right hand side of Eq. (8.2.1) can be solved term-wise with the example kernel

$$\begin{aligned}\int dE K(E_{\text{SD}}, E) \cos(k \frac{2\pi}{\Delta E_{\text{SD}}} E) &= \exp\left(-\frac{k^2 \sigma^2}{2}\right) \cos\left(k \frac{2\pi}{\Delta E_{\text{SD}}} E_{\text{SD}}\right) \\ \int dE K(E_{\text{SD}}, E) \sin(k \frac{2\pi}{\Delta E_{\text{SD}}} E) &= \exp\left(-\frac{k^2 \sigma^2}{2}\right) \sin\left(k \frac{2\pi}{\Delta E_{\text{SD}}} E_{\text{SD}}\right),\end{aligned}$$

and thus one obtains

$$\begin{aligned}g(E_{\text{SD}}) &= \int dE K(E_{\text{SD}}, E) f(E) \\ &= \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \exp\left(-\frac{k^2 \sigma^2}{2}\right) \cos\left(k \frac{2\pi}{\Delta E_{\text{SD}}} E_{\text{SD}}\right) + b_k \exp\left(-\frac{k^2 \sigma^2}{2}\right) \sin\left(k \frac{2\pi}{\Delta E_{\text{SD}}} E_{\text{SD}}\right) \\ &\stackrel{!}{=} \frac{\tilde{a}_0}{2} + \sum_{k=1}^{\infty} \tilde{a}_k \cos\left(k \frac{2\pi}{\Delta E_{\text{SD}}} E_{\text{SD}}\right) + \tilde{b}_k \sin\left(k \frac{2\pi}{\Delta E_{\text{SD}}} E_{\text{SD}}\right).\end{aligned}\quad (8.2.4)$$

The terms of the sums correspond to each other, so that the parameters of the solution  $f(E)$  are

$$a_0 = \tilde{a}_0, \quad a_k = \tilde{a}_k \exp\left(\frac{k^2 \sigma^2}{2}\right), \quad b_k = \tilde{b}_k \exp\left(\frac{k^2 \sigma^2}{2}\right). \quad (8.2.5)$$

If  $g(E_{\text{SD}})$  was a smooth function, the parameters  $\tilde{a}_k$  and  $\tilde{b}_k$  would eventually drop to zero as  $k$  increases. But because of the statistical fluctuations, they are never exactly zero, even for large  $k$ . The corresponding coefficients  $a_k$  and  $b_k$  of the unfolded solution  $f(E)$  on the other hand are blown up exponentially, as  $k$  increases. The full unfolded solution is therefore entirely dominated by random oscillations at high frequencies, generated by random fluctuations in the original data.

It is also possible to understand this intuitively. The limited detector resolution smears out small scale structures in  $f(E)$ . The inverse procedure therefore amplifies such structures. Statistical fluctuations always produce an artificial small scale structures, which are then accidentally amplified by the unfolding, too.

Simply dropping terms of the expansion above some index  $k_{\text{max}}$  does not work, as this also introduces oscillations, which are known as Gibb's phenomenon [201]. Instead, oscillations at high frequencies need to be suppressed in a smooth way. Several approaches were proposed in the past to do this. A short overview is given below.

In order to keep the following discussion more general, the distribution of the true values shall be called  $f(x)$  and the distribution of the observations  $g(y)$ , so that Eq. (8.2.1) turns to

$$g(y) = \int dx K(y, x) f(x) + \varepsilon(y). \quad (8.2.6)$$

Unfolding methods suppress the oscillations by an explicit or implicit *regularisation* of the solution  $f(x)$ .

### Iterative algorithms

The D'Agostini-algorithm [205, 207] and the Gold-algorithm [204] are examples of iterative algorithms.

Iterative algorithms operate on binned distributions, so that the transitions  $f(x) \rightarrow \{f_i\} = \mathbf{f}$  and  $g(y) \rightarrow \{g_j\} = \mathbf{g}$  are made. This turns Eq. (8.2.6) into a matrix equation

$$\mathbf{g} = \mathbf{K}\mathbf{f}. \quad (8.2.7)$$

Each element  $K_{ij}$  of the kernel matrix can now be interpreted as the probability for observing an event in bin  $i$ , if its true value was in bin  $j$ .

Iterative algorithms require, that  $\mathbf{f}$  and  $\mathbf{g}$  have the same rank. If  $\mathbf{g}$  has a larger rank than  $\mathbf{f}$  (the only allowed case), the following modification of the original matrix equation is used instead

$$\mathbf{g} = \mathbf{A}\mathbf{f} \quad \Rightarrow \quad \underbrace{\mathbf{A}^T \mathbf{g}}_{\tilde{\mathbf{g}}} = \underbrace{\mathbf{A}^T \mathbf{A}\mathbf{f}}_{\tilde{\mathbf{A}}\mathbf{f}} \quad \Rightarrow \quad \tilde{\mathbf{g}} = \tilde{\mathbf{A}}\mathbf{f}, \quad (8.2.8)$$

whereas  $\mathbf{A}^T$  is the transposed matrix of  $\mathbf{A}$ . The iterative algorithms are then applied to new variables  $\tilde{\mathbf{g}}$  and  $\tilde{\mathbf{A}}$ , which have the right rank.

The solution  $\mathbf{f}$  is then approximated in successive steps. The D'Agostini- and Gold-algorithm do not approximate  $\mathbf{f}$  directly, but the transformation matrix  $\mathbf{D}$  with the property

$$\mathbf{f} = \mathbf{D} \mathbf{g}. \quad (8.2.9)$$

The next iteration of  $\mathbf{D}$  is always calculated from the observations  $\mathbf{g}$  and the current intermediate solution of  $\mathbf{f}$ , which is usually initialised with  $\mathbf{g}$ .

Approximating  $\mathbf{D}$  instead of  $\mathbf{f}$  has the advantage, that a regularisation condition can be enforced upon  $\mathbf{D}$  by construction. The D'Agostini- and the Gold-algorithm enforce positive-definiteness of  $\mathbf{D}$ , which assures positive entries for  $\mathbf{f}$  and damps the development of oscillations with large frequency and large amplitudes.

The D'Agostini-algorithm is particularly well motivated by the theory of Bayesian statistic. It regards the intermediate solution as the prior knowledge, from which the next iteration of  $\mathbf{D}$  is inferred in regard to the observed data distribution  $\mathbf{g}$ . The Gold-algorithm is not based on such a deeper concept, but very powerful nevertheless. It is successfully applied to obtain the cosmic ray flux from fluorescence detector data in ref. [203].

Convergence of iterative algorithms can usually be proven, but the converged solution is in general not meaningful. The intermediate solution starts to diverge from a good regularised solution at a certain iteration depth.

This is a consequence of the central feature of iterative algorithms. It is possible to show [208], that they lead to faster convergence of eigen vectors of the solution  $\mathbf{f}$  which correspond to large scale structures than those which represent small scale structures. The amplitudes of the latter is small at the beginning so that undesired oscillations are avoided if the iteration is stopped at an intermediate step. This leads an implicit regularisation.

Unfortunately, the criterion of when to stop the iteration for the optimal result is not well defined. A particularly useful criterion, the “weighted mean squared error” is expensive to calculate [203]. More severe is a systematic bias, which iterative algorithms introduce to the first and last bins of the solution  $\mathbf{f}$  [203], if the data distribution  $\mathbf{g}$  does not fall off toward both ends.

### Regularised fits

The RUN-algorithm [201, 202, 209] and the SVD-based unfolding [206] are regularised fits. In these approaches, a forward folded parameterisation of  $f(x)$  is fitted to  $g(y)$  under special conditions that suppresses the oscillations of  $f(x)$ .

The parameterisation of  $f(x)$  is usually a linear one

$$f(x|\mathbf{a}) = \sum_i a_i p_i(x), \quad (8.2.10)$$

which is particularly easy to fit. Note, that the act of binning  $f(x)$  can also be formally described as a linear parameterisation, although as a very rough one.

Inserting this parameterisation into the right hand side of Eq. (8.2.6) leads to

$$\int dx K(y, x) f(x) = \int dx K(y, x) \sum_i a_i p_i(x) = \sum_i a_i \tilde{p}_i(y) \quad (8.2.11)$$

with  $\tilde{p}_i(y) = \int dx K(y, x) p_i(x).$

The unfolding is thus reduced to a fit of a system of linear functions to the observed distribution  $g(y)$ .

The fit itself is explicitly regularised by introducing an additional constraint. In the case of a fit based on the likelihood method, a regularisation term  $r(\mathbf{a})$  is added to the negative logarithm of the likelihood function

$$-\ln L(\mathbf{a}) = -\ln L_0(\mathbf{a}) + \tau r(\mathbf{a}), \quad (8.2.12)$$

whereas  $L_0(\mathbf{a})$  is the unmodified likelihood and  $\tau$  is a tuning parameter of the regularisation.

The regularisation term  $r(\mathbf{a})$  is chosen in such a way that it has a minimum at  $\mathbf{a}$  and becomes large, if the solution has fast oscillations. Since the regularisation term increases in a smooth way with the oscillations, they are gradually suppressed and Gibb's phenomenon is avoided.

An optimal choice for the regularisation term  $r(\mathbf{a})$  is not established. A common choice is to minimise the curvature of the solution [201, 206]

$$r(\mathbf{a}) = \int dx (f(x))^2, \quad (8.2.13)$$

but functions of other derivatives of  $f(x)$  or completely different approaches are also considered. The solution  $f(x)$  is a function of the tuning factor  $\tau$ . The tuning factor  $\tau$  has to be chosen carefully from problem to problem. If it is too large, the solution  $f(x)$  will be dominated by the regularisation. If it is too small, the influence of the fluctuations  $\varepsilon(y)$  are not efficiently suppressed.

The details of the parameterisation of  $f(x)$ , the fit and the regularisation term  $r(\mathbf{a})$  vary from method to method. The RUN-algorithm uses a sophisticated combination of parameterisation and regularisation. It also has a semi-automatic way of suggesting the proper tuning of  $r(\mathbf{a})$  as a function of the input data. The SVD-based unfolding offers comparable features. It was decided to use the RUN-algorithm in this study.

### 8.2.2 RUN-algorithm

The RUN algorithm is very sophisticated and was successfully applied in many experiments [201]. The RUN-algorithm has a well established reference implementation, which can be obtained online [210]. This study is based on this reference implementation.

The regularisation is the central aspect of unfolding algorithms, but introduces a systematic bias to the result  $f(x)$ . It will be shown in the following that the RUN-approach only suppresses contributions to the solution  $f(x)$  which are not statistically significant if the regularisation is tuned properly.

The RUN-algorithm operates on the binned distribution  $\{g_j\}$  of the data. The main features of the algorithm are the following:

- The detector kernel  $K(y, x)$  is built internally from a set of Monte-Carlo events, which the user has to supply.
- The unfolded solution  $f(x)$  is factorised in the following way

$$f(x) = f_{\text{user}}(x) f_{\text{corr}}(x), \quad (8.2.14)$$

whereas  $f_{\text{user}}$  is a user-supplied input, which should be as close to the final solution as possible. Only the correction  $f_{\text{corr}}(x)$  is fitted by the algorithm.

- The parameterisation of  $f_{\text{corr}}(x)$  is done with B-splines [211] of order 4. Splines have optimal approximative properties [212]. B-splines form a linear independent basis of splines, so that  $f(x)$  may be written in the form of Eq. (8.2.10).
- The fit of the parameters  $\mathbf{a}$  of the solution is performed with a likelihood method. The distribution of the bin entries of the data distribution  $\{g_j\}$  is correctly modeled as a Poisson distribution.
- The regularisation is done by minimising the curvature of the correction  $f_{\text{corr}}$ . As a special consequence of the B-spline parameterisation, the total curvature can be calculated with a constant matrix  $\mathbf{C}$  [201]:

$$r(\mathbf{a}) = \int dx (f_{\text{corr}}(x))^2 = \mathbf{a}^T \mathbf{C} \mathbf{a}, \quad (8.2.15)$$

whereas  $\mathbf{a}$  is the parameter vector.

The advantage of Eq. (8.2.14) becomes apparent now. The full solution  $f(x)$  may have a large curvature, but with a proper choice of  $f_{\text{user}}$ , the correction  $f_{\text{corr}}(x)$  will be rather smooth. In such a case, the regularisation produces the minimal bias.

The regularised unfolding is based on the maximisation of the likelihood of the parameterised solution

$$f(x) = f_{\text{user}}(x) f_{\text{corr}}(x) = \sum_i a_i p_i(x) f_{\text{user}}(x), \quad (8.2.16)$$

which is forward folded with the detector kernel  $K(y, x)$ . The folded basis functions are

$$\tilde{p}_i(y) = \int dx K(y, x) f_{\text{user}}(x) p_i(x). \quad (8.2.17)$$

The expected value in bin  $j$  of the data distribution  $\{g_j\}$  is calculated as

$$\begin{aligned} \langle g_j \rangle &= \frac{1}{y_j - y_{j-1}} \int_{y_{j-1}}^{y_j} dy K(y, x) f(x) = \sum_i A_{ji} a_i \\ \text{with } A_{ji} &= \frac{1}{y_j - y_{j-1}} \int_{y_{j-1}}^{y_j} dy \tilde{p}_i(y), \end{aligned} \quad (8.2.18)$$

the Poisson distribution is the statistical model of the distribution of observed value  $g_j$  around the true value  $\langle g_j \rangle$ .

This defines the necessary input for a likelihood function without the regularisation term. The negative logarithm of this function can be written as

$$-\ln L_0(\mathbf{a}) = -\sum_j \langle g_j \rangle(\mathbf{a}) + \sum_j g_j \ln \langle g_j \rangle(\mathbf{a}) + \text{const.} \quad (8.2.19)$$

Adding the regularisation term leads to

$$-\ln L(\mathbf{a}) = -\ln L_0(\mathbf{a}) + \frac{1}{2} \tau r(\mathbf{a}) = -\ln L_0(\mathbf{a}) + \frac{1}{2} \tau \mathbf{a}^T \mathbf{C} \mathbf{a}, \quad (8.2.20)$$

with  $\tau$  as the tuning parameter of the regularisation. The factor  $1/2$  is just a convenient definition, as will become apparent below.

A minimisation of this function yields a regularised solution as a function of  $\tau$ . For a fixed  $\tau$ , this can be done with standard methods. To see the effect of the regularisation,  $-\ln L(\mathbf{a})$  is expanded into a Taylor-series around a point  $\tilde{\mathbf{a}}$  close to the minimum:

$$-\ln L(\mathbf{a}) \simeq -\ln L(\tilde{\mathbf{a}}) - (\mathbf{a} - \tilde{\mathbf{a}}) \mathbf{h} + \frac{1}{2} (\mathbf{a} - \tilde{\mathbf{a}})^T \mathbf{H} (\mathbf{a} - \tilde{\mathbf{a}}) + \frac{1}{2} \tau \mathbf{a}^T \mathbf{C} \mathbf{a}, \quad (8.2.21)$$

with the gradient vector  $\mathbf{h}_i = -\partial(-\ln L(\mathbf{a}))/\partial a_i$  and the Hesse matrix  $H_{ij} = \partial^2(-\ln L(\mathbf{a}))/(\partial a_i \partial a_j)$ . Close to the minimum, this approximation is usually good and higher order terms of the Taylor series may be neglected. If constant contributions are omitted, the expression simplifies to

$$-\ln L(\mathbf{a}) \simeq -\mathbf{a}^T (\mathbf{h} + \mathbf{H} \tilde{\mathbf{a}}) + \frac{1}{2} \mathbf{a}^T \mathbf{H} \mathbf{a} + \frac{1}{2} \tau \mathbf{a}^T \mathbf{C} \mathbf{a} \quad (+\text{const.}) \quad (8.2.22)$$

At the minimum, the Hesse matrix is approximately equal to the inverse of the covariance matrix  $\mathbf{V}$  of the solution vector  $\tilde{\mathbf{a}}$

$$\mathbf{H} \simeq \mathbf{V}^{-1}. \quad (8.2.23)$$

The next step is to derive a transformation for  $\mathbf{a}$ , which transforms  $\mathbf{H}$  to the unitary matrix and further simplifies Eq. (8.2.22). The covariance matrix  $\mathbf{V}$  is then a unitary matrix, too, and the components of the transformed parameter vector are uncorrelated random variables with variance  $\sigma = 1$ .

Since the Hesse matrix is positive definite and symmetric at the minimum, it is possible to find an orthogonal matrix  $\mathbf{U}_1$  (a rotation in parameter space) that diagonalises  $\mathbf{H}$

$$\mathbf{U}_1^T \mathbf{H} \mathbf{U}_1 = \mathbf{D}, \quad (8.2.24)$$

whereas  $\mathbf{D}$  is diagonal. To turn  $\mathbf{D}$  into the unitary matrix, another matrix  $\mathbf{D}^{-1/2}$  is appended to the transformation, which is defined as

$$D_{ij}^{-1/2} = \begin{cases} 1/\sqrt{D_{jj}} & , i = j \\ 0 & , i \neq j, \end{cases} \quad (8.2.25)$$

so that

$$\mathbf{D}^{-1/2} \mathbf{U}_1^T \mathbf{H} \mathbf{U}_1 \mathbf{D}^{-1/2} = \mathbf{1}. \quad (8.2.26)$$

It is possible to append yet another orthogonal matrix  $\mathbf{U}_2$  without changing this result:

$$\mathbf{U}_2^T \mathbf{D}^{-1/2} \mathbf{U}_1^T \mathbf{H} \mathbf{U}_1 \mathbf{D}^{-1/2} \mathbf{U}_2 = \mathbf{U}_2^T \mathbf{1} \mathbf{U}_2 = \mathbf{1}. \quad (8.2.27)$$

This freedom will be exploited in a few steps. Comparing Eq. (8.2.27) with Eq. (8.2.22) shows, that the transformation leads to a new parameter vector  $\bar{\mathbf{a}}$  with

$$\mathbf{a} = \mathbf{U}_1 \mathbf{D}^{-1/2} \mathbf{U}_2 \bar{\mathbf{a}} \quad (8.2.28)$$

and a corresponding new set of basis functions  $\bar{\mathbf{p}}(x)$ , which are a linear combination of the original basis functions

$$\begin{aligned} f_{\text{corr}}(x) &= \mathbf{a}^T \mathbf{p}(x) = \bar{\mathbf{a}}^T \mathbf{U}_2^T \mathbf{D}^{-1/2} \mathbf{U}_1^T \mathbf{p}(x) = \bar{\mathbf{a}}^T \bar{\mathbf{p}}(x) \\ \text{with } \bar{\mathbf{p}}(x) &= \mathbf{U}_2^T \mathbf{D}^{-1/2} \mathbf{U}_1^T \mathbf{p}(x). \end{aligned} \quad (8.2.29)$$

The transformation turns Eq. (8.2.22) into

$$\begin{aligned} -\ln L(\mathbf{a}_2) &\simeq -\bar{\mathbf{a}}^T \mathbf{U}_2^T \mathbf{D}^{-1/2} \mathbf{U}_1^T (\mathbf{h} + \mathbf{H} \bar{\mathbf{a}}) + \frac{1}{2} \bar{\mathbf{a}}^T \bar{\mathbf{a}} \\ &\quad + \frac{1}{2} \tau \bar{\mathbf{a}}^T \mathbf{U}_2^T \mathbf{D}^{-1/2} \mathbf{U}_1^T \mathbf{C} \mathbf{U}_1 \mathbf{D}^{-1/2} \mathbf{U}_2 \bar{\mathbf{a}} \quad (+\text{const.}). \end{aligned} \quad (8.2.30)$$

The freedom of the last rotation matrix  $\mathbf{U}_2$  may be now used to make also the regularisation matrix  $\mathbf{C}$  in the last term diagonal. To show this, the other transformations are applied to  $\mathbf{C}$  to get a matrix  $\mathbf{C}_1$

$$\mathbf{U}_2^T \mathbf{D}^{-1/2} \mathbf{U}_1^T \mathbf{C} \mathbf{U}_1 \mathbf{D}^{-1/2} \mathbf{U}_2 = \mathbf{U}_2^T \mathbf{C}_1 \mathbf{U}_2. \quad (8.2.31)$$

The matrix  $\mathbf{C}$  is positive definite and symmetric, applying  $\mathbf{D}^{-1/2}$  or  $\mathbf{U}_1$  does not change this property. Thus, an orthogonal matrix  $\mathbf{U}_2$  may be chosen such, that

$$\mathbf{S} = \mathbf{U}_2^T \mathbf{C}_1 \mathbf{U}_2 \quad (8.2.32)$$

is a diagonal matrix. The diagonal entries in  $S_{jj}$  in  $\mathbf{S}$  may be freely arranged in increasing order  $S_{jj} \leq S_{j+1,j+1}$  by permutating pairs of columns and rows of  $\mathbf{U}_2$ .

With this final step, Eq. (8.2.22) reduces to

$$-\ln L(\mathbf{a}_2) \simeq -\bar{\mathbf{a}}^T \mathbf{U}_2^T \mathbf{D}^{-1/2} \mathbf{U}_1^T (\mathbf{h} + \mathbf{H} \bar{\mathbf{a}}) + \frac{1}{2} \bar{\mathbf{a}}^T (\mathbf{1} + \tau \mathbf{S}) \bar{\mathbf{a}} \quad (+\text{const.}). \quad (8.2.33)$$

The solution of this equation is obtained with the minimum condition  $\nabla \ln L(\bar{\mathbf{a}}) \stackrel{!}{=} 0$  as

$$\bar{\mathbf{a}} = (\mathbf{1} + \tau \mathbf{S})^{-1} \mathbf{U}_2^T \mathbf{D}^{-1/2} \mathbf{U}_1^T (\mathbf{h} + \mathbf{H} \bar{\mathbf{a}}). \quad (8.2.34)$$

The solution shows, that a regularised parameter  $\bar{a}_j$  is reduced by a factor relative to the unregularised parameter  $\bar{a}'_j$

$$\bar{a}_j = \frac{1}{1 + \tau S_{jj}} \bar{a}'_j. \quad (8.2.35)$$

The damping factor depends on the curvature of the corresponding basis function  $\bar{p}_j(x)$  of the parameter and the tuning parameter  $\tau$ .

It is thus demonstrated, that the regularisation vanishes for terms with a small curvature, while the parameters of basis functions with fast oscillations are pulled to zero in a smooth way, avoiding Gibb's phenomenon. The tuning parameter  $\tau$  defines how early the suppression sets in. Through the suppression, the regularised solution has an effective rank

$$m_0 = \sum_j \frac{1}{1 + \tau S_{jj}}, \quad (8.2.36)$$

which is smaller than the total number of basis functions and depends on the tuning parameter  $\tau$ .

To semi-automatically suggest a value for  $\tau$ , **RUN** performs a significance analysis on the unregularised parameters  $\bar{a}'_j$ . These parameters have variance 1. If a parameter  $\bar{a}'_j$  is insignificant, its mean  $\langle \bar{a}'_j \rangle$  is approximately zero and  $(\bar{a}'_j)^2$  approximately follows a  $\chi^2$  distribution. This allows to define a confidence level of the hypothesis that the parameter is insignificant, given an observed value of  $(\bar{a}'_j)^2$ . In **RUN**, a confidence level of 95 % is used which corresponds to  $(\bar{a}'_j)^2 \leq 3.84$ .

A lower limit of the effective rank  $m_0$  is given by the first  $j$  in descending order, for which this hypothesis has to be rejected. The tuning parameter  $\tau$  is then obtained by solving Eq. (8.2.36) with a given  $m_0$ .

Finally,  $m_1$  binned data points  $\{f_i\}$  can be extracted from the solution  $f(x)$

$$f_i = \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} dx \sum_j \bar{a}_j f_{\text{user}}(x) \bar{p}_j(x). \quad (8.2.37)$$

The sum and the integral commute, so that standard uncertainty propagation can be used to turn the covariance matrix  $\mathbf{V} = (1 + \tau \mathbf{S})^{-2}$  of the parameter vector  $\bar{\mathbf{a}}$  into a covariance matrix of the solution vector  $\mathbf{f}$ .

The parameters  $\bar{a}_i$  are uncorrelated, but not the final data points. The author of **RUN** recommends to extract  $m_1 \leq m_0$  bins from  $f(x)$ , to reduces the bin-to-bin correlations. If the bin-to-bin correlations are negligible, the final bins of the solution can be visualised and interpreted in the usual fashion as points with error bars.

### Systematic analysis and optimal settings

The **RUN**-algorithm automatically calculates a recommended value for effective degrees of freedom  $m_0$  of the solution, but the value still has to be supplied by the user. The algorithm only provides a lower limit for  $m_0$ . It seems to be an open question whether this lower limit is an optimal choice. Even in the numerical example of the original publication of the algorithm [201], the author uses a considerably larger value for  $m_0$ .

Another free parameter of the algorithm is the number of knots  $n_{\text{knot}}$  for the B-spline parameterisation. The author of **RUN** recommends  $n_{\text{knot}} = 2m_0 + 3$ , but it is also clearly stated, that sometimes more or less knots may lead to better results. In general, the result of **RUN** should depend only weakly on the exact choice on  $n_{\text{knot}}$  and  $m_0$ , but the question for the optimal combination remains.

The optimal settings are derived from an analysis of Monte-Carlo experiments in the following, which also serves as an end-to-end evaluation of the **RUN**-algorithm. Of particular interest is the analysis of the

Table 8.1: The table summarises the binning choice for the input and output of the unfolding algorithm. The input distribution has equi-distant bins, the binning of the output distribution is more coarse and the last two bins with small statistics are slightly larger.

distribution	bins	lower bin edges in $\lg(E/\text{eV})$								
observed	20	18.0	18.1	18.2	...	19.8	19.9	20.0		
unfolded	9	18.0	18.3	18.5	18.7	18.9	19.1	19.3	19.5	19.7

bias of the solution  $\mathbf{f}$  and the size of the bin-to-bin correlations in the unfolded distribution  $\{f_i\}$ . The binning choice of the input and output distributions is thereby fixed and summarised in Table 8.1.

In case of the input distribution, the binning has a negligible impact on the unfolding, as long as the number of bins is about twice as large as  $m_0$ . The binning choice of the output is a compromise between using as few bins as possible to keep the bin-to-bin correlations small, and yet having enough to observe the interesting features in the unfolded flux distribution.

The Monte-Carlo events are generated in a similar way as in Chapter 7, and make use of the data models derived in there.

- (1) A true energy  $E$  is drawn at random from a model of the true flux. The model is a piece-wise power law with spectral indices based on ref. [188]:

$$\frac{dJ}{dE} \propto \begin{cases} E^{-3}, & \lg(E/\text{eV}) < 18.4 \\ E^{-2.7} 10^{-0.3 \times 18.4}, & 18.4 \leq \lg(E/\text{eV}) < 19.6 \\ E^{-3.9} 10^{-0.3 \times 18.4} 10^{1.2 \times 19.6}, & 19.6 \leq \lg(E/\text{eV}). \end{cases} \quad (8.2.38)$$

The trailing factors fulfill the boundary conditions. A corresponding true zenith angle  $\theta$  is drawn at random from the distribution

$$\frac{dJ}{d\sin^2 \theta} \propto 1 \quad (8.2.39)$$

in the range  $60^\circ < \theta < 82^\circ$  for each true energy  $E$ . An ideal energy estimator  $R_\mu$  is calculated with  $R_\mu = (E/E_{\text{cal}})^{1/\gamma}$ .

- (2) The ideal energy estimator  $R_\mu$  is fluctuated according to the model of the shower-to-shower fluctuations and the model of the SD resolution:  $R_\mu \rightarrow R_\mu^{\text{sh-sh}} \rightarrow R_\mu^{\text{SD}}$ . The reconstructed energy estimator is converted back into an energy with the calibration function  $R_\mu^{\text{SD}} \rightarrow E_{\text{SD}}$ .
- (3) The true cosmic ray energy  $E$  is fluctuated according to the model of the fluorescence detector resolution:  $E \rightarrow E_{\text{FD}}$ .
- (4) The event is accepted at random according to the model of the SD reconstruction probability.

These steps are repeated until  $N = 45000$  pseudo events are collected, and, independently,  $N_{\text{MC}}$  Monte-Carlo events. Only accepted events count as pseudo events, but accepted *and* rejected events count as Monte-Carlo events. RUN uses both accepted and rejected events to calculate the detector kernel  $K(E_{\text{SD}}, E)$  internally. The rejected events are used to calculate the detector efficiency. The number  $N_{\text{MC}}$  is chosen such, that the number of accepted events in the Monte-Carlo sample is  $20 \times N$ . Both data sets together form a Monte-Carlo experiment and 20 independent Monte-Carlo experiments are generated. The Monte-Carlo experiments are then unfolded with RUN.

To assess, whether the bin-to-bin correlations of an unfolded distribution may be neglected, RUN proposes the following test [202]:

- (1) Take the full covariance matrix  $\mathbf{V}[\mathbf{f}]$  of the unfolded distribution. Generate a large number (5000 events in this case) of random samples from the full covariance matrix, with a multivariate normal distribution with zero mean

$$h(\mathbf{x}) = \frac{1}{(2\pi)^{k/2} |\mathbf{V}|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right). \quad (8.2.40)$$

- (2) Calculate the  $\chi^2$  for each sample, taking only the diagonal entries in the covariance matrix into account

$$\chi_{\text{obs}}^2 = \sum_i \frac{X_i^2}{V_{ii}} \quad (8.2.41)$$

Calculate the probability  $P(\chi^2 \geq \chi_{\text{obs}}^2) = \int_{\chi_{\text{obs}}^2}^{\infty} d\chi^2 f_{\chi^2}(\chi^2 | n_{\text{dof}})$ , whereas  $f_{\chi^2}(\chi^2 | n_{\text{dof}})$  is the theoretical  $\chi^2$  distribution for  $n_{\text{dof}}$  degrees of freedom.

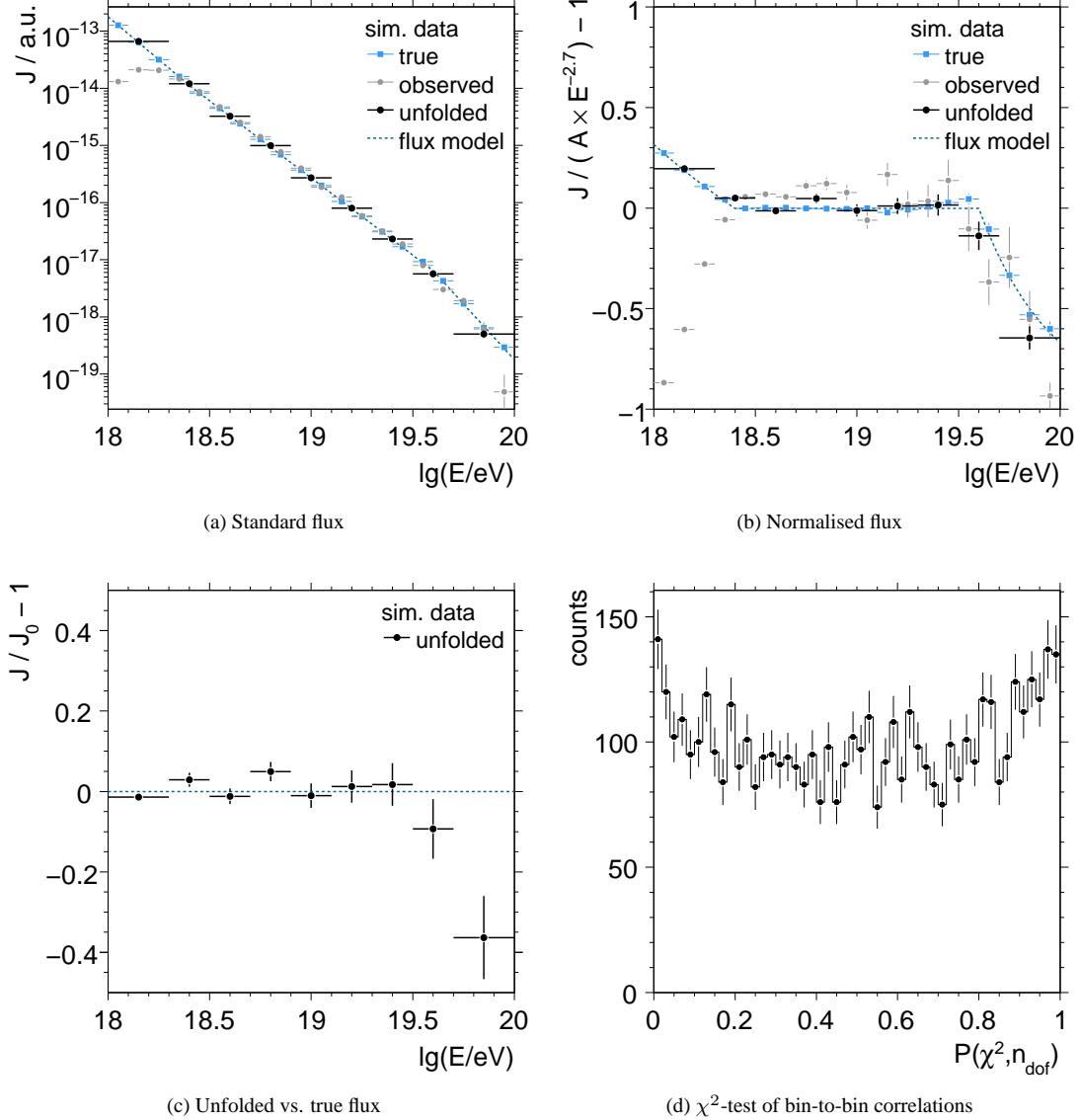


Figure 8.4: Shown is one of the Monte-Carlo experiments, which are used to test the unfolding with the RUN algorithm. This example is unfolded with  $m_0 = 9$  and  $n_{\text{knot}} = 21$ . a) The dashed line is the flux model. The true flux (blue squares) is a random sample derived from the flux model. The observed flux (gray circles) is measured in the detector. The unfolded flux (black circles) is the result of the unfolding. b) The same is shown, but the fluxes are shown relative to a reference flux  $A \times E^{-2.7}$ . c) The points show the deviation of the unfolded result from the flux model. d) The histogram shows the result of  $\chi^2$ -test of the bin-to-bin correlations, as described in the text.

If the correlations are negligible, the obtained distribution is approximately flat.

Fig. 8.4 shows an example Monte-Carlo experiment. The input distribution is well reproduced by the unfolding except in the last bin. A comparable measure of the unfolding bias is the sum

$$\sum_i \left( \frac{J_i - J_{0,i}}{J_{0,i}} \right)^2 \quad (8.2.42)$$

over all bins of the solution, whereas  $J_i$  and  $J_{0,i}$  are the unfolded and the true flux in bin  $i$ .

The distribution obtained from the  $\chi^2$ -test in the example is not exactly flat. Instead it has a “U”-shape, which is a result of the bin-to-bin correlations. A comparable measure of the non-flatness of the distribution is obtained by fitting a second order polynomial

$$p_0 + p_1 x + p_2 x^2, \quad (8.2.43)$$

the value  $p_2/\sigma[p_2]$  serves as a measure of the curvature and therefore of the size of the bin-to-bin correlations.

Fig. 8.5 show the impact of the unfolding parameters  $m_0$  and  $n_{\text{knot}}$  on these measures. The smallest bin-to-bin correlations are obtained, if the effective number of degrees of freedom in the unfolding is equal to the number of bins in the final distribution:  $m_0 = m_1 = 9$ . Increasing the number of knots  $n_{\text{knot}}$  increases the bias of the result, if a smaller value is used  $m_0$ . At  $m_0 = 9$  and above, the impact of  $n_{\text{knot}}$  becomes negligible.

Based on these results, the following unfolding parameters are selected:

$$m_0 = 9 \quad \text{and} \quad n_{\text{knot}} = 2m_0 + 3 = 21.$$

The systematics of the unfolding with these parameters are further analysed. The results are shown in Fig. 8.6. The unfolded result cannot be entirely bias-free as mentioned before. However, the RUN-algorithm should introduce only a small bias compared to the statistical uncertainty in each bin for an optimal choice of  $m_0$ , which is can be confirmed.

Only the last bin shows a considerable bias, while it is negligible in the other bins. The bias of 10 % in the last bin will be corrected in the unfolding of the data, while adding a systematic uncertainty of the same order to this bin.

The RUN-algorithm calculates the full covariance matrix  $V$  of the statistical uncertainty of the unfolded flux  $J$ . This matrix is not diagonal. If the off-diagonal entries are neglected  $\sigma_{\text{est},i}[J] = V_{ii}$ , one obtains an estimate of the real fluctuations which is about 20 % too small on average, as shown on the right hand side of Fig. 8.6.

In order to approximately compensate this, the uncertainties  $\sigma_{\text{est},i}[J]$  of the unfolded data will be increased by 20% in graphic displays. Fits of the unfolded flux should make use of the full covariance matrix, which will be given along with numerical tables of the unfolded flux.

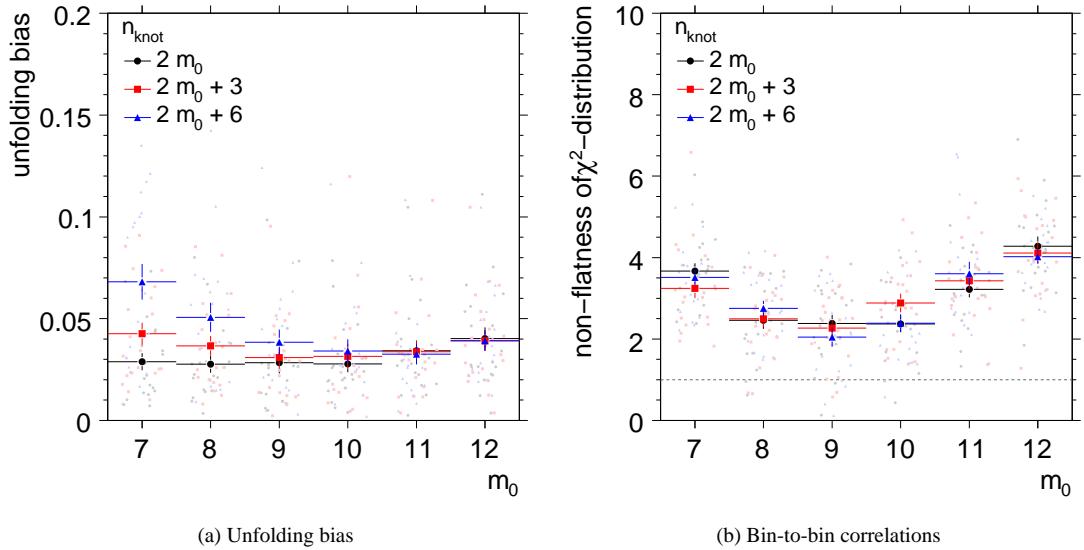


Figure 8.5: Shown is the result of the optimisation analysis of the unfolding parameters  $m_0$  and  $n_{\text{knot}}$ . The measures are defined in the text, small values are better.

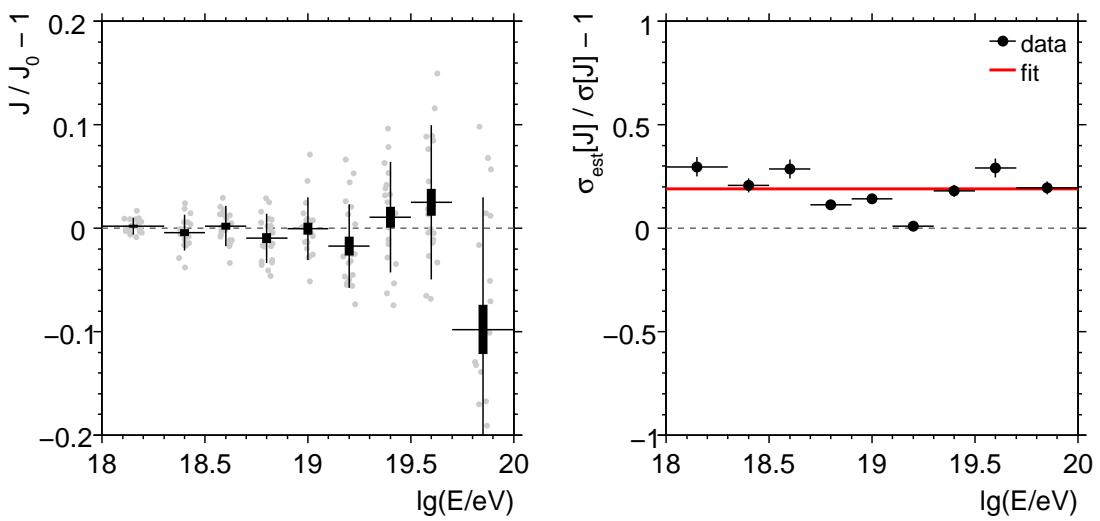


Figure 8.6: Shown is a statistical analysis of the bias and the reported statistical uncertainty of the **RUN**-unfolding in case of the parameter choice  $m_0 = 9$  and  $n_{\text{knot}} = 21$ . Left: The profile (black) shows the average bias of the unfolded flux  $J$  with respect to the true flux  $J_0$ . Thin vertical bars show the size of the statistical uncertainty  $\sigma[J]$  of the unfolded flux in each bin. Thick vertical bars indicate the statistical uncertainty of the average bias. Small gray points represent the individual bias in the 20 Monte-Carlo experiments. Right: The points represent the relative difference of the statistical uncertainty  $\sigma_{\text{est}}[J]$  calculated by **RUN** and the root of the variance  $\sigma^2[J]$  obtained from many Monte-Carlo experiments.

### 8.2.3 Application to data

The RUN-unfolding with optimal settings is now applied to 45063 SD events in the zenith angle range  $60^\circ < \theta < 82^\circ$ , which were collected between 2004/01/01 and 2009/01/01. This high quality event sample is the result of selections, which are summarised in Chapter 7. The energy  $E_{\text{SD}}$  of the SD events is calibrated with the fluorescence detector (FD), as described in Chapter 7.

The Monte-Carlo input for the detector kernel  $K(E_{\text{SD}}, E)$  used in the unfolding is generated as described above. The number of accepted events in the Monte-Carlo is 20 times larger than the number of real events. There are several systematic uncertainties to consider for the unfolded cosmic ray flux  $\{J_i\}$ , which are summarised in Fig. 8.7 and Fig. 8.8.

The FD energy scale is the largest systematic uncertainty. The FD energy has a global systematic uncertainty  $\sigma_{\text{sys}}^{\text{FD}}[E] = 22\%$ , as discussed in Chapter 6. The SD energy  $E_{\text{SD}}$  is calibrated to the FD energy  $E_{\text{FD}}$  and thus fully inherits this uncertainty. In order to use standard error propagation to derive the systematic uncertainty on the flux  $J(E)$  from this energy uncertainty, the unfolded flux  $\{J_i\}$  is approximately interpolated with the following approach

$$J(E) \simeq \exp((1-p) \ln J_i + p \ln J_{i+1}) \quad \text{for } \ln E_i \leq \ln E < \ln E_{i+1} \quad (8.2.44)$$

with  $p = (\ln E - \ln E_i)/(\ln E_{i+1} - \ln E_i)$ .

The corresponding systematic uncertainty  $\sigma_{\text{sys}}^E(J)$  of the flux is  $\sigma_{\text{sys}}^E[J] \simeq (dJ/dE) \sigma_{\text{sys}}^{\text{FD}}[E]$  and reaches 100 % in the highest energy bin  $10^{19.7} \text{ eV} < E < 10^{20} \text{ eV}$ .

Another global systematic uncertainty derives from the SD exposure. The exposure calculation was described in the first section of this chapter and has an estimated systematic uncertainty of 3 %. It affects all data points  $\{J_i\}$  in the same way and contributes an uncertainty  $\sigma_{\text{sys}}^{\text{exp}}[J] = 3\%$  to the flux.

The remaining systematic uncertainties concern the detector kernel  $K(E_{\text{SD}}, E)$ . The statistical nature of the detector kernel  $K(E_{\text{SD}}, E)$  itself introduces an uncertainty to the unfolding. The uncertainty is evaluated by repeating the unfolding 100 times with 100 independently generated Monte-Carlo inputs. The final unfolded flux  $\{J_i\}$  is actually the average of these 100 unfoldings and the variance is regarded as a systematic uncertainty. Apart from this intrinsic uncertainty of the kernel  $K(E_{\text{SD}}, E)$ , there are systematic uncertainties to consider:

- **SD energy calibration.** The energy calibration constants  $E_{\text{cal}}$  and  $\gamma$  have statistical and systematic uncertainties which generate an additional systematic uncertainty in the SD energy  $E_{\text{SD}}$  of up to 8 %, as shown in Fig. 8.7.
- **SD resolution.** The SD energy resolution also shown in Fig. 8.7 is limited partly by shower-to-shower fluctuations of the energy estimator  $R_\mu^{\text{SD}}$  and by the sampling fluctuations of the detector. The systematic uncertainty of the both resolution combined is estimated to be 5 % (absolute change, not relative change).
- **SD reconstruction probability.** The model of the SD reconstruction probability  $P_{\text{SD}}$  was fitted to the data in Chapter 7. The statistical and systematic uncertainties in the threshold parameters lead to a systematic uncertainty in the flux  $J$  below the point of full efficiency at  $10^{18.7} \text{ eV}$ .

Details about these uncertainties can be found in Chapter 7. The uncertainty of each one of these models is propagated into the flux  $J$  with a Monte-Carlo approach. The model is varied randomly according to its statistical and systematic uncertainties. Then, a new detector kernel  $\tilde{K}(E_{\text{SD}}, E)$  is generated with the varied model. The unfolding is done with the varied detector kernel  $\tilde{K}(E_{\text{SD}}, E)$ . This is repeated 100 times. The propagated uncertainty of the model is the square root of the difference between the variance of the flux  $J$  with the varied kernel  $\tilde{K}(E_{\text{SD}}, E)$  and the variance of the flux  $J$  with the unmodified kernel  $K(E_{\text{SD}}, E)$  which was called intrinsic uncertainty above.

Finally, the kernel related uncertainties are added quadratically to form the kernel uncertainty  $\sigma_{\text{sys}}^K[J]$ , which is shown in Fig. 8.8b). The kernel uncertainty  $\sigma_{\text{sys}}^K[J]$  has bin-to-bin correlations, which are neglected in the following. In all but the highest energy bin, the kernel uncertainty  $\sigma_{\text{sys}}^K[J]$  is at the level of 10 %. In the last bin, it is at the level of 50 %.

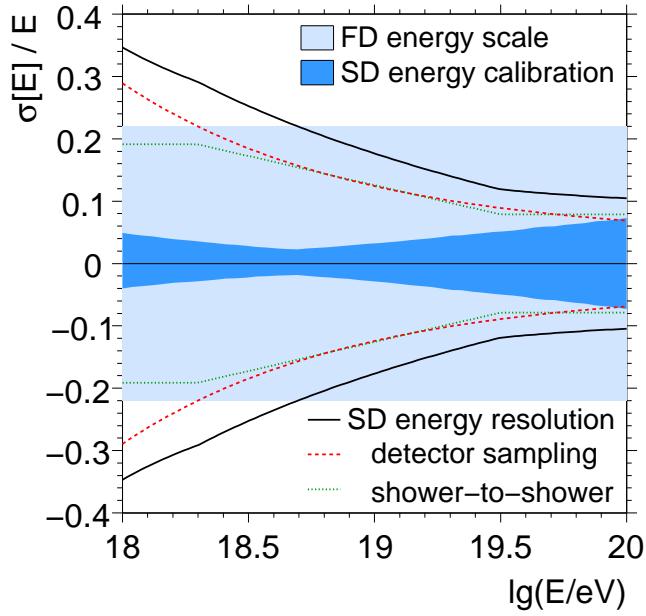


Figure 8.7: The plot compares the systematic uncertainty of the reconstructed cosmic ray energy with the energy resolution (see text). Systematic uncertainties are correlated and represented by the filled areas. The SD energy resolution and its components are shown by lines, which represent the size of the uncorrelated fluctuations of the energy from event-to-event.

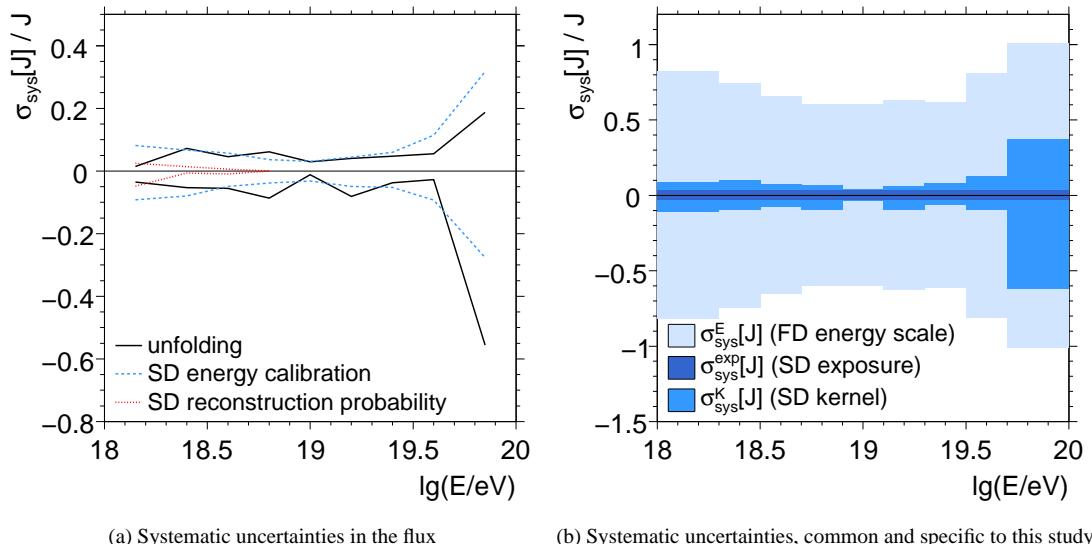


Figure 8.8: The plots show the systematic uncertainties in the derived flux as a function of the cosmic ray energy. In a), systematic uncertainties which are specific to this study are shown in detail. These uncertainties are combined in b), and compared to the common uncertainties of the SD exposure and the FD energy scale (see text).

Fig. 8.9 presents the unfolded flux  $J(E)$  its various uncertainties. Tables of the unfolded flux  $J(E)$  and its full covariance matrix are provided in Appendix G. Correlations between adjacent bins of up to  $\pm 40\%$  are found, which are negative below  $10^{18.9}$  eV and positive above. The total uncertainty of the flux is strongly dominated by the systematic uncertainties.

Two consistency checks are applied to the flux  $J(E)$ , which are also shown in Appendix G. They test for a dependency on the atmospheric profile during summer and winter months and for a possible time evolution of the data quality during the five years of data acquisition. No dependency or inconsistency is found.

The spectrum is very steep and the interesting details only show up, if the flux  $J(E)$  is normalised to an arbitrary reference flux  $J_{\text{norm}}(E)$ . The following reference flux is used in this study

$$J_{\text{norm}}(E) = A \left( \frac{E}{10^{19} \text{ eV}} \right)^{-2.8}, \quad (8.2.45)$$

whereas  $A$  is fitted to the unfolded flux  $J$  in the energy range  $10^{18.6}$  eV  $< E < 10^{19.7}$  eV, yielding

$$A = 4.029 \times 10^{-20} \text{ km}^{-2} \text{ yr}^{-1} \text{ sr}^{-1} \text{ eV}^{-1}. \quad (8.2.46)$$

The normalised result  $J/J_{\text{norm}} - 1$  is shown in Fig. 8.10.

### Effect of unfolding

Fig. 8.9 and Fig. 8.10 allow to compare the raw flux estimate which is not corrected for the SD resolution and detection efficiency effects with the unfolded flux  $J(E)$ .

The largest difference is observed at  $E < 10^{19}$  eV, where the raw flux estimate is by factor of two smaller due to the reduced SD detection efficiency  $P_{\text{SD}}$ . In the energy range  $10^{18.5}$  eV  $< E < 10^{19}$  eV, the raw flux is about 10 % larger due to bin-to-bin migration effects.

At  $E > 10^{19}$  eV, the unfolding has only a small impact on the result. This is because the random fluctuations of the SD energy  $E_{\text{SD}}$  decreases as the energy increases and eventually become smaller than the bin size. As a consequence, the bin-to-bin migration effects become negligible.

### Flux suppression

The unfolded flux  $J(E)$  becomes steeper below  $10^{18.5}$  eV and above  $10^{19.7}$  eV. Both features are in qualitative agreement with other analyses [17, 30, 188, 213–215]. The feature around  $10^{18.5}$  eV is called the ankle of the cosmic ray flux  $J(E)$ . The suppression above  $10^{19.7}$  eV is of particular interest, since it may be generated by the GZK-effect described in Chapter 3.

The GZK-effect predicts an attenuation of cosmic rays over the scale of a few tens of Mpc starting between  $10^{19.5}$  eV and  $10^{19.7}$  eV. A Mpc is the typical distance between galaxies. The cosmic rays at higher energies interact frequently with photons of the cosmic microwave background. The cosmic rays are either destroyed by photodisintegration if they are heavy nuclei or loose energy by pion production if they are protons. The observed onset of the suppression at about  $10^{19.7}$  eV fits well to the expected onset for cosmic protons, as shown in Fig. 3.3a) of Chapter 3.

The significance of the suppression above  $10^{19.7}$  eV is derived from the unfolded flux  $J(E)$  under the assumption that the flux is a simple power law between  $10^{18.5}$  eV and  $10^{19.7}$  eV

$$J(E) = A \left( \frac{E}{10^{19} \text{ eV}} \right)^{\alpha}, \quad (8.2.47)$$

analogue to Eq. (8.2.45). The power law is fitted to data points in the range  $10^{18.5}$  eV  $< E < 10^{19.7}$  eV. The prediction  $J_{\text{pred}}$  of this fit is compared with the observed flux  $J_{\text{obs}}$  in the highest energy bin  $10^{19.7}$  eV  $< E < 10^{20}$  eV. Out of the three systematic uncertainties, only the kernel uncertainty  $\sigma_{\text{sys}}^K[J]$  needs to be regarded. The other systematic uncertainties affect all data points in the same way so that the influence cancels in this analysis.

The result of the comparison is shown in Table 8.2. The flux suppression is significant at the level of three standard deviations.

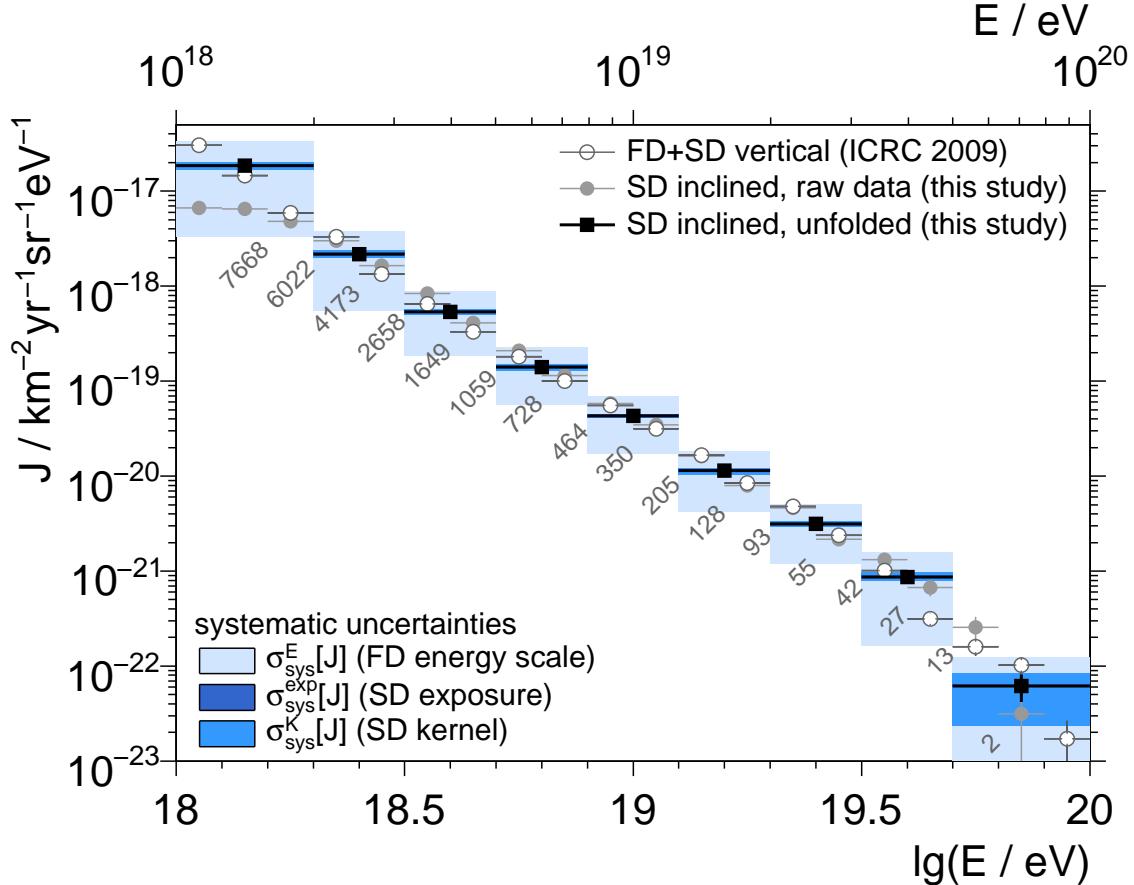


Figure 8.9: Black squares show the cosmic ray flux  $J(E)$  derived with the unfolding method from very inclined air showers ( $60^\circ < \theta < 82^\circ$ ). Gray circles represent the “raw” flux estimate, which is not corrected by the unfolding technique. White circles show the unfolded flux  $J_{\text{vert}}(E)$  from an independent analysis of FD and SD data which is based on vertical showers ( $0^\circ < \theta < 60^\circ$ ) [30]. The colored boxes represent different systematic uncertainties of the unfolded flux  $J(E)$ . The systematic uncertainties  $\sigma_{\text{sys}}^E[J]$  and  $\sigma_{\text{sys}}^{\text{exp}}[J]$  of the FD energy scale and the SD exposure apply to both  $J(E)$  and  $J_{\text{vert}}(E)$  in the same way. The systematic uncertainty  $\sigma_{\text{sys}}^K$  is specific to  $J(E)$ . The shown statistical uncertainty of  $J(E)$  uses the approximation  $\sigma_i[E] = 1.2 \times \sqrt{V_{ii}}$ , since the bin-to-bin correlations of  $J(E)$  cannot be shown. The approximation is derived from an analysis of Monte-Carlo experiments described earlier in the text.

Table 8.2: The results of the flux suppression analysis are shown, as described in the text. The constants  $A$  and  $\alpha$  of the fit of Eq.(8.2.47) are given together with their statistical uncertainties and correlation, with  $A_0 = 4.029 \times 10^{-20} \text{ km}^{-2} \text{yr}^{-1} \text{sr}^{-1} \text{eV}^{-1}$ . The values  $J_{\text{pred}}$  and  $J_{\text{obs}}$  denote the predicted and the observed flux in the energy bin  $10^{19.7} \text{ eV} < E < 10^{20} \text{ eV}$ . The last column shows the significance of the difference  $J_{\text{pred}} - J_{\text{obs}}$  in standard deviations. The fluxes are normalised to an arbitrary constant  $J_c = 10^{-34} \text{ m}^{-2} \text{s}^{-1} \text{sr}^{-1} \text{eV}^{-1}$ .

	$A/A_0$	$\alpha$	corr.	$J_{\text{pred}}/J_c$	$J_{\text{obs}}/J_c$	sigma
stat.	$1.00 \pm 0.02$	$-2.78 \pm 0.03$	0.520	$0.0578 \pm 0.0045$	$0.0196 \pm 0.0063$	5.0
stat.+sys.	$1.00 \pm 0.04$	$-2.78 \pm 0.06$	0.159	$0.0585 \pm 0.0083$	$0.0196^{+0.0096}_{-0.0137}$	3.1

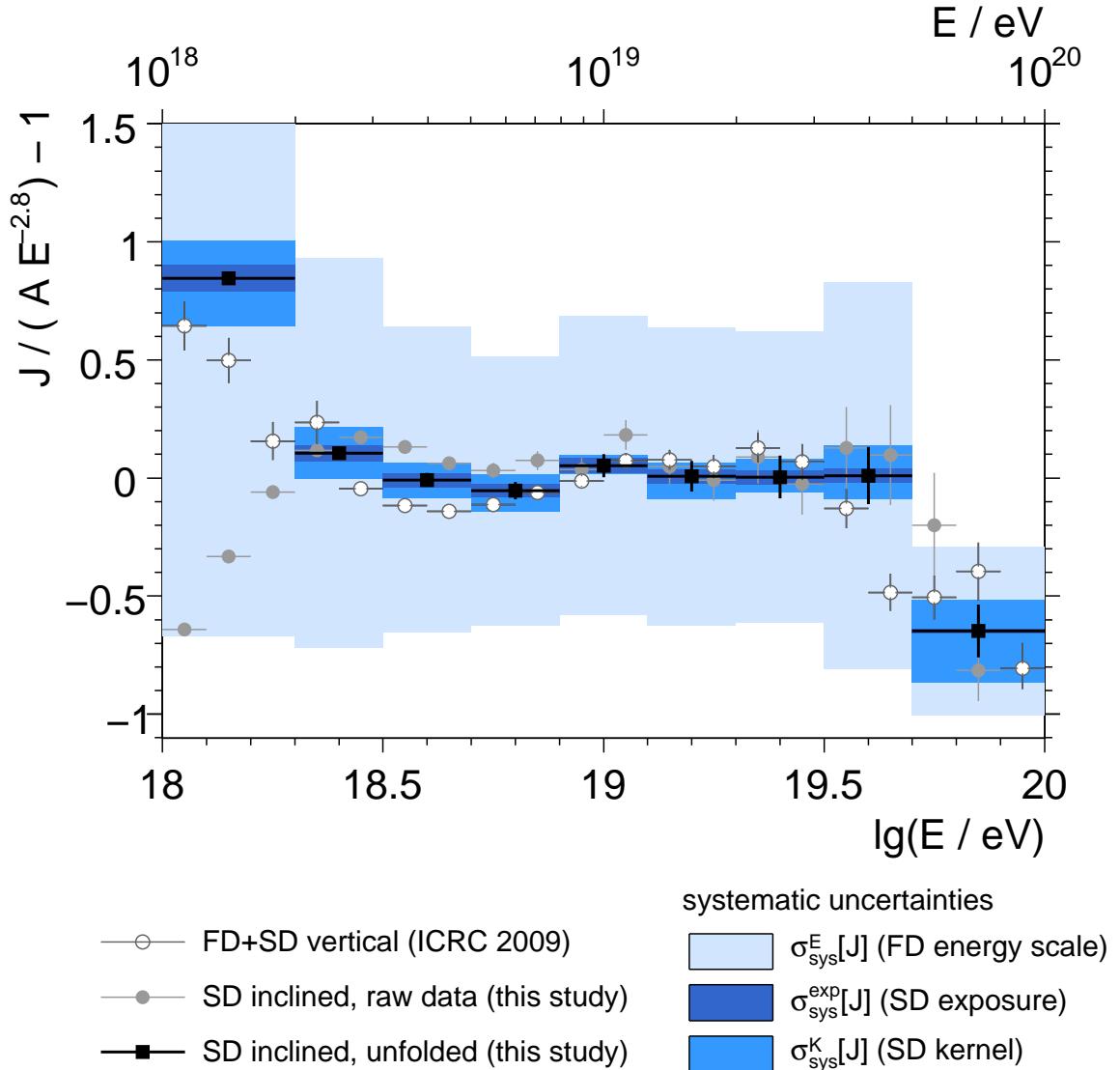


Figure 8.10: Shown are the cosmic ray fluxes from Fig. 8.9 normalised to an arbitrary reference flux  $J_{\text{norm}}(E) = A(E/10^{19} \text{ eV})^{-2.8}$ . The constant  $A = 4.029 \times 10^{-20} \text{ km}^{-2} \text{ yr}^{-1} \text{ sr}^{-1} \text{ eV}^{-1}$  is obtained from a fit of  $J_{\text{norm}}(E)$  to the unfolded flux  $J(E)$  in the energy range  $10^{18.6} \text{ eV} < E < 10^{19.7} \text{ eV}$ .

### Comparison with an independent analysis

The unfolded cosmic ray flux  $J(E)$  is derived from very inclined air showers ( $60^\circ < \theta < 82^\circ$ ) measured with the Pierre Auger Observatory. Also shown in Fig. 8.9 and Fig. 8.10 is the unfolded flux  $J_{\text{vert}}(E)$  derived from vertical showers ( $0^\circ < \theta < 60^\circ$ ) measured with the Pierre Auger Observatory [30]. The latter is the latest official result published by the Pierre Auger Observatory and shall be called *vertical flux*.

Both analyses are based on independent data because of the disjunct zenith angle ranges. They use different reconstruction and energy calibration methods and therefore have different systematics. The flux  $J_{\text{vert}}(E)$  is derived from SD and FD events. The data points between  $10^{18}$  eV and  $10^{18.4}$  eV are derived from an FD measurement of the cosmic ray flux. The data points above  $10^{18.4}$  eV are FD calibrated SD events like in the case of  $J(E)$ .

From the statistical point of view, both results are estimates of the same true flux  $J_0(E)$  and therefore expected to agree. The systematic uncertainty  $\sigma_{\text{sys}}^{\text{E}}[J] = 22\%$  of the FD energy scale and the systematic uncertainty  $\sigma_{\text{sys}}^{\text{exp}}[J] = 3\%$  of the SD exposure affect both results in the same way and therefore cancel in a direct comparison.

The flux estimates  $J(E)$  and  $J_{\text{vert}}(E)$  are indeed very similar. The agreement is particularly good in the range  $10^{18.7}$  eV  $< E < 10^{19.5}$  eV. The ankle and the flux suppression are observed in both estimates.

Some deviations are nevertheless noticeable in Fig. 8.10. The vertical flux  $J_{\text{vert}}(E)$  appears slightly shifted and tilted compared to  $J(E)$ . The flux suppression appears to set in at about  $10^{19.7}$  eV in  $J(E)$  but already at about  $10^{19.5}$  eV in  $J_{\text{vert}}(E)$ .

To quantify this deviation, the vertical flux  $J_{\text{vert}}(E)$  is re-binned to match the binning used for  $J(E)$ . Since  $J(E)$  and  $J_{\text{vert}}$  are samples of the same true flux  $J_0(E)$ , the difference

$$\Delta J_i = (J_{\text{vert},i} - J_{0,i}) - (J_i - J_{0,i}) = J_{\text{vert},i} - J_i \quad (8.2.48)$$

in bin  $i$  should have an expectation  $\langle \Delta J_i \rangle = 0$  and a variance

$$\sigma_i^2[\Delta J] = \sigma_i^2[J_{\text{vert}}] + \sigma_i^2[J]. \quad (8.2.49)$$

If this hypothesis was true and if the uncertainties  $\sigma_i^2[\Delta J]$  were Gaussian, the following sum over all  $N$  bins

$$\chi_{\text{obs}}^2 = \sum_i^N \left( \frac{\Delta J_i}{\sigma_i[\Delta J]} \right)^2 \quad (8.2.50)$$

would follow a  $\chi^2$ -distribution with  $N$  degrees of freedom and the hypothesis could be rejected at a well defined confidence level  $P(\chi^2 > \chi_{\text{obs}}^2)$ .

In reality, there are bin-to-bin correlations in the unfolded fluxes  $J_{\text{vert}}(E)$  and  $J(E)$  and part of the uncertainty of  $J_B$  is systematic, so that the calculated confidence level loses its exact meaning and can only be regarded as a measure of the agreement or disagreement.

Fig. 8.11 shows the deviation  $\Delta J/\sigma[\Delta J]$  of  $J_{\text{vert}}(E)$  and  $J(E)$ . The points with the largest deviations are found at the onsets of the flux suppression and the ankle.

If only statistical uncertainties are regarded in the comparison, the fluxes are clearly incompatible:

$$\frac{\chi_{\text{obs}}^2}{N} \approx 6.0 \quad \rightarrow \quad P(\chi^2 \geq \chi_{\text{obs}}^2) \approx 2 \times 10^{-8}. \quad (8.2.51)$$

If systematic uncertainties are included in the analysis, the results roughly agree:

$$\frac{\chi_{\text{obs}}^2}{N} \approx 1.1 \quad \rightarrow \quad P(\chi^2 \geq \chi_{\text{obs}}^2) \approx 0.39. \quad (8.2.52)$$

The analysis indicates some tension between the vertical flux  $J_{\text{vert}}(E)$  and the result of this study  $J(E)$ , which has to be pushed to the edge of its systematic uncertainties to get an agreement. This should be investigated further in the future.

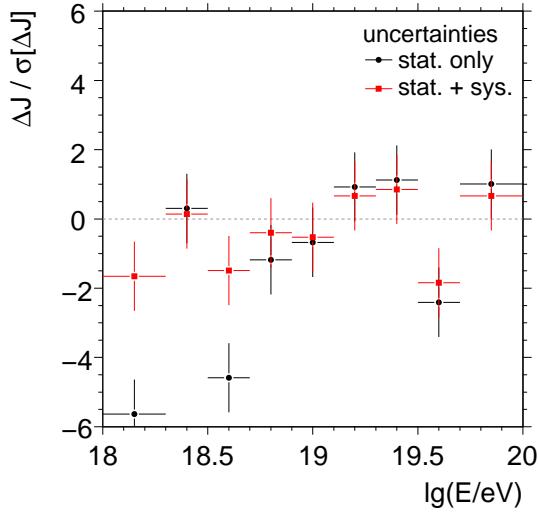


Figure 8.11: The plot shows the deviation of the flux obtained in this study from the latest reference result [30] in standard deviations. The reference flux is the same as in Fig. 8.9 and Fig. 8.10.

### 8.3 Summary

The cosmic ray flux  $J(E)$  in the energy range  $10^{18} \text{ eV} < E < 10^{20} \text{ eV}$  was derived in this chapter from 45063 selected SD events in the zenith angle range  $60^\circ < \theta < 82^\circ$ , collected between 2004/01/01 and 2009/01/01. The energy  $E_{\text{SD}}$  of these SD events was calibrated against the FD energy  $E_{\text{FD}}$ .

Detector effects were removed from the cosmic ray flux  $J(E)$  with an unfolding method, the **RUN**-algorithm. The unfolding depends on a detailed statistical models of the SD resolution and efficiency, which were taken from Chapter 8. The output of the **RUN**-algorithm was studied in detail with Monte-Carlo experiments and was found to be suitable for the unfolding of the cosmic ray flux  $J(E)$ .

The uncertainty of the cosmic ray flux is dominated by the systematic uncertainty  $\sigma_{\text{sys}}[E] = 22\%$  of the FD energy scale. It generates a corresponding systematic uncertainty  $\sigma_{\text{sys}}^E[J]$  in the cosmic ray flux  $J(E)$  between 60% and a factor of two. Uncertainties in the energy calibration and the unfolding add 10% uncertainty to the flux  $J(E)$  in most bins except the highest energy bin  $10^{19.7} \text{ eV} < E < 10^{20} \text{ eV}$ , where the uncertainty is about 50%.

A flux suppression above  $10^{19.7} \text{ eV}$  is observed despite these large uncertainties with a significance of three standard deviations. The onset of this suppression qualitatively agrees with the onset predicted by the GZK-effect for intergalactic protons. This seems to be in contradiction with the indications found in Chapter 7 that the cosmic rays above  $10^{19.7} \text{ eV}$  could be mostly heavy nuclei.

The flux  $J(E)$  derived from very inclined air showers is compared with an independent estimate  $J_{\text{vert}}(E)$  derived from vertical showers in the zenith angle range  $0^\circ < \theta < 60^\circ$ . Both are independent estimates of the same true cosmic ray flux and agree within one standard deviation.

Comparisons between the vertical flux  $J_{\text{vert}}(E)$  and the flux  $J(E)$  obtained from very inclined air showers will improve the understanding of the systematic uncertainties in both flux estimates. Eventually, the estimates will be brought into statistical agreement which will then allow all cosmic ray analyses to use the full zenith angle range of recorded SD events.

This work described a complete procedure to derive the flux  $J$  of ultra-high energy cosmic rays in the energy range  $10^{18} \text{ eV} < E < 10^{20} \text{ eV}$  from data of very inclined air showers, which were recorded with the Pierre Auger Observatory between 2004/01/01 and 2009/01/01. The analysis used the high event statistics gathered by the surface detector (SD) of the observatory. The SD measurement of the cosmic ray energy  $E$  was based on a calibration of the SD signal described by the energy estimator  $R_\mu$  with the fluorescence detector (FD), which is able to perform a calorimetric measurement of  $E$ .

All steps in this work required input about the average properties of very inclined air showers on the ground and their fluctuations, as well as the SD signal response to such showers. The input was obtained by simulating and studying 6480 very inclined air showers and the same number of SD events in the zenith angle range  $60^\circ < \theta < 88^\circ$ . The simulations were computed with the program CORSIKA [49, 89], the high-energy hadronic interaction models QGSJet-II [115–120] and EPOS [130, 131], and the low energy hadronic interaction model FLUKA [102, 103].

Particularly important for the SD event reconstruction was the lateral particle density  $n_\mu$  and the total number  $N_\mu$  of muons on the ground:

- The muon density  $n_\mu$  was parameterized. The parameterization is not based on theoretical input, but on a fit of a linear parameterization to the simulation output. The  $n_\mu$ -model agrees with the simulation at the level of 10 %.
- A systematic uncertainty of about 30 % was found in the simulated number of muons  $N_\mu$ , which is of the same order as the difference found between proton and iron induced showers.
- Shower-to-shower fluctuations of  $N_\mu$  of 3 to 4 % were observed in case of iron showers and 14 % to 21 % in case of proton showers. The ranges represent the systematic uncertainties. The fluctuations of  $N_\mu$  are approximately Gaussian with a tail towards small values of  $N_\mu$ .

The SD signal response  $f_\mu(S_\mu)$  to muons and the contribution  $\langle\epsilon\rangle$  of electromagnetic particles to the SD signal were studied with simulated SD events, which were further inputs for the SD event reconstruction:

- A model  $f_\mu$  of the signal response of SD station to muons was derived, which included signal fluctuations generated by muons and electromagnetic particles in the shower front.
- The signal contribution  $\langle\epsilon\rangle$  of electromagnetic particles to the SD signal was compared with the model from ref. [16]. Deviations between 10 % and 20 % were found. The deviations were absorbed into the  $f_\mu$ -model so that the combination of both models in this study was almost bias-free.

A new software was written within the Auger Offline-frame to perform the SD event reconstruction. The new software combined the features of two existing programs **efit** [17, 161–165] and **SdHorRec** [14] and added minor improvements. Its purpose is to reconstruct an energy estimator  $R_\mu \propto N_\mu \propto E^{1/\gamma}$  and the direction  $(\theta, \phi)$  of air showers recorded with the SD. The bias and resolution of these reconstructed variables was analyzed with simulated SD events.

- The resolution of  $R_\mu$  and  $(\theta, \phi)$  depends on the energy  $E$  and zenith angle  $\theta$  of the shower. For a  $10^{19} \text{ eV}$  shower with an inclination  $\theta = 70^\circ$ , the resolution of  $R_\mu$  is 16 % and the angular resolution  $0.4^\circ$ .
- The systematic uncertainty of  $R_\mu$  in the zenith angle range  $60^\circ < \theta < 82^\circ$  is at the level of 3 %. The systematic uncertainty of the zenith angle  $\theta$  is  $0.1^\circ$ .

The reconstruction software was designed to make the models of the muon density  $n_\mu$  and the SD signal response  $f_\mu$  easily replaceable. This feature was used to compare the models derived in this study with others from ref. [10, 11, 15, 17].

- An overall bias of 20 % in the energy estimator  $R_\mu$  is generated by the model of the electromagnetic signal component  $\langle\epsilon\rangle$ , if its bias is not implicitly or explicitly corrected.
- If overall biases are corrected, the performance of all models in the zenith angle range  $60^\circ < \theta < 82^\circ$  is comparable. The models derived in this study achieve a slightly better resolution of  $R_\mu$  in simulated events and fit better to real events in most cases.

The energy calibration was based on a fit of the power law  $E = E_{\text{cal}} \times R_\mu^\gamma$  to high quality events recorded simultaneously in the SD and the FD. Selection cuts accepted 411 events for this study. A new method was introduced to perform the fit which used a complete probability density function  $f_{\text{tot}}$  of the distribution of these events for the first time.

- The effect of the limited SD efficiency between  $10^{18}$  eV and  $10^{18.5}$  eV is included in the  $f_{\text{tot}}$ -model, which allowed to use events in this energy range to constrain the fit. Before, it had been necessary to reject such events. The new method improved the event statistics by a factor of three.
- The model  $f_{\text{tot}}$  was used to fit a model of the SD detection efficiency to the data.
- The model  $f_{\text{tot}}$  was further used to fit the shower-to-shower fluctuations of  $R_\mu$  to the data. The fluctuations of  $R_\mu$  are sensitive to the cosmic ray composition. They decrease from about 20 % at  $10^{18}$  eV to less than 10 % at  $10^{20}$  eV. This observation is compatible with a scenario where the composition of cosmic ray masses is getting heavier as the energy  $E$  increases, but not significant enough to draw definite conclusions.
- The calibration constant  $\gamma$  is sensitive to a change in the composition of cosmic ray masses. The fitted value is compatible with the scenario mentioned above, but the observation is again not very significant.
- A large excess of muons in real events is found compared to air shower simulations at the same energy. Real events show 20 % more muons at  $10^{19}$  eV than the closest simulated scenario.
- The FD measurement of the cosmic ray energy  $E$  has a systematic uncertainty of 22 %. The energy calibration transfers this uncertainty to the SD. Statistical and systematic uncertainties in the calibration fit add another 4 % to 7 %.

The cosmic ray flux  $J(E)$  was reconstructed from 45063 selected SD events in the zenith angle range  $60^\circ < \theta < 82^\circ$ . The apparent cosmic ray flux is distorted by the limited SD resolution and detection efficiency, especially between  $10^{18}$  eV and  $10^{19}$  eV. An unfolding method was used to correct these effects. The unfolding used the models of the SD resolution and detection efficiency which were derived for the energy calibration. The fitted shower-to-shower fluctuations of  $R_\mu$  were another input.

The uncertainty of the flux  $J(E)$  was dominated by the systematic uncertainty of the FD energy of 22 %. The corresponding systematic uncertainty of the flux varied between 60 % and 100 %. Uncertainties in the unfolding and the energy calibration added about 10 % uncertainty to the flux  $J(E)$  below  $10^{19.7}$  eV and about 50 % in the range  $10^{19.7} \text{ eV} < E < 10^{20}$  eV.

A flux suppression above  $10^{19.7}$  eV was observed with a significance of three standard deviations. The onset of this suppression agrees qualitatively with the predicted GZK-effect [31, 32].

The flux  $J(E)$  derived from very inclined air showers in the zenith angle range  $60^\circ < \theta < 82^\circ$  was compared with an independent estimate  $J_{\text{vert}}(E)$  derived from vertical showers in the zenith angle range  $0^\circ < \theta < 60^\circ$  [30]. Both are independent estimates of the true cosmic ray flux and agree within one standard deviation.

Comparisons between the vertical flux  $J_{\text{vert}}(E)$  and the flux  $J(E)$  obtained from very inclined air showers will improve the understanding of the systematic uncertainties in both flux estimates. Eventually, the estimates will be brought into statistical agreement which will then allow all cosmic ray analyses to seamlessly use the full zenith angle range of recorded SD events.

# Chapter 9

## Acknowledgements

A work like this is not possible without the feedback and support of many people. First of all I would like to thank my supervisor Prof. Dr. Thomas Hebbeker who has been a great inspiration to me. He kept my mind open and sharp by asking the right questions. I value his profound advice and that he allowed me to follow my own ideas.

I am grateful to my co-supervisor Prof. Dr. Martin Erdmann who offered me his valuable insight into programming, hadronic interactions, the unfolding of experimental distributions, and many other topics. I also thank Dr. Matthias Leuthold for his guidance during a major part of this thesis.

For valuable discussions and support I would like to thank my colleagues at the RWTH Aachen University, the Forschungszentrum Karlsruhe, and the Horizontal Air Shower task group within the Pierre Auger Collaboration. I am especially indebted to Prof. Dr. Pierre Billoir, Dr. Oscar Blanch Bigas, Dr. Olivier Deligny, Dr. Ralph Engel, Stefan Fliescher, Dr. Javier Gonzales, Marius Grigat, Dr. Ioana Mariş, Dr. Tanguy Pierog, Dr. Markus Roth, Dr. Victor Olmos-Gilbaja, Nils Scharf, Peter Schiffer, Talianna Schmidt, Stephan Schulte, Dr. Fabian Schüssler, Dr. Ralph Ulrich, Dr. Michael Unger, Dr. Inés Valiño, Dr. Victor Vázquez, Dr. Darko Veberič, Tobias Winchen, and Prof. Dr. Enrique Zas. For reading parts of the manuscript, I am thankful to Anna Nelles, Nils Scharf, and Tobias Winchen.

This work would have been much more difficult without the T4 selected surface detector events provided by Dr. Olivier Deligny on the one side, and the reconstructed fluorescence detector events and ADST analysis tools provided by the Auger group at the Forschungszentrum Karlsruhe on the other.

For support and helpful discussions during the planning phase of the air shower simulations I am grateful to Dr. Carla Bleve. Dr. Tanguy Pierog and Dr. Dieter Heck helped with questions about CORSIKA and the hadronic interaction models. Jean-Noel Albert provided his continuous support concerning the usage of the computing facilities in Lyon.

The air shower simulations performed in this study were supported in part by a grant of computing time from the Lyon supercomputing center CC-IN2P3 and in part by a grant of the Ohio Supercomputer Center. I am grateful to Jim Beatty and the OSC team for making the use of the latter grant possible.

Talianna Schmidt contributed to the energy calibration software and the SD event reconstruction software. The energy calibration software is partly based on code written by Dr. Ioana Mariş. The SD event reconstruction is partly based on code written by Dr. Javier Gonzales, Dr. Victor Olmos-Gilbaja, Dr. Markus Roth, Talianna Schmidt, and Dr. Inés Valiño.

This work was funded in parts by the German Federal Ministry for Education and Research BMBF and the German Research Foundation DFG.

Without the love and support of my girlfriend and my family, I would never have come this far.



Appendix **A**

# Analytical and numerical calculations

## A.1 Point of shower maximum and particle interaction lengths

Input for the calculations in Fig. 3.7 in Chapter 3 are the following.

- A parameterisation of the slant depth  $X(h)$  and the density  $\rho(h)$  in the spring atmosphere [42, 43] in Malargüe, Argentina. The altitude of the experiment is taken as 1400 m and taken into account.
- The hadronic attenuation lengths  $\Lambda_\pi \approx 120 \text{ g cm}^{-2}$  and  $\Lambda_K \approx 140 \text{ g cm}^{-2}$  for pions and kaons, taken from [21]. The attenuation lengths are approximately valid in the energy range of 10 GeV to 1000 GeV. The electromagnetic radiation length  $X_0 \approx 36.6 \text{ g cm}^{-2}$  for dry air is taken from [23].
- The following parameterisation of the observed  $X_{\max}$  in Fig. 3.2:

$$X_{\max}(E)/\text{g cm}^{-2} = 725 + \begin{cases} 71 \times [\lg(E/\text{eV}) - 18.35] & \text{if } \lg(E/\text{eV}) < 18.35 \\ 40 \times [\lg(E/\text{eV}) - 18.35] & \text{if } \lg(E/\text{eV}) \geq 18.35. \end{cases} \quad (\text{A.1.1})$$

The total atmospheric depth  $X_{\text{atm}}$  for an air shower is obtained by integrating the air density  $\rho_{\text{air}}(h)$  along the path of the shower:

$$X_{\text{atm}} = \int_{d_{\text{top}}}^0 dd \rho_{\text{air}}(h(d)), \quad (\text{A.1.2})$$

where  $d(h)$  is the distance of a point on the shower axis with the altitude  $h$  and the point where the shower axis intersects with the ground, as illustrated in Fig. 2.4. The interchanged integral limits are necessary, because  $d$  is zero at the ground, while  $X$  is zero at the top of the atmosphere.

The relation of the distance  $d$  to the altitude above the ground  $h$  is geometrical

$$d(h) = [(R + h)^2 + R^2 (\cos^2 \theta - 1)]^{1/2} - R \cos \theta, \quad (\text{A.1.3})$$

with  $R$  as Earth's radius including the local altitude of Malargüe above the sea level, and  $\theta$  as the zenith angle of the shower. This simple result is obtained by assuming that the Earth is a sphere in good approximation over the involved distances.

Eq. (A.1.3) can be inverted to yield  $h(d)$ . The value  $d_{\text{top}} = s(112 \text{ km})$  is the top end of the atmospheric parameterisation of  $\rho_{\text{air}}(h)$ . This completes the integral in Eq. (A.1.2), which is shown in Fig. 3.7a) as a function of the zenith angle  $\theta$ .

The distance  $d_{\max}$  of the electromagnetic shower maximum  $X_{\max}$  is obtained by solving the equation

$$X_{\max} = \int_{d_{\text{top}}}^{d_{\max}} dd \rho_{\text{air}}(h(d)) \quad (\text{A.1.4})$$

for  $d_{\max}$ . This is done numerically. Once  $d_{\max}$  is obtained,  $h_{\max}$  can be calculated with the inverted form of Eq. (A.1.3). Fig. 3.7c) and d) show both values as a function of the zenith angle  $\theta$ .

The interaction length  $l_{\text{int}}$  (a distance in meters) in Fig. 3.7b) is calculated from the hadronic attenuation and radiation lengths (which are slant depths):

$$\{\lambda_h, \lambda_{\text{em}}\} = \int_{l_{\text{int}}}^0 dh \rho(h) \approx \rho(h) \times l_{\text{int}} \quad \Leftrightarrow \quad l_{\text{int}} = \{\lambda_h, \lambda_{\text{em}}\}/\rho(h). \quad (\text{A.1.5})$$

## A.2 Exp-normal distribution

The exp-normal distribution turns up in the context of this work. It describes the distribution of a random variable, which is the logarithm of a variable with a normal distribution.

It can be derived from the normal distribution  $f(x)$  by the change of variables  $y = \ln x$ :

$$\begin{aligned} f(x) &\propto \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ \Rightarrow g(y) &\propto \exp\left(-\frac{(e^y-\mu)^2}{2\sigma^2} + y\right). \end{aligned} \quad (\text{A.2.1})$$

This is the inverse change of variables which is required to obtain the common log-normal distribution. The parameters  $\mu$  and  $\sigma^2$  are not equal to the mean and variance of  $g(y)$ . They are just parameters of the distribution after the change of variables.

Calculating analytically the normalisation and the first moments of  $g(y)$  turns out to be extremely difficult. Closed forms can be derived by calculating the characteristic function:

$$\begin{aligned} \phi_y(t) &= \int_{-\infty}^{\infty} dy \exp(ity) \exp\left(-\frac{(e^y-m)^2}{2\sigma^2} + y\right) \\ &= \int_{-\mu}^{\infty} dz (z+\mu)^{it} \exp\left(-\frac{z^2}{2\sigma^2}\right), \end{aligned} \quad (\text{A.2.2})$$

using the substitution  $e^z = y + \mu$  in the second line. This integral was solved with the program **Mathematica** [216] under the condition  $\mu > 0$ :

$$\phi_y(t) = 2^{\frac{1}{2}i(i+t)} \sigma^{it} \left[ \sigma \Gamma\left(\frac{1}{2} + \frac{it}{2}\right) {}_1F_1\left(-\frac{it}{2}; \frac{1}{2}; -\frac{\mu^2}{2\sigma^2}\right) \right] \quad (\text{A.2.3})$$

$$+ \sqrt{2} \mu \Gamma\left(1 + \frac{it}{2}\right) {}_1F_1\left(-\frac{1}{2}i(i+t); \frac{3}{2}; -\frac{\mu^2}{2\sigma^2}\right). \quad (\text{A.2.4})$$

The solution contains the gamma function  $\Gamma(z)$  and the confluent hypergeometric function  ${}_1F_1$ , which is defined as

$${}_1F_1(a; b; z) = \frac{\Gamma(b)}{\Gamma(a)} \sum_{k=0}^{\infty} \frac{\Gamma(a+k)}{\Gamma(b+k)} \frac{z^k}{k!}. \quad (\text{A.2.5})$$

The expansion of the characteristic function in  $t$  yields the moments of  $g(y)$  (see e.g. [147]):

$$\phi_y(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \mu_k^*, \quad (\text{A.2.6})$$

whereas  $\mu_0^*$  is the normalisation of  $g(y)$  and  $\mu_k^*/\mu_0^*$  the  $k$ -th moment.

The expansion yields:

$$\mu_0^* = \sqrt{\frac{\pi}{2}}\sigma \left[ 1 + \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right) \right] \quad (\text{A.2.7})$$

$$\begin{aligned} \mu_1^* &= \mu_0^* \left[ \frac{1}{2} \ln 2 + \ln \sigma \right] \\ &\quad + \frac{\sqrt{\pi}\sigma}{2\sqrt{2}} \left[ +\psi_0(1) \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right) + \psi_0\left(\frac{1}{2}\right) - {}_1F_1^{(1,0,0)}\left(0; \frac{1}{2}; -\frac{\mu^2}{2\sigma^2}\right) \right] \\ &\quad - \frac{1}{2}\mu {}_1F_1^{(1,0,0)}\left(\frac{1}{2}; \frac{3}{2}; -\frac{\mu^2}{2\sigma^2}\right) \end{aligned} \quad (\text{A.2.8})$$

$$\begin{aligned} \mu_2^* &= \mu_0^* \left[ \frac{(\ln 2)^2}{8} + \frac{\ln 2 \ln \sigma}{2} + \frac{(\ln \sigma)^2}{2} \right] \\ &\quad + \left( \frac{\ln 2}{2\sqrt{2}} + \frac{\ln \sigma}{\sqrt{2}} \right) \left[ \frac{1}{2}\psi_0(1)\sqrt{\pi}\sigma \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right) + \frac{1}{2}\sqrt{\pi}\sigma\psi_0\left(\frac{1}{2}\right) \right. \\ &\quad \left. - \frac{1}{2}\sqrt{\pi}\sigma {}_1F_1^{(1,0,0)}\left(0; \frac{1}{2}; -\frac{\mu^2}{2\sigma^2}\right) - \frac{\mu}{\sqrt{2}} {}_1F_1^{(1,0,0)}\left(\frac{1}{2}; \frac{3}{2}; -\frac{\mu^2}{2\sigma^2}\right) \right] \\ &\quad + \frac{1}{\sqrt{2}} \left[ \frac{1}{16}\pi^{\frac{5}{2}}\sigma + \frac{1}{8}\psi_0^2(1)\sqrt{\pi}\sigma \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right) + \frac{1}{48}\pi^{\frac{5}{2}}\sigma \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right) + \frac{1}{8}\sqrt{\pi}\sigma\psi_0^2\left(\frac{1}{2}\right) \right. \\ &\quad \left. - \frac{1}{4}\sqrt{\pi}\sigma\psi_0\left(\frac{1}{2}\right) {}_1F_1^{(1,0,0)}\left(0; \frac{1}{2}; -\frac{\mu^2}{2\sigma^2}\right) - \frac{\psi_0(1)\mu}{2\sqrt{2}} {}_1F_1^{(1,0,0)}\left(\frac{1}{2}; \frac{3}{2}; -\frac{\mu^2}{2\sigma^2}\right) \right. \\ &\quad \left. + \frac{1}{8}\sqrt{\pi}\sigma {}_1F_1^{(2,0,0)}\left(0; \frac{1}{2}; -\frac{\mu^2}{2\sigma^2}\right) + \frac{\mu}{4\sqrt{2}} {}_1F_1^{(2,0,0)}\left(\frac{1}{2}; \frac{3}{2}; -\frac{\mu^2}{2\sigma^2}\right) \right] \end{aligned} \quad (\text{A.2.9})$$

whereas  $\psi_n(z) = \frac{d^{n+1}}{dz^{n+1}} \ln \Gamma(z)$  is the polygamma function,  $\operatorname{erf}(z)$  the error function, and  ${}_1F_1(a; b; z)^{(m,0,0)}$  the  $m$ -th derivative of  ${}_1F_1(a; b; z)$  with respect to  $a$ .

### A.3 Effective area of a SD station and average muon signal

To calculate the excepted number of muons in a SD station from a given ground density of muons  $n_\mu$ , a formula for the effective ground area  $A_{\text{station}}$  of the station shall be derived. This effective area is the equivalent ground area, into which a muon of inclination  $\theta_\mu$  has to fall to enter the active detector volume of the station. The area is illustrated in Fig. A.1, a geometrical analysis leads to

$$A_{\text{station}}(\theta_\mu) = A_{\text{top}} + A_{\text{side}} = \pi r^2 + 2hr|\tan \theta_\mu|, \quad (\text{A.3.1})$$

$$\approx 10.2 \text{ m} + 4.3 \text{ m}|\tan \theta_\mu| \quad (\text{A.3.2})$$

with  $r = 1.8 \text{ m}$  and  $h = 1.2 \text{ m}$  as the radius and height of the active detector volume from Chapter 4.

Another interesting related quantity is the average signal  $\langle S_\mu \rangle$  produced by an ideal muon in the active detector volume. Ideal means here, that the muon does not decay in the volume and that it has sufficient energy to maintain a Cherenkov light production efficiency of 100 % along its track in water. This is case for muons above 1.5 GeV [16].

For such an ideal muon, the average muon signal  $\langle S_\mu \rangle$  is proportional to the average track length  $\langle l_\mu \rangle$  of the muon in the active detector volume. With the definition of the unit VEM from Chapter 4 one obtains

$$\langle S_\mu \rangle(\theta_\mu) = \frac{\langle l_\mu \rangle(\theta)}{\langle l_\mu \rangle(0^\circ)} S_\mu(0^\circ) = \frac{\langle l_\mu \rangle(\theta)}{h} \text{VEM}. \quad (\text{A.3.3})$$

It is possible to derive a closed form of the distribution of  $l_\mu$  analytically [217], which is quite complex. On the other hand, it was realised by several authors (see e.g. [16]) that an analytical formula for the average track length can be obtained by a simple argument.

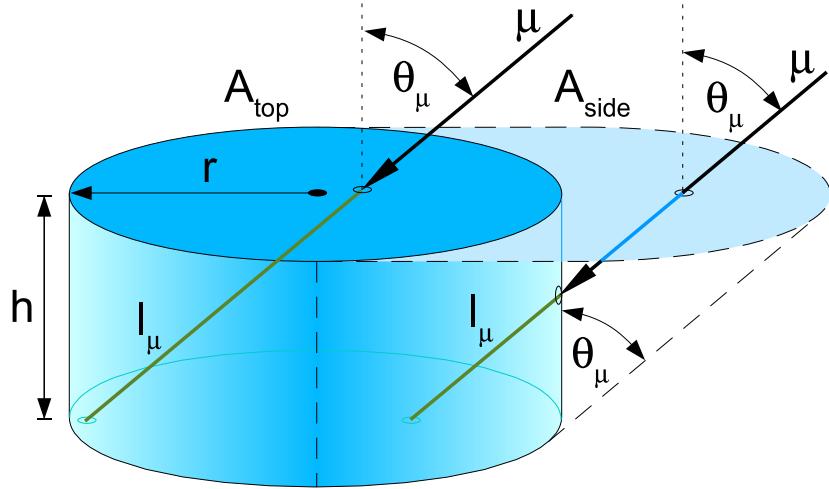


Figure A.1: The drawing illustrates the effective ground area  $A_{\text{station}} = A_{\text{top}} + A_{\text{side}}$  of a SD station and the track vector in the active detector volume  $l_\mu$  of an ideal muon (see text). Both depend on the inclination  $\theta_\mu$  of the muon.

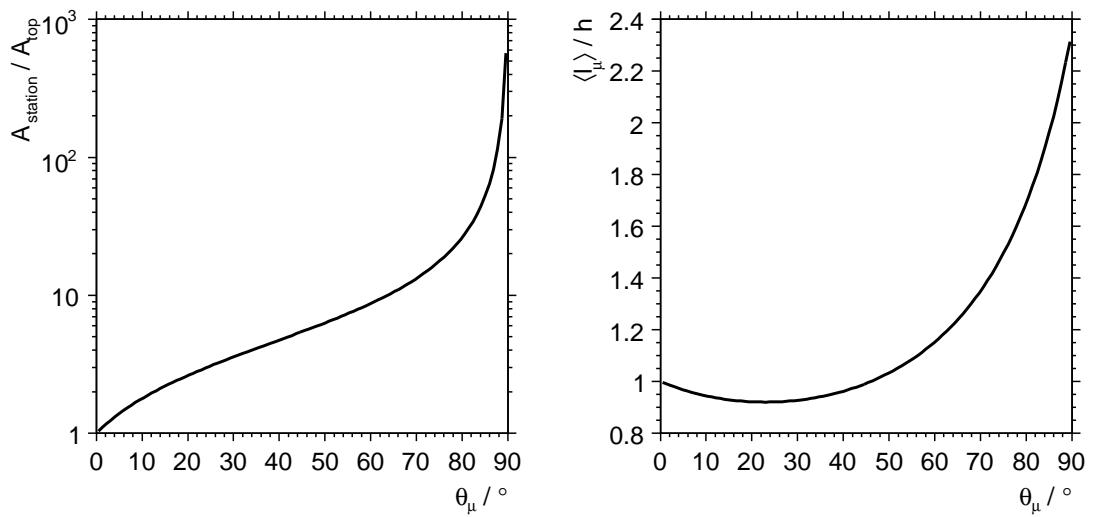


Figure A.2: The left plot shows the effective ground area  $A_{\text{station}}$  of a SD station in units of the lid area. The area diverges for  $\theta_\mu \rightarrow 90^\circ$ . The right plots shows the average track length in units of the vertical height of the detector volume  $h$ .

---

## APPENDIX A. ANALYTICAL AND NUMERICAL CALCULATIONS

---

The argument starts with a division of the effective ground area  $A_{\text{station}}$  into infinitesimal oriented area elements  $d\mathbf{A}$ . The center of each area element is a possible impact point of a muon with a corresponding track length vector  $\mathbf{l}_\mu$ . The average track length then is

$$\langle l_\mu \rangle = \frac{\int_{A_{\text{station}}} l_\mu d\mathbf{A}}{\int_{A_{\text{station}}} e_\mu d\mathbf{A}} = \frac{\int_{A_{\text{station}}} l_\mu dA_\perp}{\int_{A_{\text{station}}} \cos \theta_\mu dA} = \frac{V_{\text{station}}}{\cos \theta_\mu A_{\text{station}}}, \quad (\text{A.3.4})$$

with  $e_\mu = \mathbf{l}_\mu / l_\mu$  as the normalised muon direction,  $dA_\perp = e_\mu d\mathbf{A} = \cos \theta_\mu dA$  as the perpendicular component of the area element with respect to  $e_\mu$ , and  $V_{\text{station}}$  as the active detector volume. The last step exploits, that  $dV = dA_\perp l_\mu$  is also an infinitesimal volume element of the detector.

Inserting  $V_{\text{station}} = \pi r^2 h$  and Eq. (A.3.1) gives

$$\langle l_\mu \rangle = \frac{\pi r^2 h}{\cos \theta_\mu (\pi r^2 + 2rh \tan \theta_\mu)} = \frac{h}{\cos \theta_\mu + \frac{2h}{\pi r} \sin \theta_\mu}, \quad (\text{A.3.5})$$

and an average muon signal of

$$\langle S_\mu \rangle = \frac{1 \text{ VEM}}{\cos \theta_\mu + \frac{2h}{\pi r} \sin \theta_\mu}. \quad (\text{A.3.6})$$

Fig. A.2 summarises the numerical results.

It shall be noted, that there also exist another convention of the effective area of the station in the literature, which corresponds to a different definition of the muon density  $n_\mu$ , *e.g.* in ref. [10]. In this work, the muon density  $n_\mu$  is defined as the time integrated flux through an infinitesimal ground plane area. In the reference, it is defined as the time integrated flux through an infinitesimal area element of the shower front plane.

The effective area of the SD station in the latter case is  $A_{\text{station}} \cos \theta$ , whereas  $\theta$  is the inclination of the shower axis. For muons close to the shower axis where  $\theta_\mu \approx \theta$ , this definition has the advantage that the effective area does not diverge at  $\theta_\mu \rightarrow 90^\circ$ . But if this approximation does not hold, the other definition only makes the calculation of the effective area more complex and the definition used in this work becomes more intuitive.

## A.4 Projection of surface detector stations

In Chapter 6, it becomes necessary to project the position of a SD station located on the curved surface of the Earth into a Cartesian ground coordinate system, which is tangential to Earth's surface at the impact point of an air shower.

The apparent position of the SD station in this plane is the projection of the station position along the average trajectory of the particle hits. It is assumed that these particles are mostly muons which travel in straight lines and originate from a common point at the maximum of the shower in the atmosphere. In that case, the projection is well defined. In very inclined air showers ( $60^\circ < \theta < 90^\circ$ ), this approach works well.

Other possible projection schemes are shown in Fig. A.3. Technically, the simplest possible projection is that along the z-axis of the Cartesian ground coordinate system. Better is a projection along the direction of the shower axis. Both kinds of projections introduce a radial offset  $\Delta r$  of the SD station in the lateral shower coordinate system with respect to the correct projection. The size of this bias is calculated geometrically for different zenith angles  $\theta$ . The distance  $d_{\max}(\theta)$  between the shower impact point and the point of the shower maximum is taken from Appendix A.1.

The results are shown in Fig. A.4. The shower axis projection produces a smaller bias  $\Delta r/r$ . The bias has an impact on the shower reconstruction in which the measured signal  $S$  in a station is compared with the expected signal  $\langle S \rangle(r)$  at the apparent radial distance  $r$  of the station from the shower axis. The expected signal  $\langle S \rangle$  as a function of the radial distance  $r$  from the shower axis is roughly a power law with an index  $\alpha \approx -2$ , as shown in Chapter 3

$$\langle S \rangle \propto r^{-2}, \quad (\text{A.4.1})$$

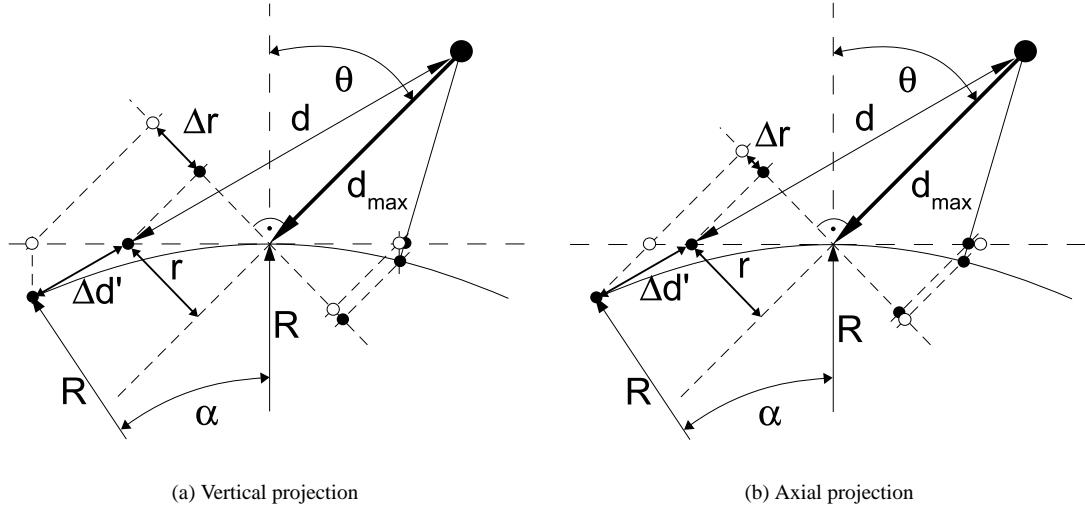


Figure A.3: The drawings show two possible projections of SD stations into a Cartesian coordinate system, which is tangential to Earth's surface at the impact point of an air shower. The shower axis is defined by the zenith angle  $\theta$  and the distance  $d$  between the impact point and the shower maximum. The real muon trajectory that leads to the SD station is expected to start at the shower maximum (thin line). The projection along the real muon trajectory (black points) has a radial distance  $r$  to the shower axis, which has an offset  $\Delta r$  to a) the vertical projection or b) the shower axis projection (both indicated with white points). The angle  $\alpha$  is the angular distance from the position of the SD station and the impact point for an observer at the center of the Earth with the radius  $R$ .

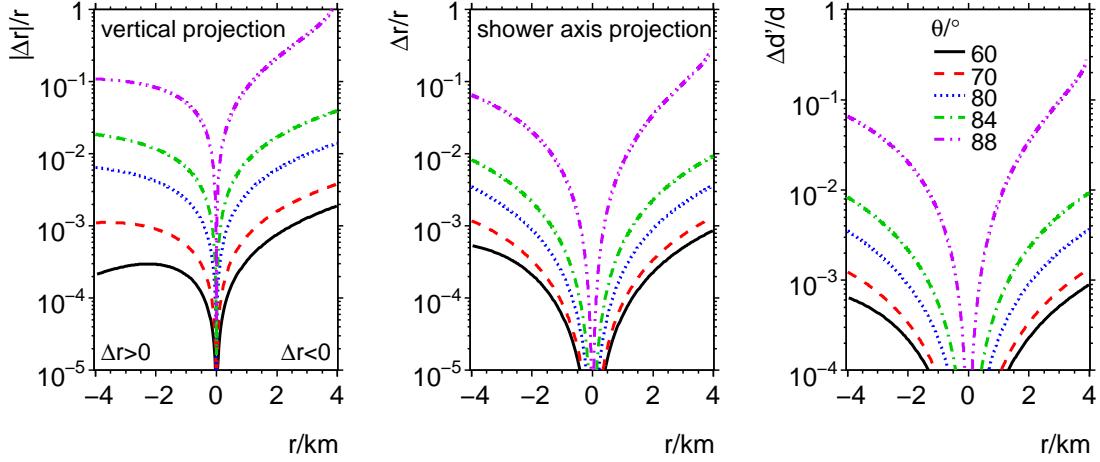


Figure A.4: Shown are geometrical calculations based on Fig. A.3. Left: The graphs show the bias  $\Delta r/r$  of the apparent position of the SD station in the lateral coordinate system introduced by the vertical projection. The bias is positive in the late part of the shower and negative in the early part. Center: The graphs show the bias  $\Delta r/r$  of the apparent position of the SD station in the lateral coordinate system introduced by the shower axis projection. The bias is always positive. Right: The graphs show the relative increase  $\Delta d'/d$  of the distance  $\Delta d' + d$  between the muon origin and the position of the SD station with respect to the distance  $d$  of its apparent position in the ideal ground plane.

therefore the bias  $\Delta S$  in the expected signal is about  $\Delta S/\langle S \rangle \approx 2\Delta r/r$ . The bias  $\Delta S$  becomes significant at  $\theta > 84^\circ$ , if the approximate projections are used instead of the projection along the average muon trajectory.

Also shown in Fig. A.4 is the additional path length  $\Delta d'$  that the muon has to travel from the apparent position of the SD station in the ideal ground plane to its true position on the ground. The prediction  $\langle S \rangle$  is calculated in the ground plane. The prediction is only comparable with the true measurement, if the apparent collection area of the station does not change significantly and if additional muon attenuation may be neglected over the additional distance  $\Delta d'$ . The relative error in the apparent collection area equals  $(\Delta d'/d)^2$  and is negligible up to  $\theta = 88^\circ$ . It is assumed that the additional muon attenuation can be neglected if  $\Delta d'/d < 0.01$ . In this case, the projection remains valid up to  $\theta \approx 84^\circ$ .

## A.5 Calculation of the data model $f_{\text{tot}}$

The following probability density function (p.d.f.) is derived in Chapter 7:

$$f'_{\text{tot}}(R_\mu^{\text{SD}}, E_{\text{FD}} | \mathbf{p}) = \int d\theta \int dE g_{\text{FD-rec}}(E_{\text{FD}} | E, \theta, \mathbf{p}) h_{\text{FD-hyb}}(E, \theta | \mathbf{p}) \times \\ \int dR_\mu^{\text{sh-sh}} P_{\text{SD}}(R_\mu^{\text{SD}}, \theta, \mathbf{p}) g_{\text{SD-rec}}(R_\mu^{\text{SD}} | R_\mu^{\text{sh-sh}}, \theta, \mathbf{p}) h_{\text{SD-sh-sh}}(R_\mu^{\text{sh-sh}} | R_\mu(E), \mathbf{p}) \quad (\text{A.5.1})$$

whereas  $g_{\text{FD-rec}}$  and  $g_{\text{SD-rec}}$  model the measurement uncertainty of the FD and the SD respectively,  $h_{\text{FD-hyb}}$  is the distribution of the true energies observed in the FD,  $P_{\text{SD}}$  is the SD reconstruction probability (which is not a p.d.f.), and  $h_{\text{SD-sh-sh}}$  is the model of the shower-to-shower fluctuations. The energy  $E$  and the zenith angle  $\theta$  are the true values of the cosmic ray,  $R_\mu$  is the ideal energy estimator corresponding to  $E$ , and  $R_\mu^{\text{sh-sh}}$  is the true energy estimator realised in the particular event. The calibration function  $R_\mu = (E/E_{\text{cal}})^{1/\gamma}$  predicts the relation between  $R_\mu$  and  $E$  which removes the remaining degree of freedom in the integral.

Furthermore,  $f'_{\text{tot}}$  has to be normalised by integrating over the input range of  $E_{\text{FD}}$  and  $R_\mu^{\text{SD}}$ , respectively:

$$f_{\text{tot}}(R_\mu^{\text{SD}}, E_{\text{FD}}) = \frac{f'_{\text{tot}}(R_\mu^{\text{SD}}, E_{\text{FD}})}{\int_{R_\mu^{\text{SD,min}}}^{R_\mu^{\text{SD,max}}} dR_\mu^{\text{SD}} \int_{E_{\text{FD,min}}}^{E_{\text{FD,max}}} dE_{\text{FD}} f'_{\text{tot}}(R_\mu^{\text{SD}}, E_{\text{FD}})}. \quad (\text{A.5.2})$$

Solving these equations requires a calculation a three-dimensional convolution integral over  $R_\mu^{\text{SD}}$ ,  $E_{\text{FD}}$ , and  $\theta$  and additionally a two-dimensional integral for the normalisation. These calculations can only be done numerically. The computation needs to be fast, because  $f_{\text{tot}}$  is recalculated several hundred times in a single maximum likelihood fit. A Monte-Carlo integration cannot be used, because typical minimisation algorithms like those in MINUIT [148] require  $f_{\text{tot}}$  to be deterministic. Some approximations are necessary to perform the computation in reasonable time. The current approach is simple and there likely is still room for improvement in the future.

The calculation time increases exponentially with the number of dimensions to integrate over in Eq. (A.5.1). It is already implicitly assumed in Eq. (A.5.1), that any dependency of the component p.d.f.s on the shower azimuth  $\phi$  is negligible. The integration over the zenith angle  $\theta$  is now carried out approximately in order to reduce the problem to two dimensions.

Most events are found in the comparably narrow zenith angle range from  $60^\circ$  to  $70^\circ$ , and all but one of the model components depend only weakly on  $\theta$  in this range:

- The FD-resolution of  $E_{\text{FD}}$  and the SD-resolution of  $R_\mu^{\text{sh-sh}}$  depend only weakly on  $\theta$  in the regarded zenith angle range. The former will be shown later in this chapter, the latter is shown in Chapter 6. This leads to the substitution:

$$g_{\text{FD-rec}}(E_{\text{FD}} | E, \theta, \mathbf{p}) \rightarrow g_{\text{FD-rec}}(E_{\text{FD}} | E, \mathbf{p}) \\ g_{\text{SD-rec}}(R_\mu^{\text{SD}} | R_\mu^{\text{sh-sh}}, \theta, \mathbf{p}) \rightarrow g_{\text{SD-rec}}(R_\mu^{\text{SD}} | R_\mu^{\text{sh-sh}}, \mathbf{p}).$$

- The shower-to-shower fluctuations  $h_{\text{SD-sh-sh}}$  are independent of the zenith angle  $\theta$ , as shown in Chapter 5.

The model of the SD reconstruction probability  $P_{\text{SD}}$  is thus the only  $\theta$ -dependent model component left. The correlation between energy  $E$  and zenith angle  $\theta$  in  $h_{\text{FD-hyb}}$  is weak and the p.d.f. approximately factorises:

$$h_{\text{FD-hyb}}(E, \theta | \mathbf{p}) \approx \tilde{h}_{\text{FD-hyb}}(E | \mathbf{p}) \times \tilde{h}_{\text{FD-hyb}}(\theta | \mathbf{p}).$$

The distribution of hybrids  $h_{\text{FD-hyb}}$  is fixed in the maximum likelihood fit. If  $P_{\text{SD}}(R_{\mu}^{\text{SD}}, \mathbf{p})$  was fixed, too, the integration over  $P_{\text{SD}}$  could be carried out just once in the beginning of every minimisation to obtain an average reconstruction probability  $\langle P_{\text{SD}} \rangle$ , which then used in the rest of the calculations

$$\int d\theta h_{\text{FD-hyb}}(E, \theta | \mathbf{p}) P_{\text{SD}}(R_{\mu}^{\text{SD}}, \theta, \mathbf{p}) = \tilde{h}_{\text{FD-hyb}}(E | \mathbf{p}) \langle P_{\text{SD}}(R_{\mu}^{\text{SD}}, \mathbf{p}) \rangle.$$

Unfortunately,  $P_{\text{SD}}(R_{\mu}^{\text{SD}}, \mathbf{p})$  has free parameters, which are fitted with the maximum likelihood method. Thus, the first approach fails and the following approximation is used instead

$$\int d\theta h_{\text{FD-hyb}}(E, \theta | \mathbf{p}) P_{\text{SD}}(R_{\mu}^{\text{SD}}, \theta, \mathbf{p}) \approx \tilde{h}_{\text{FD-hyb}}(E | \mathbf{p}) P_{\text{SD}}(R_{\mu}^{\text{SD}}, \langle \theta \rangle, \mathbf{p}),$$

which would only be exact, if  $P_{\text{SD}}$  was a linear function of  $\theta$ .

The remaining integrals over  $R_{\mu}^{\text{SD}}$  and  $E_{\text{FD}}$  in Eq.(A.5.1) are discretized and calculated on a lattice. This is the simplest form of a numerical integration, based on the formula

$$\begin{aligned} \int_{x_0}^{x_n} dx f(x) &\approx \sum_i f(\bar{x}_i) \Delta x_i \\ \text{with } \bar{x}_i &= \frac{1}{2}(x_{i+1} + x_i), \quad \Delta x_i = (x_{i+1} - x_i). \end{aligned} \tag{A.5.3}$$

The lattice uses equi-distant steps in  $\lg E_{\text{FD}}$  and  $\lg R_{\mu}^{\text{SD}}$ , because  $R_{\mu}^{\text{SD}}$  and  $E_{\text{FD}}$  span over orders of magnitude. The differentials in  $f'_{\text{tot}}$  need to be substituted accordingly

$$\begin{aligned} dE &= \ln 10 E d\lg E \rightarrow \ln 10 \bar{E} \Delta \lg E \\ dR_{\mu}^{\text{sh-sh}} &= \ln 10 R_{\mu}^{\text{sh-sh}} d\lg R_{\mu}^{\text{sh-sh}} \rightarrow \ln 10 \bar{R}_{\mu}^{\text{sh-sh}} \Delta \lg R_{\mu}^{\text{sh-sh}}. \end{aligned}$$

The optimal grid constants were found by experimenting: 70 steps in  $\lg R_{\mu}^{\text{SD}}$  in the interval  $(-1.2, 1.5)$  and 50 steps in  $\lg E_{\text{FD}}$  in the interval  $(17.5, 20.5)$  are an acceptable tradeoff between speed and quality.

The lattice calculation has the other advantage. Once it is completed, the function value of  $f_{\text{tot}}$  at a given point  $(R_{\mu}^{\text{SD}}, E_{\text{FD}})$  can be calculated very fast. This is ideal for the application of  $f_{\text{tot}}$  in the maximisation of the likelihood function  $L(\mathbf{p})$ :

$$L(\mathbf{p}) = \prod_i f_{\text{tot}}((R_{\mu}^{\text{SD}})_i, (E_{\text{FD}})_i | \mathbf{p}). \tag{A.5.4}$$

The normalisation of  $f_{\text{tot}}$  can be calculated very fast for the same reason.

The numerical computation of  $f_{\text{tot}}$  is still time consuming. With the current approach, a single minimisation step in MINUIT takes about a second on an Athlon™ 64 3700+ CPU. The same step with a  $100 \times 70$  grid already takes about 30 seconds. The numerical minimisation routines usually needs about 100 iterations, thus a single fit with a fine grid can easily take an hour and more.

# Appendix B

## Air shower simulation: Technical summary

The tables Table B.1 and Table B.2 give a technical summary of the air shower library described in Chapter 5. The showers were simulated with CORSIKA [49, 89], using QGSJet-II [126], EPOS [130, 131], and FLUKA [102, 103]. Fig. B.1 shows the computation time and disc space requirements of the showers.

The whole library was generated in two main production runs: the run IDs 4681-6480 mark the first run, run IDs 1-4680 the second. The second production run has lower run IDs than the first run due to historical reasons. The first run was generated with the U.S. standard atmosphere model, while the second uses the Malargüe spring atmosphere [41, 43], which is close to a yearly average atmosphere over Malargüe. Fixes of minor technical issues (like program crashes) led to the use of three different versions of the main program CORSIKA in the second run.

Another difference between the runs is the use the radial thinning. CORSIKA allows to remove particles with a probability  $\propto r/r_{\text{thin}}^4$  inside a cone of radius  $r_{\text{thin}}$  around the shower axis. This option was turned on in the first production run and turned off in the second run.

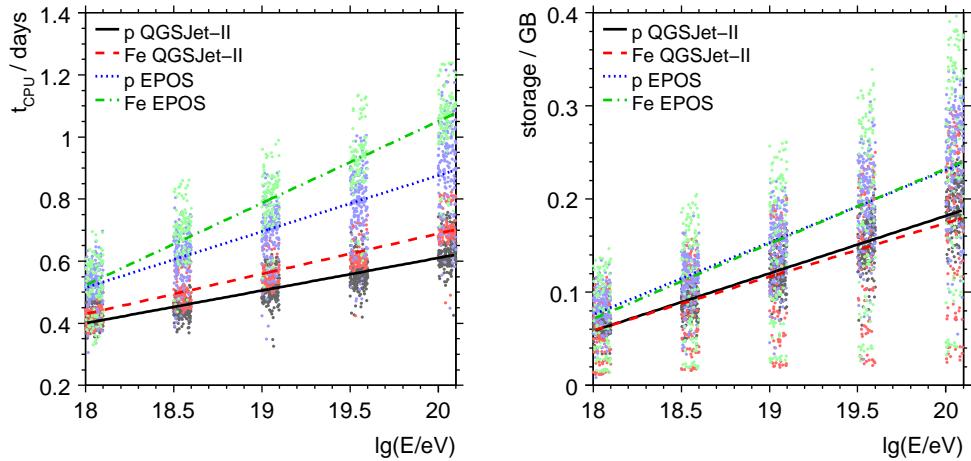


Figure B.1: The points show the CPU time and the storage requirements of the simulated air showers. The CPU time refers to the computing time on an AMD Opteron<sup>TM</sup> CPU with 2 GHz. In total, the production took approximately 4000 CPU days and 631 GByte of disc space. The lines are simple fits to the data. It is a coincidence, that the applied thinning strategy produces nearly equal average file sizes for proton and iron showers, although the latter have more muons.

---

Table B.1: The showers in the library are distributed in small finite regions of the parameter space. The table shows the starting point of each region and its width. Each region contains five proton showers and one iron shower for both QGSJet-II and EPOS, which makes 6480 showers in total.

Parameter	Distribution	Start point	Width
$\lg E/\text{eV}$	flat	18.0, 18.5, 19.0, 19.5, 20.0	0.1
$\theta/^\circ$	$\sin(\theta) \cos(\theta)$	60, 62, 64, 66, 70, 74, 78, 82, 86	2
$\phi/^\circ$	flat	0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330	10

Table B.2: The table shows a summary of the technical aspects of the simulated air shower library.

Parameter	Value	Comment
simulation program	CORSIKA	options: CURVED, UPWARD, (SLANT)
	6.511	run ID 4681-6480, no SLANT option
	6.616, 6.617, 6.720	run ID 1-4680, with SLANT option
HE hadronic model	QGSJet-II-3, EPOS-1.61	for hadrons with $E > 80 \text{ GeV}$
LE hadronic model	FLUKA2006	for hadrons with $E \leq 80 \text{ GeV}$
thinning level	$10^{-6}$	energy fraction, where thinning sets in
$w_{\max}(\text{hadrons}, \mu)$	$E/10^{15} \text{ eV}$	weight limit for hadrons, muons
$w_{\max}(e, \gamma)$	$E/10^{13} \text{ eV}$	weight limit for electrons, photons
$r_{\text{thin}}$	150 m	run ID 4681-6480, with radial thinning
	-	run ID 1-4680, no radial thinning
$p_{\text{thr}}(\text{hadrons}, \mu)$	0.1 GeV	momentum threshold for hadrons, muons
$p_{\text{thr}}(e, \gamma)$	250 keV	momentum threshold for electrons, photons
ground altitude	1425 m	above sea level
$B$	$24.6 \mu\text{T}$	magnitude of geomagnetic field
$\theta_B$	$-35.2^\circ$	inclination of geomagnetic field
$\delta_B$	$4.2^\circ$	declination of geomagnetic field
atmosphere model	U.S. standard	run ID: 4681-6480
	Malargüe spring	run ID: 1-4680

## SD station: signal response model

Technically, the model of the signal response to muons  $f_\mu(S_\mu|\theta_\mu, k)$  described in Chapter 5 is derived through the following steps:

- (1) Extract the total signal  $S^1$  from every station in the library of simulated SD events that has one muon hit. Stations with total signals below the T1 threshold are included. Proton and iron showers are treated alike, as well as the showers generated with the hadronic interaction models QGSJet-II and EPOS.
- (2) Correct the total signal response  $S^1$  to one muon with the model of average signal ratio  $\langle\epsilon\rangle$  from ref. [16], which is also used in the event reconstruction in Chapter 6.
- (3) Make histograms of the distributions of  $S^1/(1 + \langle\epsilon\rangle)$  as a function of the average local muon inclination  $\theta_\mu$  at the position of the station. Fourteen distributions are generated in steps of  $2^\circ$  between  $\theta_\mu = 60^\circ$  and  $\theta = 88^\circ$ . Each distribution has 500 bins spanning from 0 VEM to 100 VEM.

Distributions for intermediate  $\theta_\mu$ -angles are obtained via linear interpolation of the existing ones:

$$f_\mu(S_\mu^1|\theta_\mu, 1) = (1 - p) f_\mu(S_\mu^1|\theta_{\mu,1}, 1) + p f_\mu(S_\mu^1|\theta_{\mu,2}, 1), \quad (\text{C.0.1})$$

with  $p = (\theta_\mu - \theta_{\mu,1})/(\theta_{\mu,2} - \theta_{\mu,1})$ . A linear interpolation produces the correct mean and variance of the intermediate  $f_\mu$ , but not necessarily the right higher moments. Still, for small intervals in  $\theta_\mu$ , the approximation is good.

The signal distributions for  $k$  muons are derived from the distributions for one muon by auto-convolution, following Eq. (5.5.7). Performing a numerical auto-convolution is computationally expensive. It takes  $O(N^k)$  time in the typical notation of informatics with  $k$  as the number of muons and  $N = 500$  as the number of bins in the histogram. The evaluation of the  $f_\mu(S_\mu|\theta_\mu, k)$ -model has to be fast, because the p.d.f. is called up to  $10^6$  times in the reconstruction of a single event. Therefore,  $f_\mu(S_\mu|\theta_\mu, k)$  is pre-calculated up to  $k = 9$  and approximated by a normal distribution for  $k \geq 10$  with mean and width

$$\mu = k \int dx x f_\mu(x|\theta_\mu, 1) \quad (\text{C.0.2})$$

$$\sigma^2 = k \left( \int dx x^2 f_\mu(x|\theta_\mu, 1) - \left( \int dx x f_\mu(x|\theta_\mu, 1) \right)^2 \right), \quad (\text{C.0.3})$$

following the standard approach. The Central Limit theorem assures that  $f_\mu$  approaches a normal distribution for large  $k$ .

Fig. C.1 shows the result of this implementation. Introducing the approximation at  $k = 10$  is acceptable, but leads nevertheless to a visible discontinuity in the shape of  $f_\mu(S_\mu|\theta_\mu, k)$ , as the normal approximation ignores a tail towards large signals.

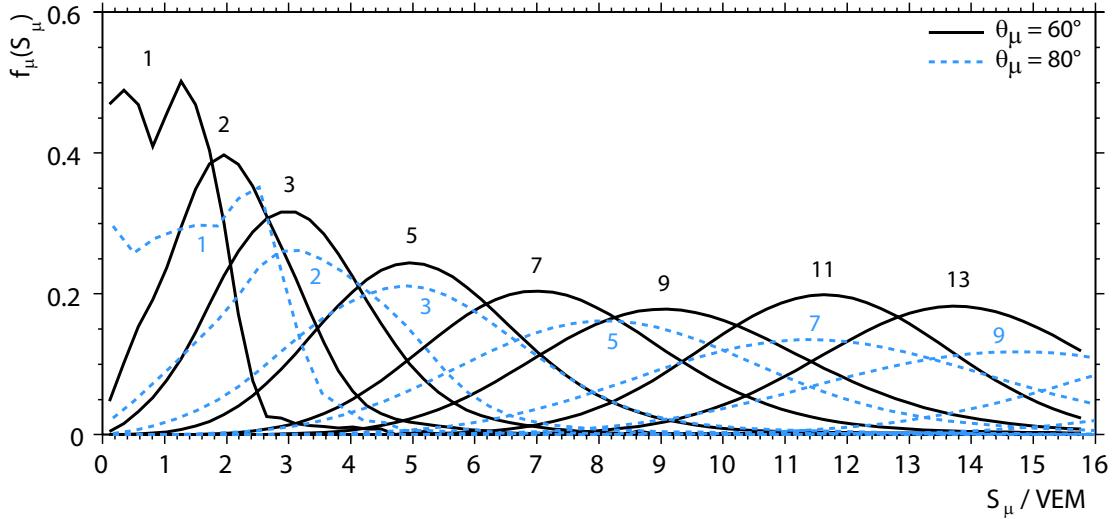


Figure C.1: The plot illustrates the auto-convolution approach to obtain  $f_\mu(S_\mu|\theta_\mu, k)$  from  $f_\mu(S_\mu|\theta_\mu, 1)$  used in this study. Depicted are the p.d.f.s  $f_\mu(S_\mu|\theta_\mu, k)$  for several values of  $k$  at two muon inclinations  $\theta_\mu$ . At  $k \geq 10$ , the auto-convolution is replaced by a Gaussian approximation, as described in the text. The approximation can be noticed by a slight change in the shape of  $f_\mu(S_\mu|\theta_\mu, k)$ .

Other authors point out [218] that the numerical computation of the auto-convolution for large  $k$  is faster, if it is performed in Fourier space. The characteristic function of  $f_\mu(S_\mu|\theta_\mu, k)$  is

$$\hat{f}_\mu(\hat{S}_\mu|\theta_\mu, k) = \int_{-\infty}^{\infty} dS_\mu e^{iS_\mu \hat{S}_\mu} f_\mu(S_\mu|\theta_\mu, k), \quad (\text{C.0.4})$$

whereas  $\hat{S}_\mu$  is the corresponding variable of  $S_\mu$  in Fourier space.

It has the useful property, that an auto-convolution is an exponentiation

$$\hat{f}_\mu(\hat{S}_\mu|\theta_\mu, k) = (\hat{f}_\mu(\hat{S}_\mu|\theta_\mu, 1))^k, \quad (\text{C.0.5})$$

see *e.g.* ref. [147]. The characteristic function  $\hat{f}_\mu(\hat{S}_\mu|\theta_\mu, 1)$  can be stored in memory. The time for the exponentiation scales like  $O(\ln k)$ , so that the full auto-convolution in Fourier space than takes  $O(N \ln k)$  time. A Fast-Fourier transformation [219] needs only  $O(N \ln N)$  time to compute  $f_\mu$  from  $\hat{f}_\mu$ . If  $k$  is large, this approach can provide a huge speed up.

The Gaussian approximation of  $f_\mu(S_\mu|\theta_\mu, k)$  at large  $k$  may thus be completely avoided. Also, a better interpolation in  $\theta_\mu$  is possible with characteristic functions. A first implementation of the Fourier-based computation is included in the Offline-HAS package [167], but still needs more refinement to run at acceptable speed and is therefore not used in this study.

# Appendix D

## Reconstruction of SD and FD events: technical details

### SD event reconstruction

Fig. D.1 shows a flow chart of the module sequence used to reconstruct the SD events described in Chapter 6. The T4 selection of the recorded SD events is done with a separate program and described in Appendix E. The rest of the reconstruction is implemented in the Auger `Offline`-framework [137, 168]. The SD reconstruction is based on `Offline-v2r5p7-Godot`. The module `SdHorizontalReconstruction` is used in the internal revision number 8254. The source code of this revision can be obtained online [167].

The `Offline`-reconstruction is based on a sequence of modules. Each module performs a specific task and may be configured with a steering card. After the preprocessing by the T4 selection program, the SD events are read by the `EventFileReaderOG` module. The raw SD event contains only FADC traces of the signals in individual stations. The module `SdCalibratorOG` performs the conversion of these FADC traces to signals in units of VEM, as described in Chapter 4. It also finds the signal start time in each trace with the trace cleaning procedure described in Appendix E.

The `SdEventSelectorOG` module is the next step in the sequence. It implements the rejection of lightning events, calculates whether the event passes the T5-Prior criterion, and marks whether events if they are in bad time period of data acquisition. The necessity to reject certain periods of data acquisition is explained in Chapter 8.

The `SdPlaneFitOG` module fits the shower direction using a model of a plane shower front moving with the speed of light. This preliminary shower direction is required in the next reconstruction steps which are implemented in the `SdHorizontalReconstruction` module. This module performs the reconstruction of the energy estimator  $R_\mu$  of the cosmic ray and the reconstruction of the shower direction using a model of a curved shower front. The reconstruction of the energy estimator is based on models of the lateral density of muons  $n_\mu$  in the shower at the ground level (`MuonProfile`), the signal response of the SD stations to such muons (`TankResponse`), and the contribution of electromagnetic particles to this signal (`EMComponent`).

The `SdEventPosteriorSelectorOG` calculates whether the event passes the strict T5-Posterior criterion. Finally, the event is written in the ADST-format [220]. Events in this format can be used in data analyses or investigated with the `EventBrowser` program.

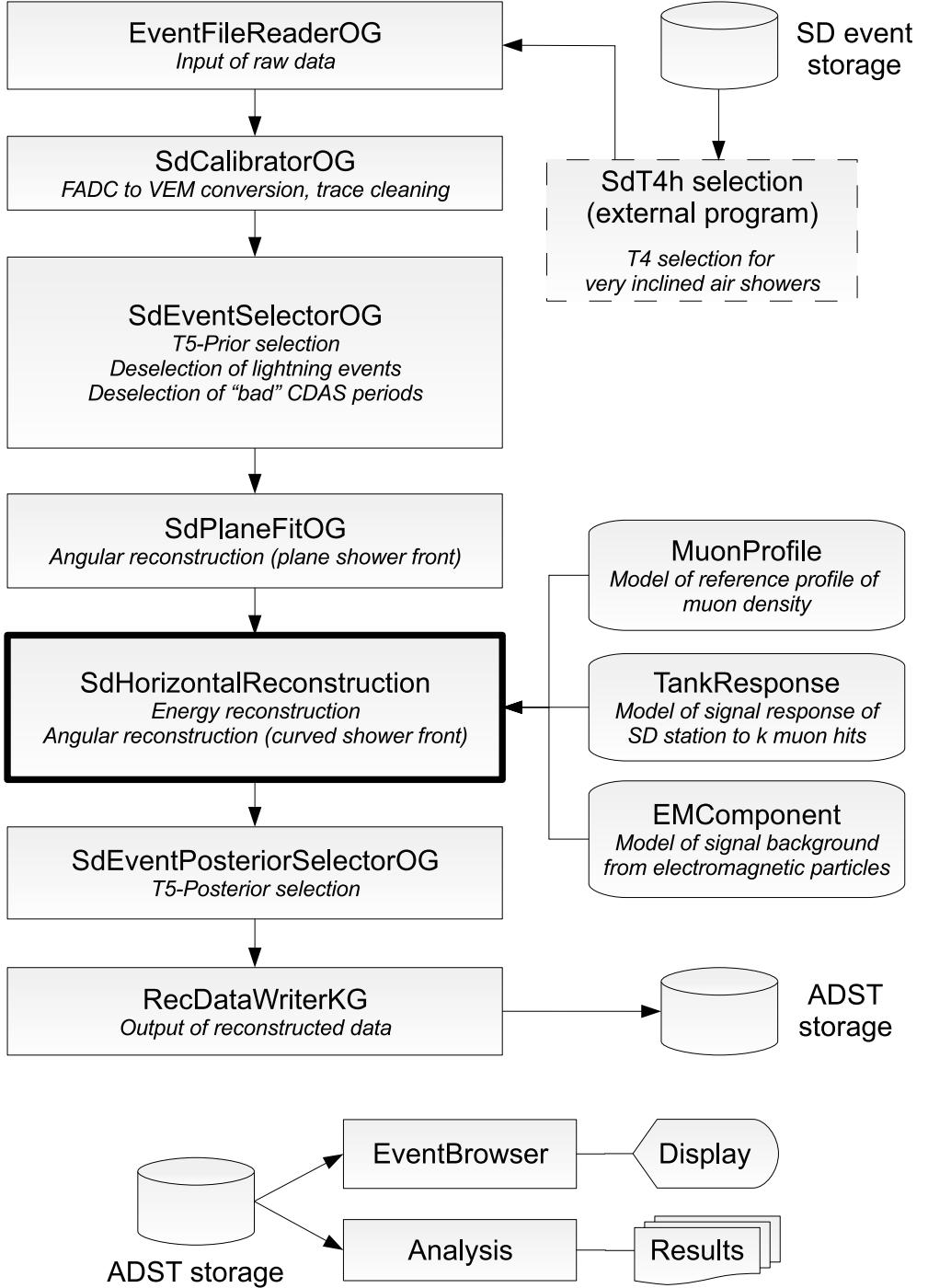


Figure D.1: Shown is a flow chart of the SD reconstruction of very inclined showers based on Offline modules. The new module **SdHorizontalReconstruction** is highlighted. The reconstructed data is written in the ADST format [220], which can be displayed with the EventBrowser program or used for analyses. The reconstruction chain uses pre-selected **SdT4h** events as input, the station selection in the module **SdEventSelectorOG** is turned off, it is only used for its other selection flags (see text).

## FD event reconstruction

The reconstructed FD events are taken from an official production [220], which is also based on `Offline-v2r5p7-Godot` [137]. The FD event reconstruction described in Chapter 6 is implemented as the following sequence of modules:

- (1) `EventFileReaderOG`: Read the raw FD and SD event data.
- (2) `SdCalibratorOG`: Find the signal start times in the triggered SD stations, which are regarded as part of the event by the FD trigger as described in Chapter 4. Perform the VEM calibration of the signals in these SD stations.
- (3) `FdEventSelectorKG`: Perform a basic event pre-selection, for example to choose only events which were recorded simultaneously with two FD buildings or which have many triggered SD stations. Optionally reject laser shots.
- (4) `FdCalibratorOG`: Convert the raw ADC signal in each camera pixel into counts of UV photons. Correct for the overall SD-FD timing offset which is generated by the different data acquisition systems of the FD and the SD.
- (5) `FdPulseFinderOG`: Locate the signal pulse in each pixel. Calculate the arrival time for the pulse and its integrated photon count.
- (6) `PixelSelectorOG`: Reject isolated pixels with no close neighbors which are generated by random noise.
- (7) `FdSDPFinderOG`: Find the shower detector plane of the event. Reject pixels if they are incompatibility with this plane.
- (8) `HybridGeometryFinderOG`: Find a SD station close the shower detector plane. Use its signal arrival time together with the signal arrival times in the camera pixels to reconstruct the distance and orientation of the shower axis.
- (9) `FdApertureLightOG`: Reconstruct the light profile of the shower at the aperture of the telescope.
- (10) `FdProfileReconstructorKG`: Reconstruct the energy loss profile of the shower, based on light profile at the aperture of the telescope, the reconstructed shower axis, models of the atmospheric scattering of light, models of the emission of Cherenkov and fluorescence light, and a model of the invisible energy of the shower. The integral of the profile yields the position of the electromagnetic shower maximum  $X_{\max}$  and the shower energy  $E$ .



# Appendix E

## SD event reconstruction: background rejection

### Trace cleaning

The current trace cleaning algorithm is based on ref. [171]. The algorithm starts at the level of the individual PMTs of a station:

- (1) Extract signal pieces from the full high gain trace of each PMT. A signal piece is defined by at least  $n \geq 2$  bins subsequent bins, where each bin has at least 3 FADC counts above the baseline. Store the start and stop index of the piece, calculate its integrated signal  $S$  and peak signal  $S_{\text{peak}}$ .
- (2) Join pieces of the first step separately for each PMT. Pieces are joined, if the gap between two pieces is smaller than  $20 + \text{length of a first piece}$ , and at least one of the two conditions is met:
  - (a) The first piece has a signal  $S_1 > 0.3 S_2$ , whereas  $S_2$  is the signal of the second piece.
  - (b) The second piece has a peak signal  $S_{\text{peak}}$  of less than 5 FADC counts above the baseline.
- (3) Merge the signal pieces of different PMTs, so that only one set of pieces remains. Overlapping pieces between different PMTs are merged by enlarging the time window of the piece and averaging the signals  $S_i$  of the overlapping pieces.
- (4) The start time of the merged trace of the station is determined from the merged piece with the largest integrated signal.

### T4 selection

The T4 selection for very inclined showers [173] is based on a top-down approach. At the time of this writing, the selection is done with a stand alone program and not within the Auger Offline framework, but work is in progress to integrate it in the near future.

The T4 algorithm is applied after a preselection and the already described trace cleaning. The preselection removes all known faulty stations and stations which are not part of the regular SD grid (like the infill described in Chapter 4). The algorithm starts with  $n$  preselected and triggered stations, and then applies the following steps:

- (1) Set the number  $k$  of rejected stations to zero.
- (2) Remove isolated stations. An isolated station has less than one neighbor in a distance  $d_1 = 4.7$  km or less than two neighbors in a distance  $d_2 = 6.2$  km. If  $n > 40$ ,  $d_1$  is reduced to 3.4 km to save computing time.

---

(3) Apply a preliminary reconstruction of the shower axis.

(a) Calculate the barycenter, according to

$$\mathbf{r}_b = \frac{\sum_{i=1}^{n-k} \mathbf{r}_i S_i^{1/3}}{\sum_{i=1}^{n-k} S_i^{1/3}}, \quad (\text{E.0.1})$$

whereas  $\mathbf{r}_i$  and  $S_i$  are the ground position and signal of the station. In the following,  $\mathbf{r}'_i$  denotes the station coordinates relative to the barycenter.

(b) Fit the model of a plane shower front moving with the speed of light to the signal start times  $t_i$ , while neglecting the individual altitude  $z'_i$  of each station with respect to the ground plane

$$c(\langle t_i \rangle - t_0) = -\mathbf{a} \cdot \mathbf{r}_i \Leftrightarrow \langle t_i \rangle = t_0 - \frac{1}{c}(ux'_i + vy'_i), \quad (\text{E.0.2})$$

with  $\mathbf{a} = (u, v, w)^T$  as the shower incoming direction, as illustrated in Fig. 6.9 a). Free parameters of the model are  $u$ ,  $v$ , and  $t_0$ . An unique and fast solution for three parameters can be obtained from the linear least squares method for parameter estimation, as described in *e.g.* ref. [147].

(c) The shower front fit is improved by approximately taking the altitude of the stations into account

$$\langle t_i \rangle = t_0 - \frac{1}{c}(ux'_i + vy'_i + w_0 z'_i), \quad (\text{E.0.3})$$

whereas  $w_0 = \sqrt{\max(0, 1 - u_0^2 - v_0^2)}$  is the fixed vertical component of the normalised shower direction from the first fit. Because  $w_0$  is not considered as a free parameter, the linear least squares method can still be applied.

(d) If  $n - k = 3$ , refine the fit further by adding an approximate term for the spherical shape of the shower front

$$\langle t_i \rangle = t_0 - \frac{1}{c}(ux'_i + vy'_i + w_1 z'_i + w_1 r_i^2 / (2d_{\max})), \quad (\text{E.0.4})$$

whereas  $r_i^2 = x'^2_i + y'^2_i - (ux'_i + vy'_i)^2$  is the radial distance to the preliminary shower axis going through the barycenter and  $w_1$  is the fixed value from the second fit. The radius of the shower front sphere is roughly approximated here by  $d_{\max} \approx 7.1 \text{ km} / \cos \theta$ .

(4) If  $n - k > 3$ , accept the current configuration under the following conditions:

(a) Physical values for  $u$  and  $v$ :  $u^2 + v^2 < 1$ .

(b) Small residuals  $\Delta t_i = |t_i - \langle t_i \rangle|$ :

$$\begin{aligned} \max(\Delta t_i) &< (n - 2) \min(w, 0.2) \times 250 \text{ ns} \\ \sqrt{\frac{\sum_i (\Delta t_i)^2}{n - 3}} &< (n - 2) \min(w, 0.2) \times 200 \text{ ns} \end{aligned}$$

The factor  $(n - 2)$  relaxes the restriction in case of large events, taking into account that the shower front model is only approximate. The factor  $\min(w, 0.2)$  tightens the restrictions as the shower inclination grows, because the natural variance of the arrival time is smaller in very inclined showers as discussed in Chapter 3.

(c) Compact spatial configuration. The radial distance  $r_i$  to the preliminary shower axis may not be too large:

$$\max(r_i) < \sqrt{n - 2} \times 1.3 \text{ km}.$$

Again, the restriction is relaxed for larger showers.

(d) Configuration may not be aligned. The principal axes of the configuration are calculated in the ground plane. The length of both axes has to be larger than 0.3 km.

(e) The configuration still has to fulfill the T3 trigger condition, see Chapter 4.

If  $n - k = 3$ , accept the current configuration under a variation of the conditions:

(a) Physical values for  $u$  and  $v$ :  $u^2 + v^2 < 1$ .

(b) At least two stations with a integrated signal-to-peak ratio  $> 1.6$ .

(c) The sides of the triangle formed by the station positions are smaller than 2.8 km. The area of the triangle is between 0.2 km and 1.2 km. The lower limit on the area efficiently rejects aligned configurations.

(d) The configuration still has to fulfill the T3 trigger condition, see Chapter 4.

(5) If the configuration is not accepted, increase  $k$  by one. Repeat (2) and (3) with all possible combinations of rejecting  $k$  stations, starting with the stations that have the lowest signals. Reject the event altogether, if  $k > n - 3$ .

Only the selected stations with signal, and the active but silent stations are considered in the following reconstruction steps. The rejected ones are marked as accidentally triggered.



Appendix **F**

## Event selection: examples of rejected events

This appendix shows some example events which are rejected by quality cuts applied to SD and FD events in Chapter 6 and Chapter 7.

Fig. F.1 shows examples of events, which are rejected by the SD selection. Example events rejected by the cuts `SdThetaMin` and `SdBadPeriod` are not shown, they do not have a special signature.

- Fig. F.1(a) shows a typical class of events which are rejected by the T4 algorithm. The footprint of the event is not compact enough for a reliable reconstruction of the shower energy and direction.
- Fig. F.1(b) shows an event, which did not pass the T5 selection, as the station with the largest signal (the bottom left) is not surrounded by six active stations. The core position of such an event is not well constrained by data, which would result in large statistical and systematical uncertainties of the reconstructed energy and direction.
- Fig. F.1(c) shows a spectacular example of a less common class of events, which are rejected by the T5 selection. Events such as this one pass the T5-Prior criteria, but not the T5-Posterior one. This particular very inclined shower ( $\theta \approx 83^\circ$ ) fell into a hole in the array, thus the stations close to the core are missing.

It happens sometimes in near horizontal showers, that the station with the largest signal is not the one closest to the shower core. In this event, the station with the largest signal is in the early part of the event above the hole and had six active neighbors, thus it was able to pass the Prior T5 requirement.

- Fig. F.1(d) shows an event which is rejected by the cut on the zenith angle. The near horizontal shower has a reconstructed zenith angle of about  $87^\circ$ . The event illustrates the butterfly structure of the muon density profile in the shower front plane well, which appears very strongly at these inclinations (see Chapter 5). Projected on the ground, the butterfly structure turns into the shape of a mirrored "S" of the footprint.

The reconstruction is biased in case of these near horizontal showers and therefore they are rejected.

Fig. F.2 and Fig. F.3 show events, which are rejected by the FD quality cuts. Fig. F.2(a)-(c), the right side of Fig. F.3(a), Fig. F.3(b), and Fig. F.3(d) show the energy loss profile of the shower as a function of the slant depth, the red curve represents the fit of the Gaisser-Hillas formula to the data. The point at the maximum shows the uncertainty of the position of the shower maximum with the horizontal error bars, and the uncertainty of the normalisation of the Gaisser-Hillas formula with the vertical error bars. The quality of the fit is indicated by the reduced  $\chi^2$  value in the top right corner of the plot.

Fig. F.2(d) shows the detected photon flux at the camera and the fraction of Cherenkov light contributing to this photon flux. The latter is estimated by the reconstruction algorithm. Several types of Cherenkov light are distinguished by different colors, important for following discussion is only the total sum.

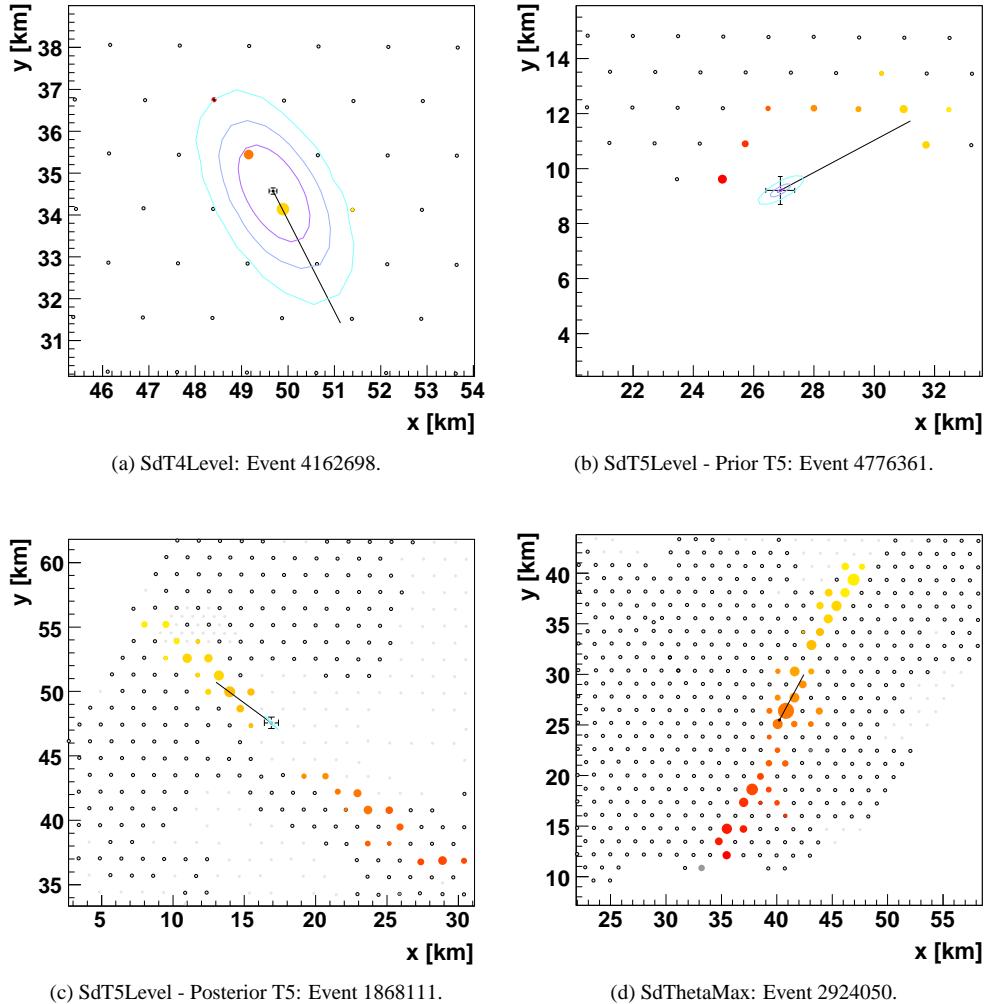


Figure F.1: The plots show examples of SD events in a top view on the SD array, which are rejected by the SD event selection used in this work. Colored circles represent triggered stations. The color indicates the arrival time of the signal, blue are early signals, red late signals. Full and open circles represent stations with recorded signals. The radii of the circles are proportional to the logarithm of the recorded signal. Dark small gray circles represent active untriggered stations, light small gray circles represent stations that are inactive at the time of the event or not yet deployed in the field. The shower core is indicated by a black point with error bars. The arrival direction of the shower projected on the ground is indicated by a black line which ends in the shower core point. The plots are taken from the EventBrowser in the ADST package [220].

## APPENDIX F. EVENT SELECTION: EXAMPLES OF REJECTED EVENTS

---

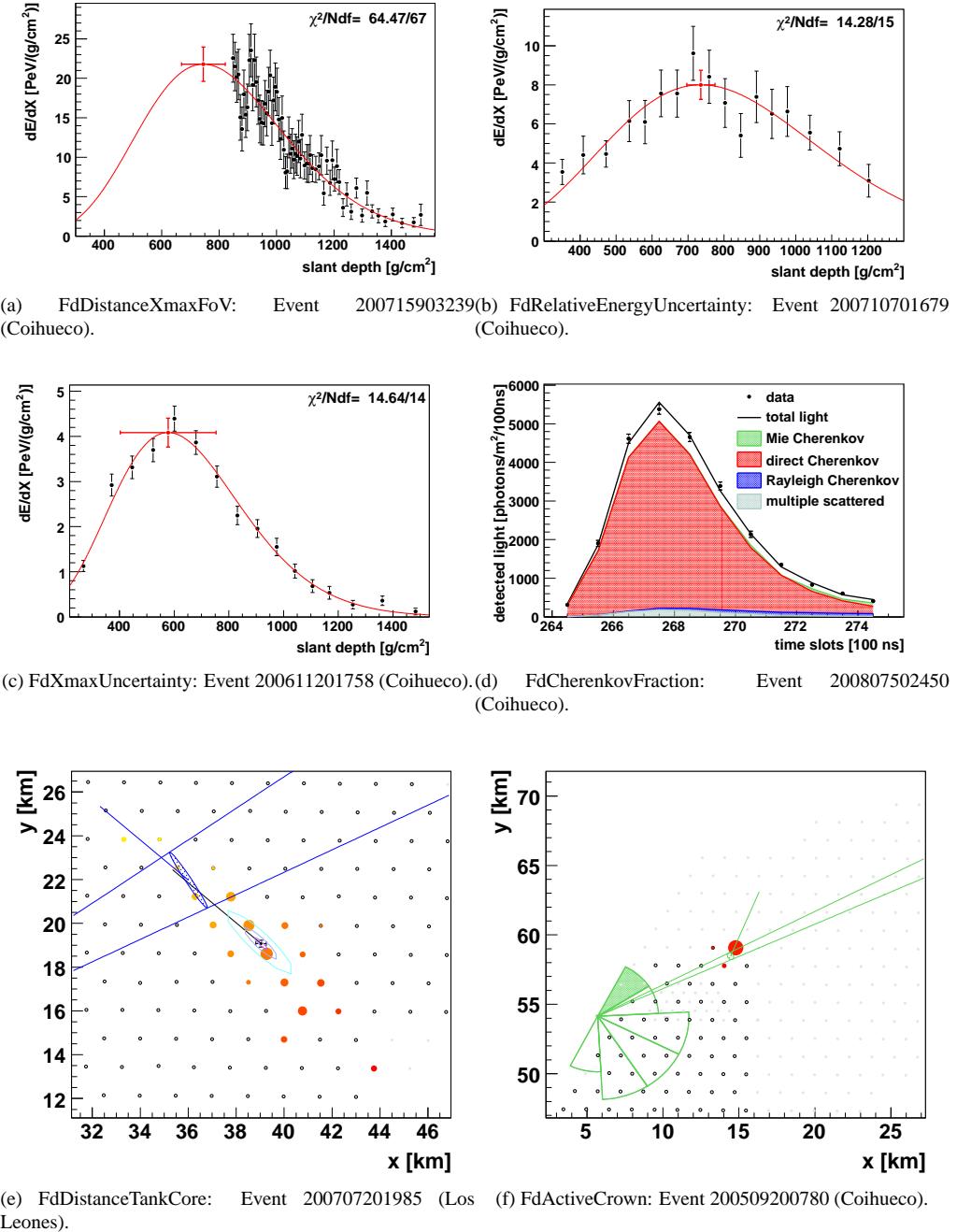


Figure F.2: First set of examples of FD events, which are rejected by the quality cuts used in this work (see text). The Auger ID is shown below each example. The plots are taken from the EventBrowser in the ADST package [220].

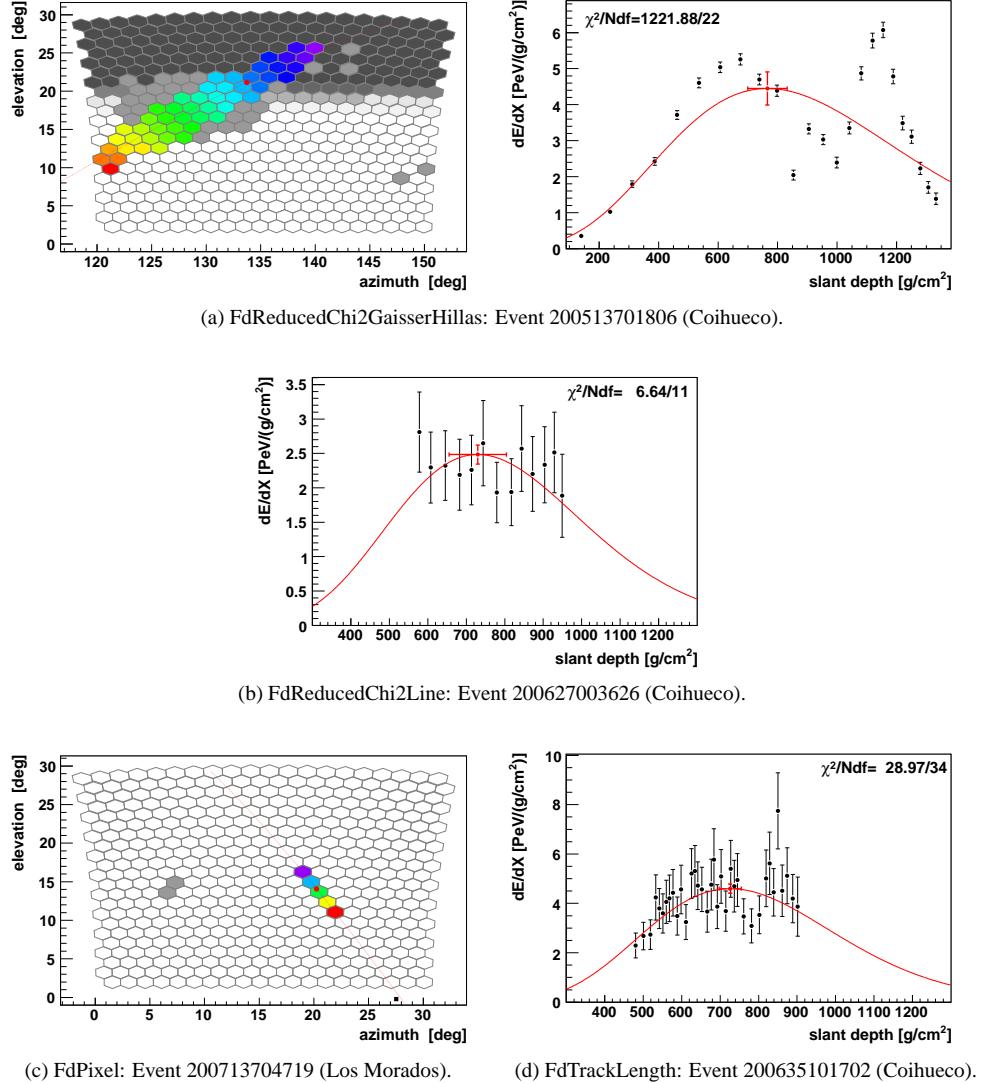


Figure F.3: Second set of examples of FD events, which are rejected by the quality cuts used in this work (see text). The Auger ID is shown below each example. The plots are taken from the EventBrowser in the ADST package [220].

## APPENDIX F. EVENT SELECTION: EXAMPLES OF REJECTED EVENTS

---

Fig. F.2(e) and Fig. F.2(f) show the top view of the SD array as in Fig. F.1. Additionally, the pattern filled ellipse indicates the reconstructed core position of the FD.

The left side in Fig. F.3(a) and Fig. F.3(c) shows a view of the camera pixels. The colored pixels are triggered pixels which survived the pixel selection algorithm, the color indicates the arrival time; blue is early, red is late. Gray pixels have triggered but are rejected by the pixel selection. The black region at the top in Fig. F.3(a) indicates the cloud level in the field of view of the camera. The red curve indicates the position of the shower detector plane fitted to the pixel data.

No example events for the cuts `FdThetaMin` and `FdMieDatabase` are included, as event rejected by these cuts do not have a special signature.

- Fig. F.2(a) shows a typical event, that has its shower maximum outside the field of view of the camera. The reconstructed parameters of such events have large statistical and systematical uncertainties, and are therefore rejected. In some way, the cut is similar to the T5 criteria in surface events.
- Fig. F.2(b) shows a comparably far shower with an energy around  $10^{18}$  eV. The photon flux at the camera is low, which propagates into a low resolution of the integrated energy.
- Fig. F.2(c) shows a high energy shower, which had its shower maximum very above the center of the array, with a direction quite close to the camera axis. In this case, the resolution of the camera pixels is not fine enough to resolve the point of the shower maximum well, as opposed to the situation of a close shower which is seen from the side. For the quality of the reconstructed energy, this is a comparably weak cut, because it mostly rejects events which are also dropped by the cuts `FdDistanceXmaxFoV` and `FdRelativeEnergyUncertainty`.
- Fig. F.2(d) shows an event, where the shower direction was very close the camera axis, and therefore most of the light in the camera is actually Cherenkov light and not fluorescence light. We trust the FD reconstruction to handle such events up to a fraction of 80%. There are not enough events with larger fractions to conclude, whether they introduce a bias or not. They are cutted away to be conservative.
- Fig. F.2(e) shows an example event, which was dropped because the station with the largest signal was too far away from the reconstructed shower core on the ground.
- Fig. F.2(f) shows an example event, which was dropped because shower fell into outside the array. To properly derive the trigger probability of the SD from FD hybrids, such events have to be excluded.
- Fig. F.3(a) shows an event, which was rejected due to a bad fit of the Gaisser-Hillas function to the energy loss profile. The camera view indicates, that the shower passed through a thin cloud. The sharp second maximum may be caused by Cherenkov light scattered by the cloud. Currently, there is no cloud rejection algorithm for the FD reconstruction, which is why such events need to be recognised and rejected through the quality of the fit of the Gaisser-Hillas formula.
- Fig. F.3(b) shows a distant event, which triggered only a few pixels. The observed energy loss profile is also compatible with the fit of a line, which means that the Gaisser-Hillas function is not well constrained by the data and therefore the event is rejected. Although the cut is well motivated, it shows only little effect on the optimisation analysis.
- Fig. F.3(c) shows an event with only five pixels. The precision of the reconstructed shower axis depends on the number of pixels, as does the resolution of the shower maximum.
- Fig. F.3(d) shows an event, which is dropped because of its short track length in the camera.



## Cosmic ray flux: tables and consistency checks

The unfolded cosmic ray flux  $J(E)$  in the energy range from  $10^{18}$  eV to  $10^{20}$  eV derived from fluorescence detector calibrated surface detector events is presented in Chapter 8. The tables Table G.1 and Table G.2 present the numerical result of that chapter.

Two cross-checks are applied to the raw flux  $J_{\text{raw}}$ , which is not unfolded and thus statistically simpler to handle because there are no bin-to-bin correlations. For the sake of these cross-checks the unfolding is irrelevant. The results are shown in Fig. G.1.

The density profile  $\rho_{\text{atm}}(h)$  of the atmosphere over the southern Pierre Auger Observatory changes during summer and winter. This can have a small impact on the produced number of muons  $N_\mu$  during the shower development and produce shifts in the SD reconstructed cosmic ray energy  $E_{\text{SD}} \propto N_\mu$ , as discussed in Chapter 5. The impact can be estimated by comparing the flux  $J_{\text{summer}}$  collected during October to March with the flux collected during April to September in each year. No summer-winter effect is found within the statistical resolution.

There could be an unknown time evolution in the data. The data quality could have improved or degraded over the five years of data acquisition between 2004/01/01 and 2009/01/01. The cosmic ray flux has to be time independent. The data is tested against a possible time dependence by comparing the flux  $J_{\text{early}}$  collected up to 2007/07/17 with the flux  $J_{\text{late}}$  after this date which divides the collected SD exposure in two equal halves. No effect is found within the statistical resolution.

Table G.1: The table shows the unfolded cosmic ray flux  $J(E)$  obtained from FD calibrated SD events in the zenith angle range  $60^\circ < \theta < 82^\circ$ . The flux is normalised to an arbitrary constant  $J_c = 10^{-34} \text{ m}^{-2} \text{s}^{-1} \text{sr}^{-1} \text{eV}^{-1}$ . Three systematic uncertainties are distinguished:  $\sigma_{\text{sys}}^{\text{E}}[J]$  caused by uncertainties in the FD energy scale,  $\sigma_{\text{sys}}^{\text{K}}[J]$  caused by the uncertainties in the unfolding, and  $\sigma_{\text{sys}}^{\text{exp}}[J]$  caused by the uncertainty in the SD exposure. The covariance matrix is shown in Table G.2.

index	$\lg(E/\text{eV})$	$J/J_c$	$\Delta J_{\text{sys}}^{\text{E}}/J_c$	$\Delta J_{\text{sys}}^{\text{K}}/J_c$	$\Delta J_{\text{sys}}^{\text{exp}}/J_c$
1	18.0 – 18.3	5912	+4854 -4854	+510 -647	177
2	18.3 – 18.5	690	+514 -514	+69 -66	21
3	18.5 – 18.7	170.28	+111.26 -111.26	+12.54 -12.80	5.11
4	18.7 – 18.9	44.75	+26.94 -26.94	+3.20 -4.24	1.34
5	18.9 – 19.1	13.71	+8.24 -8.24	+0.59 -0.47	0.41
6	19.1 – 19.3	3.61	+2.27 -2.27	+0.22 -0.34	0.11
7	19.3 – 19.5	0.992	+0.610 -0.610	+0.076 -0.064	0.030
8	19.5 – 19.7	0.275	+0.223 -0.223	+0.035 -0.027	0.008
9	19.7 – 20.0	0.0196	+0.0198 -0.0198	+0.0072 -0.0122	0.0006

Table G.2: Shown is the statistical covariance matrix of the flux  $J(E)$  in Table G.1. The diagonal entries in bold type represent the relative uncertainty  $\sigma[J]/J$ . The off-diagonal entries represent the correlation coefficients. The indices  $i$  and  $j$  correspond to the index in Table G.1.

$i \setminus j$	1	2	3	4	5	6	7	8	9
1	<b>0.0073</b>								
2	-0.4030	<b>0.0194</b>							
3	0.1098	-0.3557	<b>0.0246</b>						
4	-0.0171	0.0678	-0.2528	<b>0.0320</b>					
5	-0.0109	0.0232	-0.0186	-0.0600	<b>0.0383</b>				
6	0.0094	-0.0243	0.0451	-0.0879	0.0855	<b>0.0541</b>			
7	-0.0039	0.0095	-0.0205	0.0473	-0.1123	0.2072	<b>0.0758</b>		
8	0.0005	-0.0029	0.0045	-0.0088	0.0308	-0.1217	0.3327	<b>0.1006</b>	
9	0.0007	0.0029	-0.0002	-0.0047	0.0025	0.0082	-0.1112	0.3975	<b>0.2758</b>

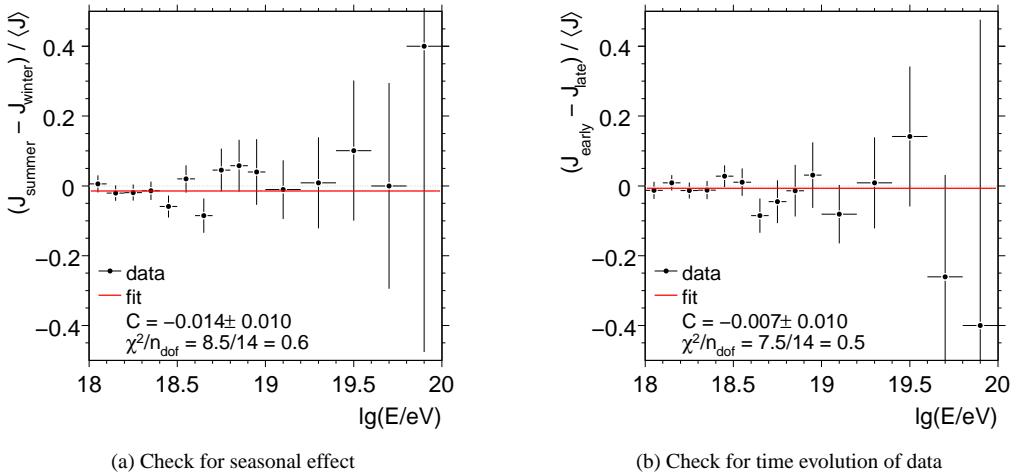


Figure G.1: The points compare the relative difference between pairs of flux estimates, which are described in the text. The solid lines are fits of a constant  $C$  to the data points. The value of  $C$  is given in the legend. The  $\chi^2$  shows the significance of a possible deviation from  $C = 0$ .

# Bibliography

- [1] V. F. Hess, Phys. Z. **13**, 1084 (1912).
- [2] C. D. Anderson, Phys. Rev. **43**, 491 (1933).
- [3] C. D. Anderson and S. H. Neddermeyer, Phys. Rev. **50**, 263 (1936).
- [4] C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini, and C. F. Powell, Nature **159**, 694 (1947).
- [5] M. H. Blewett, Rev. Sci. Instrum. **24**, 725 (1953).
- [6] P. Auger *et al.*, Rev. Mod. Phys. **11**, 288 (1939).
- [7] J. Linsley, Phys. Rev. Lett. **10**, 146 (1963).
- [8] R. M. Baltrusaitis *et al.*, Nucl. Instrum. Meth. **A240**, 410 (1985).
- [9] Lyndon Evans and Philip Bryant (editors), J. Instrum. **3**, S08001 (2008).
- [10] M. Ave, R. A. Vázquez, and E. Zas, Astropart. Phys. **14**, 91 (2000).
- [11] M. Ave *et al.*, Astropart. Phys. **14**, 109 (2000), arXiv astro-ph/0003011.
- [12] P. Billoir, *Testing the Surface Detector of the Engineering Array with Horizontal Showers*, Auger internal note GAP-2000-045 (2000).
- [13] M. Ave *et al.*, Phys. Rev. **D 65**, 63007 (2002).
- [14] P. Billoir, O. Deligny, and A. Letessier-Selvon, *A Complete Procedure for the Reconstruction of Inclined Air Showers*, Auger internal note GAP-2003-003 (2003).
- [15] G. Rodríguez, *Horizontal air showers at the Pierre Auger Observatory*, Ph.D. thesis, Universidade de Santiago de Compostela, Spain, Santiago de Compostela, Spain (2006).
- [16] I. Valiño Rielo, *Detection of horizontal air showers and neutrino induced showers with the Pierre Auger Observatory*, Ph.D. thesis, Universidade de Santiago de Compostela, Santiago de Compostela, Spain (2007).
- [17] V. Olmos-Gilbaja, *A complete method to obtain the energy spectrum of inclined cosmic rays detected with the Pierre Auger Observatory*, Ph.D. thesis, Universidad de Santiago de Compostela, Santiago de Compostela, Spain (2009).
- [18] E. Zas, New J. Phys. **7**, 130 (2005), arXiv astro-ph/0504610.
- [19] J. Abraham *et al.* (Pierre Auger Collab.), *Conventions for the Pierre Auger Observatory*, revised draft, <https://edms.cern.ch/document/317390/1> (unpublished).

- [20] M. Ave *et al.*, Phys. Rev. Lett. **85**, 2244 (2000).
- [21] T. Gaisser, *Cosmic Rays and Particle Physics* (Cambridge University Press, Cambridge, 1990).
- [22] T. Stanev, *High Energy Cosmic Rays* (Springer Verlag, Berlin, 2003).
- [23] C. Amsler *et al.* (Particle Data Group), Physics Letters **B 667**, 1 (2008).
- [24] J. Blümner, R. Engel, and J. R. Hörandel, *Cosmic Rays from the Knee to the Highest Energies*, arXiv astro-ph/0904.0725 (unpublished).
- [25] N. L. Grigorov *et al.*, Yad. Fiz **11**, 1058 (1970).
- [26] T. H. Burnett *et al.*, Phys. Rev. Lett. **51**, 1010 (1983).
- [27] J. A. Simpson, Ann. Rev. Nucl. Part. Sci. **33**, 323 (1983).
- [28] V. Berezinsky, *Transition from galactic to extragalactic cosmic rays*, Proc. 30th Int. Cosmic Ray Conf. (2007), arXiv astro-ph/0710.2750.
- [29] M. Unger *et al.* (Pierre Auger Collab.), *Study of the Cosmic Ray Composition above 0.4 EeV using the Longitudinal Profiles of Showers observed at the Pierre Auger Observatory*, Proc. 30th Int. Cosmic Ray Conf. (2007).
- [30] F. Schüssler *et al.* (Pierre Auger Collab.), *Measurement of the cosmic ray energy spectrum above  $10^{18}$  eV using the Pierre Auger Observatory*, Proc. 31st Int. Cosmic Ray Conf. (2009).
- [31] K. Greisen, Phys. Rev. Lett. **16**, 748 (1966).
- [32] G. T. Zatsepin and V. A. Kuz'min, Sov. J. Exp. Theor. Phys. Lett. **4**, 78 (1966).
- [33] A. M. Hillas, Ann. Rev. Astron. Astrophys. **22**, 425 (1984).
- [34] J. Abraham *et al.* (Pierre Auger Collab.), Astropart. Phys. **27**, 155 (2007).
- [35] J. Abraham *et al.* (Pierre Auger Collab.), Astroparticle Physics **29**, 243 (2008).
- [36] J. Abraham *et al.* (Pierre Auger Collab.), Phys. Rev. Lett. **100**, 211101 (2008), arXiv astro-ph/0712.1909.
- [37] J. Abraham *et al.* (Pierre Auger Collab.), Science **318**, 938 (2007).
- [38] J. D. Hague *et al.* (Pierre Auger Collab.), *Correlation of the Highest Energy Cosmic Rays with Nearby Extragalactic Objects in Pierre Auger Observatory Data*, Proc. 31st Int. Cosm. Ray Conf. (2009).
- [39] D. Allard *et al.*, J. Cosmol. Astropart. Phys. **2006**, 005 (2006), arXiv astro-ph/0605327.
- [40] M. Véron-Cetty and P. Véron, Astron. & Astrophys. **455**, 773 (2006).
- [41] B. Keilhauer, *Investigation of Atmospheric Effects on the Development of Extensive Air Showers and their Detection with the Pierre Auger Observatory*, Ph.D. thesis, Karlsruhe University, Karlsruhe, Germany (2003).
- [42] B. Keilhauer *et al.*, Astropart. Phys. **22**, 249 (2004).
- [43] B. Keilhauer, B. Wilczyńska, and H. Wilczyński, *Molecular Atmosphere Profiles for Malargüe*, Auger internal note GAP-2005-021 (2005).
- [44] M. Lemoine and G. Sigl, *Physics and Astrophysics of Ultra-High-Energy Cosmic Rays* (Springer Verlag, Berlin, 2001).
- [45] J. Matthews, Astropart. Phys. **22**, 387 (2005).

## BIBLIOGRAPHY

---

- [46] W. Heitler, *The Quantum Theory of Radiation*, third ed. ed. (Oxford University Press, Oxford, 1954).
- [47] H. M. Barbosa, F. Catalani, J. A. Chinellato, and C. Dobrigkeit, Astropart. Phys. **22**, 159 (2004).
- [48] J. Engel, T. K. Gaisser, P. Lipari, and T. Stanev, Phys. Rev. D **46**, 5013 (1992).
- [49] D. Heck *et al.*, *CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers*, Report FZKA 6019, Karlsruhe (1998).
- [50] T. Pierog *et al.*, *Latest Results of Air Shower Simulation Programs CORSIKA and CONEX*, Proc. 30th Int. Cosmic Ray Conf., Vol. **4**, 625–628 (2007), arXiv astro-ph/0802.1262.
- [51] G. Molière, Z. Naturforsch. A **2**, 133 (1947).
- [52] G. Molière, Z. Naturforsch. A **3**, 78 (1948).
- [53] B. Rossi, *High Energy Particles* (Prentice-Hall, Englewood Cliffs, NJ, 1952).
- [54] M. Ave, Ph.D. thesis, Universidade de Santiago de Compostela, Santiago de Compostela, Spain (2004).
- [55] X. Bertou and P. Billoir, *On the origin of the asymmetry of ground densities in inclined showers*, Auger internal note GAP-2000-017 (2000).
- [56] M. Ave, (2009), private communication.
- [57] L. Cazón, *Modelling the Muon Time Distribution In Extensive Air Showers*, Ph.D. thesis, Universidade de Santiago de Compostela, Santiago de Compostela, Spain (2004).
- [58] L. Cazón, R. A. Vázquez, A. A. Watson, and E. Zas, Astropart. Phys. **21**, 71 (2004).
- [59] H.-J. Drescher and G. Farrar, Astropart. Phys. **19**, 235 (2003).
- [60] J. L. Harton *et al.* (Pierre Auger Collab.), *Progress with the Northern Part of the Pierre Auger Observatory*, Proc. 31st Int. Cosmic Ray Conf. (2009).
- [61] I. Allekotte *et al.* (Pierre Auger Collab.), Nucl. Instrum. Meth. **A586**, 409 (2007), arXiv astro-ph/0712.2832.
- [62] J. Abraham *et al.* (Pierre Auger Collab.), Nucl. Instrum. Meth. **A523**, 50 (2004).
- [63] M. A. Lawrence, R. J. O. Reid, and A. A. Watson, J. Phys. G: Nucl. Part. Phys. **17**, 733 (1991).
- [64] C. L. Pryke and J. Lloyd-Evans, Nucl. Instrum. Meth. **A354**, 560 (1995).
- [65] X. Bertou *et al.*, Nucl. Instrum. Meth. **A568**, 839–846 (2006).
- [66] M. Aglietta *et al.* (Pierre Auger Collab.), *Response of the Pierre Auger Observatory Water Cherenkov Detectors to Muons*, Proc. 29th Int. Cosmic Ray Conf., Vol. **7**, 83-+ (2005).
- [67] P. Ghia *et al.* (Pierre Auger Collab.), *Testing the surface detector simulation for the Pierre Auger Observatory*, Proc. 30th Int. Cosmic Ray Conf. (2007).
- [68] J. Abraham *et al.* (Pierre Auger Collab.), *The aperture of the Pierre Auger Observatory surface detector for extensive air showers below 60°: from the trigger system to exposure calculation*, in preparation (unpublished).
- [69] X. Bertou, (2007), private communication.
- [70] J. Řídký *et al.*, *The Fluorescence Detector of the Pierre Auger Observatory*, to appear in Nucl. Instrum. Meth. (unpublished).
- [71] M. Ave *et al.* (AIRFLY Collab.), Nucl. Instrum. Meth. **A597**, 50 (2008), arXiv astro-ph/0711.4583.

- [72] M. Ave et al. (AIRFLY Collab.), *Astropart. Phys.* **28**, 41 (2007), arXiv astro-ph/0703132.
- [73] M. Nagano, K. Kobayakawa, N. Sakaki, and K. Ando, *Astroparticle Physics* **22**, 235 (2004), arXiv astro-ph/0406474.
- [74] R. Cester *et al.* (Pierre Auger Collab.), *Atmospheric aerosol monitoring at the Pierre Auger Observatory*, Proc. 29th Int. Cosmic Ray Conf. (2005).
- [75] M. Prouza et al. (Pierre Auger Collab.), *Systematic study of atmosphere-induced influences and uncertainties on shower reconstruction at the Pierre Auger Observatory*, Proc. 30th Int. Cosmic Ray Conf. (2007).
- [76] S. Y. BenZvi *et al.* (Pierre Auger Collab.), *Nucl. Instrum. Meth.* **A574**, 171 (2007), arXiv astro-ph/0609063.
- [77] B. Fick *et al.*, *J. Instrum.* **1**, P11003 (2006).
- [78] S. BenZvi *et al.*, *Astropart. Phys.* **28**, 312 (2007).
- [79] A. BenZvi *et al.* (Pierre Auger Collab.), *New method for atmospheric calibration at the Pierre Auger Observatory using FRAM, a robotic astronomical telescope*, Proc. 30th Int. Cosmic Ray Conf. (2007).
- [80] A. Schmidt *et al.*, *Nucl. Instrum. Meth.* **A601**, 347 (2009).
- [81] M. Kleifges *et al.* (Pierre Auger Collab.), *Extension of the Pierre Auger Observatory using high elevation fluorescence telescopes (HEAT)*, Proc. 31st Int. Cosmic Ray Conf. (2009).
- [82] M. Platino *et al.* (Pierre Auger Collab.), *AMIGA - Auger Muons and Infill for the Ground Array of the Pierre Auger Observatory*, Proc. 31st Int. Cosmic Ray Conf. (2009).
- [83] A. M. van den Berg *et al.* (Pierre Auger Collab.), *Radio detection of cosmic rays at the southern Auger Observatory*, Proc. 31st Int. Cosmic Ray Conf. (2009).
- [84] H. Dembinski, P. Billoir, O. Deligny, and T. Hebbeker, *Inferring Average Ground Profiles of the Muon Density of Inclined Air Showers from Monte-Carlo Simulations at Ultra-High Energy*, paper in preprint, arXiv astro-ph/0904.2372 (unpublished).
- [85] L. G. Dedenko, *A new method of solving the nuclear cascade equation*, Proc. 9th Int. Cosmic Ray Conf., Vol. **1**, 662 (1965).
- [86] A. M. Hillas, *Calculations on the propagation of mesons in extensive air showers*, Proc. 9th Int. Cosmic Ray Conf. (1965).
- [87] T. Bergmann *et al.*, *Astropart. Phys.* **26**, 420 (2007), astro-ph/0606564.
- [88] S. J. Sciutto, *The AIRES system for air shower simulations. An update*, Proc. 27th Int. Cosmic Ray Conf. (2001), arXiv astro-ph/0106044v1.
- [89] D. Heck and J. Knapp, *Upgrade of the Monte Carlo Code CORSIKA to Simulate Extensive Air Showers with Energies > 10<sup>20</sup> eV*, Report FZKA 6097B, Karlsruhe (1998).
- [90] H.-J. Drescher and G. R. Farrar, *Phys. Rev. D* **67**, 116001 (2003), astro-ph/0212018.
- [91] T. Pierog, presentation in collaboration meeting, Malargüe, Argentina, Nov. 2008.
- [92] S. P. Swordy *et al.*, *Astropart. Phys.* **18**, 129 (2002), arXiv astro-ph/0202159.
- [93] J. Knapp *et al.*, *Astropart. Phys.* **19**, 77 (2003), arXiv astro-ph/0206414.
- [94] J. Knapp, *High-energy interactions and extensive air showers*, Proc. 25th Int. Cosmic Ray Conf. (1997), arXiv astro-ph/9710277.

## BIBLIOGRAPHY

---

- [95] W. Nelson, H. Hirayama, and D. Rogers, Technical report, Stanford Linear Accelerator Center, Stanford, California (1985).
- [96] A. Migdal, Phys. Rev. **103**, 1811 (1956).
- [97] E. Konishi, A. Adachi, N. Takahashi, and A. Misaki, J. Phys. G: Nucl. Part. Phys. **17**, 719 (1991).
- [98] D. Heck, Nucl. Phys. B Proc. Suppl. **151**, 127 (2006), arXiv astro-ph/0410735.
- [99] H. Fesefeldt, *Report PITHA-85/02*, RWTH Aachen (1985).
- [100] S. A. Bass *et al.*, Progress in Particle and Nuclear Physics **41**, 255 (1998).
- [101] M. Bleicher *et al.*, J. Phys. G: Nucl. Part. Phys. **25**, 1859 (1999).
- [102] A. Fassó *et al.*, *The physics models of FLUKA: status and recent developments*, Computing in High Energy and Nuclear Physics 2003 Conference (CHEP2003), La Jolla, CA, USA, March 24-28, 2003, (paper MOMT005), eConf C0303241 (2003), arXiv hep-ph/0306267 (2003).
- [103] A. Fassó, A. Ferrari, J. Ranft, and P. Sala, *FLUKA: a multi-particle transport code*, CERN-2005-10 (2005), INFN/TC 05/11, SLAC-R-773 (2005).
- [104] C. Meurer, *Muon production in extensive air showers and fixed target accelerator data*, Ph.D. thesis, Karlsruhe University, Karlsruhe, Germany (2007).
- [105] R. Engel, T. K. Gaisser, P. Lipari, and T. Stanev, *Air shower calculations with the new version of SIBYLL*, Proc. 26th Int. Cosmic Ray Conf. (1999).
- [106] T. K. Gaisser and F. Halzen, Phys. Rev. Lett. **54**, 1754 (1985).
- [107] L. Durand and P. Hong, Phys. Rev. Lett. **58**, 303 (1987).
- [108] J. C. Collins and D. E. Soper, Ann. Rev. Nucl. Part. Sci. **37**, 383 (1987).
- [109] R. S. Fletcher, T. K. Gaisser, P. Lipari, and T. Stanev, Phys. Rev. D **50**, 5710 (1994).
- [110] V. N. Gribov, Sov. Phys. JETP **26**, 414 (1968).
- [111] V. N. Gribov, Sov. Phys. JETP **26**, 483 (1969).
- [112] K. A. Ter-Martirosyan, Phys. Lett. B **44**, 179 (1973).
- [113] K. A. Ter-Martirosyan, Phys. Lett. B **44**, 377 (1973).
- [114] M. Baker and K. A. Ter-Martirosyan, Physics Reports **28**, 1 (1976).
- [115] N. N. Kalmykov and S. Ostapchenko, Yad. Fiz. **56**, 105 (1993).
- [116] N. N. Kalmykov and S. S. Ostapchenko, Phys. At. Nucl. **56**, 346 (1993).
- [117] N. N. Kalmykov, S. S. Ostapchenko, and A. I. Pavlov, Bull. Russ. Acad. Sci. Phys. **58**, 1966 (1994).
- [118] N. N. Kalmykov and S. S. Ostapchenko, Yad. Fiz. **58**, 21 (1994).
- [119] N. N. Kalmykov, S. S. Ostapchenko, and A. I. Pavlov, Izv. RAN Ser. Fiz. **58**, 21 (1994).
- [120] N. N. Kalmykov, S. S. Ostapchenko, and A. I. Pavlov, Nucl. Phys. B Proc. Suppl. **52B**, 17 (1997).
- [121] J. R. Hörandel, *Untersuchung der hadronischen Komponente ausgedehnter Luftschauder und Messung der Elementzusammensetzung der kosmischen Strahlung*, Ph.D. thesis, Karlsruhe University, Karlsruhe, Germany (1997) available as Report FZKA 6015.
- [122] Hörandel, J. R. et al., *Estimation of the Cosmic Ray Composition with the KASCADE Hadron Calorimeter*, Proc. 25th Int. Cosmic Ray Conf., Vol. **6**, 93-+ (1997).

- [123] A. Capella and A. Krzywicki, Phys. Rev. D **18**, 3357 (1978).
- [124] A. Capella and J. Tran Than Van, Z. Phys. C **10**, 249 (1981).
- [125] R. J. Glauber and G. Matthiae, Nucl. Phys. **B21**, 135 (1970).
- [126] S. S. Ostapchenko, Nucl. Phys. B Proc. Suppl. **151**, 143 (2006), hep-ph/0412332.
- [127] S. S. Ostapchenko, Phys. Lett. B **636**, 40 (2006), hep-ph/0412332.
- [128] S. S. Ostapchenko, Phys. Rev. D **74**, 014026 (2006), hep-ph/0505259.
- [129] G. Bossard *et al.*, Phys. Rev. D **63**, 054030 (2001), arXiv hep-ph/0009119.
- [130] K. Werner, F.-M. Liu, and T. Pierog, Phys. Rev. C: Nucl. Phys. **74**, 11 (2006), <http://link.aps.org/abstract/PRC/v74/e044902>.
- [131] K. Werner and T. Pierog, AIP Conf. Proc. **928**, 111 (2007), arXiv astro-ph/0707.3330.
- [132] M. Risse, D. Heck, and J. Knapp, *The CORSIKA Shower Library in Lyon after One Year of Production: Status and Experiences*, Auger internal note GAP-2002-011 (2002).
- [133] R. Bruijn and J. Knapp, *An unthinned vertical proton shower at  $10^{19}$  eV: simulation details and access to data*, Auger internal note GAP-2009-011 (2009).
- [134] A. M. Hillas, *Two interesting techniques for Monte-Carlo simulation of very high energy hadron cascades*, Proc. 17th Int. Cosmic Ray Conf. (1981).
- [135] A. M. Hillas, Nucl. Phys. B Proc. Suppl. **52B**, 29 (1997).
- [136] M. Kobal, Astropart. Phys. **15**, 259 (2001).
- [137] S. Argiro *et al.*, Nucl. Instrum. Meth. **A580**, 1485 (2007).
- [138] J. Allison *et al.*, IEEE T. Nucl. Sci. **53**, 270 (2006).
- [139] P. Skelton, *Speeding up the Geant 4 Tank Simulator*, Auger internal note GAP-2004-066 (2004).
- [140] P. Billoir, Astropart. Phys. **30**, 270 (2008).
- [141] H. Dembinski and M. Leuthold, *The Inclined CORSIKA Proton Shower Library*, Auger internal note GAP-2006-094 (2006).
- [142] Wiki-page: Auger Monte-Carlo productions, <https://www.auger.unam.mx/AugerWiki/MonteCarloProduction>.
- [143] Ohio Supercomputer Center (Columbus, Ohio), <http://www.osc.edu>.
- [144] Supercomputer center CC-IN2P3 (Lyon, France), <http://cc.in2p3.fr>.
- [145] N. Olsen, T. Sabaka, and L. Toffner-Clausen, EARTH PLANETS AND SPACE **52**, 1175 (2000), <http://www.ngdc.noaa.gov/geomag>.
- [146] National Aeronautics and Space Administration (NASA), *U.S. Standard Atmosphere 1976*, Report NASA-TM-X-74335 (1976).
- [147] F. James, *Statistical Methods in Experimental Physics*, 2nd edition ed. (World Scientific Publishing, Singapore, 2006).
- [148] F. James, *Minuit Reference Manual*, CERN Program Library Office CERN-IT Division CH-1211 Geneva 23 Switzerland, 1998.
- [149] Particle Data Group, Phys. Lett. B **592**, 275 (2004), review of Particle Physics.

## BIBLIOGRAPHY

---

- [150] G. Rodríguez *et al.*, *Simulation of the Surface Detector response for inclined shower reconstruction*, Auger internal note GAP-2009-049 (2009).
- [151] I. Valiño *et al.* (Pierre Auger Collab.), *The electromagnetic component of inclined air showers at the Pierre Auger Observatory*, Proc. 31st Int. Cosmic Ray Conf. (2009).
- [152] K. Kamata and J. Nishimura, Prog. Theoret. Phys. Suppl. **6**, 93 (1958).
- [153] Greisen, in *Progress in elementary particle and cosmic ray physics*, edited by J. G. Wilson and S. A. Wouthuysen (North-Holland Publishing, Amsterdam, 1956), Vol. 3.
- [154] M. Ave *et al.*, Nucl. Instrum. Meth. **A578**, 180 (2007).
- [155] I. C. Mariş *et al.*, *Comparison of Two Methods to Infer the Cosmic Ray Spectrum at Highest Energies*, Auger internal note GAP-2007-044 (2007).
- [156] I. C. Mariş *et al.*, *Reducing the influence of the threshold events on the surface detector energy calibration*, Auger internal note GAP-2007-116 (2007).
- [157] I. C. Mariş, *Measurement of the Ultra High Energy Cosmic Ray Flux using Data of the Pierre Auger Observatory*, Ph.D. thesis, Universität Karlsruhe, Karlsruhe, Germany (2008).
- [158] I. C. Mariş, *CIC/energy calibration/spectrum*, Pierre Auger Collaboration Meeting, Malargüe (2008).
- [159] M. Ave *et al.*, Phys. Rev. **D 67**, 043005 (2003).
- [160] J. Alvarez-Muñiz, G. Rodríguez-Fernández, I. Valiño, and E. Zas, *An Alternative Method for Tank Signal Response and S(1000) Calculation*, Auger internal note GAP-2005-054 (2005).
- [161] N. Busca *et al.*, *An estimate of the cosmic ray spectrum using inclined data of the Pierre Auger Observatory*, Auger internal note GAP-2006-025 (2006).
- [162] P. Facal San Luis *et al.*, *Cosmic ray spectrum using inclined showers: Update presented in September 2006*, Auger internal note GAP-2006-107 (2006).
- [163] P. Facal San Luis *et al.*, *Cosmic ray spectrum with inclined showers: November 2006 update*, Auger internal note GAP-2007-027 (2007).
- [164] P. Facal San Luis *et al.* (Pierre Auger Collab.), *Measurement of the UHECR spectrum above  $10^{19}$  eV at the Pierre Auger Observatory using showers with zenith angles greater than  $60^\circ$* , Proc. 30th Int. Cosmic Ray Conf. (2007), arXiv astro-ph/0706.4322v1.
- [165] D. Newton *et al.* (Pierre Auger Collab.), *Selection and reconstruction of very inclined air showers with the Surface Detector of the Pierre Auger Observatory*, Proc. 30th Int. Cosmic Ray Conf. (2007).
- [166] H. Dembinski *et al.*, *Reconstruction of very inclined air showers with Offline*, Auger internal note GAP-2009-008 (2009).
- [167] Offline-HAS software, revision 8254, Subversion repository <https://svn.auger.unam.mx/svn/DPA/PhysicsModules/HASOffline/tags/trunk>.
- [168] D. Veberič and M. Roth, *Offline Reference Manual: SD Reconstruction*, Auger internal note GAP-2005-035 (2005), an updated version is shipped with the Offline software distribution.
- [169] M. Horvat and D. Veberič, *On Shower-Front Start-Time Variance*, Auger internal note GAP-2007-057 (2007).
- [170] I. C. Mariş and M. Roth, *65 Stations in One Shot. Is this lightning?*, Auger internal note GAP-2006-079 (2006).

- [171] P. Billoir, *FADC trace cleaning in Surface Detector through a segmentation procedure*, Auger internal note GAP-2005-074 (2005).
- [172] Event Selection Group, *The Official SD Event Selection as of March 2005*, Auger internal note GAP-2005-023 (2005).
- [173] P. Billoir, *Top-down Selection of Events and Stations in Surface Detector Triggers*, Auger internal note GAP-2006-072 (2006).
- [174] O. B. Bigas and J. Gonzalez, *Aiming for an official T4 above 60 degrees*, Auger internal note GAP-2007-028 (2007).
- [175] C. Bonifazi and A. Letessier-Selvon, *Angular Resolution of the Auger Surface Detector*, Auger internal note GAP-2006-016 (2006).
- [176] E. Patrizot *et al.*, *First steps towards the definition of a "quality trigger" (T5) for the SD acceptance calculations*, Auger internal note GAP-2004-023 (2004).
- [177] D. Allard *et al.*, *Aperture calculation of the Pierre Auger Observatory surface detector*, Proc. 29th Int. Cosmic Ray Conf. (2005).
- [178] D. Allard *et al.*, *The trigger system of the Pierre Auger Surface Detector: operation, efficiency and stability*, Proc. 29th Int. Cosmic Ray Conf. (2005).
- [179] I. Allekotte, H. Asorey, X. Bertou, and M. Gómez Berisso, *You thought you understood hexagons?*, Auger internal note GAP-2008-114 (2008).
- [180] P. Summers and B. Fick, *A Numerical Recipe for Fluorescence Detector Geometric Reconstruction*, Auger internal note GAP-2003-045 (2003).
- [181] M. Unger *et al.*, Nucl. Instrum. Meth. **A588**, 433 (2008), arXiv astro-ph/0801.4309.
- [182] J. Bellido, *Alignment and Synchronization of Bay4 at Los Leones and Bay3 at Coihueco using the CLF*, Auger internal note GAP-2004-059 (2004).
- [183] C. Bonifazi *et al.* (Pierre Auger Collab.), Nucl. Phys. Proc. Suppl. **190**, 20 (2009), arXiv astro-ph/0901.3138.
- [184] T. K. Gaisser and A. M. Hillas, *Reliability of the Method of Constant Intensity Cuts for Reconstructing the Average Development of Vertical Showers*, Proc. 15th Int. Cosm. Ray Conf., Vol. **8**, 353 (1977).
- [185] B. Dawson *et al.* (Pierre Auger Collab.), *Hybrid Performance of the Pierre Auger Observatory*, Proc. 30th Int. Cosmic Ray Conf. (2007).
- [186] C. Di Giulio, P. Facal San Luis, G. Rodriguez Fernández, and V. Verzi, *Estimation of the shower energy and  $X_{max}$  with the spot reconstruction method*, Auger internal note GAP-2009-040 (2009).
- [187] N. G. Busca, *The ultra high energy cosmic ray flux from the Southern Pierre Auger Observatory data*, Auger internal note GAP-2006-108 (2006).
- [188] J. Abraham *et al.* (Pierre Auger Collab.), Phys. Rev. Lett. **101**, 061101 (2008), arXiv astro-ph/0806.4302.
- [189] J. Brack, (2007), private communication.
- [190] B. Kégl and D. Veberič, *The "blunt" cut:  $S_{38}$  vs.  $E_{FD}$  calibration using a piecewise linear/power law fit*, (May, 2008), presentation in analysis meeting, Lamar, Colorado.
- [191] B. Kégl and D. Veberič, (Apr, 2009), private communication.

## BIBLIOGRAPHY

---

- [192] V. Verzi *et al.* (Pierre Auger Collab.), Nucl. Phys. B Proc. Suppl. **165**, 37 (2007).
- [193] M. Ave, P. Bauleo, and T. Yamamoto, *Signal Fluctuations in the Auger Surface Detector Array*, Auger internal note GAP-2003-030 (2003).
- [194] M. Unger, R. Engel, F. Schüssler, and R. Ulrich, *Measurement of Xmax as a function of Energy*, Auger internal note GAP-2007-005 (2007).
- [195] E. Cantoni *et al.* (KASCADE Collab.), *The sensitivity of KASCADE-Grande to the cosmic ray primary composition between  $10^{16}$  eV and  $10^{18}$  eV*, Proc. 31st Int. Cosmic Ray Conf. (2009).
- [196] M. Ave *et al.*, *Determining the SD Energy Scale and Number of Muons Using Air Shower Universality*, Auger internal note GAP-2007-021 (2007).
- [197] M. Ave *et al.*, *Monte Carlo and Hybrid Studies of Shower-to-Shower Fluctuations*, Auger internal note GAP-2007-052 (2007).
- [198] M. Ave and F. Schmidt, *Update on the Number of Muons in Auger Data*, Auger internal note GAP-2008-130 (2008).
- [199] C. Bonifazi and P. L. Ghia, *Selection of data periods and calculation of the SD geometrical acceptance*, Auger internal note GAP-2006-101 (2006).
- [200] C. Bonifazi and A. Letessier-Selvon, *Event selection using the T5 time distribution*, Auger internal note GAP-2006-042 (2006).
- [201] V. Blobel, *Unfolding methods in high energy physics experiments*, Proceedings of the 1984 CERN School of Computing (1984).
- [202] V. Blobel and E. Lohrmann, *Statistische und numerische Methoden der Datenanalyse* (Teubner Verlag, Wiesbaden, Germany, 1998).
- [203] H. Geenen, *Reconstruction of the Primary Energy Spectrum from Fluorescence Telescope Data of the Pierre Auger Observatory*, Ph.D. thesis, Universität Wuppertal, Wuppertal Germany (2007).
- [204] R. Gold, Technical Report No. ANL-6984, Argonne National Laboratory, Argonne, Illinois (1964).
- [205] G. D'Agostini, Nucl. Instrum. Meth. **A362**, 487 (1992).
- [206] A. Hoecker and V. Kartvelishvili, Nucl. Instrum. Meth. **A372**, 469 (1996), arXiv hep-ph/9509307v2.
- [207] G. D'Agostini, Technical Report No. Yellow Report 99-03, CERN, CERN, Switzerland (1999).
- [208] A. K. Louis, *Inverse und schlecht gestellte Probleme* (Teubner Verlag, Stuttgart und Leipzig, 1989).
- [209] V. Blobel, Technical Report No. TN361, OPAL, DESY, Germany (1996).
- [210] V. Blobel, reference implementation of the RUN-algorithm, available at <http://www.desy.de/~blobel>.
- [211] C. de Boor, *A Practical Guide to Splines* (Springer Verlag, New York, Heidelberg, Berlin, 1978).
- [212] J. Stoer and R. Burlirsch, *Introduction to Numerical Analysis* (Springer Verlag, New York, Heidelberg, Berlin, 1980).
- [213] M. Takeda *et al.*, Phys. Rev. Lett. **81**, 1163 (1998).
- [214] Hanlon, W. F. and Jui, C. C. H. and Sokolsky, P. V. and Cao, Z. and Thomson, G. B. for the High Resolution Fly's Eye Collaboration, *Stereoscopic Measurement of the Flux of Ultra High Energy Cosmic Rays by the High Resolution Fly's Eye*, Proc. 31st Int. Cosmic Ray Conf. (2009).

---

## BIBLIOGRAPHY

- [215] M. I. Pravdin *et al.*, *The cosmic rays energy spectrum of the Yakutsk EAS Array*, Proc. 31st Int. Cosmic Ray Conf. (2009).
- [216] Wolfram Research, Inc., *Mathematica Edition: Version 7.0* (Wolfram Research, Inc., Champaign, Illinois, 2008).
- [217] D. Supanitsky and X. Bertou, *Semi-Analytical Model of the Three Fold Charge Spectrum in a Water Cherenkov Tank*, Auger internal note GAP-2003-113 (2003).
- [218] P. Billoir and O. Deligny, (2008), private communication.
- [219] J. W. Cooley and J. W. Tukey, *Math. Comput.* **19**, 297–301 (1965).
- [220] The Auger Observer, ADST software and official event reconstruction, <http://augerobserver.fzk.de/>.