



BERGISCHE UNIVERSITÄT WUPPERTAL  
FACHBEREICH C – MATHEMATIK UND NATURWISSENSCHAFTEN

**KASCADE-Grande Measurements  
of Energy Spectra for  
Elemental Groups of Cosmic Rays**

*KASCADE-Grande Messungen  
der Energiespektren einzelner  
Elementgruppen der kosmischen Strahlung*

**Inauguraldissertation**

zur Erlangung des akademischen Grades eines  
Doktors der Naturwissenschaften (Dr. rer. nat.),

vorgelegt dem  
Fachbereich C – Mathematik und Naturwissenschaften  
der Bergischen Universität Wuppertal

von  
**Dipl.-Phys. Daniel Fuhrmann**  
aus Wermelskirchen.

März 2012

Referent: Prof. Dr. Karl-Heinz Kampert, Fachbereich C, Universität Wuppertal  
Korreferent: Prof. Dr. Markus Risse, Department für Physik, Universität Siegen

Die Dissertation kann wie folgt zitiert werden:

urn:nbn:de:hbz:468-20120515-141648-9

[<http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3Ahbz%3A468-20120515-141648-9>]

## Abstract

In 1958, 46 years after the discovery of cosmic rays by Hess, Kulikov and Khristiansen measured the “knee” of the cosmic ray spectrum, which is associated with a change of the index of the power law. Many decades passed, before in 2003 the KASCADE experiment clarified *post hoc* that this knee is caused by a change in the spectra of light cosmic ray mass groups. The knowledge about the composition of cosmic rays and the existence of specific structures in the energy spectrum around the knee is of fundamental importance for understanding things like the transition from galactic to extragalactic origin of cosmic rays, or the cosmic ray acceleration and propagation in general.

The KASCADE-Grande experiment, an extension of the original KASCADE experiment, is dedicated to measure the energy spectra for elemental groups of cosmic rays at one or two orders of magnitude higher energies than its predecessor, namely in the energy range from  $10^{16}$  eV to  $10^{18}$  eV. It thereby enables the verification of a possible second knee expected at approximately  $10^{17}$  eV.

This work’s primary objective is to determine the energy spectra of different cosmic ray mass groups by means of unfolding techniques applied to the KASCADE-Grande dataset. This analysis yields strong indications for a knee-like structure in the spectrum of the heavy component of cosmic rays (represented by iron) at about  $8 \times 10^{16}$  eV. For the first time, it is demonstrated experimentally that the sought-after second knee exists and is related to a change in the spectra of heavy cosmic ray mass groups. This insight contributes considerably to our understanding of cosmic ray physics.

## Kurzzusammenfassung

Im Jahre 1958, 46 Jahre nach der Entdeckung der kosmischen Strahlung durch Hess, detektierten Kulikov und Khristiansen das sogenannte “Knie” im Spektrum der kosmischen Strahlung, welches mit einer Veränderung im Index des Potenzgesetzes verknüpft ist. Viele Jahrzehnte vergingen bis im Jahre 2003 durch das KASCADE Experiment *post hoc* klargestellt werden konnte, dass dieses Knie durch eine Veränderung in den Spektren der leichten Massengruppen verursacht wird. Die Kenntnis von der Komposition der kosmischen Strahlung sowie der Existenz spezifischer Strukturen im Energiespektrum im Bereich des Knies ist von fundamentaler Bedeutung für das Verständnis von Dingen wie zum Beispiel dem Übergang von kosmischer Strahlung galaktischen zu extragalaktischen Ursprungs, oder deren Beschleunigung und Propagation im Allgemeinen.

Das KASCADE-Grande Experiment, eine Erweiterung des ursprünglichen KASCADE Experiments, ist der Messung der Energiespektren einzelner Elementgruppen der kosmischen Strahlung, im Bereich einer oder zweier Größenordnungen höherer Energien als dessen Vorgänger, gewidmet. Es werden dabei Energien von  $10^{16}$  eV bis  $10^{18}$  eV abgedeckt. Das Experiment ermöglicht es dadurch die Existenz eines vermuteten zweiten Knies im Bereich von  $10^{17}$  eV zu verifizieren.

Die Bestimmung der Energiespektren verschiedener Massengruppen der kosmischen Strahlung durch Anwendung von Entfaltungsmethoden auf den KASCADE-Grande Messdatensatz ist das Hauptziel dieser Dissertation. Diese Analyse ergibt deutliche Hinweise auf die Existenz einer knieähnlichen Struktur im Spektrum der schweren Massengruppe (repräsentiert durch Eisen) bei Energien von ungefähr  $8 \times 10^{16}$  eV. Zum ersten Mal wurde experimentell nachgewiesen, dass ein zweites Knie existiert und mit einer Änderung im Spektrum der schweren Massengruppen verknüpft ist. Diese Erkenntnis trägt maßgeblich zu unserem Verständnis von der Physik der kosmischen Strahlung bei.



# Contents

List of Figures	V
List of Tables	IX
<b>1 Introduction and Motivation</b>	<b>1</b>
<b>2 Cosmic Rays and Extensive Air Showers</b>	<b>5</b>
2.1 Origin and Acceleration of Cosmic Rays . . . . .	6
2.2 The Knees of the Cosmic Ray Spectrum . . . . .	7
2.3 Transition from Galactic to Extragalactic Cosmic Rays . . . . .	9
2.4 Extensive Air Showers . . . . .	13
<b>3 The KASCADE-Grande Experiment</b>	<b>17</b>
3.1 Setup of the KASCADE-Grande Experiment . . . . .	17
3.1.1 KASCADE detector stations . . . . .	18
3.1.2 Grande detector stations . . . . .	19
3.2 Air Shower Reconstruction with KASCADE-Grande . . . . .	20
3.2.1 Charged particle number reconstruction . . . . .	20
3.2.2 Muon number reconstruction . . . . .	22
3.3 Quality Cuts . . . . .	23
3.4 Quality of the Shower Reconstruction and Correction Methods . . . . .	26
<b>4 Preliminaries for Unfolding the Energy Spectra</b>	<b>29</b>
4.1 Choice of Observables . . . . .	29
4.2 Outline of the Deconvolution Analysis . . . . .	37
<b>5 The Response Matrix</b>	<b>41</b>
5.1 Investigations to the Intrinsic Shower Fluctuations. . . . .	41
5.1.1 Simulation of the air shower development . . . . .	41
5.1.2 Parametrization of the intrinsic shower fluctuations . . . . .	45
5.2 Investigations to the Experiment's Properties . . . . .	59
5.2.1 Simulation of the experiment's properties . . . . .	59
5.2.2 The experiment's efficiency . . . . .	60
5.2.3 The systematic reconstruction uncertainties . . . . .	66
5.2.4 The statistical reconstruction uncertainties . . . . .	70
5.3 Computation and Test of the Response Matrix . . . . .	79
5.3.1 Computation of the response matrix . . . . .	79
5.3.2 Conditioning of the response matrix . . . . .	81
5.3.3 Consistency checks for the response matrix . . . . .	86

<b>6</b>	<b>Unfolding Methods and Monte Carlo Tests</b>	<b>93</b>
6.1	Why Unfolding Algorithms? . . . . .	93
6.2	Unfolding Algorithms . . . . .	95
6.2.1	Regularized unfolding . . . . .	95
6.2.2	Iterative unfolding . . . . .	98
6.3	Implementation and Quality of the Unfolding Algorithms . . . . .	105
6.3.1	Definition of the “iteration depth” . . . . .	105
6.3.2	Generation of toy datasets . . . . .	107
6.3.3	Error propagation . . . . .	110
6.3.4	Stopping criterion for Gold’s unfolding algorithm . . . . .	131
6.3.5	Quality of Gold’s unfolding algorithm . . . . .	137
<b>7</b>	<b>Unfolding the Energy Spectra for Elemental Groups of Cosmic Rays</b>	<b>149</b>
7.1	Some Preparations for the Unfolding of the Measured Data . . . . .	149
7.1.1	The optimal iteration depth . . . . .	149
7.1.2	Error propagation . . . . .	151
7.2	Unfolding of the KASCADE-Grande Dataset . . . . .	152
7.2.1	Energy spectra for elemental groups of cosmic rays . . . . .	153
7.2.2	Consistency and quality of the result . . . . .	156
7.2.3	Preliminary result based on EPOS 1.99 . . . . .	162
7.3	Analysis of the Energy Spectra – Physical Conclusions . . . . .	171
7.3.1	Significance of the knee-like structure in the flux of iron nuclei . . . . .	171
7.3.2	Comparison of knee positions . . . . .	173
7.3.3	Spectral indices of the all-particle spectrum . . . . .	176
7.4	Comparison with Other Results . . . . .	178
7.4.1	Comparison with spectra obtained by other unfolding analyses of our collaboration . . . . .	178
7.4.2	Comparison with spectra obtained by other analysis methods of our collaboration . . . . .	182
7.4.3	Comparison with spectra obtained by other experiments . . . . .	185
<b>8</b>	<b>Summary, Final Discussion, and Outlook</b>	<b>189</b>
	<b>Appendices</b>	<b>201</b>
<b>A</b>	<b>Results of the Parametrization of the <math>\log_{10}N_{\text{ch}}^{\text{true}}</math> Distribution</b>	<b>201</b>
A.1	Parameters of the Parametrization of the $\log_{10}N_{\text{ch}}^{\text{true}}$ Distribution . . . . .	201
A.2	Energy Dependence of the Parameters of the Parametrization of the $\log_{10}N_{\text{ch}}^{\text{true}}$ Distribution – Parameters of the Interpolation . . . . .	203
<b>B</b>	<b>Results of the Parametrization of the <math>\log_{10}N_{\text{p}}^{\text{true}}</math> Distribution</b>	<b>205</b>
B.1	Parameters of the Parametrization of the $\log_{10}N_{\text{p}}^{\text{true}}$ Distribution . . . . .	205
B.2	Energy Dependence of the Parameters of the Parametrization of the $\log_{10}N_{\text{p}}^{\text{true}}$ Distribution – Parameters of the Interpolation . . . . .	207
<b>C</b>	<b>Results of the Parametrization of the Efficiency</b>	<b>209</b>
<b>D</b>	<b>Results of the Parametrization of the Systematic Bias in the Recon- struction Procedure</b>	<b>210</b>

---

<b>E</b>	<b>Results of the Parametrization of the Statistical Reconstruction Uncertainties for Charged Particles</b>	<b>211</b>
E.1	Parameters of the Parametrization of the Statistical Reconstruction Uncertainties for Charged Particles . . . . .	211
E.2	Charged Particle Number Dependence of the Parameters of the Parametrization of the Statistical Reconstruction Uncertainties for Charged Particles – Parameters of the Interpolation . . . . .	213
<b>F</b>	<b>Results of the Parametrization of the Statistical Reconstruction Uncertainties for Muons</b>	<b>216</b>
F.1	Parameters of the Parametrization of the Statistical Reconstruction Uncertainties for Muons . . . . .	216
F.2	Muon Number Dependence of the Parameters of the Parametrization of the Statistical Reconstruction Uncertainties for Muons – Parameters of the Interpolation . . . . .	218
<b>G</b>	<b>Entries of the Diagonal Matrix of the Response Matrix Factorization by SVD</b>	<b>221</b>
<b>H</b>	<b>Values of the Elemental Energy Spectra</b>	<b>224</b>
	<b>Bibliography</b>	<b>229</b>



# List of Figures

1.1	Cosmic ray all-particle spectrum, <i>status quo</i> . . . . .	2
1.2	Motivation for this work . . . . .	3
2.1	Comparison between different models for the transition from galactic to extragalactic cosmic rays . . . . .	12
2.2	Extensive air shower progeny . . . . .	14
2.3	Extensive air showers – muons and electrons . . . . .	15
3.1	Setup of the KASCADE-Grande experiment . . . . .	18
3.2	KASCADE detector station . . . . .	19
3.3	Grande detector station . . . . .	20
3.4	LECF and lateral distribution function for charged particles . . . . .	21
3.5	LECF and lateral distribution function for muons . . . . .	23
3.6	Fiducial area . . . . .	25
3.7	Quality of the shower reconstruction . . . . .	27
4.1	Sketch of the convolution of the cosmic ray spectrum and the air shower detection with KASCADE-Grande . . . . .	30
4.2	Advantage of using two instead of one observable for the unfolding . . . . .	31
4.3	Lateral distribution of charged particles . . . . .	32
4.4	Correlation between fluctuations of the charged particle densities and distance to shower core . . . . .	34
4.5	Charged particle lateral distributions for S(500) method . . . . .	34
4.6	Separation quality between different primaries in case of different possible observables as basis for a deconvolution . . . . .	35
4.7	The two-dimensional shower size distributions as possible basis for a deconvolution . . . . .	36
4.8	Measured shower size distribution used for the unfolding analysis . . . . .	38
5.1	Air pressure distribution whilst the KASCADE-Grande measurements . . . . .	44
5.2	The effect of intrinsic shower fluctuations . . . . .	45
5.3	Parametrization of the charged particle distribution with a Gaussian . . . . .	48
5.4	Final parametrization of the charged particle distribution . . . . .	48
5.5	Energy dependence of the parameters of the final parametrization of the charged particle distribution . . . . .	49
5.6	Final parametrization of the charged particle distribution with interpolated parameters . . . . .	50
5.7	Chi-squares of the final parametrization of the charged particle distribution with interpolated parameters . . . . .	50

5.8	Ratio of showers exhibiting a muon number above a given threshold to all showers as a function of the charged particle number . . . . .	52
5.9	Relation between the muon threshold value and the inflection point parameter of the error function . . . . .	53
5.10	Final parametrization of the muon distribution . . . . .	55
5.11	Energy dependence of the parameters of the final parametrization of the muon distribution . . . . .	55
5.12	Final parametrization of the muon number distribution with interpolated parameters . . . . .	57
5.13	Chi-squares of the final parametrization of the muon number distribution with interpolated parameters . . . . .	57
5.14	Final parametrization of the correlated $\log_{10}N_{\text{ch}}^{\text{true}}-\log_{10}N_{\mu}^{\text{true}}$ distribution	58
5.15	Trigger efficiency vs. electron or muon number . . . . .	62
5.16	Number of triggered stations in dependence on the true shower sizes . . .	62
5.17	Efficiency vs. charged particle or muon number . . . . .	64
5.18	Efficiency vs. charged particle and muon number . . . . .	64
5.19	Cross-check to the parametrization of the efficiencies . . . . .	65
5.20	Parametrization of the bias in the charged particle reconstruction . . . . .	68
5.21	Parametrization of the bias in the muon reconstruction . . . . .	69
5.22	Distribution of the charged particle reconstruction uncertainty . . . . .	72
5.23	Distribution of the charged particle reconstruction uncertainty (slices) . .	72
5.24	Interpolation of the parameters of the parametrization of the charged particle reconstruction uncertainty . . . . .	73
5.25	Distribution of the charged particle reconstruction uncertainty (slices, interpolated parameters) . . . . .	74
5.26	Distribution of the muon reconstruction uncertainty . . . . .	76
5.27	Distribution of the muon reconstruction uncertainty (slices) . . . . .	76
5.28	Interpolation of the parameters of the parametrization of the muon reconstruction uncertainty . . . . .	77
5.29	Distribution of the muon reconstruction uncertainty (slices, interpolated parameters) . . . . .	78
5.30	The response matrix . . . . .	82
5.31	The conditioned response matrix . . . . .	86
5.32	Comparison between the shower size distributions based on direct Monte Carlo simulations and on the parametrizations . . . . .	88
5.33	Comparison between the shower size distributions based on direct Monte Carlo simulations and on the parametrizations (slices) . . . . .	88
5.34	Comparison between the shower size distributions based on direct Monte Carlo simulations and on the parametrizations (slices, pure H or Fe) . . .	89
5.35	Comparison between the measured shower size distribution and the parametrizations (slices) . . . . .	90
5.36	Comparison between the measured shower size distribution and the parametrizations (maxima) . . . . .	91
5.37	Comparison between the measured shower size distribution and the parametrizations (contours) . . . . .	91
6.1	Why unfolding techniques? . . . . .	94
6.2	Chi-square value as a function of the iteration depth . . . . .	106
6.3	Generation of toy datasets . . . . .	109

6.4	Computation of the “true” uncertainties . . . . .	112
6.5	Computation of the estimated statistical uncertainties . . . . .	114
6.6	Comparison between “true” and estimated absolute statistical uncertainties	115
6.7	Computation of the estimated bias . . . . .	117
6.8	Comparison between “true” and estimated absolute bias . . . . .	118
6.9	Comparison between “true” and estimated relative bias . . . . .	119
6.10	Comparison between the intrinsic shower fluctuations and the experiment’s resolution . . . . .	122
6.11	Modification of the parametrization of the intrinsic shower fluctuations . .	124
6.12	Modification of the parametrization of the reconstruction accuracies . . .	127
6.13	Influence of the modified parametrizations on the solution . . . . .	128
6.14	Comparison of all sources of uncertainties . . . . .	130
6.15	The <i>WMSE</i> as well as the <i>RVB</i> in dependence on the iteration depth . . .	133
6.16	Comparison between the true and the estimated <i>WMSE</i> or <i>RVB</i> . . . . .	136
6.17	Test unfolding based on the realistic toy dataset . . . . .	138
6.18	Detection probability . . . . .	139
6.19	Test unfolding based on the realistic toy dataset, more statistics . . . . .	141
6.20	Unfolding an equidistribution . . . . .	142
6.21	Problematic toy spectra. . . . .	144
6.22	The <i>WMSE</i> in case of the “problematic” toy dataset. . . . .	145
6.23	Reducing the number of shown primaries in case of the “problematic” toy datasets . . . . .	145
6.24	Comparison between different unfolding algorithms . . . . .	147
7.1	Estimated <i>WMSE</i> for real data . . . . .	150
7.2	Estimated relative bias for real data . . . . .	152
7.3	Energy spectra for elemental groups of cosmic rays . . . . .	154
7.4	Mean logarithmic mass distribution for real data . . . . .	155
7.5	Comparison between the solutions of different unfolding algorithms . . . .	158
7.6	The quality of the data description by the solution . . . . .	159
7.7	Comparison between measured and reconstructed shower size distributions	160
7.8	Comparison between measured and reconstructed shower size distribu- tions, excesses . . . . .	161
7.9	Comparison between the measured shower size distribution and the param- etrizations (maxima), EPOS 1.99 . . . . .	164
7.10	Comparison between the measured shower size distribution and the param- etrizations (contours), EPOS 1.99 . . . . .	164
7.11	Energy spectra for elemental groups of cosmic rays, EPOS 1.99 . . . . .	165
7.12	All-particle spectrum for EPOS 1.99 and QGSJET-II-02 . . . . .	167
7.13	The quality of the data description by the solution, EPOS 1.99 . . . . .	168
7.14	Comparison between the measured shower size distribution and the param- etrizations (maxima), EPOS 1.99 vs. QGSJET-II-02 . . . . .	169
7.15	Judging the significance of the knee-like feature in the flux of iron nuclei .	172
7.16	Comparison between the knee positions of the light and the heavy com- ponent of cosmic rays . . . . .	174
7.17	The all-particle spectrum fitted by a double power law . . . . .	177
7.18	Comparison between different unfolding results, individual primaries . . .	180
7.19	Comparison between different unfolding results, all-particle spectra . . . .	181

---

7.20	Comparison between results of different analysis methods, individual primaries . . . . .	183
7.21	Comparison between results of different analysis methods, all-particle spectra	185
7.22	Comparison of all-particle cosmic ray energy spectra obtained by diverse experiments . . . . .	186
8.1	Comparison with other experiments' mean logarithmic masses . . . . .	193
8.2	Mean $X_{\max}$ measured with the Pierre Auger Observatory . . . . .	194
8.3	Other experiments' all-particle spectra . . . . .	195
8.4	Comparison of all-particle cosmic ray energy spectra obtained by diverse experiments and elemental spectra . . . . .	197

# List of Tables

5.1	Simulation statistics used for the parametrization of shower fluctuations . . . . .	42
5.2	Parameters and options set in CORSIKA . . . . .	43
6.1	Parameters of the realistic trial spectra . . . . .	107
A.1	Values of the parameters $p_3$ and $p_5$ of the $\log_{10}N_{\text{ch}}^{\text{true}}$ distribution . . . . .	201
A.2	Values of the parameters $p_1$ , $p_2$ and $p_4$ of the $\log_{10}N_{\text{ch}}^{\text{true}}$ distribution . . . . .	202
A.3	Results of the interpolation of the energy dependence of the parameters used in the $\log_{10}N_{\text{ch}}^{\text{true}}$ distribution fit . . . . .	203
A.4	Chi-squares per degree of freedom of the fits to the $\log_{10}N_{\text{ch}}^{\text{true}}$ distribution with interpolated parameters . . . . .	204
B.1	Values of the parameters $p_7$ , $b_1$ and $b_2$ of the $\log_{10}N_{\mu}^{\text{true}}$ distribution . . . . .	205
B.2	Values of the parameters $p_6$ and $b_0$ of the $\log_{10}N_{\mu}^{\text{true}}$ distribution . . . . .	206
B.3	Results of the interpolation of the energy dependence of the parameters used in the $\log_{10}N_{\mu}^{\text{true}}$ distribution fit. . . . .	207
B.4	Chi-squares per degree of freedom of the fits to the $\log_{10}N_{\mu}^{\text{true}}$ distribution with interpolated parameters . . . . .	208
C.1	Parameters of the parametrization of the efficiency . . . . .	209
D.1	Parameters of the parametrization of the systematic bias in the charged particle reconstruction . . . . .	210
D.2	Parameters of the parametrization of the systematic bias in the muon reconstruction . . . . .	210
E.1	Values of the parameters $p_1$ , $p_2$ and $p_3$ of the parametrization of the charged particle reconstruction uncertainty . . . . .	212
E.2	Results of the interpolation used in the parametrization of the charged particle reconstruction uncertainties ( $p_1$ ) . . . . .	213
E.3	Results of the interpolation used in the parametrization of the charged particle reconstruction uncertainties ( $p_2$ ) . . . . .	213
E.4	Results of the interpolation used in the parametrization of the charged particle reconstruction uncertainties ( $p_3$ ) . . . . .	214
E.5	Chi-squares per degree of freedom of the fits with interpolated parameters to the distributions of the charged particle reconstruction uncertainties . . . . .	215
F.1	Values of the parameters $p_1$ , $p_2$ and $p_3$ of the parametrization of the muon reconstruction uncertainty . . . . .	217

---

F.2	Results of the interpolation used in the parametrization of the muons reconstruction uncertainties ( $p_1$ ) . . . . .	218
F.3	Results of the interpolation used in the parametrization of the muons reconstruction uncertainties ( $p_2$ ) . . . . .	218
F.4	Results of the interpolation used in the parametrization of the muons reconstruction uncertainties ( $p_3$ ) . . . . .	219
F.5	Chi-squares per degree of freedom of the fits with interpolated parameters to the distributions of the muons reconstruction uncertainties . . . . .	220
H.1	Values of the elemental energy spectra, H . . . . .	224
H.2	Values of the elemental energy spectra, He . . . . .	225
H.3	Values of the elemental energy spectra, C . . . . .	225
H.4	Values of the elemental energy spectra, Si . . . . .	226
H.5	Values of the elemental energy spectra, Fe . . . . .	226
H.6	Values of the elemental energy spectra, all-particle . . . . .	227

# Chapter 1

## Introduction and Motivation

Cosmic rays have been discovered by Hess<sup>1</sup> in 1912 [90] during several free balloon ascents. He observed an increase of the ionization rate of air with increasing altitude, and concluded that there must be a hitherto unknown radiation penetrating into the Earth's atmosphere from the outer space. This important discovery was confirmed *ex post facto* by Kolhörster in 1913/14 [115] by means of improved electrometers during balloon ascents to even higher altitudes.

In 1929, Bothe and Kolhörster [43] proved the corpuscular nature of cosmic rays. This was confirmed by Clay [52], who revealed an interrelation between cosmic rays and the magnetic latitude, and hence concluded that cosmic rays are charged particles.

In 1938, Kolhörster et al. [116] and, independently, Auger et al. [18] discovered that cosmic rays are individual charged particles, which interact with a nucleus of the Earth's atmosphere, and consequently induce a cascade of secondary particles. These so-called *extensive air showers* were detected with a setup of cloud chambers and Geiger-Müller tubes, separated by some tens of meters, and operated in coincidence. This finding pioneered a new approach to investigate cosmic rays at the highest energies<sup>2</sup>: the measurement of extensive air showers by means of large ground based detector arrays.

In the subsequent years, insights about the shape of the energy spectrum of cosmic rays have been attained. The flux of cosmic rays follows a power law over many orders of magnitude in energy, overall appearing rather featureless (cf. Fig. 1.1). However, there are a few structures observable. In 1958, Kulikov and Khristiansen [119] discovered a distinct steepening in the measured electron shower size spectrum for particle numbers larger than  $8 \times 10^5$ , what corresponds roughly to a primary energy of the shower inducing cosmic ray particle of  $E \approx 8 \times 10^{15}$  eV. This should be the first measurement of the so-called *knee* of the cosmic ray spectrum. Three years later, Peters [148] concluded that the position of this knee will depend on the atomic numbers of the cosmic ray particles, if their acceleration is correlated to magnetic fields. He interpreted the knee “as a magnetic rigidity cut-off in the source which contributes the bulk of primaries below  $10^{15}$  eV,” and, more precisely, to be at “a magnetic rigidity corresponding to that of protons with about  $10^{15}$  eV.” However, this would mean that also the spectra of heavier cosmic ray primaries ought to exhibit such knee structures successively.

For many decades, the detection of changes in the mass composition of primary cosmic rays just beyond the first knee was the attempt of many air shower experiments. Round about half a century after the discovery of Kulikov and Khristiansen, the KASCADE

---

<sup>1</sup>“For his discovery of cosmic radiation,” Hess has been awarded the Nobel Prize in Physics in 1936.

<sup>2</sup>Where the fluxes of cosmic rays are very low, and hence statistically not accessible for balloon or satellite based experiments that measure the primary cosmic rays immediately.

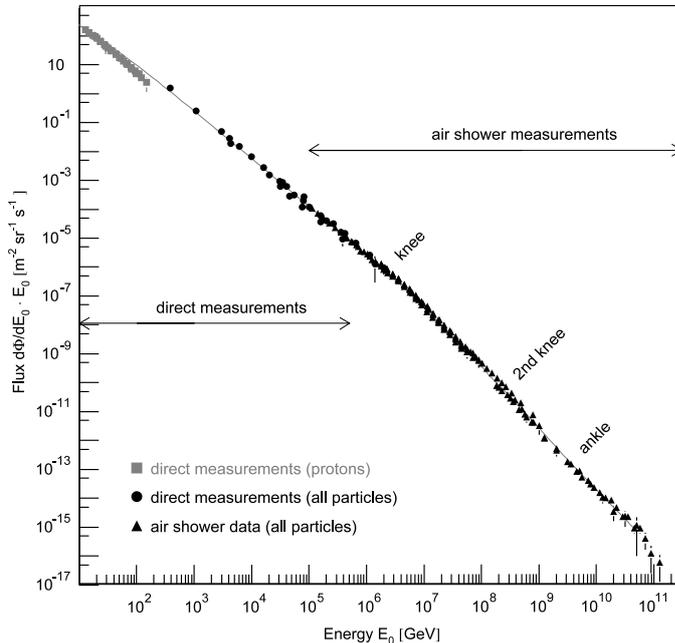


Figure 1.1: The all-particle energy spectrum of cosmic rays obtained by direct measurements above the Earth’s atmosphere (e.g. with balloons or satellites), or by indirect ones based on air shower detectors at ground level. Additionally shown is the flux of primary protons for direct measurements (figure taken from [42]).

experiment [10] clarified that the change in spectral index at the knee is indeed caused by a decrease of the so far dominating light mass group of cosmic rays [9]. This result was achieved by means of an unfolding analysis disentangling the manifold convoluted energy spectra of five mass groups from the measured two-dimensional shower size distribution of electrons and muons at observation level.

Nowadays, there are numerous theories about the origin and acceleration of cosmic rays. Concerning the knee position, some of them predict in contrast to the rigidity dependence considered by Peters a correlation with the mass of the particles. Hence, it is of great interest to verify whether or not also the spectra of heavy components exhibit analogous structures, and if, at what energies. The KASCADE-Grande experiment [14] extends the accessible energy range of KASCADE to higher energies up to  $10^{18}$  eV and allows by this to investigate the composition of cosmic rays at energy regions where a “*second knee*” is expected, and where the transition from galactic to extragalactic cosmic rays could occur. The discovery of the so-called “*iron knee*” would enable the validation or falsification of the various theoretical models.

Following this purpose, in this thesis, the KASCADE-Grande measurements will be analysed similar to the aforementioned studies [9] of the KASCADE data. As motivation, in Fig. 1.2, left panel, some contours of the two-dimensional shower size distribution of charged particles and muons measured with KASCADE-Grande are depicted. Based on simulations, the dashed lines mark the most probable shower size combinations generated in air showers initiated by primaries of constant energy (for different angle of incidences, bracketed by the two shown extreme assumptions for  $0^\circ$  and  $40^\circ$ ). To guide the eyes, the course of the light and heavy edges<sup>3</sup>, belonging to the light respectively heavy mass

<sup>3</sup>As can be seen from Fig. 5.36, the upper edge (called “light edge”) of the distribution corresponds

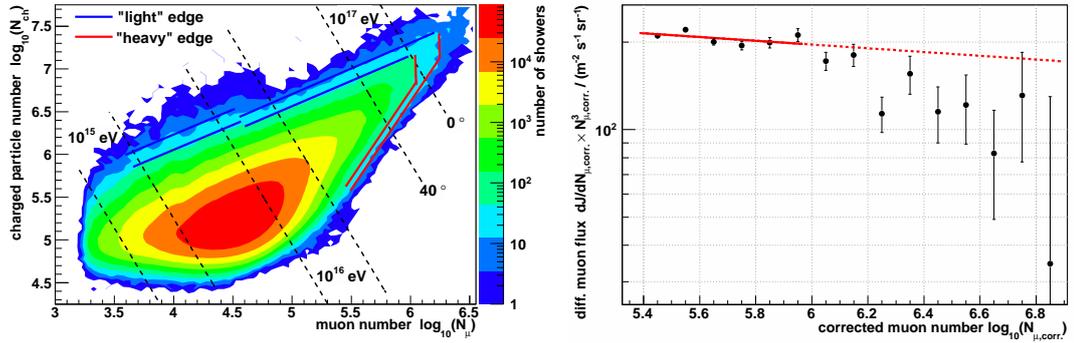


Figure 1.2: Left panel: Contours of the two-dimensional shower size distribution of charged particles and muons measured with KASCADE-Grande. Based on simulations, the dashed lines mark the most probable shower size combinations generated in air showers initiated by primaries of constant energy (for different angle of incidences). To guide the eyes and better point out discontinuities, the course of the light and heavy edges, belonging to the light respectively heavy mass groups, is visualized by some hand-drawn lines along some borders of the outer contours (see text for details). Right panel: The differential muon flux measured with KASCADE-Grande. A slight change in the spectral index can be guessed at muon numbers at about  $\log_{10}(N_{\mu,\text{corr.}}) \approx 6.0$ , what roughly corresponds to  $\sim 10^{17}$  eV primary energy. The spectrum is fitted up to muon numbers of  $\log_{10}(N_{\mu,\text{corr.}}) = 6.0$  with a single power law function (red line). The fit function is extrapolated (dashed red line) to higher muon numbers/energies in order to point out the deviation from this single power law at higher energies.

groups, is visualized by some hand-drawn lines along some borders of the outer contours. There is a discontinuity<sup>4</sup> discernible in the light edge at about  $5 \times 10^{15}$  eV (vertical showers). This corresponds exactly to the energy range where the knees in the spectra of light primaries have been observed. Hence, based on this simplified, and for sure scientifically barely stressable first inspection of the measurements, also in the KASCADE-Grande data slight indications for a knee-like structure in the lighter mass groups are given. Furthermore, the heavy edge shows such a discontinuity, too: At energies at about  $10^{17}$  eV (vertical showers), exactly where the iron knee is expected in the framework of a rigidity dependent cosmic ray acceleration and/or propagation. Basically, the remaining exercise would be to assign an appropriate scale for mass and energy to the measured distribution. On the right panel, the measured differential muon<sup>5</sup> fluxes are depicted. The spectrum is fitted up to muon numbers of  $\log_{10}(N_{\mu,\text{corr.}}) = 6.0$  with a single power law function (red line), which is extrapolated (red dashed line) to higher muon numbers, corresponding to higher energies. In a first and rough guess, a change in the power law index beyond  $\log_{10}(N_{\mu,\text{corr.}}) \approx 6.0$  can be observed, what corresponds to about  $\sim 10^{17}$  eV primary energy. Hence, also based on this basic observable there are indications for a second knee in the cosmic ray spectrum. As simplified these first discussions are, as motivating they are for the unfolding analysis presented in this work.

rather to the lighter mass groups, the lower edge (called “heavy edge”) to the heavy ones.

<sup>4</sup>This discontinuity is not an effect of the experiment’s efficiency, since in that case it would be oriented towards smaller numbers of showers with decreasing energy.

<sup>5</sup>Shown are the results for the corrected muon number; see Chapter 3 for details about the muon correction.

In the following, after a short introduction to the theory of cosmic rays (Chapter 2) and the KASCADE-Grande experiment (Chapter 3), the first task will be to find an appropriate observable set, which can serve as basis for the analysis (Chapter 4). In this context, the procedure of the unfolding method will be outlined. Thereafter, the response matrix that describes the convolution of the sought-after elemental energy spectra into the measured dataset will be computed and its reliability will be reviewed accurately (Chapter 5). What follows is the determination of an optimal unfolding algorithm, and its implementation will be discussed (Chapter 6). This includes also the development of strategies to estimate the uncertainties the final solution will suffer from, and extensive tests of the unfolding algorithms based on toy datasets. Finally, the dataset measured with KASCADE-Grande will be unfolded, and the obtained results will be analysed and discussed in detail, including the physical interpretation, as well as the comparison with energy spectra of other experiments (Chapter 7).

## Chapter 2

# Cosmic Rays and Extensive Air Showers

After the brief historical overview over the development of insights gained in the field of cosmic ray research given in the previous chapter, we shall now focus on the theory of cosmic rays that is related to the objectives of this thesis. Thereby, especially galactic cosmic rays are outlined, while extragalactic ones are skimmed over only peripherally and solar cosmic rays are omitted completely.

Below energies per nucleon of  $\sim 10^9$  eV, solar cosmic rays are dominating<sup>1</sup>. Cosmic rays between this threshold and energies of about  $\sim 10^{17}$  eV, or, depending on the model, up to several  $\sim 10^{18}$  eV are usually assumed to originate from our Galaxy, and to be extragalactic beyond these energies.

The most fundamental question related to cosmic rays, that is where they exactly come from and how they are accelerated, is still unanswered. There are many theoretical attempts at an explanation, however an overwhelming evidence for the reliability of one of these models is still missing. However, the analyses performed in this work will help to get a better understanding of the composition of cosmic rays beyond the first knee, and will answer the question about the existence of a knee-like structure in the spectra of heavy primaries. By this, the thesis will contribute a further piece of the puzzle towards the solution of this long-persisting mystery.

There are several publications, which give an overview over the *status quo* of the field of cosmic ray research (cf. for instance the publications [86, 42, 182], or the books [164, 158, 74, 128, 129]), such that it will be renounced in this thesis to go into detail. However, the theory that is important for the understanding and motivation of this thesis will at least be touched briefly. This includes especially the most important attempts of explanation for the origin and the acceleration of galactic cosmic rays (Section 2.1), possible interpretations of the one or several knees (Section 2.2), as well as a summary of promising models that are describing the transition from galactic to extragalactic cosmic rays (Section 2.3). Thereafter, the development of extensive air showers will be outlined, as, therein, the secondary particles are generated that are measured with KASCADE-Grande (Section 2.4).

---

<sup>1</sup>The “standard” solar wind, while, at highest energies next to the threshold of  $\sim 10^9$  eV, the solar cosmic rays are correlated with solar flares and coronal mass ejections at the sun. Solar cosmic rays are primarily protons and electrons (95%), but also alpha particles (4%) and isotopes of e.g. helium, neon, and argon (1%). For galactic cosmic rays of lower energies of some tens of GeV, their flux at the top of the Earth’s atmosphere can be modulated by this solar flux (and the Earth’s magnetic field), while beyond these energies it is less or not disturbed.

## 2.1 Origin and Acceleration of Cosmic Rays

The spectrum of cosmic rays follows over many magnitudes in energy power laws, what is the first indication for a non-thermal acceleration process. There is a huge amount of publications (cf. for instance [42, 86, 63, 35]) introducing and discussing possible acceleration scenarios for galactic cosmic rays, like supernova remnant shocks [20, 21, 186], superbubbles that have been generated by manifold supernova explosions of massive stars formed in giant OB associations [125], gamma ray bursts [60], microquasars [89], and more. Especially the supernova remnants are often considered as most favoured candidates for the acceleration of galactic cosmic rays, and hence will briefly be outlined in the following.

Already in 1934, Baade and Zwicky [20, p. 258] computed that “the total energy emitted during the existence of the super-nova [...] is of the order of [...]  $10^{54}$  ergs [...to...]  $10^{51}$  ergs”. Considering that the energy density of cosmic rays is about  $1\text{eV}/\text{cm}^3$ , the power required to sustain that intensity is  $\sim 10^{41}$  erg/s [42, p. 6]. Taking into account that there are round about three supernovae per century in our Galaxy, the required energy could already be provided if only  $\sim 10\%$  of the released kinetic energy is transferred to cosmic rays.

Today, it is theoretically established that cosmic rays can be accelerated in shock waves of supernova remnants by means of the diffusive shock acceleration mechanism<sup>2</sup> [118, 19, 26, 27, 37] up to the energies where the knees of the cosmic ray spectrum are discernible. Especially new and extended<sup>3</sup> calculations [25] indicate that it is theoretically possible to accelerate nuclei up to energies of  $Z \times 10^{17}$  eV, where  $Z$  is the atomic number (charge) of the particle. The scenario of particle acceleration in supernova remnants based on the diffusive shock acceleration mechanism is in agreement with the results of diverse experiments. Already in the early fifties, an evidence for such accelerations in supernova remnants was found by radio telescopes, which detected synchrotron radiation emitted by accelerated electrons that are gyrating in the magnetic fields of the remnant [162, 138].

Nowadays, with modern Cherenkov telescopes,  $\gamma$ -ray photons of TeV energies coming from supernova remnants have been detected, what gives strong indications for an efficient acceleration of charged particles to energies beyond 100 TeV, in consistency with the models predicting an acceleration of particles in young supernova remnant shocks [2]. Another indication for TeV particles inside supernova remnants is given by observations of non-thermal X-rays [22, 146]. While the non-thermal X-rays are certainly synchrotron emission from high energetic electrons [154], the origin of the  $\gamma$ -ray photons of TeV energies is still uncertain. Two scenarios are possible: The  $\gamma$ -ray photons could have been produced in leptonic processes<sup>4</sup>, or in hadronic ones<sup>5</sup> [61]. An experimental validation of the latter scenario would be the eagerly awaited evidence for an acceleration of hadrons in supernova remnants. What complicates the situation is the fact that the accelerated

---

<sup>2</sup>Also referred to as “first order Fermi process”. The diffusive shock acceleration mechanism bases principally on the mechanism proposed by Fermi [69], but assume that the acceleration occurs in the neighbourhood of a shock front and that the energy gains are caused by the systematic compression in the shock front (and not by small random velocities of the scatterers in magnetized clouds), in combination of an escape of particles from the acceleration region by advection downstream. It is a fast process, and leads to a power law spectrum in agreement with experimental observations.

<sup>3</sup>While the “classical” calculations, which consider a pure diffusive shock acceleration mechanism only, predict maximal energies of only  $Z \times 10^{14}$  eV [120].

<sup>4</sup>Inverse Compton effect between the same electrons, which are already producing the non-thermal X-rays, and the ambient photon field.

<sup>5</sup>Decays of  $\pi^0$  produced in hadron-hadron collisions between accelerated hadrons and the interstellar material.

electrons, which are verifiably there since they produce the observable non-thermal X-rays, can also cause a TeV  $\gamma$ -ray flux by inverse Compton scattering on the microwave background. Mastichiadis [134] concluded that this process could produce a TeV  $\gamma$ -ray flux of the same order as the flux produced from the  $\pi^0$  decay. Consequently, those  $\gamma$ -ray fluxes could also be the result of two processes. Hence, precise measurements of the fluxes are necessary, since electrons would result in flatter photon power laws, and thereby would enable a distinction between different scenarios.

The precise measurement of the  $\gamma$ -ray fluxes originating from supernova remnants is the objective of several experiments, and new results are currently published steadily; however, with sometimes varying conclusions. For instance, the LAT collaboration favoured, based on observations of the supernova remnant RX J0852.0–4622 with the Fermi Large Area Telescope [170], the production of  $\gamma$ -rays in hadronic processes, while the leptonic ones are considered less likely, however, cannot be ruled out at all. Same conclusions were drawn in many other publications of that experiment. However, in another publication [1] concerning the supernova remnant RX J1713.7–3946<sup>6</sup>, based on an analysis of Fermi Large Area Telescope observations in comparison with the measurements [3] of the H.E.S.S. experiment, the leptonic processes were favoured. The measurements do not agree with the fluxes predicted by most of the hadronic models published so far. However, it is conceivable to modify these hadronic process models in order to get a better agreement, or to assume mixed models basing on both the hadronic as well as the leptonic generation of TeV  $\gamma$ -rays. The LAT collaboration stated that further data studies, particularly based on a better modelling of the Galactic diffuse emission, are needed to give final answers whether or not hadrons are accelerated in supernova remnants.

In this context, a precise knowledge about the composition of high energy cosmic rays and the knees in the elemental spectra will enable a better understanding about the plausibility of diverse theoretical model, too. And thereby, the results obtained in this work can help prospectively to throw light on the question about the origin of high energy cosmic rays.

## 2.2 The Knees of the Cosmic Ray Spectrum

In 1958, Kulikov and Khristiansen [119] discovered a distinct steepening in the measured electron shower size spectrum for particle numbers larger than  $8 \times 10^5$ , what corresponds roughly to a primary energy of the shower inducing cosmic ray particle of  $E \approx 8 \times 10^{15}$  eV. This should be the first measurement of the so-called *knee* of the cosmic ray spectrum. Three years later, Peters [148] concluded that the position of this knee would depend on the atomic numbers of the cosmic ray particles, if their acceleration is correlated to magnetic fields. He interpreted the knee “as a magnetic rigidity cut-off in the source which contributes the bulk of primaries below  $10^{15}$  eV,” and, more precisely, to be at “a magnetic rigidity corresponding to that of protons with about  $10^{15}$  eV.” However, this would mean that also the spectra of heavier cosmic ray primaries ought to exhibit such knee structures successively.

About half a century after the discovery of Kulikov and Khristiansen, the KASCADE experiment [10] clarified that the change in spectral index at the knee is caused by a decrease of the so far dominating light mass group of cosmic rays [9]. In this thesis, based on KASCADE-Grande measurements, it will be shown that there is very likely<sup>7</sup> a

<sup>6</sup>Which is quite similar to the aforementioned remnant RX J0852.0–4622.

<sup>7</sup>If the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68] used whilst the

second knee, which is caused by a knee-like feature in the spectrum of heavy cosmic ray primaries.

There are diverse theoretical approaches used to explain the cause of the knees (see e.g. the compilations in [94, 65] or [161, p. 365 et seqq.]). A comprehensive discussion of all these models will not be possible in this thesis; and hence, only the most favoured explanation attempts will be outlined, but only briefly. Basically, the models can be divided into two categories: The *astrophysical* models that attribute the knees to a change in the elemental energy spectra of primary cosmic rays (references will be given later), or the *interaction* models that assume a change of the nature of the hadronic interactions between the high energy particles of cosmic rays and the nuclei of the Earth's atmosphere (see e.g. [107, 108, 141]). However, d'Enterria et al. [59, p. 6] stated that their insights gained based on LHC data “give[...] a strong support to the interpretation of the results in the CR “knee“ energy range ( $E_{lab} = 10^{15.5}$  eV) in terms of conventional primary spectrum and nuclear mass composition and disfavours some proposed speculative ideas that the change of the CR spectral slope could be due to a sudden change in the hadronic interaction mechanism above 2 TeV c.m. energy”, such that the *interaction* models appear rather unlikely. They will not be discussed further. The *astrophysical* models seem to be more plausible however. They can roughly be divided into three subclasses:

### **Knee caused by acceleration mechanism**

Very promising models try to explain the knees in the cosmic ray spectrum as the result of a maximal reachable energy of the diffusive shock acceleration mechanism in supernova remnant shocks (cf. e.g. [28, 114, 165, 168]). There are also models predicting a single source, e.g. a single, nearby and recent supernova, as cause of the knee [65]: The spectrum of that source is overlaid over the structure-less cosmic ray background spectrum, that originates from many other unspecified sources. However, also other possible scenarios are discussed in publications, like re-acceleration of cosmic rays by spiral shocks in the galactic wind [187], accelerations of cosmic rays in ultra-relativistic cannonballs ejected into the galactic halo [150, 189], and many more.

While the models based on the shock acceleration mechanism predict a rigidity dependent maximal reachable cosmic ray energy, alternatives like the cannonball model yield a cut-off for individual particles in dependence on their atomic mass caused by effects of relativistic beaming in the jets. By this, an exact determination of the position of the knees of individual cosmic ray mass groups, what is one objective of this thesis, will help to judge about the plausibility of the different models. However, one has to keep in mind that an observation of a rigidity dependence would not automatically rule out the cannonball models, since it is still possible that the latter is the dominant source of acceleration, and the knee is only an effect of a rigidity dependent leakage of cosmic rays from our Galaxy (discussed in the following).

### **Knee caused by leakage from our Galaxy – Propagation effects**

Whilst their propagation through the Galaxy, cosmic rays will be deflected many times in randomly oriented magnetic fields. Thereby, with increasing energy, it will be more and more difficult to confine the cosmic ray particles to our Galaxy. Hence, there will be a rigidity dependent cut-off caused by a leakage of cosmic rays from the Galaxy. For instance, Swordy [169] suggested a combination of the diffusive shock acceleration

---

unfolding analysis describe the physics of hadronic interactions reliably.

mechanism with an energy dependent propagation path length in the *Leaky Box model*. There are many other approaches to describe the propagation process, like e.g. that in [152, 122, 142, 157].

However, Erlykin and Wolfendale [64] remarked that the measured knee structure is too sharp for pure leakage models. Maybe the models have to be modified, or it is simply a combination of both, the propagation and the acceleration mechanism at the source. Final answers are still eagerly awaited. Also in this context, a precise determination of the elemental spectra of cosmic rays and possible knee structures could yield a better understanding of the reliability of these models.

### **Knee caused by interaction with the interstellar medium**

Another possible explanation for the knee is an interaction of cosmic rays with background particles of the interstellar medium whilst the propagation process. Debated are e.g. interactions with massive neutrinos (e.g. [62, 190]), photo disintegration in dense photon fields (e.g. [106, 48]), etc. However, these models would mean that whilst these interactions further secondary protons would be produced. Candia et al. [48] remarked that the cosmic ray composition should become lighter beyond  $\sim 10^{16}$  eV, what is in contradiction to the results of current air shower experiments, as well as to the results obtained in this thesis. Furthermore, the massive neutrino can be excluded, as Hannestad [82] has deduced based on WMAP and 2dFGRS measurements. Nevertheless, at the present moment such influences on the knee by interactions of cosmic rays with other particles whilst the propagation process cannot be ruled out entirely. Again, a precise measurement of the elemental cosmic ray spectra can help to improve the actual knowledge.

## **2.3 Transition from Galactic to Extragalactic Cosmic Rays**

The question about a possible transition from galactic to extragalactic cosmic rays is closely related to that about the origin of the knees and the composition of cosmic rays beyond the first knee. Cosmic rays beyond the first and the second knee, so called ultrahigh energy cosmic rays, are expected to be of extragalactic origin. This is reasoned with arguments like e.g. the large gyro-radius of the primaries at these energies, such that the magnetic fields inside our Galaxy would presumably not be strong enough to keep the particles as long in the Galaxy that they could be randomly deflected to a isotropic distribution. Hence, if the ultrahigh energy cosmic rays are accelerated within our Galaxy, it should be able to trace them back to their source; however, such powerful galactic sources are not observed so far. But, there are also further arguments, which cannot be discussed entirely in this thesis.

There are a couple of models (see e.g. compilations in [51, 31, 57, 105]) dedicated to describe the transition process, and thereby also to explain the observable ankle of cosmic rays (the ankle is denoted in Fig. 1.1). In the following, the discussion will be restricted to three exemplary models that are considered to be promising in many publications.

### **The ankle model**

In the traditional *ankle model*, the ankle-like structure in the all-particle spectrum at energies of about  $\sim 5 - 10 \times 10^{18}$  eV is interpreted as an effect of the transition from galactic to extragalactic cosmic rays (see e.g. [91, 92]). The extragalactic cosmic rays are

assumed to be subjected to a flat generation spectrum  $\propto E^{-2}$ , what could be explained based on a non-relativistic shock acceleration mechanism<sup>8</sup>, and which provides appropriate luminosities of the sources. The assumed galactic cosmic ray spectrum is comparably steep.

However, this model does not conform immediately with most of the favoured models that are describing the elemental cosmic ray spectra, and thereby especially the knees, by means of rigidity dependences: The upper edge of this “standard” galactic cosmic ray spectrum predicted by the rigidity dependence should occur already one, or even two orders of magnitude in energy below the assumed transition from galactic to extragalactic cosmic rays in the ankle model. Nevertheless, there are approaches that try to fill the gap between the possible second knee, which marks the upper edge of the usual galactic cosmic ray spectrum and the ankle by e.g. introducing an additional galactic component, a “component B” as denoted by Hillas [91, 92]. Another solution ansatz bases on the idea of Bell and Lucek [25], where the magnetic fields of supernova shocks are amplified non-linearly by the cosmic rays itself, and hence resulting in higher reachable acceleration energies as assumed in the usual models. There are further explanation attempts, which, however, cannot be treated comprehensively in this thesis.

One basic property of the *ankle model* is that it assumes a dominance of heavy particles in the transition region at around  $\sim 5 - 10 \times 10^{18}$  eV. However, this does not seem to be compatible with the results obtained by the Pierre-Auger-Observatory by measurements of the mean  $X_{\max}$  [66], where a dominance of lighter nuclei in the energy range from about  $10^{18}$  eV to  $10^{19}$  eV is concluded. Furthermore, the results obtained in this thesis give slight indications (however not statistically significant) that the spectrum of protons recovers beyond energies of  $10^{17}$  eV, while the flux of iron nuclei seems to decrease steeply. Thus, the composition at around  $10^{18}$  eV would be dominated by lighter primaries, and not by heavy ones. If this is right, this would mean an agreement with the results of the Pierre-Auger-Observatory, and hence a disagreement with the scenario predicted by the “ankle model”, too (cf. Section 7.2.1, and discussion in Chapter 8).

### The dip model

A different explanation attempt for the transition from galactic to extragalactic cosmic rays is given in the *dip model* (cf. e.g. [32]). There, the dip in the energy range of  $\sim 1 \times 10^{18}$  eV to  $\sim 4 \times 10^{19}$  eV is assumed to be related to electron-positron pair production of extragalactic protons on cosmic microwave background photons. A flattening of the dip at energies  $\gtrsim 1 \times 10^{19}$  eV is reproducing the ankle, while there is another flattening at energies  $\lesssim 1 \times 10^{18}$  eV, which is related to the transition<sup>9</sup> from galactic to extragalactic cosmic rays. This transition occurs where the adiabatic energy losses due to the expansion of the Universe and the ones due to pair-production are of equal order.

Hence, the transition of galactic to extra-galactic cosmic rays is expected to happen already at energies close/or equal to that where the second knee is observed, in agreement with a pure rigidity model for galactic cosmic rays. An additional component B like in case of Hillas model is not necessary. Furthermore, the pair production mechanism necessitates a dominance of protons<sup>10</sup> (only  $\sim 10\%$  to  $\sim 15\%$  contamination of

<sup>8</sup>Some models also predict extragalactic generation spectra up to  $\propto E^{-2.5}$ , what would rather agree with ultra-relativistic shock acceleration (cf. [123]).

<sup>9</sup>At energies  $\lesssim 1 \times 10^{18}$  eV, the extragalactic spectrum of protons computed in the dip model becomes flatter such that it will intersect the steeper galactic cosmic ray spectrum (at  $\sim 1 - 5 \times 10^{17}$  eV).

<sup>10</sup>Since the threshold for pair production is proportional to the energy/nucleon, only protons are rele-

helium nuclei) in the composition of extragalactic cosmic rays. This agrees well with the insights obtained by the Pierre-Auger-Observatory by measurements of the mean  $X_{\max}$  [66], where a domination of light nuclei in the energy range from about  $10^{18}$  eV to  $10^{19}$  eV, where the dip occurs, is concluded<sup>11</sup>. Furthermore, also the results in this thesis will give slight indications for a return of the dominance of the cosmic ray composition by lighter primaries beyond the “iron-knee” (see discussion in Chapter 8).

### The mixed composition model

In the intermediate *mixed composition model* [7, 6, 5], a mixed composition for both the galactic as well as the extragalactic cosmic rays is assumed at the transition region. In this model, the transition from galactic to extragalactic cosmic rays is assumed to happen at energies of about  $\sim 2 - 3 \times 10^{18}$  eV, and hence below the energy the transition takes place in the *ankle model*. That is why the necessity of an additional “component B” for galactic cosmic rays is reduced. The composition could<sup>12</sup> be heavy<sup>13</sup> dominated below energies of  $\sim 5 - 10 \times 10^{17}$  eV, such that the predictions would be in agreement with most of the experimental findings, as e.g. that in this thesis, where a dominance of heavy particles is observed at around  $\sim 10^{17}$  eV. One possible difficulty could be that in the *mixed composition model* the elemental spectra are subsequently suppressed at energies above  $\gtrsim A \times 10^{18}$  eV, where  $A$  is the atomic mass number [182]; hence, a mass dependent cut-off is expected, instead of an rigidity dependent one.

Figure 2.1 comprises a comparison between different models for the transition from galactic to extragalactic cosmic rays: the *ankle model* (top panel), the *mixed composition model* (middle panel), and the *dip model* (bottom panel). The predicted galactic (denoted “galactic”) and extra-galactic (denoted “CRPropa” respectively “mixed egal.”) cosmic ray spectra are illustrated by the dashed black lines and the red lines respectively. In addition, the sum of both components (denoted “galactic+extragalactic”) is given by the black lines and is compared to the all-particle spectra determined by diverse experiments. All pictures are taken from [182]; hence, see there for further details. Concerning the *dip model*, one remark has to be given: Due to the steep injection spectrum at the sources of extragalactic cosmic rays, the extrapolation to lower energies would be problematic for reasons of a too large emissivity. Therefore, usually, an “artificial” suppression of the extragalactic spectrum below a certain energy is introduced, what is, however, not shown in the figure.

According to the explanations to the different approaches given afore, the transition from the galactic to the extragalactic component varies from model to model. Due to the specific requirements on the composition and the behaviour of the elemental energy spectra attributed to certain models, an exact measurement of these properties, as done in this thesis, can help to judge about the validity of the different approaches.

---

vant in the context of energies at the dip.

<sup>11</sup>However, beyond energies of about  $10^{19}$  eV, the results of the Pierre-Auger-Observatory indicate (under suffering from larger uncertainties) a recover of the heavier particles what would contradict the dip model expectation of a steadily light composition for energies  $\gtrsim 10^{18}$  eV.

<sup>12</sup>Depending on the realization of the model.

<sup>13</sup>And – again depending on the realization of the model – probably proton dominated at energies  $\gtrsim 3 \times 10^{19}$  eV due to photodisintegration of nuclei. This could possibly be incompatible with the findings of the Pierre-Auger-Observatory [66], where (suffering from larger uncertainties) a recover of the heavy particles is observed beyond energies of  $\sim 10^{19}$  eV.

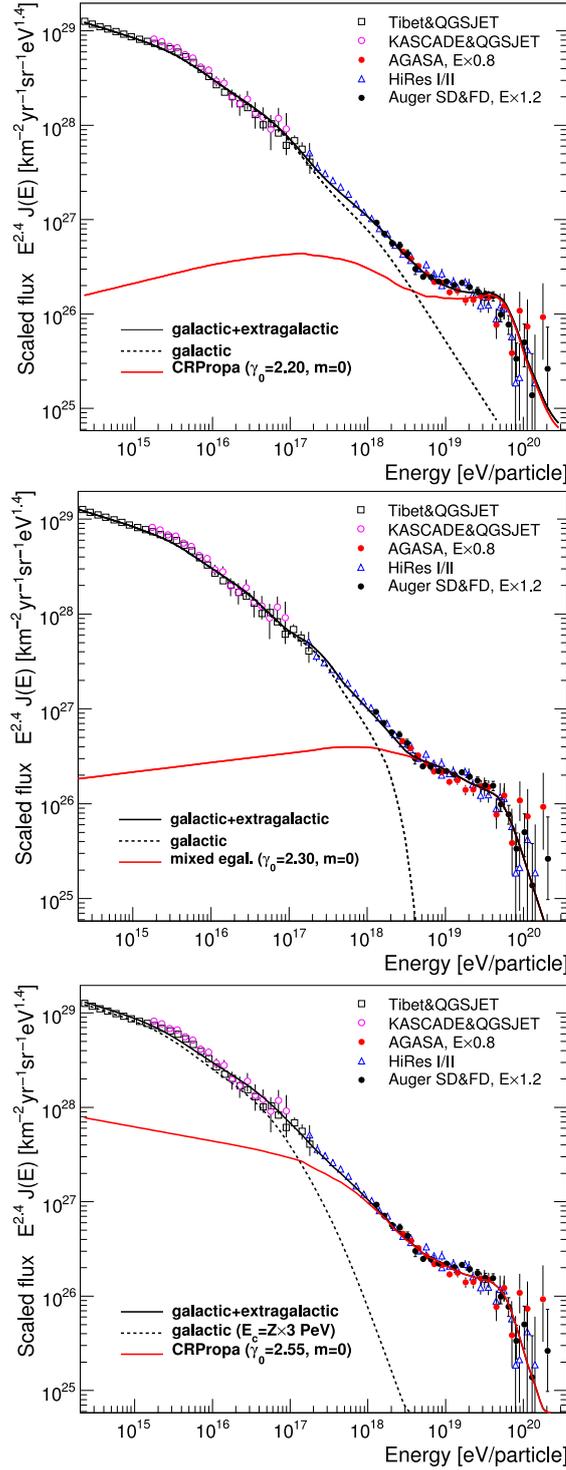


Figure 2.1: Comparison between different models for the transition from galactic to extragalactic cosmic rays: the *ankle model* (top panel), the *mixed composition model* (middle panel), and the *dip model* (bottom panel). The predicted galactic (denoted “galactic”) and extra-galactic (denoted “CRPropa” respectively “mixed egal.”) cosmic ray spectra are illustrated by the dashed black lines and the red lines respectively. In addition, the sum of both components (denoted “galactic+extragalactic”) is shown and compared to the all-particle spectra determined by diverse experiments. All pictures are taken from [182]; hence, see there for further details.

## 2.4 Extensive Air Showers

The flux of cosmic rays at energies above  $\sim 10^{14}$  eV is as small that it gets impossible to detect the cosmic ray particles by means of direct observation techniques, which are based on satellite or balloon experiments, with sufficient statistics. However, the high energy cosmic rays are interacting with the Earth's atmosphere, and thereby induce a large cascade of secondary particles. Hence, the atmosphere serves as large calorimeter for indirect high energy cosmic ray studies.

Detailed information about extensive air showers is given e.g. in [86, 42, 51], and references therein. A detailed introduction will be omitted in this thesis. The main properties of hadron<sup>14</sup> initiated showers, as far as they are important for the understanding of this work, will be discussed briefly.

If a cosmic ray nucleus penetrates into the Earth's atmosphere, it will interact with the nuclei therein. The basic principle of the development of the initiated extensive air shower is sketched in Fig. 2.2. Most of the secondary particles are pions and kaons. Before interacting further, they can also decay for instance into muons or neutrinos. The muonic component and the neutrinos are usually passing through the remaining atmosphere, and can be detected with appropriate experiments. The most intense electromagnetic component of the shower, i.e. electrons<sup>15</sup> and photons, is primarily caused by fast decaying neutral pions. They decay into photons, which initiate the electromagnetic shower. This shower component bases on two alternating processes: bremsstrahlung of electrons, and pair production by photons. The multiplication continues until the energy loss by bremsstrahlung is of the same order as that caused by ionization. Nevertheless, there will be a certain amount of electrons and photons that reach the observation level, and can hence be detected. The hadronic component can be regarded as the backbone of the shower, since it feeds the electromagnetic and muonic components. The electromagnetic component can further cause Cherenkov or fluorescence light, what is unimportant in context of this thesis however. In addition, things like radio emission etc. will be neglected in the following.

A detailed quantitative discussion of the dependence of the shower sizes like muon and electron number on the energy and type of the primary particle will be omitted, since the shower development will be simulated by means of Monte Carlo methods<sup>16</sup>, such that it can be considered as a kind of black box in context of this thesis. However, in the following, the properties of the air shower development that are important in the framework of this thesis will at least be discussed qualitatively.

What is of importance in context of this thesis is that the inelastic cross-section between the primary nucleus and the nuclei of the Earth's atmosphere depends on the type of the primary particle. For example, at primary energies of  $1 \times 10^{15}$  eV, the cross-section for iron nuclei is round about six times larger than for protons. That means that light primaries usually can penetrate deeper into the atmosphere, before a first interaction will occur. To put it simply, a nucleus with atomic mass number  $A$  can roughly be considered as superposition of  $A$  independent nuclei. Following this simplified consideration, the air shower induced by a particle with atomic mass  $A$  and energy  $E$  can be compared to the superposition<sup>17</sup> of  $A$  proton induced showers, where each has a primary energy of  $E/A$ .

<sup>14</sup>Since only the hadronic cosmic rays are of interest for this thesis, the showers induced by electrons or photons will not be discussed.

<sup>15</sup>More precisely, electrons and positrons.

<sup>16</sup>Based on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68].

<sup>17</sup>The superposition model is suitable since the typical energies of the primary exceed the binding

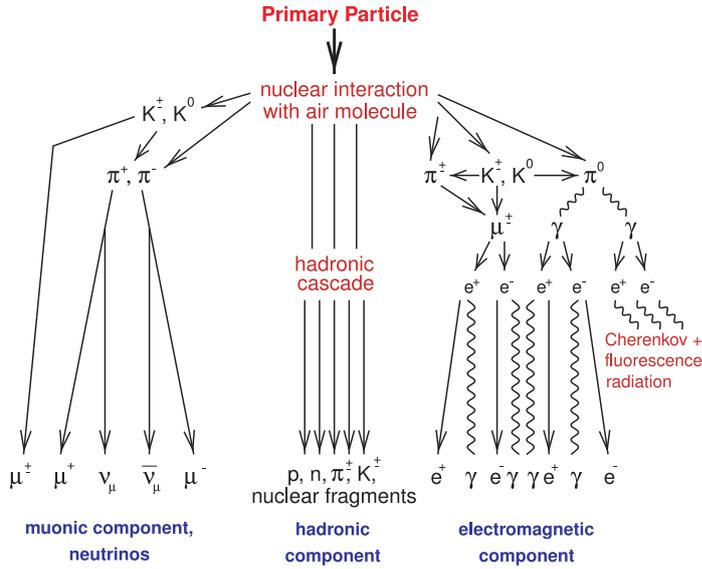


Figure 2.2: Schematic of the development of an extensive air shower initiated by a cosmic ray nucleus by a nuclear interaction with the Earth’s atmosphere (taken from [86]).

Hence, in air showers initiated by heavy primaries, more secondary particles are produced at sea level than in case of those induced by protons; however, they have smaller energies. Consequently, the number of muons at observation level is larger than in case of proton induced showers: due to the larger number of secondaries on the one hand, and on the other hand due to the concomitant on average smaller energy<sup>18</sup> per pion (kaon) causing them to decay earlier into muons, instead of further interacting hadronically; and all this in combination with the relatively small interaction probability of muons whilst the propagation through the Earth’s atmosphere. However, the smaller energies of secondary particles result in a faster attenuation of the electrons during the air shower development<sup>19</sup>. In summary, the larger the atomic mass of the air shower inducing nucleus is, the larger the muon number, and the smaller the electron number is at sea level in comparison to lighter primaries. As a conclusion, the correlation between the number of electrons and muons can serve as appropriate starting point for a composition analysis like that in this thesis.

Another property that varies between cosmic ray primaries of different mass groups are the intrinsic shower fluctuations. In general, the intrinsic shower fluctuations of heavier primaries with atomic mass  $A$  are smaller, since they can be considered as a superposition of  $A$  hydrogen nuclei, and by this the statistical fluctuations roughly<sup>20</sup> decrease by  $\sqrt{A}$ , since the primary energy is split in  $A$  independent cascades.

Last but not least, it is trivial and obvious that higher primary energies will result in higher numbers of electrons and muons detectable at observation level.

energy of nucleons by far.

<sup>18</sup>The hadronic interactions are not dominating over the decay processes for smaller charged pion energies (smaller than a critical energy of some tens of GeV, i.e. the connected Lorentz factors are comparatively small) or at high atmospheric altitudes with smaller particle densities (and the point of first interaction is located at higher altitudes for heavy particles).

<sup>19</sup>In combination with the smaller depth of the shower maximum for heavy primaries and the relatively fast attenuation of electrons, what results in smaller electron numbers at sea level, too.

<sup>20</sup>In more precise computations, the intrinsic shower fluctuations for heavy primaries are decreasing a little bit less than  $\sqrt{A}$ ; however, they are still significantly smaller than that in case of protons.

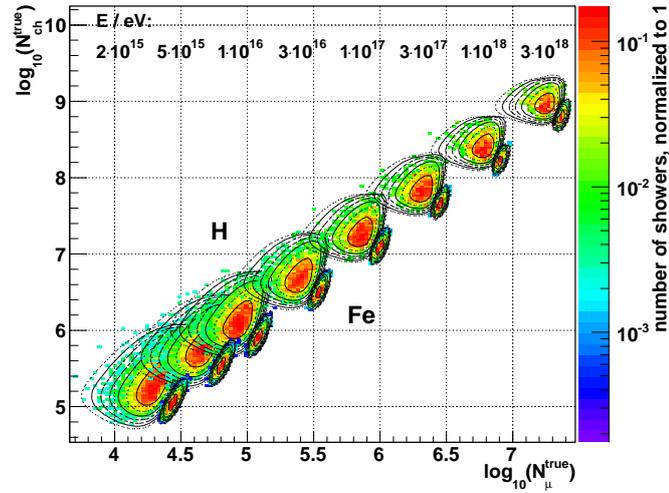


Figure 2.3: The correlation between charged particles and muons at observation level computed based on Monte Carlo simulations (histograms; based on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68]). The charged particles are primarily electrons, such that one could also speak about “electrons” instead of “charged particles”, and hence about the correlation of electrons and muons. The simulations are performed at several fixed primary energies (denoted above the respective distributions), and for primary protons (“H”) and iron nuclei (“Fe”). Additionally shown are some contour lines of the parametrizations determined in Chapter 5 (for details, see Chapter 5, and information given there to Fig. 5.14).

In Fig. 2.3, the correlation between the charged particles, respectively electrons since they are dominating the charged particles by far, and muons is depicted based on Monte Carlo simulations. As predicted afore, the larger the atomic mass of the air shower inducing nucleus is, the larger the muon number, and the smaller the charged particle, respectively electron number is at sea level. Furthermore, the distributions belonging to protons are significantly wider than that for iron nuclei, hence meaning larger intrinsic shower fluctuations. And for sure, the higher the primary energy is, the more charged particles/electrons and muons will be generated.

To summarize, the measurement of the electron and the muon number, as done by KASCADE-Grande, allows to draw inferences about the energy as well as the mass of the primary cosmic ray nucleus that initiated the air shower; and by this, enables the determination of energy spectra for elemental groups of cosmic rays. This allows answering questions like that about the origin of the second knee, as well as about the transition from galactic to extragalactic cosmic rays, etc. In the next chapter, the KASCADE-Grande experiment will be introduced briefly.



## Chapter 3

# The KASCADE-Grande Experiment

In the previous chapters, it has been clarified why the knowledge about the elemental spectra of cosmic rays is an important cornerstone of our understanding about the origin of the knees of the cosmic ray spectrum and the transition from galactic to extragalactic cosmic rays. The objective of this thesis is to provide insights that can help to put together one or more pieces of the puzzle of cosmic rays, like acceleration mechanisms, propagation processes, etc. As also explained, due to their very low fluxes at the knee and beyond, cosmic rays can only be detected indirectly by measuring the secondary particles in extensive air showers initiated by the primary cosmic ray particle. This is the purpose of the KASCADE-Grande Experiment [14], based on whose measurements the composition analysis in this thesis is carried out.

In the following, the KASCADE-Grande Experiment will be introduced briefly (Section 3.1). Thereby, as far as they are important for the understanding of this thesis, the detector components will be explained. Afterwards, it will be focused on the reconstruction procedure of the shower observables that are used in the final unfolding analysis of this work (Section 3.2). Thereafter, the comprehensive set of quality cuts necessary to guarantee reliable results will be outlined (Section 3.3). Finally, the reconstruction accuracy of the two observables the main unfolding analysis will be based on will be discussed, and appropriate means to improve especially the muon number determination will be introduced (Section 3.4).

### 3.1 Setup of the KASCADE-Grande Experiment

The KASCADE-Grande experiment measures extensive air showers initiated by cosmic ray particles with primary energies from  $10^{16}$  eV to  $10^{18}$  eV. It is located<sup>1</sup> at the Karlsruhe Institute of Technology (KIT). In Fig. 3.1, the basic setup is sketched and overlaid over a satellite picture of the KIT area (satellite picture taken from [97]). A detailed introduction to KASCADE-Grande can be found in [14], and will hence be omitted. Instead, only the main components of the experiment important for this thesis will be outlined briefly; to a certain extent basing on the introduction in [72].

In KASCADE-Grande, the extensive air showers are measured by means of 37 scintillator stations (marked with the blue quadratic markers in Fig. 3.1; cf. Section 3.1.2).

---

<sup>1</sup>Observation level of 110 m a.s.l., corresponding to an average atmospheric depth of  $1022 \text{ g/cm}^2$ , and located at the coordinates  $49.1^\circ \text{ N}$ ,  $8.4^\circ \text{ E}$ .



Figure 3.1: Sketch of the setup of the KASCADE-Grande experiment overlaid over a satellite picture of the KIT area (satellite picture taken from [97]). Additionally, the overlaid grid represents the internal KASCADE-Grande array coordinates. See text for further details.

They are optimized to measure the charged particles of the air shower, i.e. primarily electrons, but also muons. By means of the outer detectors of the co-located smaller array of the former KASCADE experiment (marked with the red shaded area in Fig. 3.1; cf. Section 3.1.1), the electromagnetic component can be measured separately from the muonic one due to iron/lead absorbers above the muon detectors. Roughly speaking, the subtraction of the muon number measured with the KASCADE detectors from the charged particle number measured with the Grande stations yields the electron number. Hence, for every shower detected with KASCADE-Grande, the total charged particle number, as well as the pure electron and muon numbers can be reconstructed.

### 3.1.1 KASCADE detector stations

The scintillators of the former KASCADE detector array cover an area of  $200 \times 200 \text{ m}^2$ . They are housed in 252 stations on a grid with 13 m spacing. While the inner stations are only equipped with liquid scintillators measuring primarily electrons and gammas, the 192 outer stations (marked with the red shaded area in Fig. 3.1) are also containing plastic scintillators underneath a shielding of 10 cm lead and 4 cm iron, which allows to measure muons separately from electrons and gammas. The absorber corresponds to round about 20 radiation lengths, resulting in a muon threshold of 230 MeV for vertically incident muons. In Fig. 3.2, a schematic of such an outer KASCADE detector station is illustrated (from [104]).

The muon detectors, which are of interest in the framework of the KASCADE-Grande experiment, consist of four plastic scintillators per station. The scintillators are of 3 cm

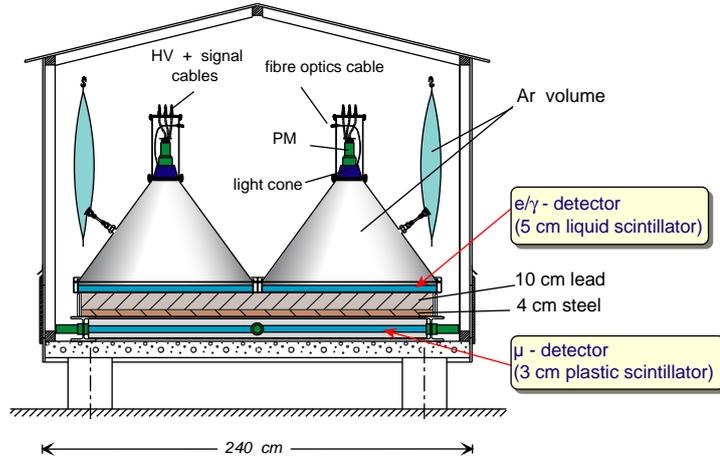


Figure 3.2: Schematic of an outer KASCADE detector station. Shown are the electron/gamma detector, the shielding, as well as the muon detector (from [104]).

thickness and their surface area is  $90 \times 90 \text{ cm}^2$  each. The light is coupled out by wavelength shifters and read out by 1.5 inch photomultipliers. The energy resolution has been determined to be of the order of 10% at 8 MeV, what corresponds to the mean energy deposit of a minimum ionizing particle (MIP).

Further details about the KASCADE experiment can be found in [10].

### 3.1.2 Grande detector stations

The scintillator array of KASCADE-Grande (the so-called Grande stations, marked with the blue quadratic markers in Fig. 3.1) covers a collecting area of approximately  $700 \times 700 \text{ m}^2$ . The 37 Grande stations are located on a hexagonal grid with an average mutual distance of 137 m. They measure the energy deposits of the charged particles of the air shower, without distinguishing between electrons and muons. Each Grande station houses  $10 \text{ m}^2$  of plastic scintillators, segmented into 16 individual scintillators of 4 cm thickness and with a surface area of  $80 \times 80 \text{ cm}^2$ , each read out with a high gain photomultiplier (for timing and particle density measurements). The four central modules are equipped additionally with low gain photomultipliers (for measurements in case of large particle densities). In Fig. 3.3, a sketch of a Grande scintillation detector station is depicted.

The 37 Grande stations are organized in 18 hexagonal trigger cells (cf. Fig. 3.1, where one trigger cell is outlined exemplarily). Each cell consists of seven stations: a central one, and six surrounding stations<sup>2</sup>. The most analyses based on KASCADE-Grande data, as also that in this thesis, only consider air showers that are capable to trigger a seven-fold coincidence<sup>3</sup>, i.e. all seven stations of an arbitrary cluster, since only then the trigger is transmitted to KASCADE for read-out of the muon detectors there.

<sup>2</sup>However, station No. 29 has only five surrounding stations, such that this cluster consists of six stations only (see Fig. 3.1).

<sup>3</sup>While the Grande data acquisition itself is already triggered by a four-fold coincidence.

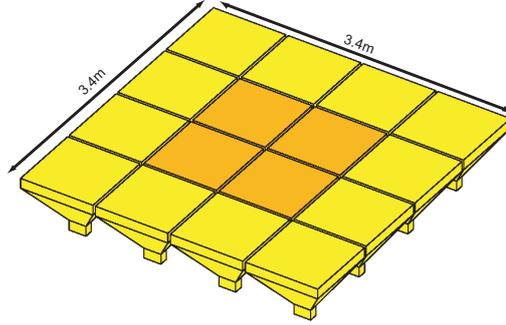


Figure 3.3: Sketch of the Grande scintillation detector. Shown are the 16 individual modules, each read out with a high gain photomultiplier. The four central modules are equipped additionally with low gain photomultipliers (taken from [14]).

## 3.2 Air Shower Reconstruction with KASCADE-Grande

The air shower reconstruction with KASCADE-Grande is discussed in detail in [14]. However, since the numbers of charged particles and muons will serve as basis for the unfolding analysis in this thesis, the reconstruction of these two sizes will be outlined briefly, whereas things like e.g. the determination of the shower core or the angle of incidence will be omitted.

### 3.2.1 Charged particle number reconstruction

The Grande detector stations allow measuring the charged particle number of the air shower. Therefore, the energy deposited in the scintillation detectors by the crossing charged particles has to be converted to the charged particle number. For this purpose, a so-called *lateral energy correction function (LECF)* is used, which describes the mean energy deposit per charged particle as a function of the distance from the shower core  $r$  correcting for the energy dependence of the stopping power and the energy release originating from photon conversion in the detector or its surroundings:

$$\frac{\text{energy deposit}}{\text{charged particle}}(r/\text{m})/\text{MeV} = \begin{cases} e^{1-0.1 \times r} + 7.51 + 0.02 \times r + 5.5 \times 10^{-5} \times r^2 + 5.4 \times 10^{-8} \times r^3 & \text{if } r \leq 450 \text{ m} , \\ \frac{\text{energy deposit}}{\text{charged particle}}(450) & \text{else .} \end{cases} \quad (3.1)$$

The parameters have been determined by means of Monte Carlo simulations. The Grande *LECF* is depicted in Fig. 3.4, left panel. For further details, and an explanation of the specific structure, see [14].

Since for every triggered Grande station the number of charged particles can be computed by means of the *LECF*, the last remaining step will be to estimate, based on the randomized measurements of the local charged particle densities at those individual stations, the total number of charged particles  $N_{\text{ch}}$  exhibited by the shower at observation level. For this purpose, the lateral distribution of charged particles has to be known,

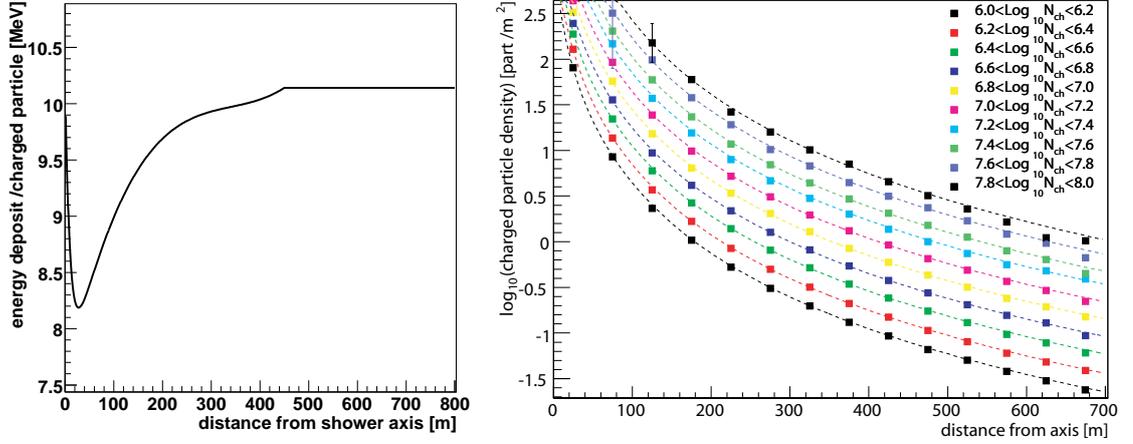


Figure 3.4: Left panel: The *LECF* according to Eq.(3.1) for the conversion of the energy deposits in the Grande detector stations into charged particle numbers. Right panel: The average experimental charged particle lateral distributions for different bins of shower sizes, and angle of incidences below  $18^\circ$ . The lines represent the predictions of the *LDF* according to Eq.(3.2) by using the average value of  $N_{\text{ch}}$  and  $s$  in the given interval of  $N_{\text{ch}}$ . Both figures are taken from [14].

and hence was determined based on Monte Carlo simulations. It was found that this distribution, i.e. the dependence of the charged particle density  $\varrho_{\text{ch}}$  on the core distance  $r$ , can be described by means of a modified NKG *lateral distribution function (LDF)* [103, 81]:

$$\varrho_{\text{ch}}(r/m)/\text{m}^{-2} = N_{\text{ch}} \times C(s) \times \left(\frac{r}{r_0}\right)^{s-\alpha} \times \left(1 + \frac{r}{r_0}\right)^{s-\beta} \quad (3.2)$$

$$\text{with } C(s) = \frac{\Gamma(\beta - s)}{2\pi r_0^2 \times \Gamma(s - \alpha + 2) \times \Gamma(\alpha + \beta - 2s - 2)}$$

Based on the simulations, the parameters were found to be  $\alpha = 1.6$ ,  $\beta = 3.4$ , and  $r_0 = 30$ . The slope of the distribution is given by the free parameter  $s$ , which is often referred to as “shower age”.

Now, a maximum likelihood function can be computed, and the parameters  $s$  and  $N_{\text{ch}}$ , i.e. the sought-after total number of charged particles, can be derived for every individual air shower event by fitting the charged particle densities locally measured at the triggered Grande stations by the afore-given lateral distribution function. For further details, see again [14].

In Fig. 3.4, right panel, the experimental average lateral distributions of charged particles for different bins of shower sizes, and angle of incidences below  $18^\circ$ , are depicted. Additionally shown are the predictions of the *LDF* according to Eq.(3.2) by using the average value of  $N_{\text{ch}}$  and  $s$  in the given interval of  $N_{\text{ch}}$ . In general, the experimental distribution is described very well by the assumed *LDF*, enabling a high quality reconstruction of the charged particle number on a event-by-event basis.

### 3.2.2 Muon number reconstruction

The muon number reconstruction with KASCADE-Grande is explained and discussed comprehensively in [73, 14], such that the basic steps will now be summarized only briefly.

Analogous to the procedure in case of the charged particles, also in case of the muons the measured energy deposits in the KASCADE muon detectors are converted to muon numbers by means of a muon *LECF*:

$$\frac{\text{energy deposit}}{\text{muon}}(r/m)/\text{MeV} = 7.461 + e^{1.762-0.0166 \times r} + 0.0002886 \times r \quad ; \quad (3.3)$$

where  $r$  is again the distance to the shower core. The energy deposit per muon is depicted in Fig. 3.5, left panel.

The lateral distribution of muons is described by means of a *LDF* based on that proposed by Lagutin and Raikin [121] for the electron component:

$$\begin{aligned} \varrho_{\mu}(r/m)/\text{m}^{-2} &= N_{\mu} \times f_{\mu}(r/m) \\ &= N_{\mu} \times \frac{0.28}{r_0^2} \times \left(\frac{r}{r_0}\right)^{p_1} \times \left(1 + \frac{r}{r_0}\right)^{p_2} \times \left(1 + \left(\frac{r}{10 \times r_0}\right)^2\right)^{p_3} \end{aligned} \quad , \quad (3.4)$$

where  $N_{\mu}$  is the total number of muons at observation level, and the parameters  $p_1 = -0.69$ ,  $p_2 = -2.39$ ,  $p_3 = -1.0$ , and  $r_0 = 320$  have been determined based on Monte Carlo simulations.

Based on a maximum likelihood estimation, the total number of muons  $N_{\mu}$  exhibited by an individual air shower at observation level can be computed as follows:

$$N_{\mu} = \frac{\sum_{i=1}^k n_i}{\sum_{i=1}^k \left( f_{\mu}(r_i/m) \times A_i/\text{m}^2 \times \cos(\Theta) \right)} \quad , \quad (3.5)$$

where  $n_i$  is the number of particles measured at a core distance  $r_i$  in one of the  $k$  muon detectors within an area  $A_i$ ,  $\Theta$  is the angle of incidence of the shower, and  $f_{\mu}$  the distribution function according to Eq.(3.4).

In Fig. 3.5, right panel, the experimental average lateral distributions of muons for different bins of muon sizes, and angle of incidences below  $40^\circ$ , are depicted. Additionally shown are the predictions of the *LDF* according to Eq.(3.4) by using the average value of  $N_{\mu}^{\text{rec}}$  in the given interval of  $N_{\mu}^{\text{rec}}$ . In general, the experimental distribution is described well by the assumed *LDF*; however, not as good as in case of the charged particles<sup>4</sup> (see Fig. 3.4, right panel). In case of intervals of smaller muon numbers, corresponding on average to smaller primary energies, the used *LDF* overestimates the local muon densities at smaller core distances, and underestimates them further afar. In case of larger muon numbers, it is the opposite. These deficits in the description of the measured densities are caused by the fixed shape of the *LDF*. In case of the charged particles, it was free (parameter  $s$ ). Due to the position of the muon detectors often quite far away from the shower core, the muon densities that are present at the KASCADE muon detectors are usually very small; in addition to the anyway small number of muons in comparison to

<sup>4</sup>Also the limitation to smaller angle of incidences below  $18^\circ$ , as it was the case in the depictions for the charged particles, does not improve the situation for the muon reconstruction significantly.

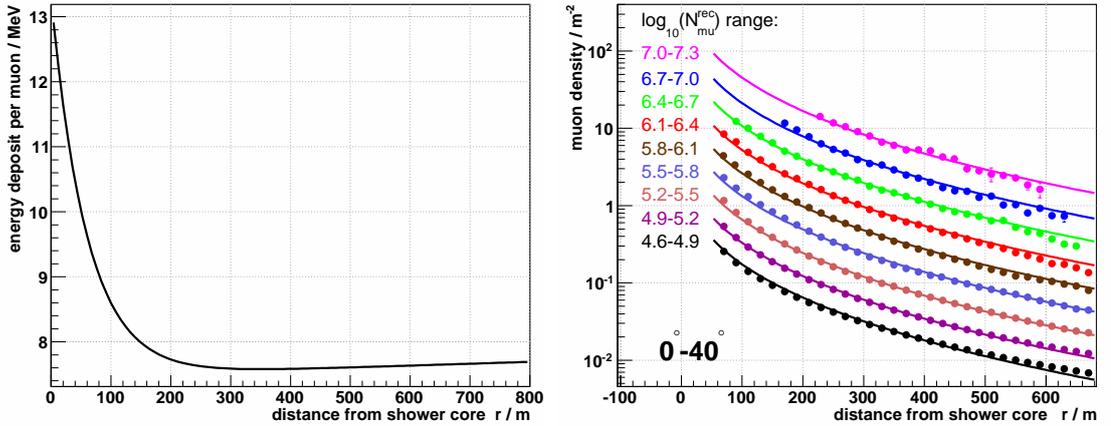


Figure 3.5: Left panel: The  $LECF$  according to Eq.(3.3) for the conversion of the energy deposits in the KASCADE muon detector stations into muon numbers. Right panel: The average experimental muon lateral distributions for different bins of muon sizes, and angle of incidences below  $40^\circ$ . The lines represent the predictions of the  $LDF$  according to Eq.(3.4) by using the average value of  $N_\mu^{\text{rec}}$  in the given interval of  $N_\mu^{\text{rec}}$ .

that of electrons, which are the main contributors to the charged particle number. For many air shower events, only in a very few muon detectors sufficient energy deposits are detectable. It was found that a fit to the measured muon density distribution on an event-by-event basis is often very unstable, if the slope of the  $LDF$  is variable. Hence, it was decided that it is more adequate to accept the systematic effects caused by the non-perfect description of the experimental muon lateral distribution by the used  $LDF$ , but, as a *quid pro quo*, ensure more stable fit results.

Since in this thesis the muon number will be used as second observable beside the total number of charged particles as basis for an unfolding analysis, it is important to reconstruct these shower sizes with the highest quality possible. Hence, in the following, the quality of the air shower reconstruction will be discussed briefly, and a method for the correction of the reconstructed muon number will be introduced. But, firstly, the standard quality cuts that are applied to guarantee a certain value of data and reconstruction quality will be outlined briefly.

### 3.3 Quality Cuts

Parallel to the comprehensive improvements of the reconstruction procedure, which finally resulted in an overall already very precise air shower reconstruction, a set of quality cuts has been developed. These cuts are necessary to eliminate those air shower events whose proper reconstruction cannot be guaranteed despite all endeavours.

The first group of cuts is related to the “hardware” of the experiment. The following conditions have to be fulfilled in order that the respective air shower event will be considered in later analyses:

- Only air showers that are capable to trigger a seven-fold<sup>5</sup> coincidence are consid-

<sup>5</sup>However, station No. 29 has only five surrounding stations, such that this trigger cluster consists of six stations only. Hence, in that case, a six-fold coincidence is demanded.

ered<sup>6</sup>, i.e. all seven Grande stations of an arbitrary cluster. The reason is that only in this case the trigger is transmitted to KASCADE for read-out of the detectors there; and hence, only then, the muon number can be measured.

- All 18 Grande trigger clusters must have been active and functioning properly. That means further that periods with technical errors in the data acquisition electronics are omitted.
- The KASCADE detector array must have been active, since only then the additional measurement of a pure muon number is possible.
- Fake events that are caused by the synchrotron radiation source ANKA, also located at the Karlsruhe Institute of Technology, are filtered out. This is realized by a query of a database that comprises information about the ANKA operation time.
- Air shower events whose maximal energy deposit is located at the border of the Grande array are omitted, since it could be possible that the reconstructed places of the shower centres will not be equal to the true ones. If the true shower core is located outside of the detector field, only the tails of the lateral distribution are hitting the Grande detectors. Nevertheless, the shower would be considered to have had his core located inside the array, what will finally cause misreconstructed shower sizes.
- In case of air showers initiated by primaries below the KASCADE-Grande energy threshold, the reconstruction procedure is unreliable. In order to discard these so-called “small events”, events that triggered less than 12 Grande stations are neglected.

After the reconstruction process, there are additional quality cuts applied that are depending on the reconstructed shower observables itself:

- The reconstruction procedure must have passed without failure. That means especially that the minimization procedures performed by means of MINUIT [99] must have been successful. Such a basic failure of the reconstruction procedure is rather seldom however (only for  $\sim 0.3\%$  of the events).
- Since the reconstruction procedure has been optimized for a reconstruction of air showers with angle of incidences below  $40^\circ$ , events with larger ones are omitted. Otherwise, a reliable reconstruction cannot be guaranteed in all cases.
- The shower core distribution is not homogeneous over the whole Grande array. This is, amongst others, caused by the hardware cut that demands a minimum of 12 Grande stations to be triggered by the respective air shower. Many extensive tests have shown that a homogeneous shower core distribution as well as a reliable shower size reconstruction can only be guaranteed if only air shower events are considered whose core-coordinates fulfil the following conditions:

$$\begin{aligned} -420 \text{ m} < x\text{-coordinate} < -50 \text{ m}, \\ -550 \text{ m} < y\text{-coordinate} < -30 \text{ m}, \text{ and} \\ 250 \text{ m} < \sqrt{x\text{-coordinate}^2 + y\text{-coordinate}^2} < 600 \text{ m} \quad . \end{aligned}$$

In Fig. 3.6, the fiducial area finally used is marked by the red coloured area.

---

<sup>6</sup>While the Grande data acquisition itself is basically already triggered by a four-fold coincidence.

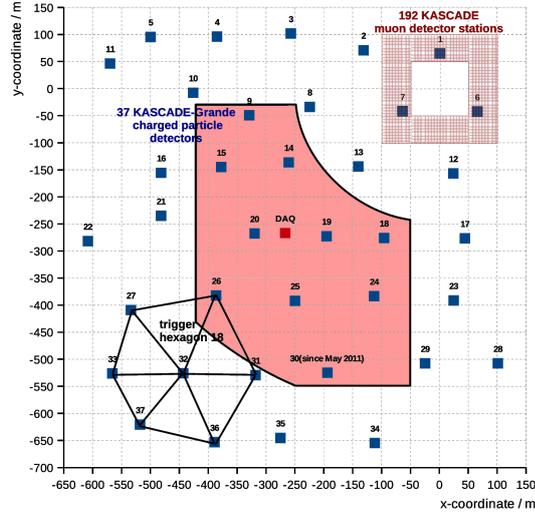


Figure 3.6: Sketch of the KASCADE-Grande setup. The fiducial area used in most of the KASCADE-Grande analyses, including that in this thesis, is marked by the red coloured area.

- If the slope parameter<sup>7</sup>  $s$  of the electron size lateral distribution fit (cf. Eq.(3.2)) is close to its outer limits, the reconstruction of the shower sizes is often unreliable. It was found that a sufficient stability can only be ensured if the slope parameter, determined whilst the fit, fulfils the following condition:  $-0.385 < s < 1.485$ .
- In some cases, the reconstructed shower sizes are too large in comparison to the actual energy deposits caused by the particles in the detectors. Hence, the shower sizes are overestimated. If  $N_c^{\text{tot}}$  is the sum of the energy deposits in the Grande scintillators by the particles of the air shower, this can roughly be converted to a particle number by  $N_c^{\text{tot}}/8.5$ , since the most probable energy deposit by a single particle in a Grande detector is 8.5 MeV. If the deviation of this roughly estimated<sup>8</sup> number of particles hitting the detectors and the reconstructed total particle number of the shower is unphysically too large, a problem in the reconstruction procedure has occurred. In that case, the reconstructed shower sizes are unreliable, and hence those showers are omitted in further analyses. Practically, the sum of the energy deposits  $N_c^{\text{tot}}$  in the Grande stations is compared on the one hand to the reconstructed charged particle number  $N_{\text{ch}}$ , on the other hand to the pure electron number<sup>9</sup>  $N_e$ . The conditions that air shower events have to fulfil in order to be considered in the analyses are:

$$\log_{10}(N_c^{\text{tot}}/8.5) > \frac{1}{3.5} \times (2.9 \times \log_{10}(N_{\text{ch}}) - 9.5) \quad ,$$

and

$$\log_{10}(N_c^{\text{tot}}/8.5) > \frac{1}{4.2} \times (2.9 \times \log_{10}(N_e) - 8.4) \quad .$$

<sup>7</sup>Respectively “shower age” parameter.

<sup>8</sup>A roughly estimated size, however, very directly measured, since no lateral distribution assumptions are made so far, and no fitting routine is necessary.

<sup>9</sup>What allows further to include possible influences from a misreconstructed muon number, since the electron number is, roughly speaking, determined by subtracting the muons measured with the KASCADE muon detectors from the number of charged particles measured with the Grande stations.

The discrimination power of all these conditions was tested extensively by means of Monte Carlo data, but was also reviewed in the context of measurements. The application of these cuts to the measured data will enable a high quality reconstruction of the respective shower sizes, and hence will pave the way for successful analyses. Nevertheless, especially the muon reconstruction has still some weaknesses, which will be discussed in the next section.

### 3.4 Quality of the Shower Reconstruction and Correction Methods

In Fig. 3.7, the charged particle (left panel) and muon (right panel) reconstruction accuracies are depicted. They have been determined based on Monte Carlo simulations, exemplarily for angle of incidences below  $18^\circ$  (however, the situation is similar in case of the other angle intervals). All necessary quality cuts introduced in the previous section have been applied. The shown relative deviations are medians, the error bars represent the quantiles at 84% and at 16%, and thus the central 68% distribution. The range of full trigger and reconstruction efficiency<sup>10</sup> is marked by the dashed vertical lines.

While in case of the charged particles the reconstruction accuracy is very good, in case of the muons the resolution is worse. It goes along with a significant bias especially at muon numbers close to the threshold of full efficiency. The muon reconstruction was analysed in detail, since the muon number is the second important observable used in the unfolding analysis presented in this thesis.

It would go considerably beyond the scope of this thesis to discuss these preparatory work in detail. However, the main facts will be summarized briefly. The by far largest source of systematic deficits is the fixed shape of the used muon lateral distribution function (cf. Eq.(3.4)). As already discussed in Section 3.2.2 and observable in Fig. 3.5, right panel, the fixed slope<sup>11</sup> causes local under- or overestimations of the muon densities in dependence on the distance from the shower core. Since in Fig. 3.7 no distinction between different core distances is made, these systematic distortions are hidden in the statistical error bars to a certain degree. Many attempts to improve the situation have been performed. The most successful solution is to simply correct the reconstructed muon number by means of a correction matrix<sup>12</sup>, which contains appropriate correction values to correct for a systematic deviation from the true muon number in dependence on the angle of incidence, the distance of the shower core to the centre of the KASCADE array, and the reconstructed muon number itself. The corrected muon number, determined by application of this matrix, is also depicted in Fig. 3.7, right panel (red hollow circles). On the one hand, the bias is now in agreement with zero. On the other hand, also the resolution seems to be improved, what is caused by the fact that the correction matrix considers also the shower core distance, and hence correct for some systematic effects still hidden in the statistical error bars of the uncorrected muon number.

This muon correction matrix is used in almost all current analyses of KASCADE-Grande data. Also in the unfolding analysis of this thesis, the muon correction<sup>13</sup> is applied, however in a slightly different way. Based on the knowledge gained in the extensive

<sup>10</sup>Most of the KASCADE-Grande analyses take into account only charged particle and muon numbers for that the experiment has full efficiency.

<sup>11</sup>As discussed in Section 3.2.2, a variable slope can cause unstable fit results, whose unreliability is often not discernible.

<sup>12</sup>Whose entries have been determined based on Monte Carlo data.

<sup>13</sup>As well as a correction of the charged particles.

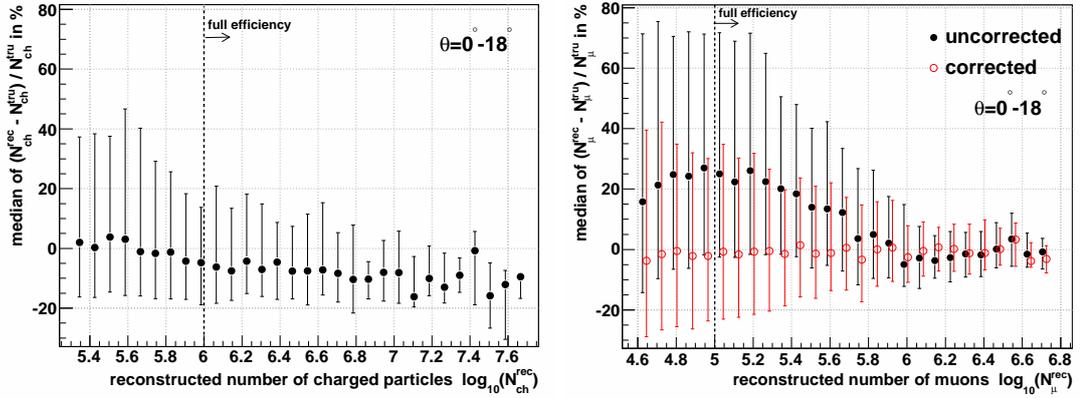


Figure 3.7: Charged particle (left panel) and muon (right panel) reconstruction accuracies derived from Monte Carlo simulations, exemplarily for angle of incidences below  $18^\circ$ . The shown relative deviations are medians, the error bars represent the quantiles at 84% and at 16%, and thus the central 68% distribution. The range of full trigger and reconstruction efficiency is marked by the dashed vertical lines. In case of the muons, also the situation after applying a correction method is shown (slightly shifted along the x-axis for a better discriminability. See text for details about the correction).

preparatory work of this thesis, the bias in the muon reconstruction will be parametrized by appropriate functions in order to correct for it (see Section 5.2.3). Instead of using a matrix, this allows easier modifications of things like bin widths etc. The resolutions of the shower sizes are taken into account, by means of appropriate parametrizations, in the response matrix for the unfolding problem, too (see Section 5.2.4). Hence, after all extensive improvements of the reconstruction procedure, and under application of the quality cuts and correction methods, the unfolding analysis can draw on a very stable dataset with well reconstructed shower sizes. This enables the determination of energy spectra for elemental groups of cosmic rays with highest quality. However, before that, some preparatory considerations are performed in the next chapter.



## Chapter 4

# Preliminaries for Unfolding the Energy Spectra

This work’s objective is to determine the composition and the energy spectrum of cosmic rays outside the Earth’s atmosphere. Whereas these observables are not immediately accessible for the ground based KASCADE-Grande experiment, it is possible to measure the particles produced in the cascade initiated by an interaction of a cosmic ray particle with the large-scale atmospheric target volume. Primarily, electrons and muons, in the following referred to as *shower sizes*, are detected. The basic procedure is sketched in Fig. 4.1.

The air shower development is subject to fluctuations; thus, the number of electrons and muons hitting the experiment’s detection area at ground varies from event to event. Moreover, these quantities will likely be reconstructed divergently from the true ones due to limited reconstruction resolution, detection efficiency etc. Consequently, the distributions of the reconstructed shower sizes are generated by a convolution<sup>1</sup> of the true primary energy spectrum with the afore indicated influences. The primary goal of this work will be to disentangle the measurement by means of a deconvolution procedure, and by this to compute the individual energy spectra of certain mass groups.

The next question to be answered is which observables should be used as basis for a deconvolution analysis (Section 4.1). Thereafter, the basic principle of the unfolding procedure will be outlined (Section 4.2).

### 4.1 Choice of Observables

In this section, an appropriate set of observables shall be determined, based on which the final unfolding analysis will be carried out. As explained in Chapter 3, the directly measured observables are not the shower sizes, e.g. the number of electrons or muons, but their energy deposits in the detectors. Hence, one could argue that it is possibly better to use these direct measured values. However, the shower sizes are the standard observables in the majority of the analyses of our collaboration, and are thus well tested and ready to use. The uncertainties are known or at least accessible through already available Monte Carlo simulations.

In [184] was figured out that “the muon number measured by KASCADE-Grande can be used for unfolding analyses, leading to an [all-particle] energy spectrum”. A distinction between different primaries was not possible due to the limited resolution of the detector

---

<sup>1</sup>The terms *convolution* and *folding* will be used synonymously in this work.

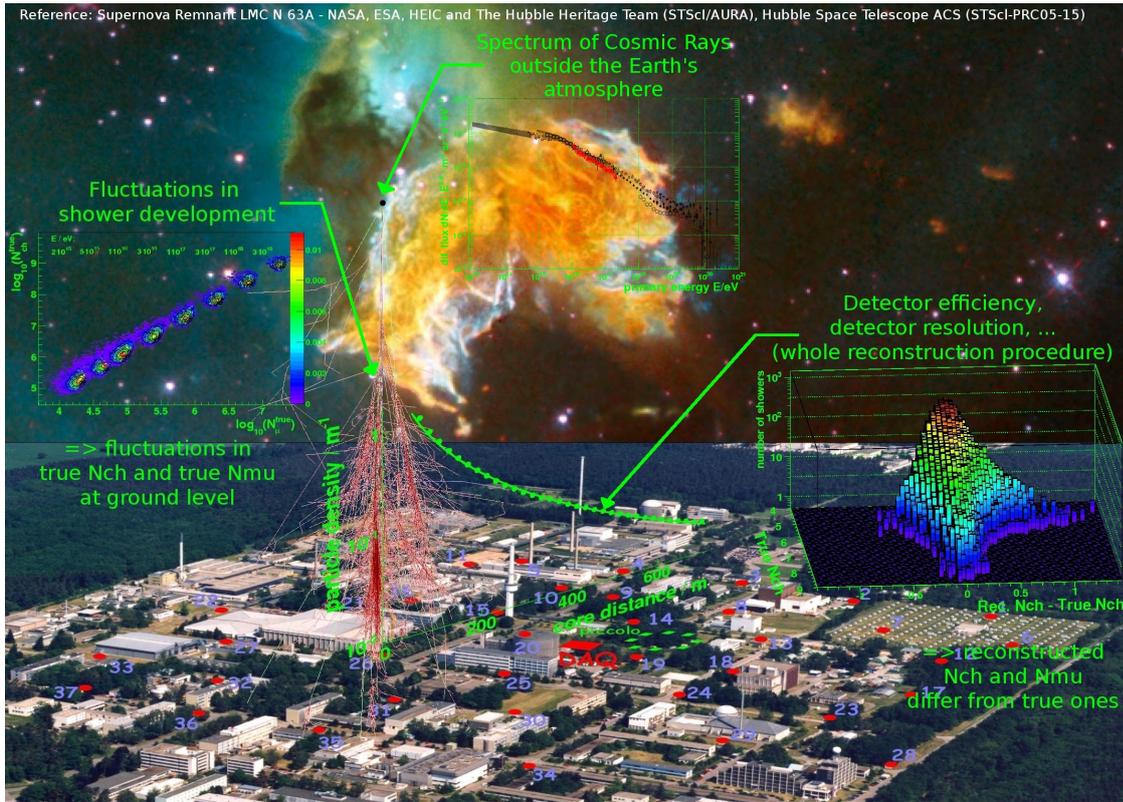


Figure 4.1: Simplified sketch of the convolution of the primary cosmic ray energy spectrum with the shower fluctuations as well as with the detector characteristics, and of the principle underlying the KASCADE-Grande shower size reconstruction.

and the limited data statistics at the time of that analysis. Now, in the present approach, two observables will be used in the analysis, instead of only one in [184], such that the analysis becomes *multivariate*. Figure 4.2 illustrates how the separation quality<sup>2</sup> is improved when using two instead of one observable. Shown is the two-dimensional logarithmic shower size distribution of charged particles  $\log_{10} N_{\text{ch}}^{\text{true}}$  vs. muons  $\log_{10} N_{\mu}^{\text{true}}$  at ground level generated using CORSIKA [88], with which the shower development in case of primary protons and iron nuclei with an energy of  $2 \times 10^{15}$  eV was simulated. In case of considering only one shower size, the separation quality is represented by the respective projection of the two-dimensional distribution into a one-dimensional along the corresponding axis. The geometric distances of the distribution's maxima of different primaries, in this case protons and iron nuclei, will get smaller each time such a projection is performed, and hence resulting in a worse separability: geometric distances  $\Delta N_{\text{ch}}$  or  $\Delta N_{\mu}$  instead of  $\sqrt{(\Delta N_{\text{ch}})^2 + (\Delta N_{\mu})^2}$ . Another advantage of the two-dimensional case is that the whole distributions are less overlapping at all.

As explained in Section 2.4 the correlation between the total number of electrons and of muons measured at a certain stage of the shower development is a promising quantity, which is not only sensitive to the primary particle's energy but also to its atomic mass. To summarize, the larger the atomic mass of the air shower inducing nucleus is, the larger

<sup>2</sup>While in this simplified illustration the separation quality is given by the geometric distance of the maxima in the distributions of the different primaries, more precisely one has also to take into account their shapes and widths. Thus, this should only be understood as a kind of first guess.

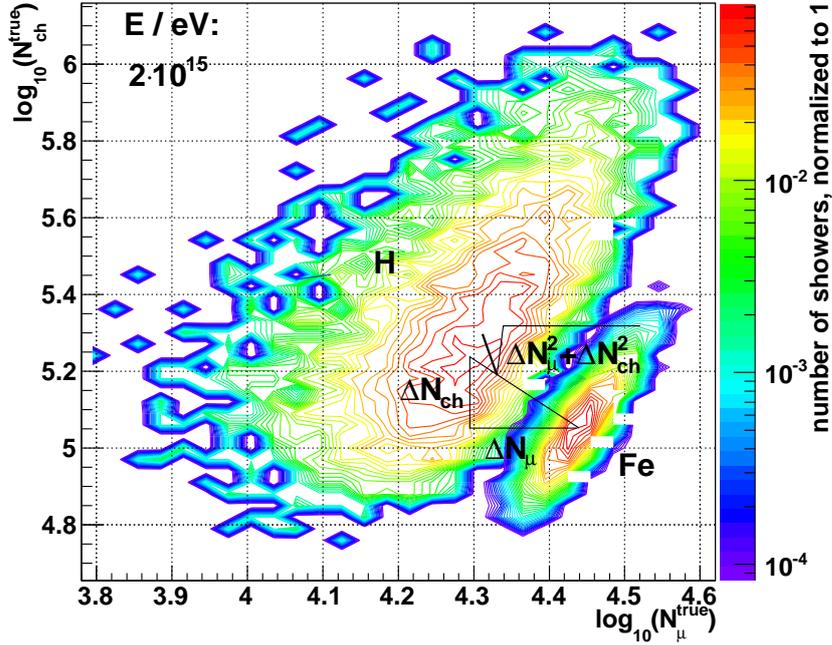


Figure 4.2: The two-dimensional  $\log_{10} N_{\text{ch}}^{\text{true}} - \log_{10} N_{\mu}^{\text{true}}$  shower size distribution for proton (H) or iron (Fe) induced showers with primary energy of  $2 \times 10^{15}$  eV. One can see that using two instead of one observable improves the separation quality between different primaries.

the muon number, and the smaller the electron number is at sea level in comparison to lighter primaries. As a conclusion, the correlation between the number of electrons and muons can serve as appropriate starting point for a composition analysis.

In case of the used KASCADE-Grande experiment [14], primarily the total number of charged particles, rather than the pure electron number<sup>3</sup>, is measured in combination with the muon number. However, the number of muons in the air shower is very small compared to that of electrons, such that the number of charged particles is approximately equivalent to the number of electrons. The composition sensitive correlation between the electrons and muons is comparably distinct in case of the correlation between all charged particles and muons, as will be shown in the following. Now, possible sets of KASCADE-Grande observables will be discussed and rated concerning their usability for a composition analysis.

The three KASCADE-Grande observables, which can serve optionally as a basis for an unfolding, are the number of muons ( $N_{\mu}$ ), the number of electrons ( $N_e$ ) and the number of charged particles ( $N_{\text{ch}}$ , containing in the vast majority electrons  $N_e$ , but also muons  $N_{\mu}$ ). All these variables are well tested by the collaboration and directly accessible via the experiment's database of reconstructed events. As possible candidates for the two-dimensional shower size plane, serving as the basis of the deconvolution, these combinations are surveyed:  $N_{\text{ch}} - N_{\mu}$  and  $N_e - N_{\mu}$ . Furthermore, a third attempt is presented using the locally measured densities of charged particles  $\rho_{\text{ch}}(r_0)$  at a certain distance  $r_0$  from the shower core, again with the muon number as second partner of the set.

<sup>3</sup>Nevertheless, the electron number can be derived by subtracting the number of muons from the total number of charged particles. However, this further reconstruction step induces additional uncertainties.

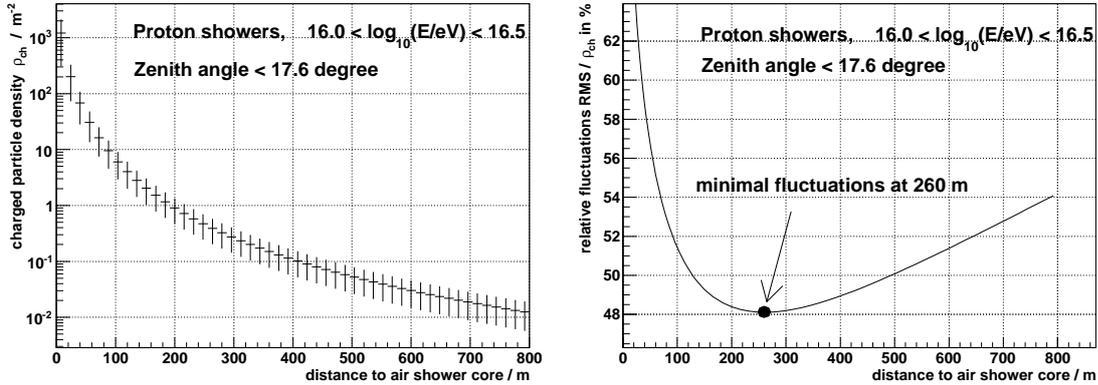


Figure 4.3: Left panel: The lateral distribution of charged particles based on Monte Carlo simulations. Right panel: The relative fluctuations in the densities of charged particles show a minimum – in this specific case at around 260 m distance from the core of the air shower.

The charged particles density  $\rho_{ch}(r_0)$  is a more native, but less tested observable. Moreover, this variable is not determined by the KASCADE-Grande standard reconstruction procedure, and therefore has to be treated supplementary.

In the following, firstly some investigations concerning the question at which core distance the local density of charged particles should be used. The lateral distribution of charged particles is determined using CORSIKA [88] 6.307 shower simulations, which are based on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68]. The experiment's response is simulated using CRES<sup>4</sup> 1.16/07, which bases on the GEANT 3.21 [46, 76] detector description and simulation tool. In the left panel of Fig. 4.3 this distribution is shown in case of proton induced air showers with primary energy  $16.0 < \log_{10}(E/\text{eV}) < 16.5$  and angle of incidence below  $17.6^\circ$ . The densities in each distance bin are averaged over some hundred events, the error bars represent the RMS. As one can see, the error bars, and thus the shower to shower fluctuations, seem to vary with the distance to the core. In Fig. 4.3, right panel, the relative fluctuations in the densities of charged particles are shown in dependence on the core distance. There can be seen a clear minimum at around 260 m. The distance with the smallest relative RMS in the density varies for different primary particle energies and types, confer Fig. 4.4, left panel. Conservatively, the different optimal distances, substituted by those of protons and iron nuclei, are averaged resulting in a common one of 235 m, represented by the horizontal dashed line. The RMS of the charged particle densities of proton or iron induced showers at the optimal distance, in comparison to that at the assumed common distance of 235 m, is depicted in Fig. 4.4, right panel. There is no large difference between both results; thus, it is possible to neglect these particle type and energy depending influences and to take into account the densities at a fixed distance of 235 m in the following. In case of measured data, this has to be done anyway because of the lack of knowledge of the primary particle's characteristics. Hillas et al. [93] stated that there will be a distinct distance from the air shower core at which the lateral distributions of all primaries overlap. This position would be a good energy discriminator. But if the relative abundances of the components of cosmic rays shall also be reconstructed, i.e. one would

<sup>4</sup>Cosmic Ray Event Simulation, a program package developed for the KASCADE [10] detector simulation.

like to distinguish between different particle types, it is not adequate to work with this specific distance at which the local densities are not sensitive to different primaries. In case of KASCADE-Grande, the lateral distributions coincide at around 500 m away from the core [44, 153, 177, 178], as shown in Fig. 4.5. The afore chosen common distance of 235 m is to a certain extent apart from this overlap area, but also not too close to the core, what would result in a worse reconstruction performance of the detectors (e.g. due to punch-through effects). Going beyond 500 m would in turn mean very low particle densities, and thus also a bad reconstruction. If the densities will be used as basis for the unfolding analysis, those at 235 m are the most appropriate ones.

In order to select the most qualified set of observables for the deconvolution, i.e.  $N_e - N_\mu$ ,  $N_{\text{ch}} - N_\mu$ , or  $\rho_{\text{ch}}(235\text{m}) - N_\mu$ , the combinations can roughly be judged by their weighted separation quality in  $y$ -direction:

$$Q_s(E) = \frac{\langle y_{\text{H}} \rangle(E) - \langle y_{\text{Fe}} \rangle(E)}{\sigma_{y,\text{H}}(E) + \sigma_{y,\text{Fe}}(E)}, \quad (4.1)$$

where  $\langle y_i \rangle$  is the mean value of the ordinates  $N_e$ ,  $N_{\text{ch}}$  or  $\rho_{\text{ch}}(235\text{ m})$  of the two-dimensional shower size plots for protons respectively iron nuclei in case of different primary energies  $E$  (the numerator of Eq.(4.1) is e.g. comparable to  $\Delta N_{\text{ch}}$  in Fig. 4.2), and  $\sigma_{y,i}$  are the RMS of the distributions in  $y$ -direction. The larger the separation quality is, the better the differentiation between the primaries will be. As depicted in Fig. 4.6 the separation is worse in case of the charged particle density while there is more or less no difference between the total number of charged particles or electrons. The fluctuations in the charged particle density are too large, due to the fact that the density at a certain distance will only be measured with a small amount of detectors. Using the integral of the lateral distribution, i.e. the total number of particles, takes into account also the distances where the distribution is insensitive to the composition (at around 500 m), but cancels out the fluctuations to a certain degree. Altogether, the weighted separation quality improves in case of taking into account more detectors and using the total number of particles, instead only the local densities.

The afore examined separation quality bases only on the one-dimensional projection of the distributions to the  $y$ -axis. Thus, the correlation between the variables stays unconsidered. However, the method can at least be pursued to have a kind of first guess, since the second contributor to the set, the muon number, was the same in all three constellations. As a more correct approach taking into account all correlations, one has to review the entire two-dimensional shower size distribution. This is shown in Fig. 4.7. The two-dimensional histograms are the measured distributions<sup>5</sup>, while the markers represent the simulation ascertained maxima in the associated probability distributions of proton as well as iron induced air showers in case of different primary energy ranges. The error ellipses are derived based on the 1-sigma uncertainties<sup>6</sup> and taking into account the correlation coefficients. Shown are three constellations: while in every case the number of muons is used, the other contributors to the shower size plot are the number of electrons (Fig. 4.7, top panel), the number of all charged particles (Fig. 4.7, middle panel) or the density of charged particles at the common core distance of 235 m (Fig. 4.7, bottom panel). The two first-mentioned instances do not differ significantly. There could be a slightly better separation between the primaries in the mid-level energy range for the second case. The results in the bin of highest energy have to be interpreted under

<sup>5</sup>The distributions, i.e. the axis ranges, are already cut to domains of full trigger and reconstruction efficiency – how it will also be claimed in the final deconvolution analysis later on.

<sup>6</sup>Originating from the shower fluctuations, the efficiencies, as well as the reconstruction accuracies.

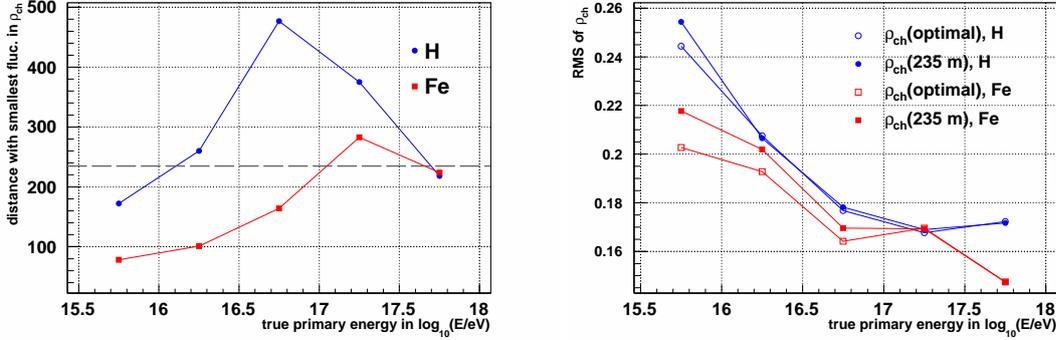


Figure 4.4: Left panel: The core distances with smallest relative fluctuations in the charged particle density in case of proton or iron induced air showers as a function of the primary energy. The dashed line represents the mean value, 235 m. Right panel: The RMS of the charged particle density at either the optimal (particle and energy specific) distance, or at 235 m away from core in dependence on the energy. The increment in the RMS due to the non-ideal distance is rather small.

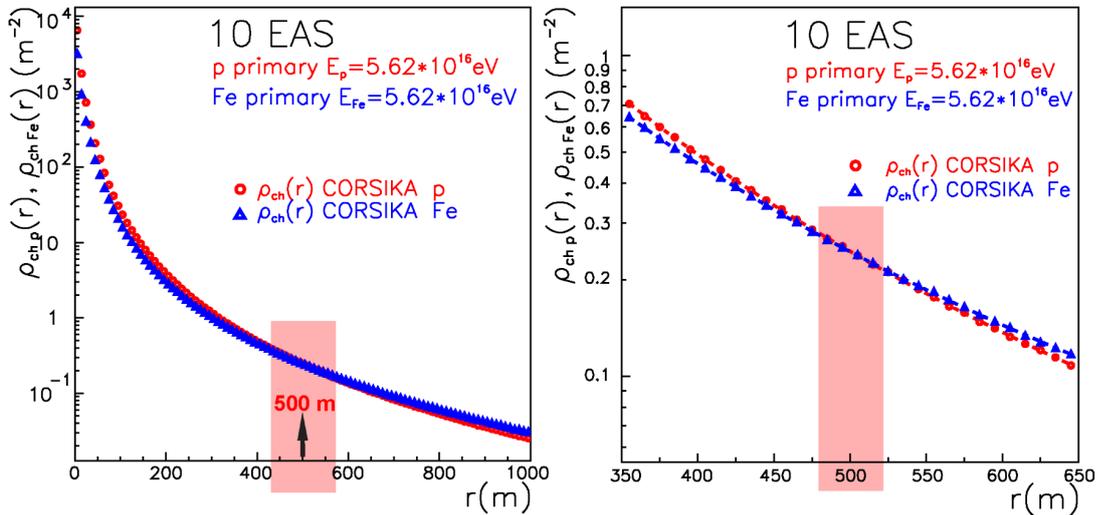


Figure 4.5: The lateral distribution of charged particles based on CORSIKA simulations for proton or iron induced air showers with a primary energy of  $5.62 \times 10^{16}$  eV. The distributions overlap at around 500 m distance from shower core. On the right panel the radial range is zoomed in and the simulations are fitted by a Linsley [127, 126] function (Figures taken from [177]).

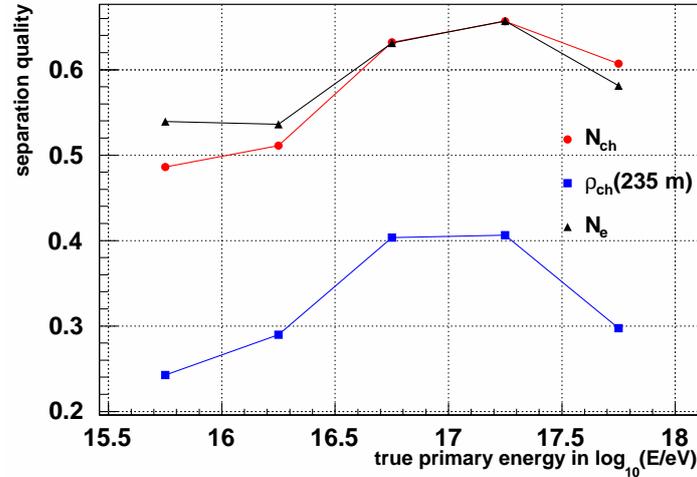


Figure 4.6: The separation quality according to Eq.(4.1) in case of  $N_e$ ,  $N_{ch}$ , or  $\rho_{ch}(235\text{ m})$  as possible second observable of the shower size set, beside the muon number, which could be used as a basis for a deconvolution analysis. The separation quality is illustrated in dependence on the primary energy.

consideration of the large uncertainties in the estimation of the maximal probabilities and in the error ellipses due to a very limited simulation statistics, and hence have to be handled with caution. The local densities do not appear to be a good choice: especially in the mid-level energy range the separation quality is worse. Even though in the highest energy bin, at a first glance, the separation seems to be somewhat better. But, again, one has to take into account the small simulation statistics. Nevertheless, the particle densities at a certain core distance get higher with increasing energy; thus, the number of particles hitting a detector station gets larger and their ascertainment will be better. The initial weakness in the approach based on the local densities, i.e. that the small numbers of particles hitting a detector result in large, primarily Poissonian uncertainties, might get unimportant in case of higher energies. But, certainly, one would have to test this with larger simulation statistics. If these checks confirm the expectation, the local particle density could be used as basis for a deconvolution in the highest energy range of KASCADE-Grande in a manner of a cross-check.

Summarizing the results of this section, as second contributor to the two-dimensional shower size plot, beside the muon number, both the number of electrons and of charged particles can likewise be utilized for an unfolding analysis. The charged particle number shows a negligible better separation quality in the middle of the examined energy range at around  $10^{16.5}\text{ eV}$  to  $10^{17.0}\text{ eV}$ . A deconvolution based on the charged particle density seems only to be promising in the second half of the energy range covered by KASCADE-Grande. Hence, an analysis using this variable can only serve as a kind of cross-check. Taking into account the obtained results, the two-dimensional shower size distribution of charged particles and muons will be used in the final unfolding analysis.

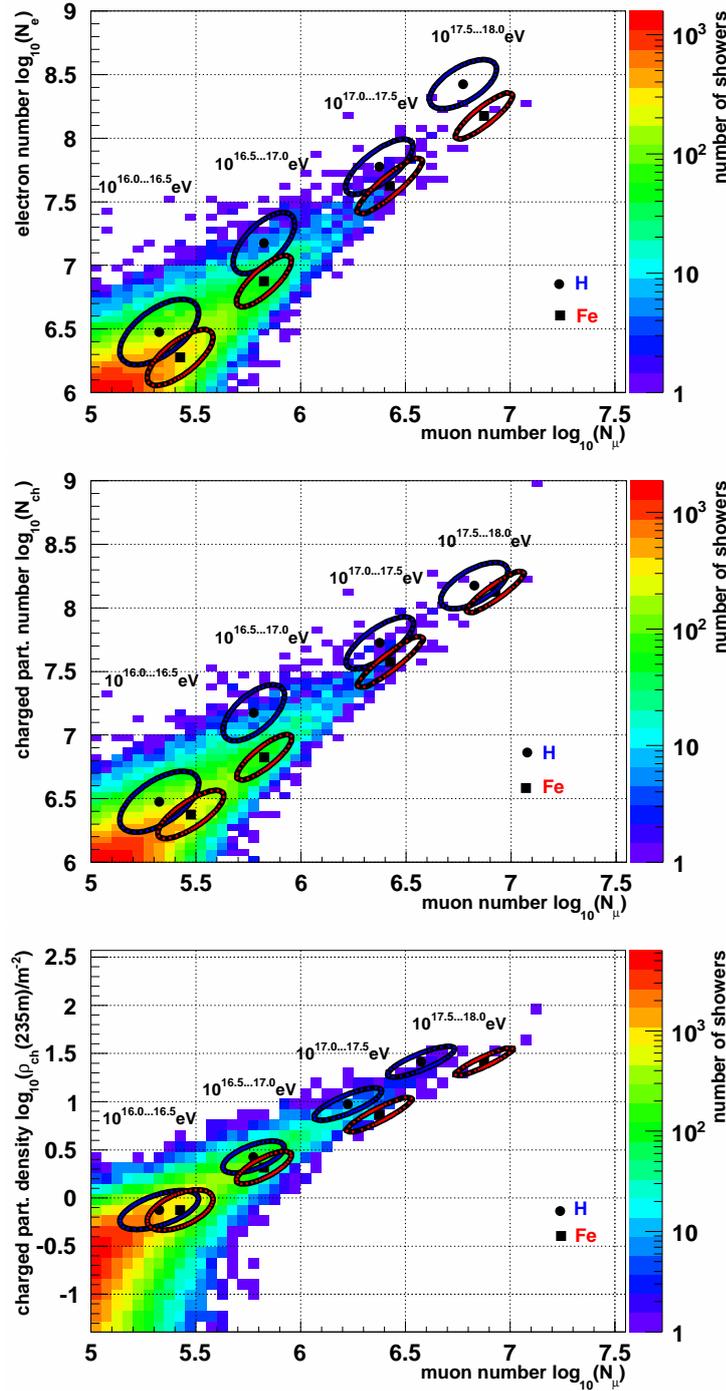


Figure 4.7: Different constellations as possible basis for an unfolding analysis: Number of electrons and of muons (top), number of all charged particles and of muons (middle), or the local density of charged particles 235 m away from core and the number of muons (bottom). The histograms are the distributions measured with KASCADE-Grande, while the markers represent the simulation ascertained maxima in the associated probability distributions of proton or iron induced air showers in case of different primary energy ranges (stated above each marker). The error ellipses are the one-sigma uncertainties regarding the shower fluctuations, the efficiencies, as well as the reconstruction accuracies.

## 4.2 Outline of the Deconvolution Analysis

The basic principle of the deconvolution procedure will be introduced now. In general, endeavours are performed to keep the analysis as close as possible to the one based on the former measurements of the KASCADE-experiment carried out by Ulrich [180, 9], in order to get better comparable results between both experiments.

In Section 4.1, a set of observables, respectively the total number of charged particles and of muons, have been determined as an appropriate basis for a deconvolution analysis. The two-dimensional shower size distribution<sup>7</sup> used for the final analysis is shown in Fig. 4.8. The axes' limits are already set such that the shown ranges only include particle numbers for that full trigger and reconstruction efficiency<sup>8</sup> is expected (cf. Section 5.2.2), realized by the following cuts:

$$\log_{10}(N_{\text{ch}}) \geq 6.0 \quad \text{and} \quad \log_{10}(N_{\mu}) \geq 5.0 \quad . \quad (4.2)$$

All events have passed the quality cuts outlined in Section 3.3, and an angle of incidence selection<sup>9</sup> of  $0^{\circ} \leq \theta \leq 18^{\circ}$ . The measurement time covers approximately 1318 days resulting in 78000 quality cuts passed and used air shower events, and an exposure of 164709 m<sup>2</sup> sr yr. The chosen bin width is 0.07 in logarithmic scale for both the charged particle and the muon numbers. In order to find this appropriate binning, the studies of the reconstruction accuracies are taken into account in such a way that it is not adequate to use a bin width beneath the achievable detector resolution (cf. Section 5.2).

As mentioned at the beginning of Chapter 4, the energy spectra of interest for different mass groups of cosmic rays are folded with the shower fluctuations and the detector characteristics resulting in the measurable distribution of charged particle and muon numbers, depicted in Fig. 4.8. Thus, the analysis' object is to disentangle the convoluted information. In order to get an easier mathematical access, the cells of the two-dimensional distribution are incrementally numbered from the left to the right, beginning from the lower left bottom corner (bin of smallest charged particle and muon number). The last cell is the most right bin at the top. The convolution of the differential fluxes  $dJ_n/d\log_{10}E$  of the primary cosmic ray nuclei  $n$  having an energy  $E$  into the measured number of showers  $N_i$  contributing to the cell  $i$ , and thus to the content of this specific charged particle and muon number bin  $(\log_{10}(N_{\text{ch}}), \log_{10}(N_{\mu}))_i$ , can be described by an integral equation:

$$N_i = \sum_{n=1}^{N_{\text{nucl}}} \int_{T_m} \int_{\Omega_{\text{tot}}} \int_{A_f} \int_E \frac{dJ_n}{d\log_{10}E} p_n d\log_{10}E \cos\theta dA d\Omega dt \quad , \quad (4.3)$$

with  $p_n = p_n((\log_{10}N_{\text{ch}}, \log_{10}N_{\mu})_i | \log_{10}E)$  .

One has to sum over all  $N_{\text{nucl}}$  nuclei contributing to the all particle cosmic ray spectrum<sup>10</sup>. The differential fluxes, given in units of energy, area, solid angle and time, have to be

<sup>7</sup>I.e. the number of measured showers as a function of the number of charged particles and muons.

<sup>8</sup>Although the efficiencies will be parametrized, there can be unexpected incidents like e.g. a failure of a single photomultiplier in a detector, possibly remaining unrecognized for a certain amount of time. Such problems can especially distort the efficiency curve near the detection threshold. Hence, the particle numbers that belong to primary energies below the threshold are omitted.

<sup>9</sup>This angle of incidence interval corresponds to a change in the atmospheric depth of  $(1 - \cos 18^{\circ}) \sim 5\%$ .

<sup>10</sup>Due to the limited separation quality between different primary particles the current analysis is restricted that way that the sum will actually only be carried out for some selected primaries representative for different mass groups.

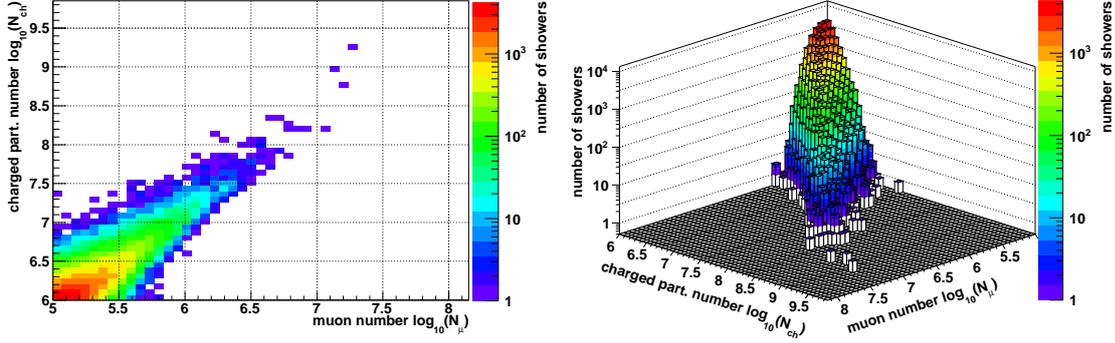


Figure 4.8: The measured shower size distribution of charged particles vs. muon number, as used for the deconvolution analysis. On the left panel as two-dimensional, on the right one as three-dimensional depiction. The axes' limits are set such that the shown ranges only include particle numbers for that full trigger and reconstruction efficiency is expected. All events have passed the quality cuts outlined in Section 3.3.

integrated over the measurement time  $T_m$ , the total solid angle  $\Omega_{\text{tot}}$  accessible for the experiment and used for the analysis, the chosen fiducial area  $A_f$  and all physically possible energies<sup>11</sup>  $E$ . The term  $p_n$  represents the conditional probability to reconstruct a certain combination of charged particle and muon number, respectively to get an entry in the cell  $(\log_{10}(N_{\text{ch}}), \log_{10}(N_{\mu}))_i$ , if the air shower inducing particle was of the type  $n$  and had an energy  $E$ . The cosine term in  $\cos \theta$   $dA$  accomplishes the transformation from the horizontal surface element to the effective detection area. In [184, p. 33 et seqq.] is proved that, in case of full trigger and reconstruction efficiency, there is no significant dependence of the shower sizes on the azimuth angle  $\phi$ . Hence, the integration over the solid angle can be carried out partly utilizing  $d\Omega = \sin \theta$   $d\theta$   $d\phi$  and considering the used zenith angle range  $0^\circ \leq \theta \leq 18^\circ$ , resulting in a coefficient  $2\pi$  and a remaining integral over the zenith angle<sup>12</sup>. Furthermore, the integration over time and area can be solved:

$$N_i = 2\pi A_f T_m \sum_{n=1}^{N_{\text{nucl}}} \int_{0^\circ}^{18^\circ} \int_{-\infty}^{+\infty} \frac{dJ_n}{d \log_{10} E} p_n \sin \theta \cos \theta d \log_{10} E d\theta \quad , \quad (4.4)$$

$$\text{with } p_n = p_n((\log_{10} N_{\text{ch}}, \log_{10} N_{\mu})_i | \log_{10} E) \quad .$$

The probabilities  $p_n$  are originating from a convolution merging the intrinsic shower fluctuations  $s_n$ , the trigger and reconstruction efficiency  $\varepsilon_n$ , and the reconstruction characteristics  $r_n$ :

$$p_n((\log_{10} N_{\text{ch}}, \log_{10} N_{\mu})_i | \log_{10} E) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} s_n \varepsilon_n r_n d \log_{10} N_{\text{ch}}^{\text{true}} d \log_{10} N_{\mu}^{\text{true}} \quad , \quad (4.5)$$

<sup>11</sup>Since the possible energies as well as the shower sizes can attain values spread over some orders of magnitude, these variables are used in logarithmic scale.

<sup>12</sup>Some extensive tests have shown that the shower fluctuations are negligibly sensitive to changes in the zenith angle within the chosen zenith angle range, which in total corresponds to a change in the atmospheric depth of only 5% (as shown in Footnote 9). Furthermore, the reconstruction properties also do not depend significantly on the zenith angles out of that range. Nevertheless, the integration over the zenith angle is still shown for reasons of clarity.

$$\begin{aligned}
\text{with } s_n &= s_n \left( \log_{10} N_{\text{ch}}^{\text{true}}, \log_{10} N_{\mu}^{\text{true}} \mid \log_{10} E \right) , \\
\varepsilon_n &= \varepsilon_n \left( \log_{10} N_{\text{ch}}^{\text{true}}, \log_{10} N_{\mu}^{\text{true}} \right) , \\
r_n &= r_n \left( \log_{10} N_{\text{ch}}, \log_{10} N_{\mu} \mid \log_{10} N_{\text{ch}}^{\text{true}}, \log_{10} N_{\mu}^{\text{true}} \right) .
\end{aligned} \tag{4.6}$$

More precisely,  $s_n$  is the probability that a nucleus  $n$ , having an energy  $E$ , induces an air shower containing a specific number of charged particles  $\log_{10} N_{\text{ch}}^{\text{true}}$  and muons  $\log_{10} N_{\mu}^{\text{true}}$  when arriving the experiment's detection plane. The probability to reconstruct, due to possible reconstruction uncertainties, a certain number of charged particles  $\log_{10} N_{\text{ch}}$  and muons  $\log_{10} N_{\mu}$  instead the true ones  $\log_{10} N_{\text{ch}}^{\text{true}}$  and  $\log_{10} N_{\mu}^{\text{true}}$  is described by  $r_n$ .

Equation (4.4) can mathematically be understood as a system of coupled integral equations and is named Fredholm [71, 17, 16, 83, 149] integral equation of the first kind:

$$\int_a^b K(y, s) x(s) ds = f(y) , \tag{4.7}$$

with  $c \leq y \leq d$  .

$K(y, s)$  is the so-called kernel, response, or transfer function<sup>13</sup>, which encodes the underlying physics that describes the relation between the model  $x$  and the observation  $f$ ,  $[a, b]$  is the domain of  $s$ , while  $[c, d]$  is the domain of  $y$ . Transferred to the actual analysis, the kernel function relates to the probabilities given in Eq.(4.6). The observation  $f$  is represented by the measured shower size plane (Fig. 4.8). The mentioned model  $x$  is the sought-after flux of cosmic rays. Hence, one has to solve Eq.(4.7) for  $x$ .

One has to keep in mind that, usually,  $f$  is not known exactly since it bases on a measurement with limited statistics only. More precisely, instead of a rather smooth  $f$  that bases on a forward folding of  $x$  under the assumption of infinite measurement statistics, a limited data sample  $f^*$  will be measured, which suffers from statistical fluctuations. Depending on  $K$ , the solution  $x$  can be extremely sensitive even to small changes in  $f$ . Integral equations, i.e. also the forward folding from  $x$  to  $f$ , in general tend to damp especially higher frequencies. That means that high frequency fluctuations in  $x$  are damped during the integration with  $K$ ; hence,  $f$  is smoother than  $x$ . But, the reverse is of interest: while solving the inverse problem exactly, higher frequencies present in  $f$  will be amplified. The higher the frequency is, the larger the amplification will be. Even small random fluctuations in  $f$  can introduce large perturbations in the solution  $x$  (for further details confer [83, p. 7 et seqq.]). And, since in this analysis only  $f^*$  is known, which suffers from fluctuations, an exact solution of Eq.(4.7) will result in a cosmic ray energy spectrum with unphysical large oscillations. Hence, such an exact solution of the problem is not desired, apart from the question if it would be mathematical achievable anyway. Instead of an exact and unbiased solution, rather an ‘‘appropriate’’ one is wanted that does not reproduce every single fluctuation originating from the limited measurement time. Such an optimal solution is biased, but less oscillating. Deconvolution strategies that regard these circumstances have to be used and will be introduced in Chapter 6.

In summary, in order to determine the fluxes of different cosmic ray particles, one has to solve a convolution problem. Due to the unavailability of a universal approach to solve such an integral equation under consideration of all circumstances, one has to find a method that yields an appropriate solution (cf. Chapter 6). All possible approaches have in common that the kernel function has to be known *a priori*. Accordingly, the first

<sup>13</sup>More precisely, in most analyses it is rather a matrix than a function.

following steps will be to compute the probabilities given in Eq.(4.6). The contributors to the kernel function can either be determined separately from each other based on simulations. Another approach would be to simulate everything directly from the beginning to the end, what means that particular particles with certain fixed energies would be assumed, which induce some hundreds or thousands of air showers, while the complete chain of CORSIKA shower simulation, CRES detector simulation, and reconstruction procedure is joined together. The resulting distribution of charged particles and muons would be taken as basis for the kernel function. The shower fluctuations, the efficiencies, and the reconstruction properties would be taken into account in only one combined simulation chain. For instance, this ansatz was adopted in [184]. However, analysing and ascertaining all contributors to the kernel function separately have the advantage that the different influences on the results can be examined independently. Moreover, the different sources of uncertainties can be kept separated. Furthermore, some contributors can be replaced without simulating the whole chain again; e.g. the part describing the shower development could be exchanged by a new one based on another interaction model than QGSJET-II-02. In contrast to these strong advantages, primarily only the not negligible additional effort of the separating method has to be mentioned. In the current analysis, balancing the pros and cons, the approach of analysing every contributor to the response function separately will be pursued. The effects of the intrinsic shower fluctuations will be parametrized in Section 5.1, while the efficiencies and properties of the reconstruction process will be analysed in Section 5.2.

## Chapter 5

# The Response Matrix

In Chapter 4, a mathematical formalism of the convolution problem was derived, in which the intrinsic shower fluctuations as well as the detector properties were considered in a so-called response matrix. Now, the probability distributions contributing to that response matrix will be parametrized (Sections 5.1 and 5.2). Subsequently, the response matrix will be computed and its reliability will be checked (Section 5.3).

### 5.1 Investigations to the Intrinsic Shower Fluctuations.

The goal of this section is to derive a parametrization for the intrinsic shower fluctuations, i.e. for  $s_n$  in Eq.(4.6). More precisely, the probability that an air shower exhibits a specific number of charged particles  $\log_{10}N_{\text{ch}}^{\text{true}}$  and muons  $\log_{10}N_{\mu}^{\text{true}}$  at the experiment's detection plane, if the air shower inducing particle is of a certain type and had a particular energy, shall be determined. For this purpose, one can<sup>1</sup> use the means of Monte Carlo simulations. In the following, the simulation strategy will be outlined (Section 5.1.1). Thereafter, the intrinsic shower fluctuations will be parametrized (Section 5.1.2).

#### 5.1.1 Simulation of the air shower development

The distributions that are required to parametrize the shower fluctuations will be determined using mono-energetic CORSIKA simulations that are based on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68]. Since the simulations are mono-energetic, no *a priori* assumptions of an underlying power law spectrum have to be made. In order to account for the problem of finite simulation statistics, the results are parametrized using appropriate functions, what induces a smoothing and extrapolates the rather unknown tails of the distributions. Moreover, these parametrizations allow to solve most of the occurring integrals analytically. The energy dependence of the parameters used in the parametrizations will be interpolated using appropriate functions. Hence, the mono-energetic results can be applied also to continuous spectra, or at least to a finer energy binning.

As mentioned in Footnote 10 of Chapter 4, it will not be possible to deconvolute the fluxes of any number of primaries. Some preliminary tests have shown that the distributions that belong to the respective primary particles are overlapping<sup>2</sup> in the two dimensional shower size plane. At a certain number of used primaries, this cross-over will be as large that the unfolded result will not be reasonable any longer. Some

---

<sup>1</sup>In case of KASCADE-Grande, one *has* to use simulations.

<sup>2</sup>The overlap results from the intrinsic shower fluctuations and the experiment's limited resolution.

Energy / PeV	Number of showers
2	6400
5	4800
10	3200
31.6	2400
100	1600
316	1200
1000	800
3160	400

Table 5.1: Listed are the numbers of simulated air showers for a certain primary energy (same numbers for all primaries).

tests, summarized in Section 5.3.2, have shown that the consideration of five primaries, each representing a different mass group, yield the best results of the deconvolution analysis. Only in a few exceptional circumstances problems occur, which can be solved by combining the deconvoluted fluxes of two or three primaries at the very end, such that the effective number of particles is four or three. These experiences conform with that of Ulrich [180], where also the spectra of five mass groups have been ascertained. For purpose of comparability, in case of the current analysis the same cosmic ray elements are chosen: protons (representing a hydrogen nucleus, and hence sometimes referred to as “hydrogen”), as well as helium, carbon (representing the carbon-nitrogen-oxygen group), silicon (representing the intermediate elemental group), and iron (representing the heavy component of cosmic rays) nuclei<sup>3</sup>.

In order to get a sufficient simulation statistics in a certain amount of time, the thinning option [87] of CORSIKA was enabled. This option’s objective is to save computing time. Below a selectable fraction of the primary energy, all particles produced in a single interaction will be represented by only one particle, to which suitable weights are assigned. The extensive tests performed by Ulrich [180] are applicable to the current simulation process. It was shown that a thinning level of  $10^{-6}$ , or even of  $10^{-5}$  can be used without significant distortion of the distributions of charged particles and muons. This conforms with the results of Hansen et al. [84], carried out at higher energies of  $10^{19}$  eV. It was stated that there is no significant dependence of the mean value as well as of the relative fluctuations of the distributions of electrons and muons on the chosen thinning level, if it is equal or less than  $10^{-6}$ . Accordingly, to be on the safe side, the chosen thinning level in this analysis will be  $10^{-6}$ . The number of air showers simulated for each primary of a certain energy is listed in Table 5.1. The relevant parameters and options that have been set in the CORSIKA program whilst the simulation process are summarized in Table 5.2.

The U.S. standard atmosphere parametrized by J. Linsley (cf. [88]) is used as atmospheric model. This atmosphere’s air has an assumed fractional-volume composition of 78.084% nitrogen, 20.9476% oxygen, 0.934% argon, and 0.0314% carbon dioxide at sea-level [183, p. 3]. At 100 m geometric altitude (close to the observation level of KASCADE-Grande of 110 m) the assumed air pressure amounts to 1001.2 mbar [183, p. 52]. As depicted in Fig. 5.1 the mean air pressure whilst the measurements is 1003 mbar with a RMS of 8.6 mbar. Taking into account the RMS, the mean conforms well with the

<sup>3</sup>The word “nuclei”/“nucleus” is sometimes omitted in this thesis. However, e.g. “iron” will every time stand for an “iron nucleus”.

Option or parameter	Setting
Atmospheric model	U.S. standard atmosphere
Magnetic field	Central Europe
Observation level	110 m a.s.l.
Azimuth angle	0° to 360°, uniformly distributed
Zenith angle	0° to 18°, distributed $\propto \sin(\vartheta) \cos(\vartheta)$
Low energy interaction model	FLUKA 2002.4
High energy interaction model	QGSJET-II-02
Transition between low and high energy interaction model	200 GeV
Energy threshold muons	100 MeV
Energy threshold electrons, positrons, gammas	3 MeV
Energy threshold hadrons	100 MeV
NKG option	Active: electromagnetic sub-cascade treated analytically.
EGS option	Active: all interactions of electrons, positrons and photons are calculated explicitly.
LONGI option	Switched on: longitudinal development of gammas, positrons, electrons, muons, hadrons, all charged nuclei and Cherenkov photons is sampled.
Thinning option	$10^{-6}$ ( $10^{-4}$ is default), what is the fraction of primary energy below which thinning becomes active.
Hadronic fragmentation flag	2 (default)

Table 5.2: The most important parameters and options set inside CORSIKA whilst all simulations.

values used in the U.S. standard atmosphere model. Hence this model can be used in the simulation procedure<sup>4</sup>. However, it has to be verified whether or not the changes of the pressure within the RMS can influence the deconvolution result. Furthermore, possible variations in the temperature have to be taken under consideration.

Firstly, the influence on the muon number will be examined. Maier [131] has found that the muon number at observation level does not depend significantly on the air pressure at observation level. On the contrary, there is a dependence on the altitude of the 200 hPa isobar, on the variation in the height difference between the 300 hPa and 500 hPa isobars, and on the effective temperature [131, p. 85]. The variation in the muon number – in case of a fixed primary energy – due to the aforementioned influences is less than 2.5%. This correspond to a change in the logarithmic muon number of  $\Delta \log_{10} N_{\mu} \approx 0.01$ . The logarithmic muon bin width used in the deconvolution process is 0.07 and by this seven times larger than the influences possibly caused by the considered effects; hence, they are negligible. This was also stated by Ulrich et al. [181] based on simulations.

<sup>4</sup>From all the properties of the atmospheric models the pressure (and possibly the temperature) is the one with the most important influence on the measured observables, as will be shown immediate in the next steps, while the other parameters are either the same in the U.S. and the European atmosphere or only with insignificant influence. The pressure is comparable in both atmospheres, such that one can use the U.S. atmosphere implemented in CORSIKA instead of the true European one.

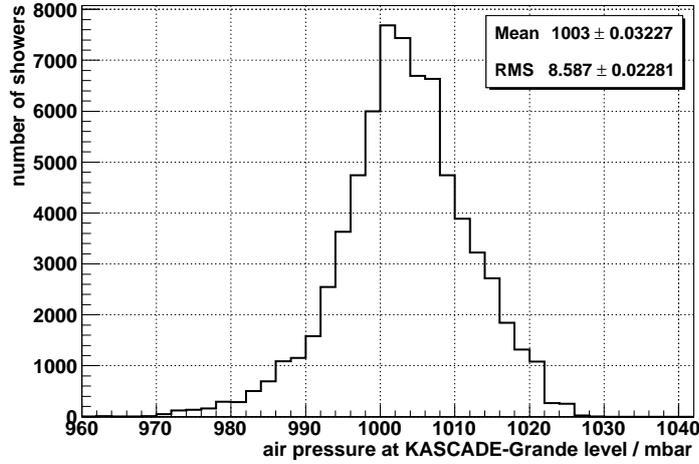


Figure 5.1: The air pressure distribution at the KASCADE-Grande level at time of data acquisition.

Now, the possible variations in the number of charged particles will be investigated. Based on measurements, Maier [131] and Antoni et al. [8] have shown that the number of electrons depends directly on the air pressure at observation level, as opposed to the muon number. Ulrich [179] has confirmed this with simulations. These results are applicable to the number of charged particles, because the electrons are by far the main contributors to this number. Furthermore, the contributing muons do not show a significant dependence as already stated afore. Consequently, it will not be distinguished between charged particles and electrons in what follows. The pressure whilst the measurements is distributed with an RMS of 8.6 mbar around the mean value of 1003 mbar. Maier [131, p. 61 et seq.] figured out that at fixed primary energy a spread of 8.8 mbar in the pressure induces a variation in the number of electrons, respectively charged particles, of 6% to 10%, mainly caused by changes in the attenuation length. Ulrich et al. [181] found effects in the order of 10% for typical atmospheric variations. More exactly, the number of charged particles arises mainly from a folding of the intrinsic shower fluctuations and the variations induced by the changes in the air pressure. The given percentage range corresponds to a modification of the logarithmic number of charged particles of  $\Delta \log_{10} N_{\text{ch}} = 0.025$  to  $\Delta \log_{10} N_{\text{ch}} = 0.041$ . Even assuming the worse scenario, the mentioned folding will be effective only inside the respective charged particle bin, which will have a logarithmic width of 0.07 in the final analysis. As a result, the influences of the variations in the air pressure can be neglected.

In summary, an appropriate simulation strategy to investigate the shower development is elaborated. The specific conditions present at the experiment's location are taken into account. Furthermore, influences originating from variations in the real atmosphere in comparison to the assumptions made in the atmospheric model are evaluated and found to be without importance for the final deconvolution analysis with respect to the used bin widths. Based on this experience, the intrinsic shower fluctuations will be analysed now.

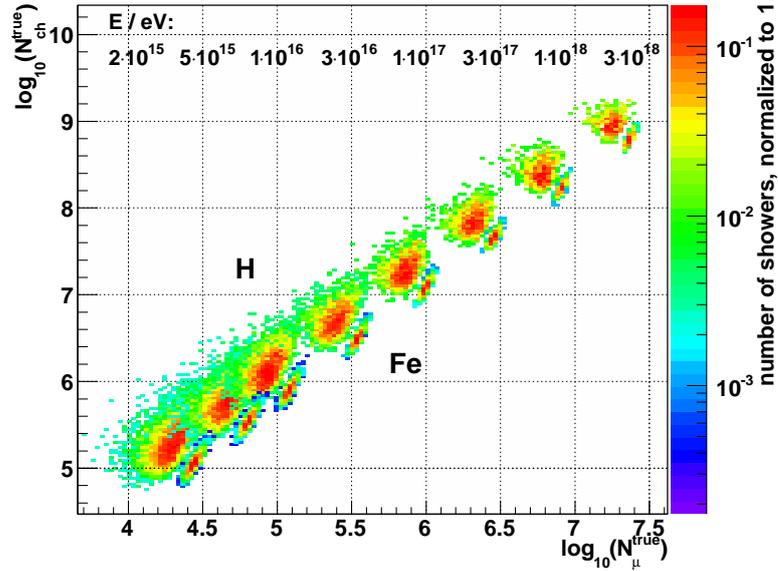


Figure 5.2: The effect of the intrinsic shower fluctuations on the two-dimensional shower size distribution  $\log_{10}N_{\text{ch}}^{\text{true}}$  vs.  $\log_{10}N_{\mu}^{\text{true}}$ . Shown are the simulated distributions for proton (upper distributions) and iron (lower distributions) induced air showers at certain primary energies, labelled above each distribution.

### 5.1.2 Parametrization of the intrinsic shower fluctuations

An air shower developing in the Earth’s atmosphere underlies fluctuations. These intrinsic shower fluctuations have to be known when deriving the fluxes of cosmic rays by means of deconvolution (cf. Eqs.(4.4) to (4.6)). Using the simulation method explained in Section 5.1.1 yields to the two-dimensional shower size plane<sup>5</sup>  $\log_{10}N_{\text{ch}}^{\text{true}}$  vs.  $\log_{10}N_{\mu}^{\text{true}}$  depicted in Fig. 5.2. Whereas only the results in case of proton and iron induced air showers are shown, the simulations are actually performed for protons, as well as for helium, carbon, silicon, and iron nuclei. The used primary energies are given in the figure as well as in Table 5.1.

The spread in the distributions originates from the intrinsic shower fluctuations, and is smaller in case of iron induced showers since they can be understood as a superposition of 56 proton induced ones, as explained in Section 2.4. An iron induced air shower represents – roughly speaking – an average over 56 proton induced ones, and by this the statistical fluctuations are suppressed in comparison to those appearing whilst the development of a shower caused by a single proton. Furthermore, it is obvious that the number of charged particles and the number of muons are correlated, as the distributions are elliptical.

In a first step, the one-dimensional distribution of charged particles (corresponding to a projection of the two-dimensional distribution on the y-axis) will be analysed. In a second step, the complete two-dimensional distribution will be parametrized including all correlations.

<sup>5</sup>In some Figures of this section the superscript “true” will be omitted, since it is contextually obvious that true sizes are used.

### Parametrization of the $\log_{10} N_{\text{ch}}^{\text{true}}$ distribution

In Fig. 5.3, exemplarily, the spread in the number of charged particles that are exhibited at observation level whilst the development of proton induced air showers with a primary energy of 2 PeV is depicted. The distribution bases on 6400 showers simulated following the strategy outlined in Section 5.1.1. The shown Gaussian fit describes the distribution only poorly ( $\chi^2/ndf = 376/63 \approx 6$ ). This is also reflected by the skewness of 0.4, what means an asymmetry in the tails: in case of that positive skew a shift to the right in comparison to a Gaussian distribution. The kurtosis of -0.1 indicates shorter tails in comparison to those of a Gaussian. Actually, that the intrinsic shower fluctuations are not simply Gaussian can be understood rather easily: the first randomized property of the shower development, the depth of the first interaction, depends on the interaction mean free path via an exponential law. The distance between the depth of first interaction and the maximum in the shower development is the next property varying by chance, now in rough approximation underlying a Gaussian distribution. Hence, already these two circumstances mean a convolution of a normal distribution and an exponential function. Therefore, it appears less likely that the resulting distributions at ground level are simply Gaussian. Instead, a convolution of different functional relations is expected.

There are some physically motivated approaches to describe the shower development, and by this the number of particles produced in this shower. For instance confer the Gaisser-Hillas-function, described in [75]. They all describe primarily the average longitudinal development, i.e. the most likely particle number rather than the complete probability distribution.

In order to find an appropriate parametrization, a phenomenological approach will be pursued instead of an analytical one. The left side of the distribution is characterized by a fast increase and an asymptotic convergence to the maximum. This behaviour is well known from the Gauss error function:

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-\tau^2} d\tau \quad , \quad z \in \mathbb{C} \quad . \quad (5.1)$$

It is related to the Gaussian cumulative distribution function  $F$  via:

$$F(x) = \frac{1}{2} \left( 1 + \text{erf} \left( \frac{x - \mu}{\sigma\sqrt{2}} \right) \right) \quad , \quad (5.2)$$

with the standard deviation  $\sigma$  and the expectation value  $\mu$ . The right tail is rather reminiscent of a simple power function. A fit based on a combination of the function given by Eq.(5.2) and a power function yielded better results. However, in some detailed tests it turned out that the result, especially in the region around the maximum, can still be improved<sup>6</sup> by an additional – however, only slight – variation through an exponential term. Consequently, a combination based on an error function, a power function, and an exponential function will be applied.

---

<sup>6</sup>In most (but not all) cases, the improvement was significant. In this context, one has to take into account that the exponential term modifies the function especially in regions around the distribution's maximum, and hence in a statistically relevant region.

The parametrization finally used is depicted in Fig. 5.4. The underlying probability density function is composed as follows:

$$\begin{aligned}
 p\left(\log_{10}N_{\text{ch}}^{\text{true}} \mid \log_{10}E\right) &= p_0 \\
 &\times \left(0.5 + 0.5 \times \operatorname{erf}\left(\frac{\log_{10}N_{\text{ch}}^{\text{true}} - p_1}{p_2}\right)\right) \\
 &\times \exp\left(\left(\log_{10}N_{\text{ch}}^{\text{true}} - p_4\right)p_3\right) \\
 &\times \left(p_4 - \log_{10}N_{\text{ch}}^{\text{true}}\right)^{p_5} .
 \end{aligned} \tag{5.3}$$

The parameters depend on the primary energy and vary with the shower inducing particle. A complete list of all determined parameters in case of the eight simulated energies and the five primaries are given in Appendix A.1.

The parameters are on their part parametrized, for each primary separately, as a function of the primary energy. By this the results obtained based on the individual mono-energetic CORSIKA showers are interpolated such that the charged particle distribution is now computable at every energy out of the observed interval. The mono-energetic simulations' advantage that one does not have to make a *ad hoc* supposition about the underlying cosmic ray spectrum continues to exist when transferring the results as mentioned to arbitrary energies by means of an interpolation. The parameters  $p_3$  and  $p_5$  were found to be independent of the energy in the sense that they can be kept constant to an average value without worsening the quality of the description. In addition to it,  $p_5$  is also insensitive to the type of the primary particle within the parameters' uncertainties. In that context, a good fit quality is assumed in case of chi-squares per degree of freedom at around 1, i.e. values out of the interval [0.6, 1.4], except for a few cases where such a quality is not achievable despite all the care taken. However, one has to keep in mind that the chi-square test gets meaningless in case of very low statistics; hence, in those cases the quality of the description is additionally judged by means of a Kolmogorov-Smirnov test [117, 163]. The energy dependence of the parameters  $p_1$ ,  $p_2$  and  $p_4$  of Eq.(5.3) is interpolated using polynomial functions of a certain degree<sup>7</sup>  $d$ :

$$f(x) = a_0 + a_1x^1 + \dots + a_{d-1}x^{d-1} + a_dx^d, d \in \mathbb{N}, \tag{5.4}$$

where  $f$  represents the parameter  $p_i$  that depends on the energy  $x$ . This is shown exemplarily in case of protons in Fig. 5.5.

The resulting parametrization of the  $\log_{10}N_{\text{ch}}^{\text{true}}$  distribution based on Eq.(5.3), where the energy dependence of the parameters is polynomial interpolated according to Eq.(5.4), is depicted in Fig. 5.6 in case of protons or iron nuclei with 2 PeV primary energy. The chi-squares per degree of freedom of the fits are shown in Fig. 5.7. In comparison with Fig. 5.4, there is no significant difference discernible between the optimal fit with free parameters and the fit based on an interpolation of the parameters. All ascertained parameter values of the polynomial interpolation based on Eq.(5.4) and the fits' chi-squares per degree of freedom are shown in Appendix A.2. The parametrization based on this interpolation will finally be used in the further steps.

<sup>7</sup>The degree was chosen such that the chi-square per degree of freedom of the polynomial fit is around 1.

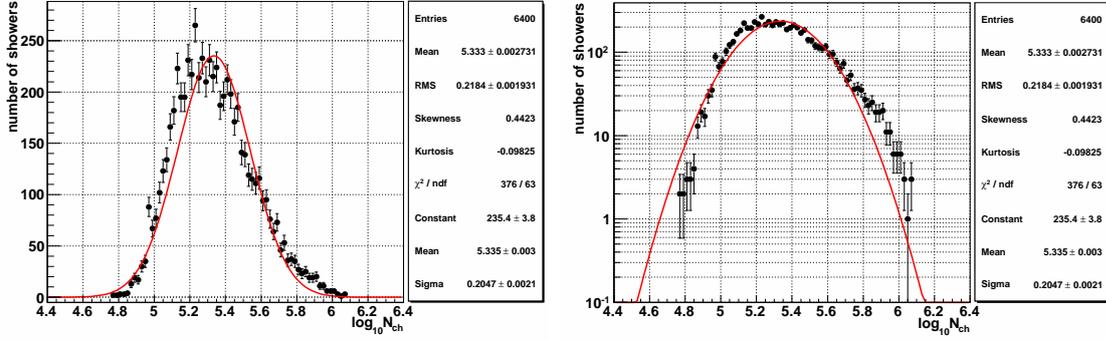


Figure 5.3: The charged particle distribution at observation level based on the simulation of 6400 proton induced air showers with a primary energy of 2 PeV. On the right panel in logarithmic scale of y-axis. The fit based on a Gaussian function describes the distribution only poorly.

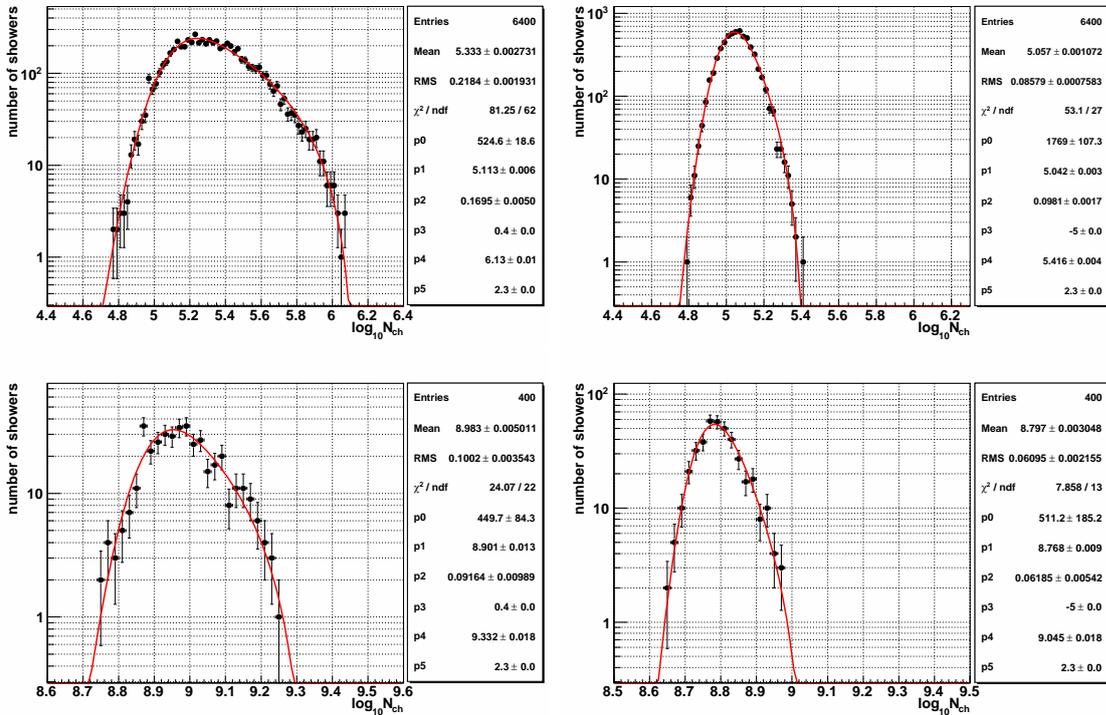


Figure 5.4: Final parametrization of the charged particle distribution based on Eq.(5.3). On the left panels in case of proton induced air showers, on the right panels in case of iron induced ones. The primary energy amounts 2 PeV (top) respectively 3.16 EeV (bottom).

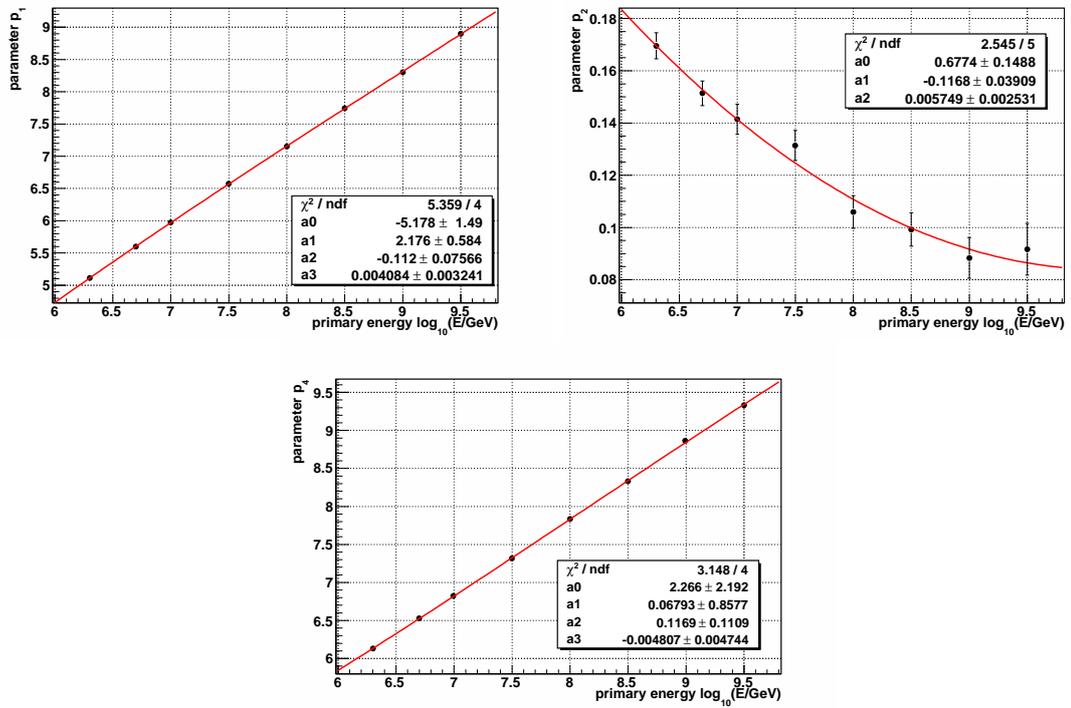


Figure 5.5: The interpolation of the energy dependence of the parameters  $p_1$ ,  $p_2$  and  $p_4$  of Eq.(5.3) with polynomials according to Eq.(5.4) of degree 3, 2, and 3 respectively, exemplarily in case of protons.

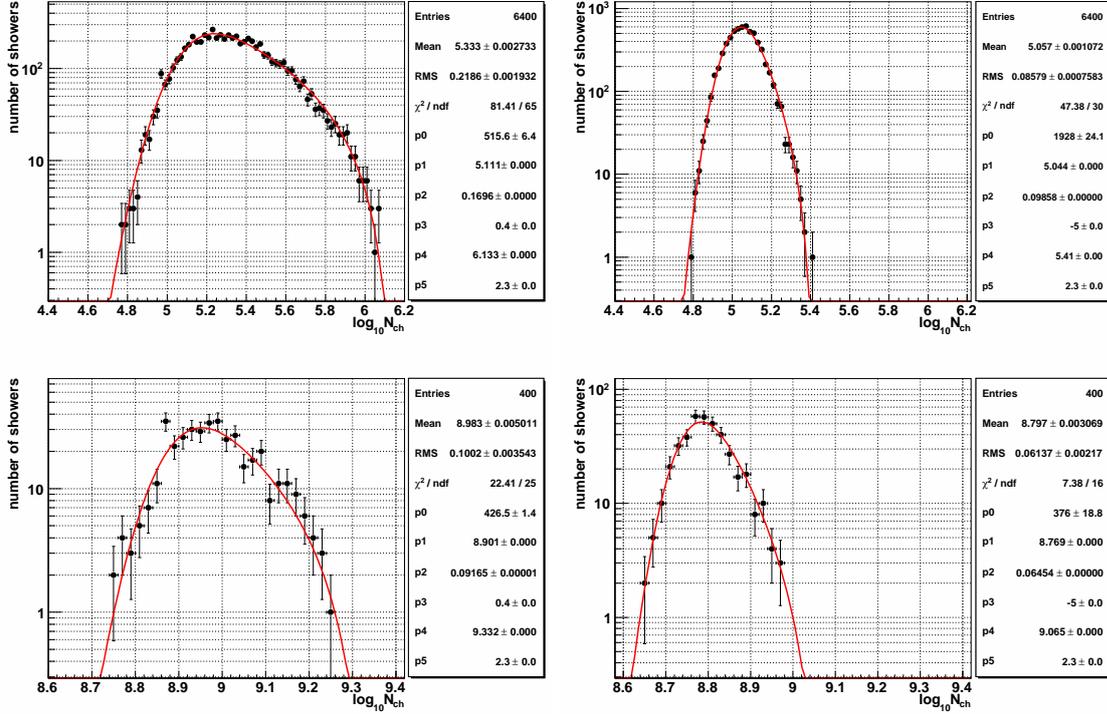


Figure 5.6: Final parametrization of the charged particle distribution based on Eq.(5.3), where the parameters themselves are parametrized with polynomials as a function of the primary energy according to Eq.(5.4). On the left panels exemplarily in case of proton induced air showers, on the right panels in case of iron induced ones. The primary energy amounts 2 PeV (top) respectively 3.16 EeV (bottom).

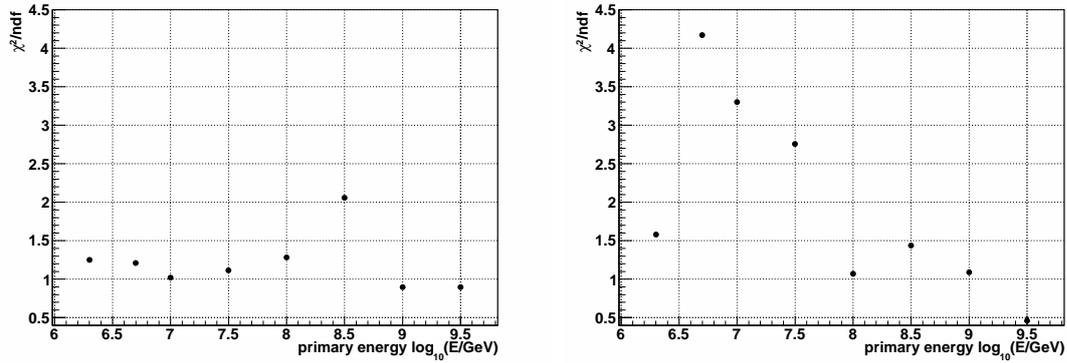


Figure 5.7: Chi-squares per degree of freedom – for all simulated energies – of the final parametrization of the charged particle distribution based on Eq.(5.3), where the parameters themselves are parametrized with polynomials as a function of the primary energy according to Eq.(5.4), exemplarily in case of proton (left panel) or iron (right panel) induced air showers.

### Parametrization of the $\log_{10}N_{\text{ch}}^{\text{true}}-\log_{10}N_{\mu}^{\text{true}}$ distribution

As already mentioned in the introduction to this section, there is a significant correlation between the numbers of charged particles  $\log_{10}N_{\text{ch}}^{\text{true}}$  and muons  $\log_{10}N_{\mu}^{\text{true}}$ . The strategy was to parametrize in a first step the  $\log_{10}N_{\text{ch}}^{\text{true}}$  distribution omitting all correlations, while in a second step the correlated  $\log_{10}N_{\text{ch}}^{\text{true}}-\log_{10}N_{\mu}^{\text{true}}$  distribution should be analysed. The latter step will be shown in what follows.

The muon number of an air shower induced by a certain cosmic ray particle depends above all on the primary energy, as can be seen in Fig. 5.2. The dependence on the charged particle number in context of a correlation between both observables is a secondary impact on the muon number distribution. Focusing on the first mentioned energy dependence and taking into account the message of Fig. 5.2, it is obvious that the muon number gets larger with increasing primary energy. To analyse the second dependence, it is assumed that the primary energy of a certain cosmic ray particle is fixed. Due to the intrinsic shower fluctuations, the muon number varies from shower to shower according to a probability distribution function. Nevertheless, there will be a most probable muon number specific for this combination of particle type and primary energy. For instance, in case of the 6400 simulated proton induced air showers with an energy of 2 PeV (cf. Table 5.1) the possible muon numbers range roughly from  $\log_{10}N_{\mu}^{\text{true}} = 3.6$  to  $\log_{10}N_{\mu}^{\text{true}} = 4.6$  around a maximum at about  $\log_{10}N_{\mu}^{\text{true}} = 4.3$ , as can be seen from Fig. 5.2. At an approximate estimate half of the simulated showers exhibit a muon number larger than or equal to the most probable one. More general, the fraction  $R$  of showers exhibiting muon numbers  $\log_{10}N_{\mu}^{\text{true}}$  above a specific threshold  $\log_{10}N_{\mu,t}^{\text{true}}$  in relation to all simulated ones can be described mathematically:

$$R\left(\log_{10}N_{\mu,t}^{\text{true}} \mid \log_{10}E\right) = \int_{\log_{10}N_{\mu,t}^{\text{true}}}^{+\infty} p\left(\log_{10}N_{\mu}^{\text{true}}\right) d\log_{10}N_{\mu}^{\text{true}} \quad , \quad (5.5)$$

where  $p\left(\log_{10}N_{\mu}^{\text{true}}\right)$  is the aforementioned normalized probability density function of the muon number, while any correlation with the charged particle number is still neglected.

In order to regard the positive<sup>8</sup> correlation between both observables, one has to take into account that an increase in the muon number is equivalent to an increase in the charged particle number. That means that, on the supposition that the threshold  $\log_{10}N_{\mu,t}^{\text{true}}$  is fixed to a certain value, there are more showers with muon numbers above this threshold if the average charged particle number is large, than in case of small charged particle numbers correlating with rather small muon numbers. Hence, the ratio  $R$  will increase with growing numbers of charged particles. The ratio of showers exhibiting muon numbers above a specific threshold  $\log_{10}N_{\mu,t}^{\text{true}}$  in relation to all simulated ones, now under consideration of the correlation with the number of charged particles  $\log_{10}N_{\text{ch}}^{\text{true}}$ , is given by:

$$R\left(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu,t}^{\text{true}} \mid \log_{10}E\right) = \frac{\int_{\log_{10}N_{\mu,t}^{\text{true}}}^{+\infty} p\left(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu}^{\text{true}}\right) d\log_{10}N_{\mu}^{\text{true}}}{\int_{-\infty}^{+\infty} p\left(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu}^{\text{true}}\right) d\log_{10}N_{\mu}^{\text{true}}} \quad , \quad (5.6)$$

<sup>8</sup>That the correlation is positive can be concluded from the distributions' shapes depicted in Fig. 5.2.

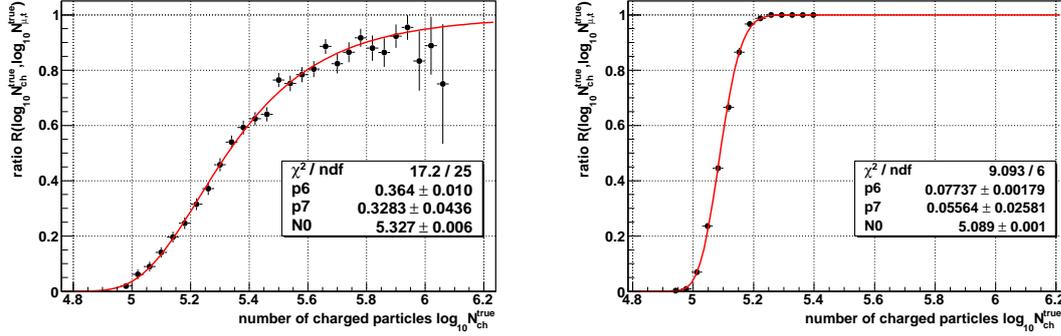


Figure 5.8: The ratio  $R$  of the number of showers exhibiting muon numbers above a certain threshold to the number of all showers as a function of the charged particle number, on the left panel in case of protons, on the right in case of iron induced air showers with a primary energy of 2 PeV. The threshold was kept fixed to  $\log_{10} N_{\mu,t}^{\text{true}} = 4.30$  in case of protons and  $\log_{10} N_{\mu,t}^{\text{true}} = 4.45$  for iron nuclei, such that the ratio  $R$  depends only on the number of charged particles now. The ratio is fitted by a function according to Eq.(5.7).

where  $p(\log_{10} N_{\text{ch}}^{\text{true}}, \log_{10} N_{\mu}^{\text{true}})$  is again the probability density function of the muon number, but now considering also the correlated charged particle number.

In Fig. 5.8, exemplarily for proton or iron showers with 2 PeV primary energy, the ratio  $R$  is depicted as a function of the charged particle number, while the threshold is kept constant to  $\log_{10} N_{\mu,t}^{\text{true}} = 4.30$  respectively  $\log_{10} N_{\mu,t}^{\text{true}} = 4.45$ . These values correspond to the respective muon numbers with maximal probability roughly estimated based on Fig. 5.2. Due to the fact that the threshold is a constant in that case,  $R$  can only vary with the charged particle number; and it does so. This is again a proof for the correlation between both observables, because otherwise the ratio would be a constant.

The functional relationship between  $R$  and  $\log_{10} N_{\text{ch}}^{\text{true}}$  is reminiscent of a Gauss error function (cf. Eq.(5.1)) because of the evident asymptotic tails. However, the ratio converges rather fast to the marginal values  $R = 0$  and  $R = 1$ , while in the middle the derivative is comparatively small. To account for this, the error function that is used to parametrize the ratio is given a varying width by extending the denominator in the error function's argument by an appropriate term depending on the variable  $\log_{10} N_{\text{ch}}^{\text{true}}$ :

$$R(\log_{10} N_{\text{ch}}^{\text{true}}, \log_{10} N_{\mu,t}^{\text{true}} | \log_{10} E) = \left( 0.5 + 0.5 \times \text{erf} \left( \frac{\log_{10} N_{\text{ch}}^{\text{true}} - \log_{10} N_0}{p_6 - p_7 \times (\log_{10} N_0 - \log_{10} N_{\text{ch}}^{\text{true}})} \right) \right) \quad (5.7)$$

While varying the threshold  $\log_{10} N_{\mu,t}^{\text{true}}$  the parameters  $p_6$  and  $p_7$  can be kept constant without worsening the fit result, whereas the parameter  $\log_{10} N_0$ , which determines the inflection point of the error function, has to be adapted. The underlying functional relation is exemplarily<sup>9</sup> shown in case of 2 PeV proton showers in Fig. 5.9 and is fitted by a quadratic function:

$$\log_{10} N_0(\log_{10} N_{\mu,t}^{\text{true}}) = b_0 + b_1 \times \log_{10} N_{\mu,t}^{\text{true}} - b_2 \times (\log_{10} N_{\mu,t}^{\text{true}})^2 \quad (5.8)$$

<sup>9</sup>In case of other primaries and energies a comparable behaviour can be seen.

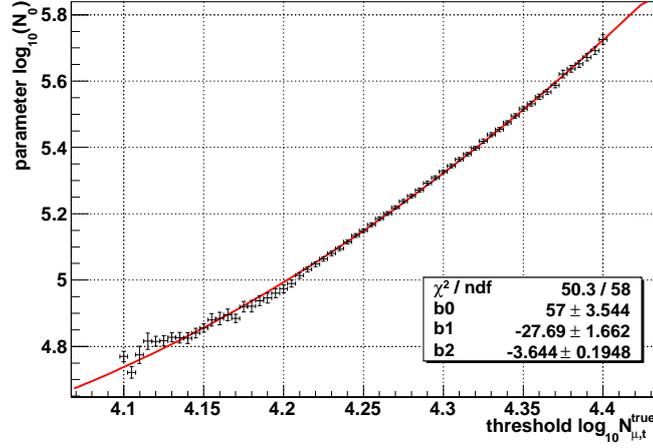


Figure 5.9: The dependence of the parameter  $\log_{10}N_0$  of Eq.(5.7) on the threshold value  $\log_{10}N_{\mu,t}^{\text{true}}$ , exemplarily shown in case of 2 PeV proton induced showers, is fitted by a quadratic function according to Eq.(5.8).

Taking into account the already parametrized probability density of charged particles  $p(\log_{10}N_{\text{ch}}^{\text{true}} | \log_{10}E)$ , given by Eq.(5.3), and the last gained insights, the probability that an air shower deposits  $\log_{10}N_{\text{ch}}^{\text{true}}$  charged particles and more than  $\log_{10}N_{\mu,t}^{\text{true}}$  muons can be written as:

$$P(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu}^{\text{true}} \geq \log_{10}N_{\mu,t}^{\text{true}} | \log_{10}E) = R(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu,t}^{\text{true}} | \log_{10}E) \times p(\log_{10}N_{\text{ch}}^{\text{true}} | \log_{10}E) d\log_{10}N_{\text{ch}}^{\text{true}}. \quad (5.9)$$

Assuming an infinitesimal small muon interval  $d\log_{10}N_{\mu}^{\text{true}}$  and setting up the equation twice, the muon number can actually be “bracketed” to one specific probability value instead of only giving limits for it. The resulting probability that a shower with primary energy  $E$  exhibits shower sizes  $\log_{10}N_{\text{ch}}^{\text{true}}$  and  $\log_{10}N_{\mu}^{\text{true}}$  is given by:

$$\begin{aligned} P(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu}^{\text{true}} | \log_{10}E) &= \\ &\left( R(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu}^{\text{true}} | \log_{10}E) \right. \\ &\quad \left. - R(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu}^{\text{true}} + d\log_{10}N_{\mu}^{\text{true}} | \log_{10}E) \right) \\ &\quad \times p(\log_{10}N_{\text{ch}}^{\text{true}} | \log_{10}E) d\log_{10}N_{\text{ch}}^{\text{true}} \\ &= s_n \times d\log_{10}N_{\text{ch}}^{\text{true}} d\log_{10}N_{\mu}^{\text{true}}, \end{aligned} \quad (5.10)$$

where  $s_n = s_n(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu}^{\text{true}} | \log_{10}E)$  are the intrinsic shower fluctuations defined in Eq.(4.6), whose parametrization is the primary objective of this section.

The remaining last step will be to determine the parameters of Eq.(5.10). The parameters of the probability density of charged particles  $p(\log_{10}N_{\text{ch}}^{\text{true}} | \log_{10}E)$ , given by Eq.(5.3), are already known. Instead of fitting a function according to Eq.(5.10) to the two-dimensional shower size distributions, it appeared more suitable to perform a fit to the one-dimensional distribution of muons. As all parameters of the charged particle probability density are known, the dependence on this observable can be eliminated from Eq.(5.10) by means of integration:

$$\begin{aligned}
 P(\log_{10}N_{\mu}^{\text{true}} | \log_{10}E) &= \int_{-\infty}^{+\infty} \frac{P(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu}^{\text{true}} | \log_{10}E)}{d\log_{10}N_{\text{ch}}^{\text{true}}} d\log_{10}N_{\text{ch}}^{\text{true}} \\
 &= \int_{-\infty}^{+\infty} \left( R(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu}^{\text{true}} | \log_{10}E) \right. \\
 &\quad \left. - R(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu}^{\text{true}} + d\log_{10}N_{\mu}^{\text{true}} | \log_{10}E) \right) \\
 &\quad \times p(\log_{10}N_{\text{ch}}^{\text{true}} | \log_{10}E) d\log_{10}N_{\text{ch}}^{\text{true}} .
 \end{aligned} \tag{5.11}$$

This function, going along with Eqs.(5.7) and (5.8), is used to fit the distribution of muons and by this to determine the remaining<sup>10</sup> parameters  $p_6$ ,  $p_7$ ,  $b_0$ ,  $b_1$  and  $b_2$ . The  $\log_{10}N_{\mu}^{\text{true}}$  distribution and the fit is illustrated exemplarily in Fig. 5.10 in case of proton and iron induced air shower having a primary energy of 2 PeV and 3.16 EeV.

In order to judge the fit quality, the chi-squares per degree of freedom are regarded. However, the remarks made in the previous section about the applicability of this test in case of rather low statistics still persist. For example, the iron distributions (Fig. 5.10 right panels) are fitted based on only a few sampling points. Hence, the chi-squares per degree of freedom should be interpreted with caution. As already mentioned, in those cases an additional Kolmogorov-Smirnov test [117, 163] is used. For instance, in case of the parametrization of the muon number distribution of iron induced air showers having an energy of 3.16 EeV the maximal Kolmogorov distance is computed to be 0.0118, while the number of effective entries counts 400. The null hypothesis assumes that the distribution sample comes from the hypothesized distribution given by the parametrization. In this example, the null hypothesis is rejected under a significance level<sup>11</sup> of 99% if the critical distance value of 0.0296 is exceeded, what is actually not the case. Thus, the parametrization is assumed to describe the muon distribution very well. Comparable tests are performed to the remaining primaries and energies, demanding a significance level of 95%. Moreover, all fits are verified subjectively by eye. A complete compilation of all parameters derived for the muon number distribution is given in Appendix B.1.

The parameters  $p_7$ ,  $b_1$  and  $b_2$  were found to be independent of the energy in the sense that they can be kept fixed to an average value without worsening the quality of the fit.

<sup>10</sup>Whilst the integration performed in Eq.(5.11), the parameters of the probability density of charged particles  $p(\log_{10}N_{\text{ch}}^{\text{true}} | \log_{10}E)$  are kept constant to the already derived ones from the previous section, listed in Appendix A.1. The infinitesimal small muon interval was set to  $d\log_{10}N_{\mu}^{\text{true}} = 0.02$ , what was found to be sufficiently small enough.

<sup>11</sup>A significance level of 99% means, roughly speaking, that one allows an 99% error probability whilst rejecting a hypothesis. There is a 99% chance that the hypothesis is falsely rejected although it is actually a right one. Nevertheless, despite this large allowed error probability the null hypothesis is not rejected in this example, what strongly confirms the properness of the fit.

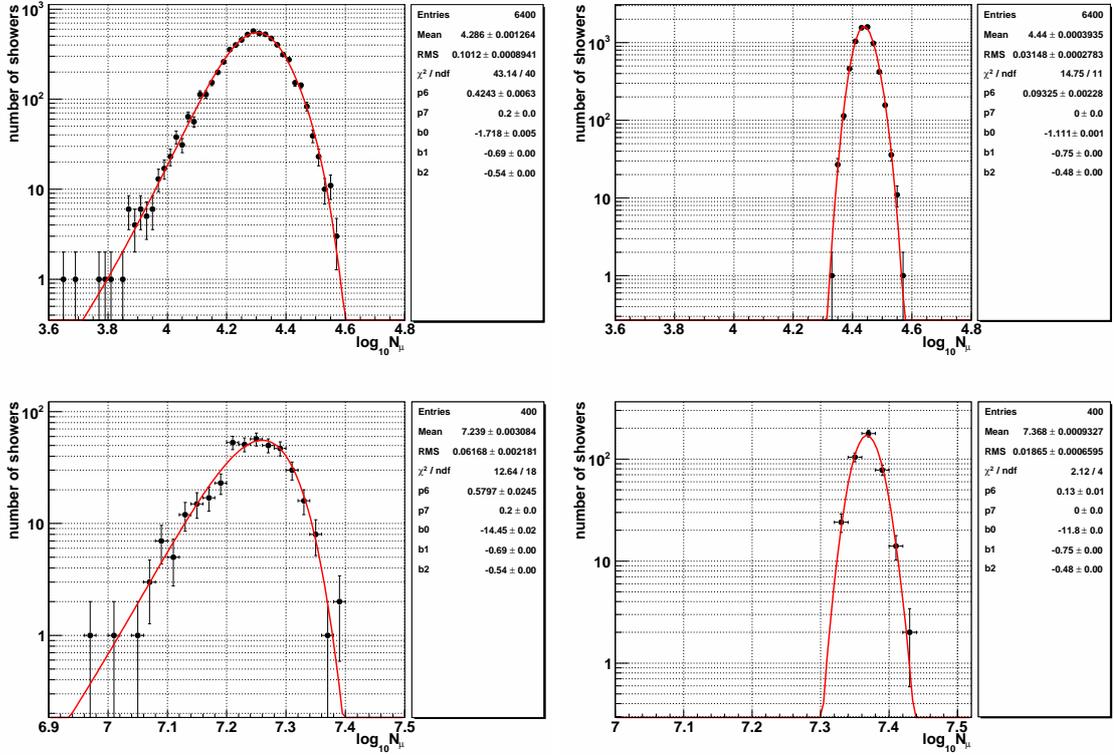


Figure 5.10: Final parametrization of the muon number distribution based on Eq.(5.11) going along with Eqs.(5.7) and (5.8). On the left panels in case of proton induced air showers, on the right panels in case of iron induced ones. The primary energy amounts 2 PeV (top) respectively 3.16 EeV (bottom).

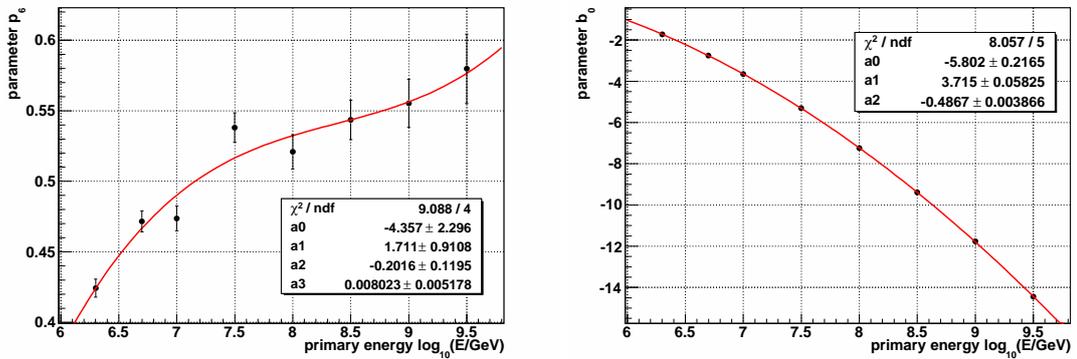


Figure 5.11: The interpolation of the energy dependence of the parameters  $p_6$  and  $b_0$  of Eq.(5.11) going along with Eqs.(5.7) and (5.8) with polynomials according to Eq.(5.4) of degree 3 and 2, exemplarily in case of protons.

The energy dependence of the parameters  $p_6$  and  $b_0$  is interpolated using polynomials according to Eq.(5.4), as already done in case of the energy depending parameters of the charged particle distribution (cf. previous section). The polynomials' constant coefficients are labelled with  $a_i$ , where  $i = 0 \dots d$ , and  $d \in \mathbb{N}$  is an appropriate degree. The results are illustrated exemplarily in case of protons in Fig. 5.11. The resulting parametrization of the  $\log_{10}N_{\mu}^{\text{true}}$  distribution based on Eq.(5.11) going along with Eqs.(5.7) and (5.8), where the energy dependence of the parameters is polynomial interpolated according to Eq.(5.4), is depicted in Fig. 5.12 in case of protons or iron nuclei with 2 PeV primary energy. The chi-squares per degree of freedom of the fits are shown in Fig. 5.13. In comparison with Fig. 5.10, there is no significant difference discernible between the optimal fit with free parameters and the fit based on an interpolation of the parameters. All ascertained parameter values of the polynomial interpolation based on Eq.(5.4) and the fits' chi-squares per degree of freedom are shown in Appendix B.2.

In Fig. 5.14, the quality of the final parametrization of the two-dimensional  $\log_{10}N_{\text{ch}}^{\text{true}} - \log_{10}N_{\mu}^{\text{true}}$  shower size distribution is reviewed. In the top panels, exemplarily the distributions, more precisely the isolines of same number of showers, exhibited by 2 PeV proton (left panel) or iron (right panel) showers are depicted ("bumpy" lines), overlaid with the final parametrization according to Eq.(5.10) (smooth lines). At the bottom panel, the distributions at all simulated energies are shown. The simulated distributions are very well described by the deduced parametrizations.

Recapitulating, the objective of this section, the mathematical access to the intrinsic shower fluctuations  $s_n = s_n(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu}^{\text{true}} | \log_{10}E)$  defined in Eq.(4.6), is achieved by the parametrization given in Eq.(5.10), whose parameters are determined by now. In order to complete the calculation of the response function introduced via Eq.(4.7), the computation of the experiment's properties, namely the efficiency  $\varepsilon_n(\log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu}^{\text{true}})$  and the reconstruction accuracy  $r_n(\log_{10}N_{\text{ch}}, \log_{10}N_{\mu} | \log_{10}N_{\text{ch}}^{\text{true}}, \log_{10}N_{\mu}^{\text{true}})$ , is still open. This will be carried out in the next section.

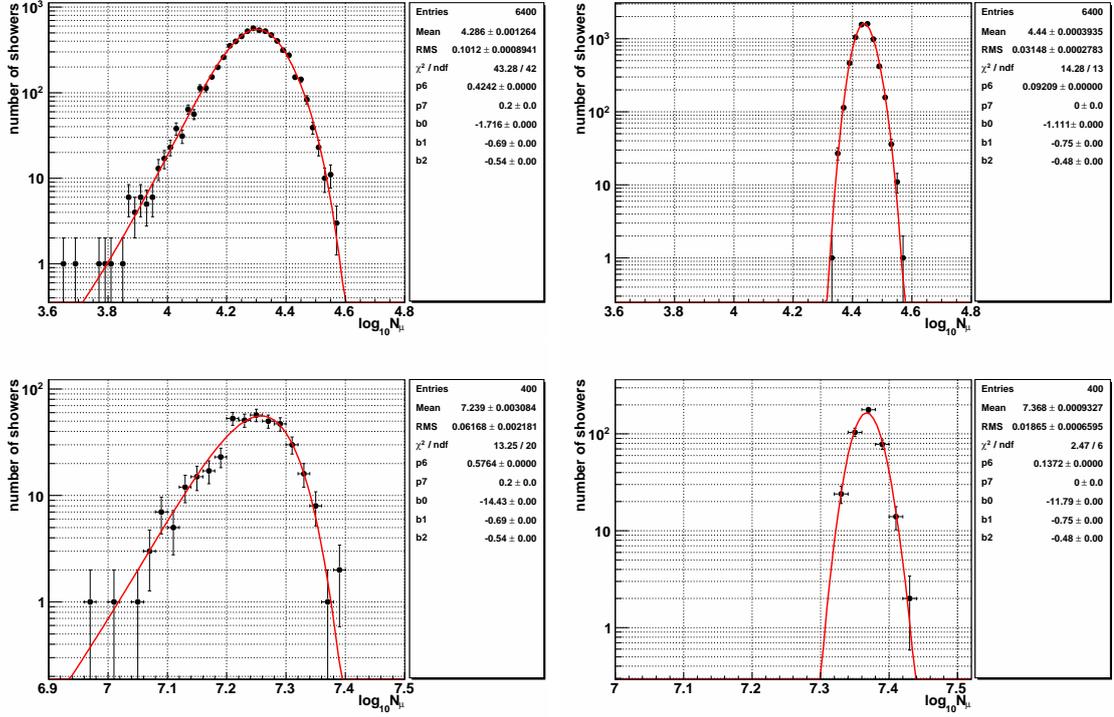


Figure 5.12: Final parametrization of the muon number distribution based on Eq.(5.11) going along with Eqs.(5.7) and (5.8), where the parameters themselves are parametrized with polynomials as a function of the primary energy according to Eq.(5.4). On the left panels exemplarily in case of proton induced air showers, on the right panels in case of iron induced ones. The primary energy amounts 2 PeV (top) respectively 3.16 EeV (bottom).

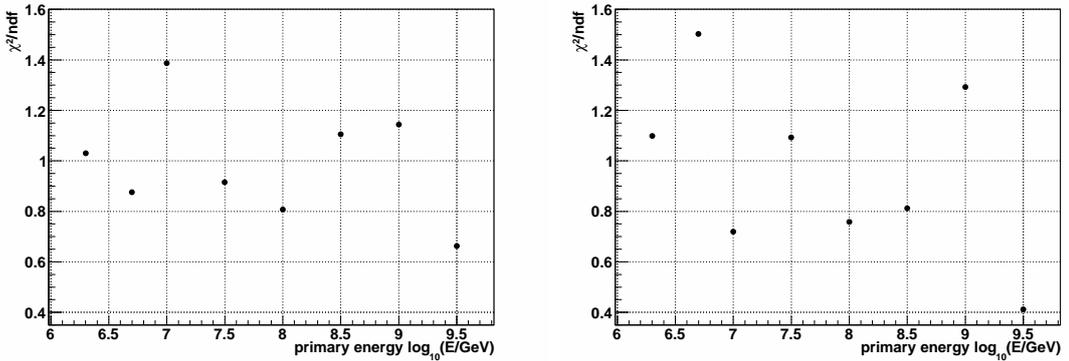


Figure 5.13: Chi-squares per degree of freedom – for all simulated energies – of the final parametrization of the muon number distribution based on Eq.(5.11) going along with Eqs.(5.7) and (5.8), where the parameters themselves are parametrized with polynomials as a function of the primary energy according to Eq.(5.4), exemplarily in case of proton (left panel) or iron (right panel) induced air showers.

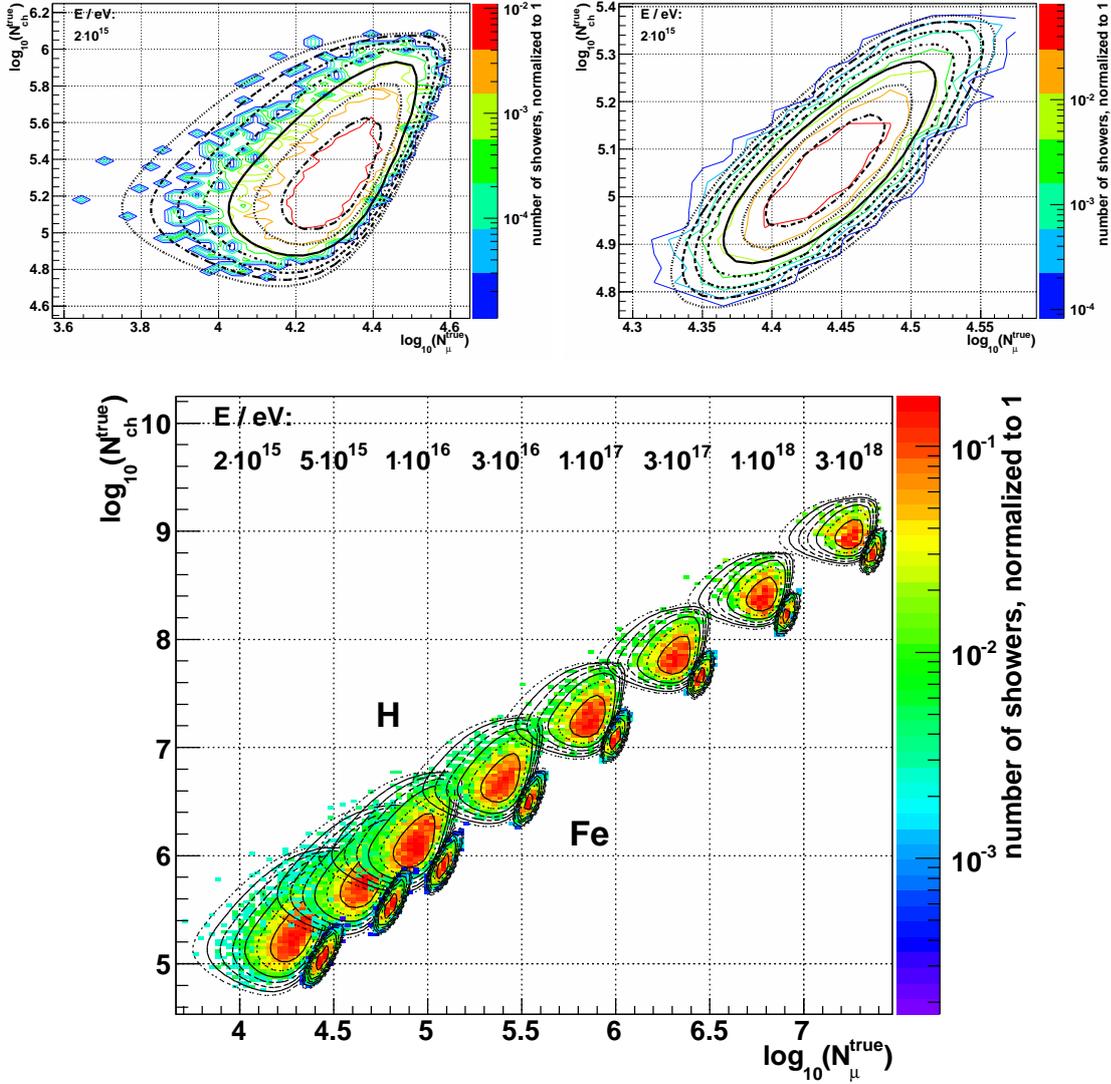


Figure 5.14: The contour lines of the simulated two-dimensional  $\log_{10}N_{ch}^{true}-\log_{10}N_{\mu}^{true}$  shower size distribution in case of proton (top left panel) and iron (top right panel) induced air showers with a primary energy of 2 PeV (“bumpy”, coloured lines). In addition, the parametrization according to Eq.(5.10) is shown (smoother lines), which describes the distribution well. The simulated shower size distributions at all used energies, labelled above each distribution, are illustrated at the bottom panel in case of proton (upper distributions) in comparison with iron (lower distributions) induced air showers. Again the parametrization according to Eq.(5.10) is overlaid (smooth lines).

## 5.2 Investigations to the Experiment's Properties

The object of this section will be to compute the remaining contributors to the response function introduced via Eqs.(4.6) and (4.7), i.e. the experiment's properties, namely the efficiency  $\varepsilon_n \left( \log_{10} N_{\text{ch}}^{\text{true}}, \log_{10} N_{\mu}^{\text{true}} \right)$  (see Section 5.2.2) and the reconstruction accuracy  $r_n \left( \log_{10} N_{\text{ch}}, \log_{10} N_{\mu} \mid \log_{10} N_{\text{ch}}^{\text{true}}, \log_{10} N_{\mu}^{\text{true}} \right)$  (see Section 5.2.3 and 5.2.4). Prior to this an appropriate strategy for the computation has to be stated: the respective experiment's properties will be derived by means of simulations, as will be outlined in Section 5.2.1.

### 5.2.1 Simulation of the experiment's properties

In order to determine the experiment's properties like efficiency and reconstruction accuracy, a set of CORSIKA [88] 6.307 shower simulations based on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68] is used. The main properties of the simulations are comparable to the ones used to parametrize the intrinsic shower fluctuations, as outlined in Section 5.1.1. However, the actually used simulations are not mono-energetic, but instead assuming a continuous energy spectrum following a power law with a differential index<sup>12</sup> of  $-2$ . The simulated primary energies range from  $1 \times 10^{14}$  eV to  $3.16 \times 10^{18}$  eV. Another difference is that the showers are entirely simulated, i.e. without using the thinning option explained in Section 5.1.1, since thinned showers cannot be treated by the used detector simulation program. The experiment's response to the CORSIKA showers is simulated using CRES<sup>13</sup> 1.16/07, which bases on GEANT 3.21 [46, 76] detector description and simulation tool. By this means, all experiment specific properties, for instance detector characteristics and electronics features, are regarded. At the very end, the respective air shower observables are reconstructed, starting from the simulated signals in the detectors, with the experiment's standard reconstruction software<sup>14</sup>. From all reconstructed events only those are used that pass the quality cuts summarized in Section 3.3.

The reliability of this simulation procedure was intensively tested by the KASCADE-Grande collaboration within many analyses. While the statistical effects like the detector resolution can be simulated by the detector simulation, *a priori* systematic effects can possibly remain unconsidered. For instance, a systematic distortion affecting the calculation of the angles of incidence could yield to preferred angle ranges, and by this could influence the cosmic rays flux determination if angles at the upper limit of the used zenith angle interval  $[0^\circ, 18^\circ]$  are e.g. pulled to values above the upper threshold. However, in a model independent comparison [14] between the results of the KASCADE and the KASCADE-Grande experiment, no systematic problems could be identified in case of the zenith angle and the shower core reconstruction. The overall conclusion is that the available simulation set yields the same detector signals and data structures as in case of measured events, and by this it is an appropriate means for the investigations that are object of this section.

<sup>12</sup>The bin widths whilst the analyses in this section are chosen as small that slight deviations from the right index of the power law spectrum do not have any significant influence on the results.

<sup>13</sup>Cosmic Ray Event Simulation, a program package developed for the KASCADE [10] detector simulation.

<sup>14</sup>Internally named KRETA, KASCADE REconstruction for extensive Air showers.

### 5.2.2 The experiment's efficiency

In the final unfolding analysis, the charged particle and muon number ranges are set such that the combined trigger and reconstruction efficiency for the respective air showers is next to 100%. However, for the calculation of the response function, also smaller shower sizes below these limits are considered in order to account for the shower fluctuations and the resulting migration effects. Thus, a parametrization of the combined efficiency is necessary. Firstly, some technical things will be discussed, then it will be focused on the trigger condition, and thereafter on the combined trigger and reconstruction efficiency.

#### Some technical things about the calculation of efficiencies

The commonly used technique for the computation of efficiencies seems often to be applied somewhat imprecisely concerning the consideration of uncertainties. Hence, it will be focused on that problem in the following. The general idea to derive for example the efficiency vs. energy<sup>15</sup> is to produce two histograms. While in one histogram the energy distribution is plotted for all air showers, in the second histogram only the energies of events are plotted that pass the respective selection conditions or cuts. By bin wise dividing the latter histogram by the first one, one gets an estimate for the efficiency. Mathematically, the efficiency in a bin  $i$  is given by  $\varepsilon_i = c_i/a_i$  if  $a_i$  is the number of entries in the histogram containing all showers, while  $c_i$  refers to the histogram after cuts.

While the procedure itself is consistent, the error calculation is sometimes carried out in a wrong way. The standard access seems to base on Poisson errors. By this, assuming a large sample statistics, the error is given by the square root of the number or entries, respectively  $\sigma_{c_i} = \sqrt{c_i}$  and  $\sigma_{a_i} = \sqrt{a_i}$ . By means of Gaussian error propagation law, the resulting uncertainty of the efficiency would be:

$$\sigma_\varepsilon = \sqrt{\left(\frac{1}{a_i}\sqrt{c_i}\right)^2 + \left(\frac{c_i}{a_i^2}\sqrt{a_i}\right)^2} = \sqrt{\frac{c_i(a_i + c_i)}{a_i^3}}. \quad (5.12)$$

This yields in some cases wrong errors as shown by this example: If in a certain bin  $i$  no event passes the cuts, i.e.  $c_i = 0$ , while there are in total more than zero events, i.e.  $a_i > 0$ , the resulting efficiency would be  $\varepsilon_i = 0 \pm 0$  without any uncertainty! Another problem is that e.g. in case of  $c_i = a_i$  the efficiency would be  $\varepsilon_i = 1 \pm \sqrt{2/a}$ . If additional  $a_i = 1$ , the resulting efficiency would be  $\varepsilon_i = 1 \pm 1.4$ . Hence, this would allow physically not possible efficiencies above 1 or below 0.

An alternative access would be to assume binomial errors. Whether or not an individual event passes the cuts is underlying a binomial probability distribution; the efficiency is the probability to succeed a cut. The number of events  $c_i$  passing the selection conditions is a random variable, while the total number of events  $a_i$  and the efficiency  $\varepsilon_i$  are parameters. The standard deviation of the binomially distributed number of entries  $c_i$  in a certain bin of the histogram is given by  $\sigma_{c_i} = \sqrt{a_i \varepsilon_{i,t} (1 - \varepsilon_{i,t})}$ , where  $\varepsilon_{i,t}$  is the true efficiency. Based on this, the statistical error of the efficiency can be calculated by means of Gaussian error propagation as follows:

$$\sigma_{\varepsilon_i} = \sqrt{\left(\frac{1}{a_i}\sqrt{a_i \varepsilon_{i,t} (1 - \varepsilon_{i,t})}\right)^2} = \frac{1}{a_i}\sqrt{c_i \left(1 - \frac{c_i}{a_i}\right)}, \quad (5.13)$$

<sup>15</sup>Or, instead, vs. the charged particle or muon number.

where the true efficiency was replaced by the estimated one  $\varepsilon_{i,t} \approx \varepsilon_i = c_i/a_i$ . Although the true distribution of the efficiency is not necessarily symmetric, the afore-calculated standard deviation is a sufficient approximation for the real uncertainty. The problem of the Poissonian approach that the errors can extend the efficiency to unphysical ranges is not present here. However, also the binomial error is zero, and by this predicts perfect certainty in the computation of the efficiency, if  $c_i = a_i$  or  $c_i = 0$ .

In contrast to the aforementioned approaches that base on frequentist statistics, there are also methods that are based on the Bayes' theorem resulting in all over correct errors. However, the final results, namely the parameters of the efficiency's parametrization, were found to be the same as in case of the binomial approach. Hence, the remaining problem of the binomial method will not worsen the final result, whereas the problems of the Poisson approach would indeed allow unphysical regions and by this could have a negative impact.

Consequently, the errors of the efficiencies will be computed in a binomial way according to Eq.(5.13). The estimated efficiencies will be fitted based on the least squares method in order to obtain a parametrization. The Bayesian approach for the error calculation will only be used as cross-check, what can easily be done with the ROOT [45] routine `TGraphAsymmErrors::BayesDivide`, which assumes a binomial model in combination with a uniform prior, but carries out the computation considering the Bayes' theorem. Furthermore, the functions used for the parametrizations are chosen such that only physical reasonable codomain ranges are covered.

### The trigger efficiency

The condition for a KASCADE-Grande detector station to trigger itself is a four-fold or seven-fold coincidence in the corresponding trigger cluster. There are overall 18 such trigger cells, each consisting of a hexagon of seven<sup>16</sup> stations. The condition used in this analysis is the seven-fold coincidence, internally also named "7/7 trigger". Additionally, there must be an energy deposit in at least one charged particle detector (Grande detectors) and one muon detector (KASCADE array). While the first condition is automatically fulfilled when a 7/7 trigger is given, the second one implies that also the KASCADE array has detected the shower, what is necessary for the measurement of the muon number, and is hence a second important trigger condition in the sense of the current unfolding analysis. Since the Grande stations do not discriminate between electrons and muons, the total number of charged particles is a criterion for a successful trigger.

In Fig. 5.15, the trigger efficiency as a function of the electron (left panel) or the muon number (right panel) is depicted. It is obvious that there is a dependence on both shower sizes. As explained at the beginning of Section 4.1, the muon to electron ratio in iron induced air showers is higher than in proton induced ones with the same primary energy. At same electron shower size, there will be more muons if the primary is an iron nucleus than in case of a proton. Since the KASCADE-Grande trigger is sensitive to both the electron and the muon number, a higher muon number means a higher trigger probability. Consequently, the trigger efficiency for iron nuclei is shifted to smaller electron numbers in comparison to the one of protons, confer Fig. 5.15, left panel. The other way around, in case of a fixed muon number, proton induced air showers will exhibit more electrons than iron induced ones. Hence, the trigger efficiency for iron nuclei is shifted to higher muon numbers, confer Fig. 5.15, right panel.

<sup>16</sup>One of the outer clusters consists only of 6 stations.

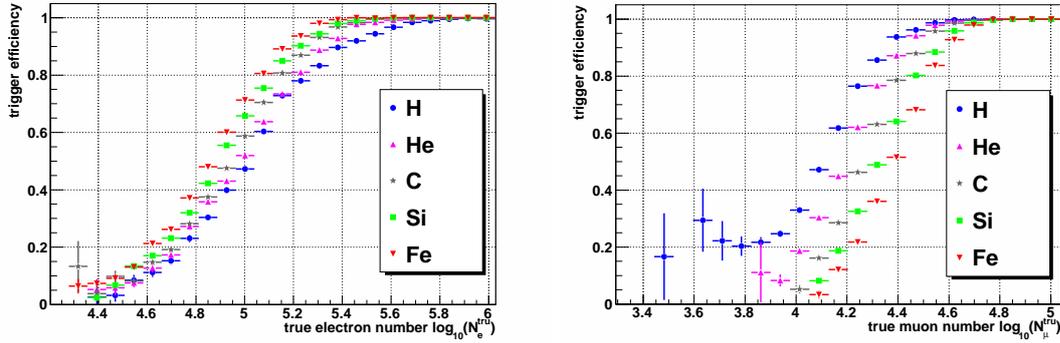


Figure 5.15: The trigger efficiency as a function of the true electron number (left) and the true muon number (right).

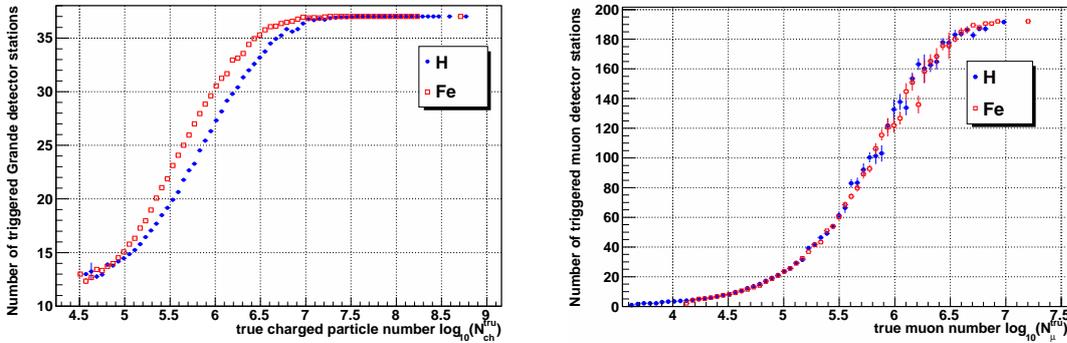


Figure 5.16: The number of triggered KASCADE-Grande detector stations as a function of the true number of charged particles (left) as well as the number of triggered KASCADE muon detector stations as a function of the true muon number (right).

In Fig. 5.16, the number of triggered KASCADE-Grande charged particle detector stations as a function of the true number of charged particles is illustrated (left panel), as well as the number of triggered KASCADE muon detector stations as a function of the true muon number (right panel). As one would expect, due to the earlier development of iron induced air showers in the atmosphere, they are laterally more extended and by this hitting more stations at ground level than similar proton induced ones. Due to this, at a fixed number of charged particles, iron induced showers hit more Grande stations than in case of protons. Hence, in case of iron nuclei, the 7/7 trigger condition will be fulfilled earlier. Figure 5.16, right panel, is only shown for the sake of completeness. The trigger condition for showers considered in the unfolding analysis only depends on the KASCADE-Grande charged particle detector stations. However, as mentioned afore, it is additionally demanded that there must be an energy deposit in at least one KASCADE muon detector station, such that the dependence shown in the right panel can impact indirectly on the trigger efficiency, too. Nevertheless, no significant difference can be seen between proton or iron induced showers<sup>17</sup>. For fixed muon numbers, the proton induced

<sup>17</sup>The proton curve starts at smaller muon numbers, but also stops at smaller ones in comparison to iron. This is an effect of the energy range chosen for the simulations: e.g. in case of the smallest simulated energy, proton showers have on average smaller muon numbers than in case of iron showers, such that the shown proton curve reaches smaller muon numbers than the iron one. Hence, this has a technical

air showers have on average higher primary energies than iron induced ones. Also the energy per particle is on average higher, what is important since the muon detectors have a threshold energy of 230 MeV. This all seems to compensate the effects of the laterally wider distribution of the iron induced air showers, such that the number of muon detector stations measuring a signal does not depend on the primary particle in case of equal total muon numbers.

### The combined trigger and reconstruction efficiency

The goal will now be to parametrize the combined trigger and reconstruction efficiency  $\varepsilon$  (such that it can be considered in the response function calculation, confer Eq.(4.6)), in the following simply called “efficiency” in contrast to the pure “trigger efficiency” discussed before. The additional reconstruction efficiency means both that an air shower has been reconstructed at all as well that the event passes all quality cuts summarized in Section 3.3. Since the two-dimensional distribution of charged particles and muons is used as basis for the unfolding analysis, these two sizes have to be treated in the computation of the response function. Hence, the efficiency in dependence on these two shower sizes have to be parametrized.

In Fig. 5.17, the efficiency as a function of the true charged particle (left panel) or muon number (right panel) is illustrated. The efficiency is similar to the pure trigger efficiency depicted in Fig. 5.15. The additional contribution of the reconstruction efficiency seems to be rather slight. The quality cuts or conditions are manifold and can be correlated. It will be difficult to analyse all these influences in a way that one could in detail interpret the impact on the efficiency. Hence, the depicted efficiency will not be discussed further than already done.

As shown, the efficiency depends on both contributors to the two-dimensional shower sizes plane, more precisely on the charged particle and muon number. In Fig. 5.18, top, the efficiency as a function of these two observables at the same time is shown, exemplarily for protons (left panel) and iron nuclei (right panel). The correlation between the shower sizes was found to be rather unimportant in case of the efficiency determination in the sense that there will not be significant impacts on the final response function. Shower sizes below full efficiency are only considered to account for possible migration effects. Consequently, a 100% correct description of the efficiency is anyway not necessary.

Considering Fig. 5.17, the dependence of the efficiency on the two shower sizes reminds of the graph of a Gauss error function, confer Eq.(5.1). Treating both observables  $\log_{10}(N_{\text{ch}}^{\text{tru}})$  and  $\log_{10}(N_{\mu}^{\text{tru}})$  as uncorrelated, the efficiency  $\varepsilon$  as a function of both these variables at the same time can be parametrized by a product of the two individual error functions:

$$\begin{aligned} \varepsilon \left( \log_{10}(N_{\text{ch}}^{\text{tru}}), \log_{10}(N_{\mu}^{\text{tru}}) \right) = & \\ & \left( 0.5 + 0.5 \times \text{erf} \left( p_0 \left( \log_{10}(N_{\mu}^{\text{tru}}) - p_1 \right) \right) \right) \\ & \times \\ & \left( 0.5 + 0.5 \times \text{erf} \left( p_2 \left( \log_{10}(N_{\text{ch}}^{\text{tru}}) - p_3 \right) \right) \right) . \end{aligned} \quad (5.14)$$

The parametrization is drawn in Fig. 5.18, bottom. All derived parameters are listed in Appendix C.

---

background and is not implicating a physically important result.

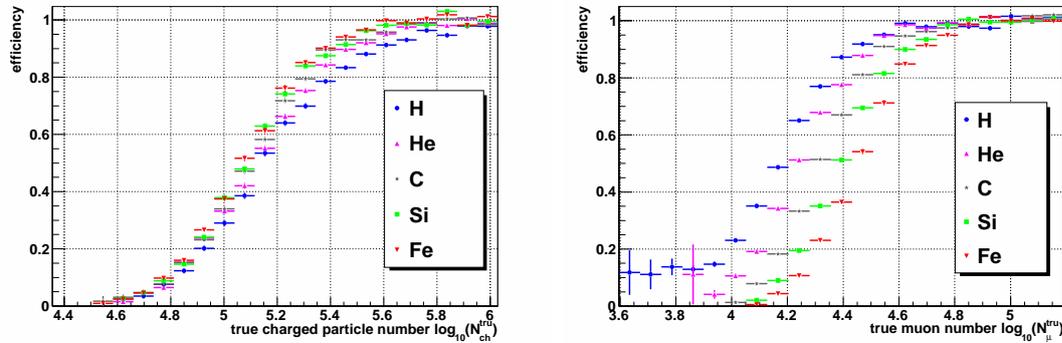


Figure 5.17: The efficiency as a function of the true charged particle number (left) and the true muon number (right).

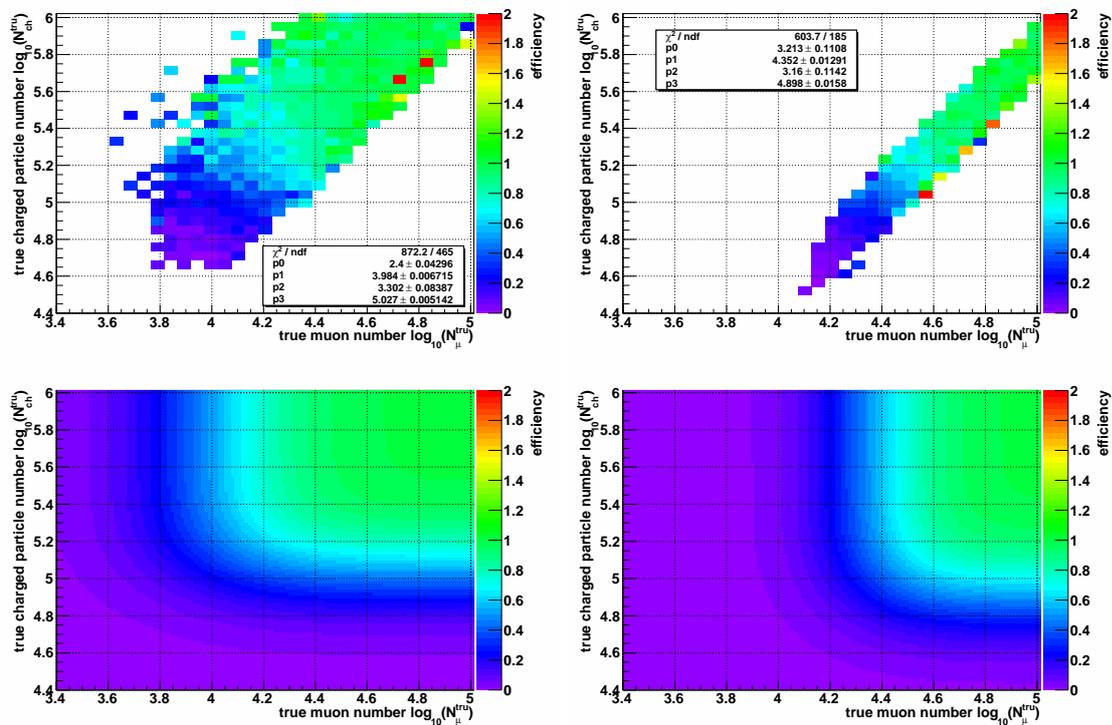


Figure 5.18: The efficiency as a function of the true charged particle and muon number (top), and the derived parametrization given by Eq.(5.14) (bottom). On the left panels in case of proton induced showers, on the right panels for iron induced ones.

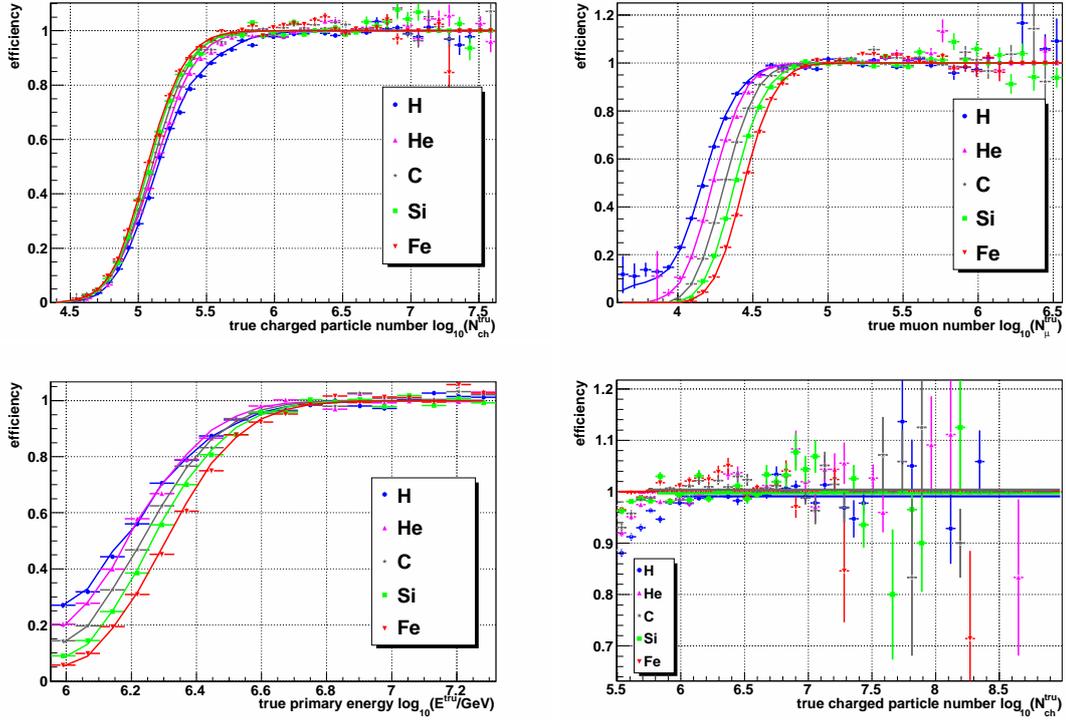


Figure 5.19: Cross-check between the efficiency computed based on the parametrization after Eq.(5.14) (lines) and the one directly simulated (markers), both as a function of the true numbers of charged particles (top left), of muons (top right), or of the true primary energy (bottom left). On the bottom right panel the simulated efficiencies, exemplarily as a function of the charged particle number, are fitted by a constant over the charged particle number range for that the efficiencies are close to 100%.

While in general there seems to be a good agreement between the efficiency (top panels of Fig. 5.18) and its parametrization (bottom panels of Fig. 5.18), the quality of the description will further be tested. The simulated (without any cuts or trigger conditions) two-dimensional shower size distribution of charged particles vs. muons, or, alternatively, the one-dimensional distribution of energies are now forward folded with the efficiency given by Eq.(5.14). The folding is carried out under either an integration of the two-dimensional shower size distribution over the charged particle respectively the muon number, or, alternatively, by integrating the one-dimensional energy distribution over both the numbers of charged particles and of muons. By this, the one-dimensional distributions of either the numbers of muons or of charged particles, or in the latter case of the primary energy, are computed as they are expected to appear for showers passing all trigger and reconstruction conditions. By dividing these computed distributions by the initial ones, which were just forward folded with the efficiency parametrization, one gets the efficiency curve as a function of the number of muons, of charged particles, or of the primary energy. In Fig. 5.19, this computed efficiency (lines) is compared to the one directly derived by applying trigger conditions and reconstruction cuts to the simulated shower events (markers, already shown in Fig. 5.17). In all three cases, there is a good agreement between the simulated and the computed efficiency discernible. Furthermore, on the bottom right panel the simulated efficiencies, exemplarily as a function of the

charged particle number, are separately fitted for all primaries by a constant for charged particle number ranges for that the efficiencies are close to 100%. The purpose of this cross-check was to check, whether or not there is a systematic bias to 100% efficiency, which could not be considered by the Gauss error function, because it converges only to one (for the right tail, to zero for the left one). The constant lines agree, considering the fit parameters' uncertainties, with 100% efficiency. Hence, above a certain threshold, the efficiencies converge in average to 100%, as also the Gauss error function does. Hence, the restrictive convergence behaviour of the Gauss error function does not induce a remarkable systematic bias.

In summary, the combined trigger and reconstruction efficiency  $\varepsilon$ , which will be used in the computation of the response function (confer Eq.(4.6)), can sufficiently be described by the parametrization given in Eq.(5.14). Depending on the true charged particle or muon numbers, full efficiency starts at around  $\log_{10}(N_{\text{ch}}^{\text{tru}}) = 6.0$  or  $\log_{10}(N_{\mu}^{\text{tru}}) = 5.0$ . In the final deconvolution analysis, only air showers exhibiting shower sizes equal or larger than these limits are considered. Nevertheless, whilst the computation of the response function also smaller values are regarded to account for possible migration effects caused by the intrinsic shower fluctuations. On average, all air showers with true primary energies equal or larger than approximately  $E^{\text{tru}} = 10^{16}$  eV are triggering KASCADE-Grande and pass the reconstruction process as well as all quality cuts. However, that an air shower is detected and all shower sizes are reconstructed does not imply that they are necessarily equal to the true ones, or at least close to them. Hence, the question about the reconstruction quality will be object of the next sections.

### 5.2.3 The systematic reconstruction uncertainties

Even if air showers are triggered and reconstructed with 100% efficiency, the determined shower sizes can differ from the true ones due to statistical fluctuations or systematic shifts. The latter ones are analysed in the following. A systematic bias in the reconstructed shower sizes can be caused by many possible sources. The main problem is an insufficient description of the true particle lateral distribution by the fits used in the reconstruction procedure. These functions have been chosen under the criterion of a balanced equilibrium between an on average accurate reconstruction of the respective shower size and in general stable fit result.

The systematic shifts will now be parametrized, supported by the experience and knowledge gained in Chapter 3, where appropriate functions for the correction of a systematic bias in the muon reconstruction have been determined. However, the bias in the reconstruction of the shower sizes will be described by functions instead of applying correction matrices in order to simplify the computation of the response function. The tiny but important difference between the thoughts made in Chapter 3 and here is that in the first case the measured shower sizes are corrected to more accurate ones, which could be used as input for further analyses, while in this section the bias caused whilst the reconstruction process shall be parametrized in order to account for it in the computation of the response functions. The measured data, finally used as input for the deconvolution analysis, will not be corrected at all, since all errors are included in the response function, and hence are considered in the unfolding process itself.

### Systematic uncertainties of the $\log_{10}N_{\text{ch}}^{\text{rec}}$ ascertainment

The systematic deviation between the true and reconstructed number of charged particles will be parametrized as a function of the true charged particle number. The determined function describes how the true size will systematically be distorted whilst the reconstruction process. This systematic bias is, beside the statistical reconstruction uncertainty discussed in Section 5.2.4, one of the contributors to the term representing the reconstruction characteristics  $r_n$  (cf. Eq.(4.6)), which for its part again contributes to the response function. As will be shown, the bias is very small. A common parametrization for all primaries yields appropriate results, such that one does not have to treat all particles separately and one can in lieu join them together with equal weights. This joining will be labelled with “All” in the following figures.

In Fig. 5.20, top left panel, the deviation between the true and the reconstructed charged particle number as a function of the true charged particle number is depicted for individual primaries as well as for the aforementioned combination of “All” primaries. One can see that there is not a significant difference between the primaries, such that they can be considered as an unity. The parametrization of the systematic shift bases on a phenomenological approach. The left trajectory of the bias is parabolic, while on the right a linear behaviour is visible. Hence, as parametrization  $C$  of the deviation a combination of a quadratic and a linear function is used:

$$C\left(\log_{10}(N_{\text{ch}}^{\text{tru}})\right) = \begin{cases} p_2 + p_1 \times (\log_{10}(N_{\text{ch}}^{\text{tru}}) - p_0)^2 & \text{if } \log_{10}(N_{\text{ch}}^{\text{tru}}) < p_0, \\ p_2 + p_3 \times (\log_{10}(N_{\text{ch}}^{\text{tru}}) - p_0) & \text{else.} \end{cases} \quad (5.15)$$

The fit is shown in Fig. 5.20, top left panel, black line. The fit parameters are additionally listed in Appendix D. In the top right panel, the reconstructed number of charged particles is corrected with  $C$ . In general, the remaining bias is very small for all primaries, and hence the common parametrization also works well when applied to individual particles. Because of the large number of markers, the remaining deviation for each individual primary and for the combination “All” is fitted by a linear function in order to get a better impression how the remaining mean bias varies with increasing particle numbers. In case of the combination of all particles, the remaining mean bias is zero (black line), but also for individual elements it is close to that value. On the bottom panel again the remaining bias is shown, but now only for the all particle case. The error bars represent the symmetric<sup>18</sup> spread instead of the error of the mean that is shown in the other plots. In comparison to the large statistical uncertainties or to the charged particles' channel width<sup>19</sup>, which is 0.07 in logarithm in the unfolding analysis, the remaining bias is negligible small. Nevertheless, the parametrization of the statistical uncertainties, which will be determined in Section 5.2.4, also allows to consider<sup>20</sup> a bias, such that the remaining one will finally also be regarded this way in the computation of the response function.

<sup>18</sup>The distribution of the reconstruction uncertainties is *de facto* not symmetric. Thus, the symmetric spread is only shown to give a rough estimate of the statistical uncertainties. A more detailed analysis will be carried out in Section 5.2.4.

<sup>19</sup>Whilst the choice of the optimal bin width for the reconstructed shower sizes used in the unfolding, primarily the width of the distribution of the deviations between the reconstructed and the true sizes above the threshold of full efficiency is considered. However, also different widths around that value have been tested and yield the final chosen width of 0.07 in the logarithm of the charged particle number. The same procedure is also applied to the muon number, resulting in the same channel width of 0.07 in logarithm of the muon number.

<sup>20</sup>However, only a bias with rather linear trajectory or small and smooth derivation can be considered, such that a first parametrization with more variable functions, as done here, is necessary.

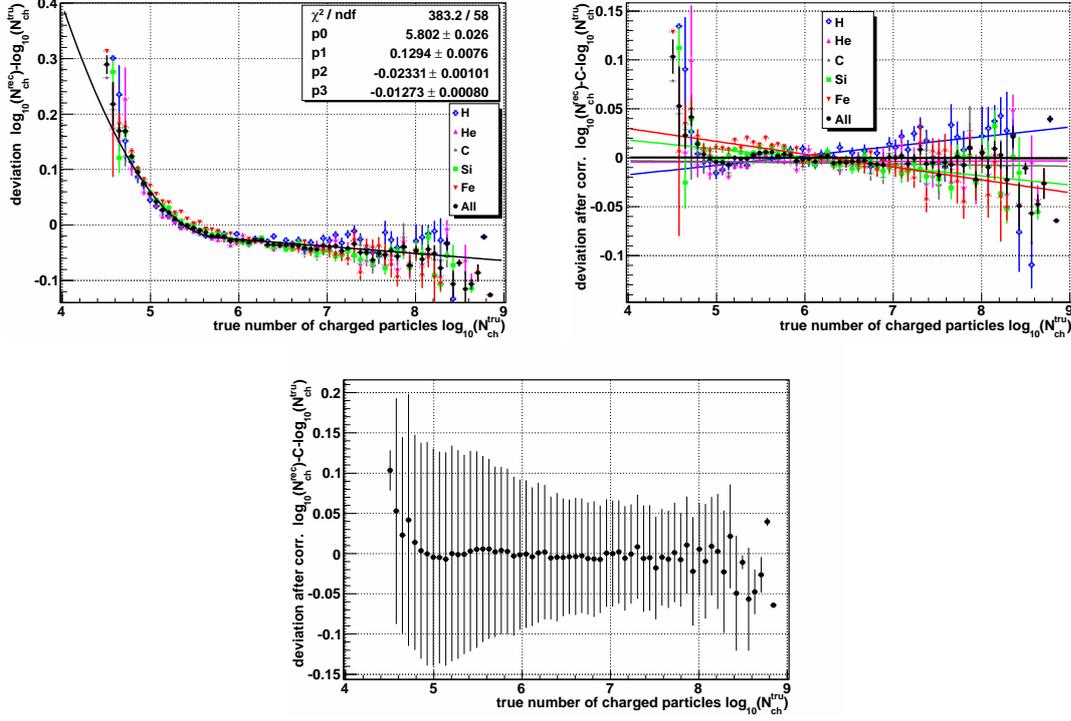


Figure 5.20: Parametrization  $C$  (cf. Eq.(5.15)) of the deviation between the true and the reconstructed charged particle number as a function of the true charged particle number in case of a combination of “All” primaries with equal weights (top left). The bias in the reconstruction in case of all contributing individual particles is also shown, but not parametrized, for purpose of comparison. The remaining bias after correction with the derived parametrization  $C$  is illustrated on the top right panel, as well as on the bottom panel. In the latter case, the error bars represent the spread instead of the error of the mean. The linear fits give an impression of the mean trajectory of the remaining bias.

In summary, a function describing the systematic bias in the reconstruction of the charged particle number has been derived, such that it can be considered whilst the computation of the response function. The remaining systematic shift that is not covered by this parametrization is already negligible small in comparison to the channel widths used in the final analysis or to the statistical fluctuations of the reconstruction, but can nevertheless be considered additionally in the parametrization of the last-mentioned uncertainties.

### Systematic uncertainties of the $\log_{10}N_{\mu}^{rec}$ ascertainment

The parametrization of the systematic deviation between the reconstructed muon number and the true one proceeds in general analogous to the one introduced afore in case of the charged particles. The keynotes are still valid and will not be repeated. Again, all primaries can be combined (labelled with “All” in the following figures) without worsening the final result. In Fig. 5.21, top left panel, the deviation between the true and the reconstructed muon number as a function of the true muon number is depicted. There is no significant difference observable between the primaries. The trajectory of the bias can be described by a polynomial function up to muon numbers of  $\log_{10}(N_{\mu}^{tru}) = 6.0$ . Beyond

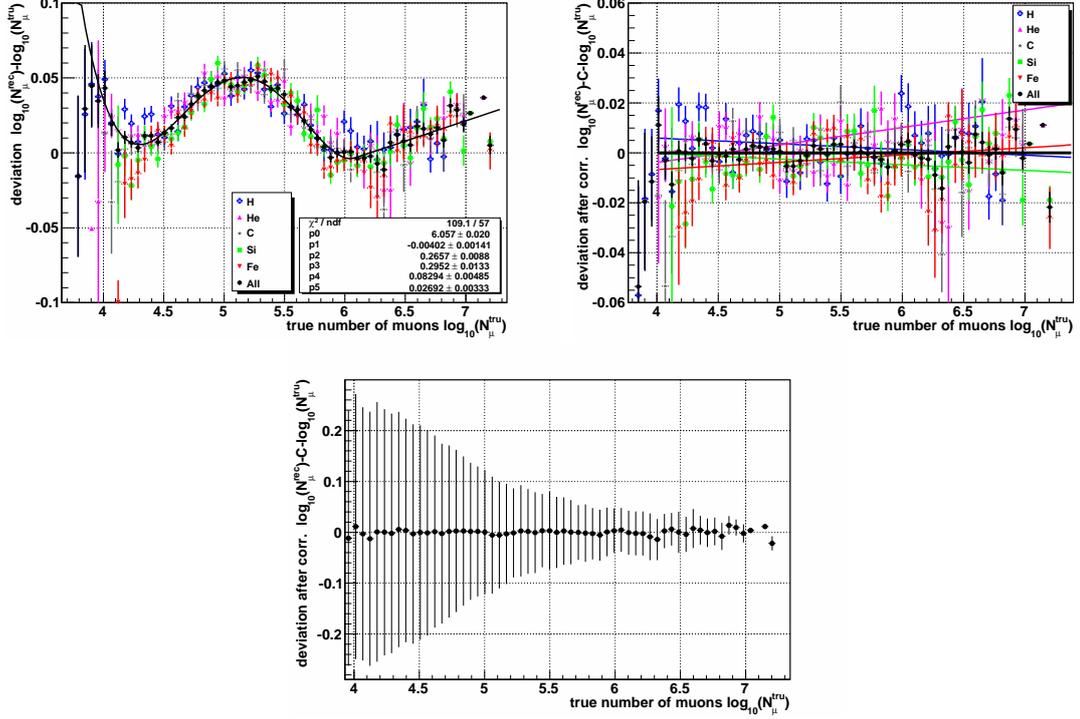


Figure 5.21: Parametrization  $C$  (cf. Eq.(5.16)) of the deviation between the true and the reconstructed muon number as a function of the true muon number in case of a combination of “All” primaries with equal weights (top left). The bias in the reconstruction in case of all contributing individual particles is also shown, but not parametrized, for purposes of comparison. The remaining bias after correction with the derived parametrization  $C$  is illustrated on the top right panel, as well as on the bottom panel. In the latter case, the error bars represent the spread instead of the error of the mean. The linear fits give an impression of the trajectory of the remaining bias.

this value, a simple linear fit can be used:

$$C\left(\log_{10}(N_{\mu}^{\text{tru}})\right) = \begin{cases} p_1 + p_2 \times \Delta^2 + p_3 \times \Delta^3 + p_4 \times \Delta^4 & \text{if } \log_{10}(N_{\mu}^{\text{tru}}) < p_0, \\ p_1 + p_5 \times \Delta & \text{else,} \end{cases} \quad (5.16)$$

$$\text{with } \Delta = \log_{10}(N_{\mu}^{\text{tru}}) - p_0.$$

The fit parameters are listed in Appendix D. In the top right panel, the remaining deviation after correction with  $C$  is presented. In general, it is very small and more symmetric around zero than before. The deviation for each individual primary and for the combination “All” is again fitted by a linear function. In case of the combination of all particles, the remaining mean bias is zero (black line), but also for individual elements it is close to that value. On the bottom panel, the remaining bias is shown for the all particle case with error bars representing the symmetric<sup>21</sup> spread, and not the error of the mean value. In comparison to the statistical reconstruction uncertainty and to the used bin width of 0.07 in logarithm of the reconstructed muon number, how it will be used

<sup>21</sup>However, in truth, the fluctuation around the mean value is not symmetric, such that the shown symmetric spread only serves as rough estimate for the statistical muon reconstruction uncertainties (which will be analysed more accurately in Section 5.2.4).

in the final unfolding analysis, the remaining bias is negligible. However, this systematic shift can additionally be regarded to a certain degree whilst the parametrization of the statistical uncertainties, such that it can also be considered this way in the calculation of the response function.

Summarizing, a function describing the systematic bias in the reconstruction of the muon number has been derived, such that it can be considered whilst the computation of the response function. The remaining systematic shift that is not covered by this parametrization is negligible small in comparison to the channel widths used in the final analysis or to the statistical fluctuations of the reconstruction, but can nevertheless additionally be considered in the parametrization of the last-mentioned uncertainties.

#### 5.2.4 The statistical reconstruction uncertainties

The main contributor to the statistical uncertainty of the shower size reconstruction is the to a certain extent random distribution of the air shower particles over the experiment's observation area. Due to these fluctuations, the number of particles exhibited by a shower at a certain detector station can differ from the one expected on average for this very events. The lateral distribution functions, based on which the total number of particles will be computed whilst the reconstruction process, are parametrized using mean values. Every time when the particle densities detected at a station differ from the average, the total number of particles will be determined deviating from the actually true one. The parametrization of those uncertainties of the shower size reconstruction will be object of this section. By this, they can be considered, besides the systematic bias already determined in Section 5.2.3, as second contributor to the term  $r_n$  that represents the reconstruction characteristics in Eq.(4.6), which for its part again contributes to the response function.

#### Statistical uncertainties of the $\log_{10}N_{\text{ch}}^{\text{rec}}$ ascertainment

In Fig. 5.22, the distribution of the deviations between the reconstructed charged particle number, whose systematic bias was corrected with  $C$  according to Eq.(5.15), and the true one is depicted in dependence on the true charged particle number. The distribution is a combination of the individual ones of the five primary particles, what increases the statistics, but does not worsen the results as will be discussed later. In general, due to the applied correction  $C$ , the average deviation is arranged around zero. However, the distribution is obviously not symmetric. With increasing number of true charged particles, the width decreases. This is due to the fact that larger particle numbers also mean higher particle densities at the detector stations, what results in a better description of the lateral profile of the shower, and hence in a more accurate reconstruction of the total charged particle number.

In order to parametrize the distribution, some slices along the "deviation" axis, each corresponding to a specific true charged particle number range, are analysed. By this the problem is transferred to a two-dimensional one. Each slice corresponds to a true charged particle number bin of width 0.1 in logarithm, what was found to be a good compromise between an adequate statistics for a fit and a sufficiently small true charged particle number range, such that the uncertainty distribution varies only slightly from one slice to another. Every slice will be fitted separately using appropriate functions. At the very end, all derived fit parameters, which up to now only allow to describe the specific situation in one slice, will be parametrized as a function of the true charged

particle number. Due to this interpolation, the statistical reconstruction uncertainties will be determinable for any interesting true charged particle number, and hence can be considered in the response function. Because of the small simulation statistics, all distributions of the five simulated primaries are add together in order to improve the available statistics for the fits, as it was also done in case of the parametrization of the bias in the reconstruction (cf. Section 5.2.3). Unlike the intrinsic shower fluctuations, the statistical reconstruction uncertainties were found not depend significantly on the particle type, such that this means to enhance the statistics is suitable. All following plots in this section will show these combined distribution, if nothing else is stated.

In Fig. 5.23, exemplarily the slices covering the true charged particle number ranges  $6.6 < \log_{10} N_{\text{ch}}^{\text{tru}} < 6.7$  (top, left) and  $7.3 < \log_{10} N_{\text{ch}}^{\text{tru}} < 7.4$  (top, right) of the distribution of the charged particle reconstruction uncertainty after correction (what was depicted in Fig. 5.22) are shown. As mentioned above, and especially for slices containing small true charged particle numbers, an asymmetric distribution can be observed. While the left part is reminiscent of a Gaussian distribution, the right one shows rather an exponential decrease. At small true charged particle number intervals, the probability density function will be parametrized by a combined Gaussian and exponential function, whereas at higher values above  $\log_{10} N_{\text{ch}}^{\text{tru}} > 7.3$  the distribution is compatible with a simple Gaussian distribution. Hence, the parametrization is given by:

$$p(\Delta N_{\text{ch}}^{\text{tru}}) = \begin{cases} p_0 \times e^{-0.5(\frac{\Delta N_{\text{ch}}^{\text{tru}} - p_1}{p_2})^2} & \text{if } \Delta N_{\text{ch}}^{\text{tru}} \leq p_1 - \frac{p_2^2}{p_3}, \\ p_0 \times e^{\frac{\Delta N_{\text{ch}}^{\text{tru}}}{p_3}} \times \underbrace{e^{-\frac{p_1}{p_3} + 0.5(\frac{p_2}{p_3})^2}}_{=\text{const. (for cont. differentiability)}} & \text{else,} \end{cases} \quad (5.17)$$

with  $\Delta N_{\text{ch}}^{\text{tru}} = \log_{10} N_{\text{ch}}^{\text{rec}} - C - \log_{10} N_{\text{ch}}^{\text{tru}},$

in case of small true charged particle numbers, i.e.  $\log_{10} N_{\text{ch}}^{\text{tru}} \leq 7.3$ , or

$$p(\Delta N_{\text{ch}}^{\text{tru}}) = p_0 \times e^{-0.5(\frac{\Delta N_{\text{ch}}^{\text{tru}} - p_1}{p_2})^2}, \quad (5.18)$$

with  $\Delta N_{\text{ch}}^{\text{tru}} = \log_{10} N_{\text{ch}}^{\text{rec}} - C - \log_{10} N_{\text{ch}}^{\text{tru}},$

in case of larger ones. The two fit functions conform, since<sup>22</sup> for  $\lim_{p_3 \rightarrow 0; p_3 < 0}(p)$  the second case term of Eq.(5.17), what describes mainly the “right” part of the distribution, converges to zero. Hence, in that case the distribution will be all over fitted by the “left” part, a simple Gaussian function, what is in agreement with Eq.(5.18). The parameter  $p_3$  can be understood to some extent as a weight assigned to the exponential part. In the limit of  $p_3 \rightarrow 0$ , the parameters  $p_1$  and  $p_2$  represent the mean value and the standard deviation of the used Gaussian function. By this, but only in the limit  $p_3 \rightarrow 0$ , the parameter  $p_1$  accounts for the remaining bias after the correction of the systematic shift (cf. Fig. 5.20, top right panel), while  $p_2$  represents the statistical reconstruction uncertainties, which are the primary object of this section and were already hinted at in Fig. 5.20, bottom panel.

As already mentioned at the beginning of Section 4.2 and discussed in Footnote 19, the chosen bin width for the final unfolding analysis is 0.07 in logarithmic scale for the charged particle number. In order to find this appropriate binning, the studies of the shower size specific reconstruction accuracies, analysed in this section here, are taken into account

<sup>22</sup>It will be shown later on in Fig. 5.24 (bottom) that  $p_3$  converges to zero (with  $p_3 < 0$ ) when  $\log_{10} N_{\text{ch}}^{\text{tru}} \rightarrow 7.3$ .

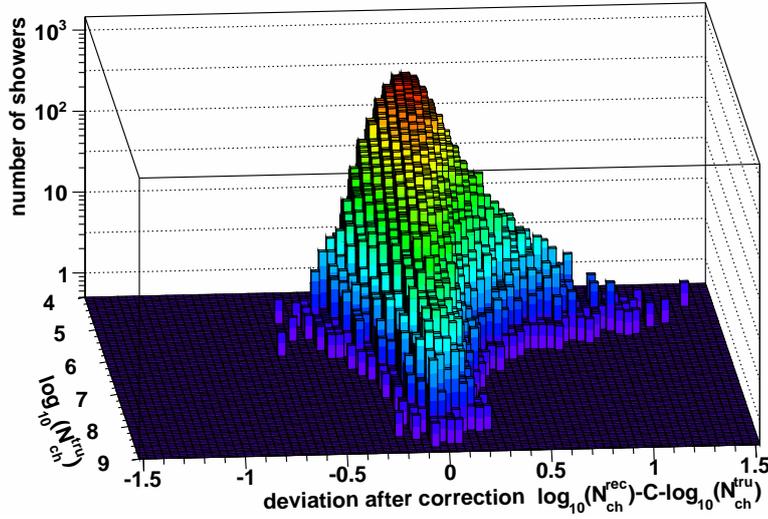


Figure 5.22: The distribution of the charged particle reconstruction uncertainty after correcting the bias in the reconstruction with  $C$ , which is given by Eq.(5.15), in dependence on the true charged particle number. The results of all five individual primaries are added together for this plot in order to increase statistics (see text).

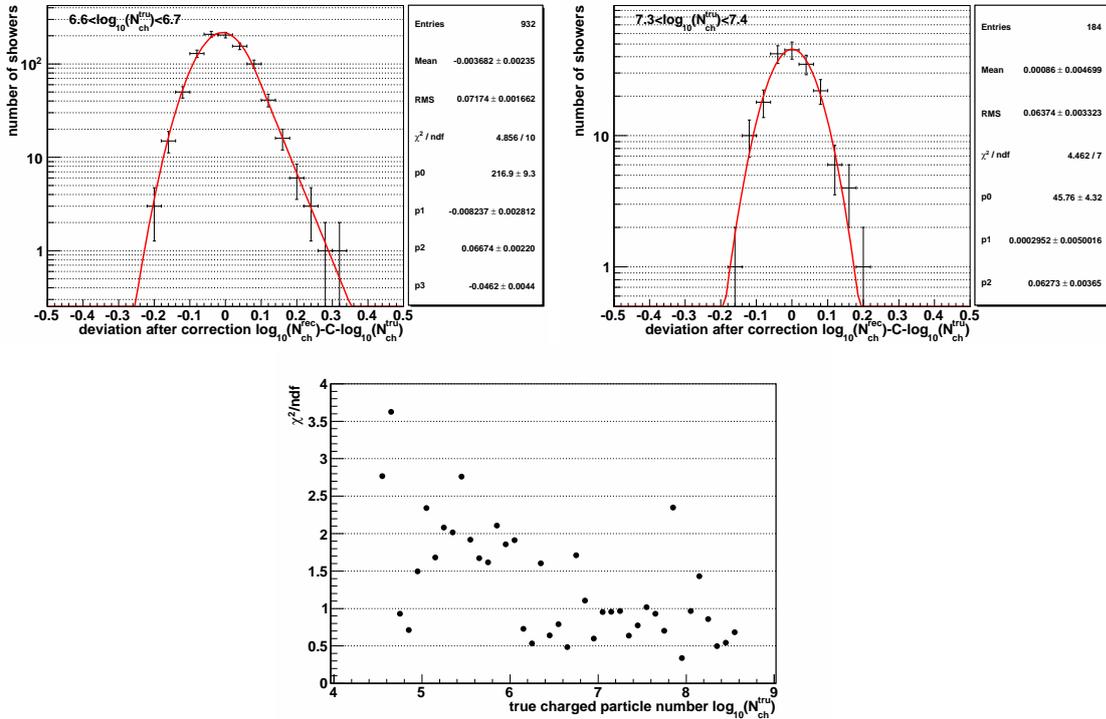


Figure 5.23: Slices  $6.6 < \log_{10} N_{\text{ch}}^{\text{tru}} < 6.7$  (top, left) and  $7.3 < \log_{10} N_{\text{ch}}^{\text{tru}} < 7.4$  (top, right) of the distribution of the charged particle reconstruction uncertainty after correction (depicted in Fig. 5.22). The distributions are fitted according to Eq.(5.17) (left) respectively Eq.(5.18) (right). At the bottom, the reduced chi-squares of all fits as a function of the true charged particle number are illustrated.

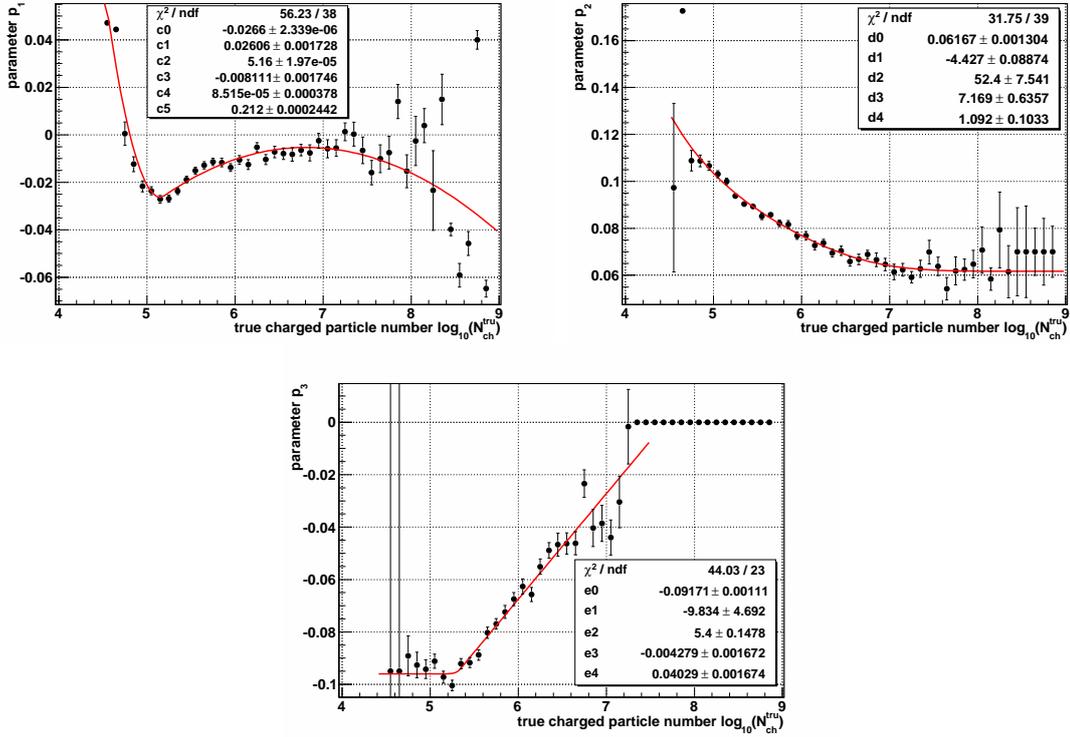


Figure 5.24: Interpolation (based on the parametrizations given by Eq.(5.19)) of the true charged particle number dependence of the parameters  $p_1$ ,  $p_2$  and  $p_3$  used in the parametrization of the charged particle reconstruction uncertainty (cf. Eq.(5.17) and Eq.(5.18)).

in such a way that it is not adequate to use a bin width beneath the achievable detector resolution. As can be derived from Fig. 5.20, bottom panel, as well as from Fig. 5.24, top right panel, in case of 100% efficiency ( $\log_{10}N_{\text{ch}}^{\text{tru}} \geq 6.0$ ) the experiment's resolution is comparable to the bin width of 0.07, which was chosen for one thing under consideration of this resolution, for another thing based on preliminary tests using different values.

The parameter  $p_0$  finally corresponds to the arbitrary normalization of the distributions. The two slices shown in Fig. 5.23, top left panel respectively the top right one, are fitted by the functions according to Eq.(5.17) respectively Eq.(5.18). The regarding reduced chi-squares, in case of all used true charged particle number intervals, are depicted at the bottom panel. In addition, as already done whilst the parametrization of the intrinsic shower fluctuations (cf. Section 5.1.2), the fit quality is judged by means of a Kolmogorov-Smirnov test and checked by eye. It was found that the chosen functions describe the uncertainty distributions very well. All derived fit parameters are listed in Appendix E.

As mentioned in the introducing thoughts at the beginning of this Section, the dependence of the fit parameters  $p_1$ ,  $p_2$  and  $p_3$  (jointly used in Eqs.(5.17) and Eq.(5.18)) on the true charged particle number has to be interpolated in order to get a functional relationship what can be used in the computation of the response function. Again, a

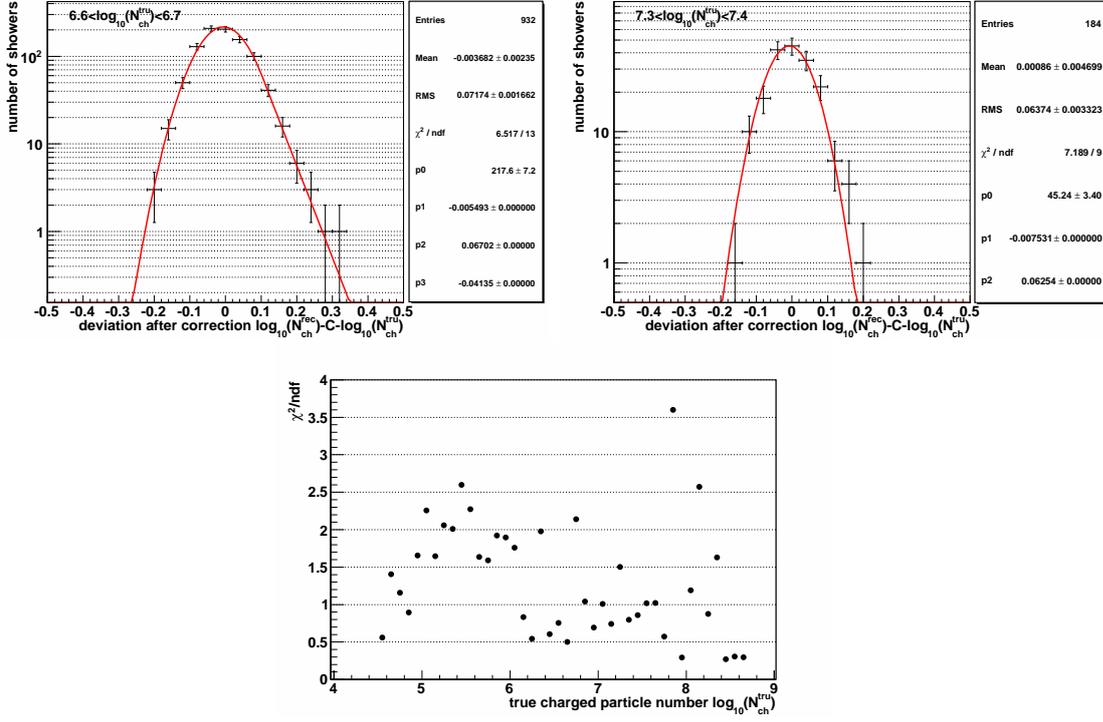


Figure 5.25: Slices  $6.6 < \log_{10} N_{\text{ch}}^{\text{tru}} < 6.7$  (top, left) and  $7.3 < \log_{10} N_{\text{ch}}^{\text{tru}} < 7.4$  (top, right) of the distribution of the charged particle reconstruction uncertainty after correction (depicted in Fig. 5.22). The distributions are fitted using Eqs.(5.17) (left panel) and (5.18) (right panel), while the parameters are fixed according to the interpolation based on Eq.(5.19). At the bottom, the reduced chi-squares of all fits, with parameters fixed according to the interpolation, are illustrated as a function of the true charged particle number.

phenomenological approach<sup>23</sup> will be pursued:

$$\begin{aligned}
 p_1(x) &= \begin{cases} c_0 + c_5(x - c_2)^2 & \text{if } x \leq c_2, \\ c_0 + c_1(x - c_2) + c_3(x - c_2)^2 + c_4(x - c_2)^3 & \text{else,} \end{cases} \\
 p_2(x) &= d_0 + d_2 x^{d_1} \times (0.5 + 0.5 \times \text{erf}(-d_4 x + d_3)), \\
 p_3(x) &= \begin{cases} e_0 + e_3 \times \text{erf}(e_1(x - e_2)) & \text{if } x < e_2, \\ e_0 + e_4(x - e_2) & \text{else,} \end{cases} \\
 &\text{with } x = \log_{10} N_{\text{ch}}^{\text{tru}}.
 \end{aligned} \tag{5.19}$$

The fits are illustrated in Fig. 5.24. Additionally, the derived values of the used parameters  $c_i$ ,  $d_i$  and  $e_i$  are listed in Appendix E. In Fig. 5.25, the fits (using Eqs.(5.17) and (5.18)) to the reconstruction uncertainty distributions are shown, now with parameters fixed according to the interpolations given by Eq.(5.19). Again, the reduced chi-squares are shown (bottom), and its values are listed in Table E.5 in Appendix E.

Comparing the results shown in Fig. 5.25 to those in Fig. 5.23, one can conclude that the additional interpolation does not worsen the quality of the fits. Hence, the

<sup>23</sup>Primarily based on polynomials, but if needed using the convergence properties of the Gauss error function.

parametrizations given by Eqs.(5.17) and (5.18), with parameters interpolated according to Eq.(5.19), will allow to account for the statistical charged particle reconstruction uncertainties whilst the computation of the response function.

### Statistical uncertainties of the $\log_{10}N_{\mu}^{\text{rec}}$ ascertainment

The parametrization of the statistical muon reconstruction uncertainties proceeds in general analogous to the one introduced afore in case of the charged particles. The keynotes are still valid and will not be repeated. Unlike the intrinsic shower fluctuations of muons, the statistical muon reconstruction uncertainties were found to not depend significantly on the particle type. Hence, in order to improve the statistics for the fits, all primaries can again be combined without worsening the final result. In Fig. 5.26, the distribution of the deviations between the reconstructed muon number, whose systematic bias was corrected with  $C$  according to Eq.(5.16), and the true one is depicted in dependence on the true muon number. In general, due to the applied correction  $C$  the average deviation is arranged around zero. However, the distribution is obviously not symmetric. With increasing number of true muons, the width decreases. This is due to the fact that larger particle numbers also mean higher particle densities at the detector stations, what results in a better description of the lateral profile of the shower, and hence in a more accurate reconstruction of the total muon number.

In order to parametrize the distribution, it is divided into slices of width 0.1 in logarithm of the true muon number, which are fitted separately. At the very end, all derived fit parameters, which up to now only allow to describe the specific situation in one slice, will be parametrized as a function of the true muon number. In Fig. 5.27, exemplarily the slices covering the true muon number ranges  $4.9 < \log_{10}N_{\mu}^{\text{tru}} < 5.0$  (top, left) and  $5.8 < \log_{10}N_{\mu}^{\text{tru}} < 5.9$  (top, right) of the distribution of the muon reconstruction uncertainty after correction (what was depicted in Fig. 5.26) are shown. As mentioned above, and especially for slices containing small true muon numbers, an asymmetric distribution can be observed. While the right part is reminiscent of a Gaussian distribution, the left one shows rather an exponential decrease. At small true muon number intervals, the probability density function will be parametrized by a combined Gaussian and exponential function, whereas at higher values above  $\log_{10}N_{\mu}^{\text{tru}} > 5.75$  the distribution is compatible with a simple Gaussian distribution. Hence, the parametrization is given by:

$$p\left(\Delta N_{\mu}^{\text{tru}}\right) = \begin{cases} p_0 \times e^{-0.5\left(\frac{\Delta N_{\mu}^{\text{tru}} - p_1}{p_2}\right)^2} & \text{if } \Delta N_{\mu}^{\text{tru}} \geq p_1 - \frac{p_2^2}{p_3}, \\ p_0 \times e^{\frac{\Delta N_{\mu}^{\text{tru}}}{p_3}} \times \underbrace{e^{-\frac{p_1}{p_3} + 0.5\left(\frac{p_2}{p_3}\right)^2}}_{=\text{const. (for cont. differentiability)}} & \text{else,} \end{cases} \quad (5.20)$$

with  $\Delta N_{\mu}^{\text{tru}} = \log_{10}N_{\mu}^{\text{rec}} - C - \log_{10}N_{\mu}^{\text{tru}}$ ,

in case of small true muon numbers, i.e.  $\log_{10}N_{\mu}^{\text{tru}} \leq 5.75$ , or

$$p\left(\Delta N_{\mu}^{\text{tru}}\right) = p_0 \times e^{-0.5\left(\frac{\Delta N_{\mu}^{\text{tru}} - p_1}{p_2}\right)^2}, \quad (5.21)$$

with  $\Delta N_{\mu}^{\text{tru}} = \log_{10}N_{\mu}^{\text{rec}} - C - \log_{10}N_{\mu}^{\text{tru}}$ ,

in case of larger ones. The two fit functions conform, since<sup>24</sup> for  $\lim_{p_3 \rightarrow 0; p_3 > 0}(p)$  the second case term of Eq.(5.20), what describes mainly the ‘‘left’’ part of the distribution,

<sup>24</sup>It will be shown later on in Fig. 5.28 (bottom) that  $p_3$  converges to zero (with  $p_3 > 0$ ) when  $\log_{10}N_{\mu}^{\text{tru}} \rightarrow 5.75$ .

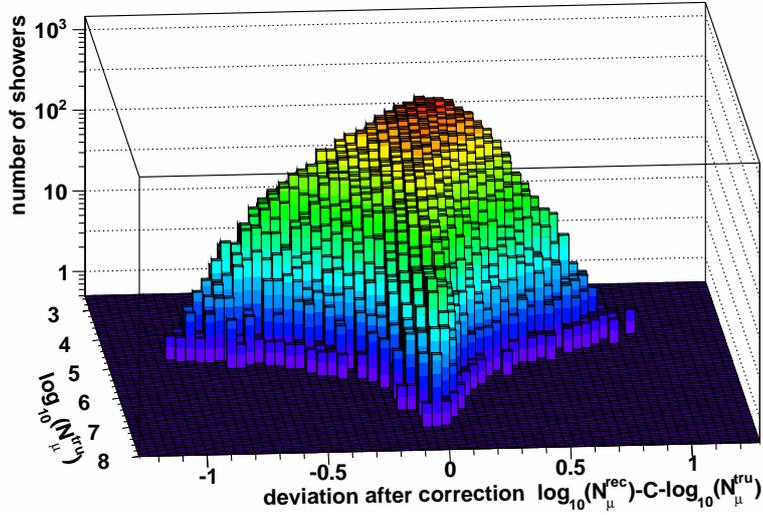


Figure 5.26: The distribution of the muon reconstruction uncertainty after correcting the bias in the reconstruction with  $C$ , which is given by Eq.(5.16), in dependence on the true muon number. The results of all five individual primaries are add together for this plot in order to increase statistics (see text).

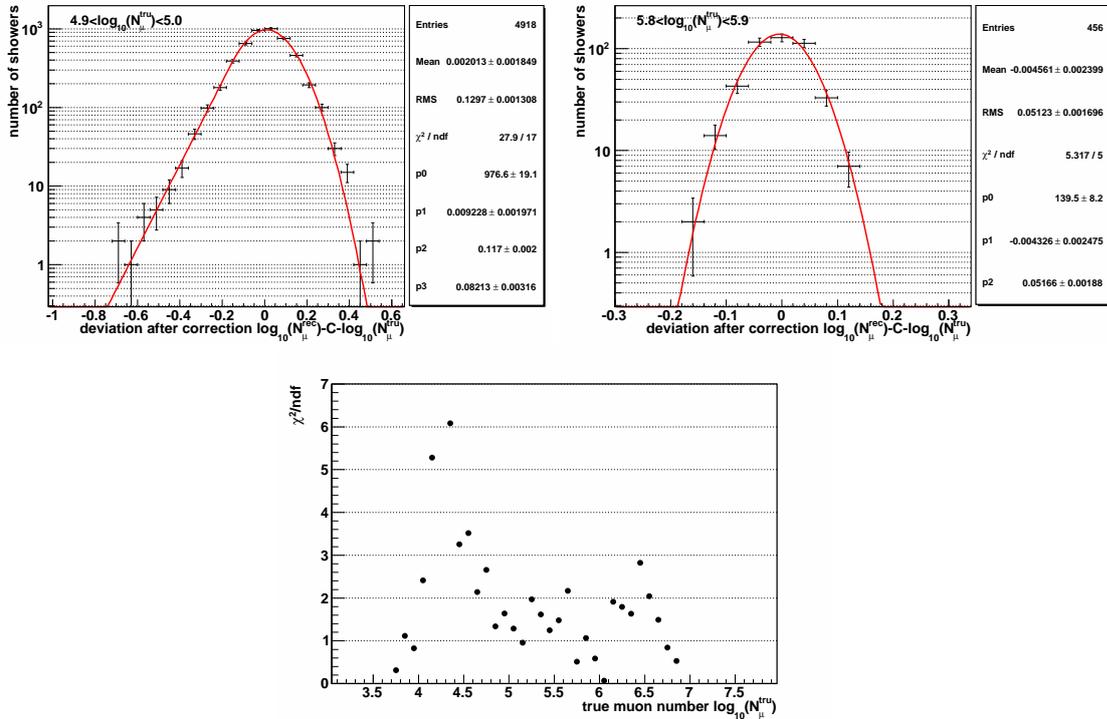


Figure 5.27: Slices  $4.9 < \log_{10} N_{\mu}^{\text{tru}} < 5.0$  (top, left) and  $5.8 < \log_{10} N_{\mu}^{\text{tru}} < 5.9$  (top, right) of the distribution of the muon reconstruction uncertainty after correction (depicted in Fig. 5.26). The distributions are fitted according to Eq.(5.20) (left) respectively Eq.(5.21) (right). At the bottom, the reduced chi-squares of all fits as a function of the true muon number are illustrated.

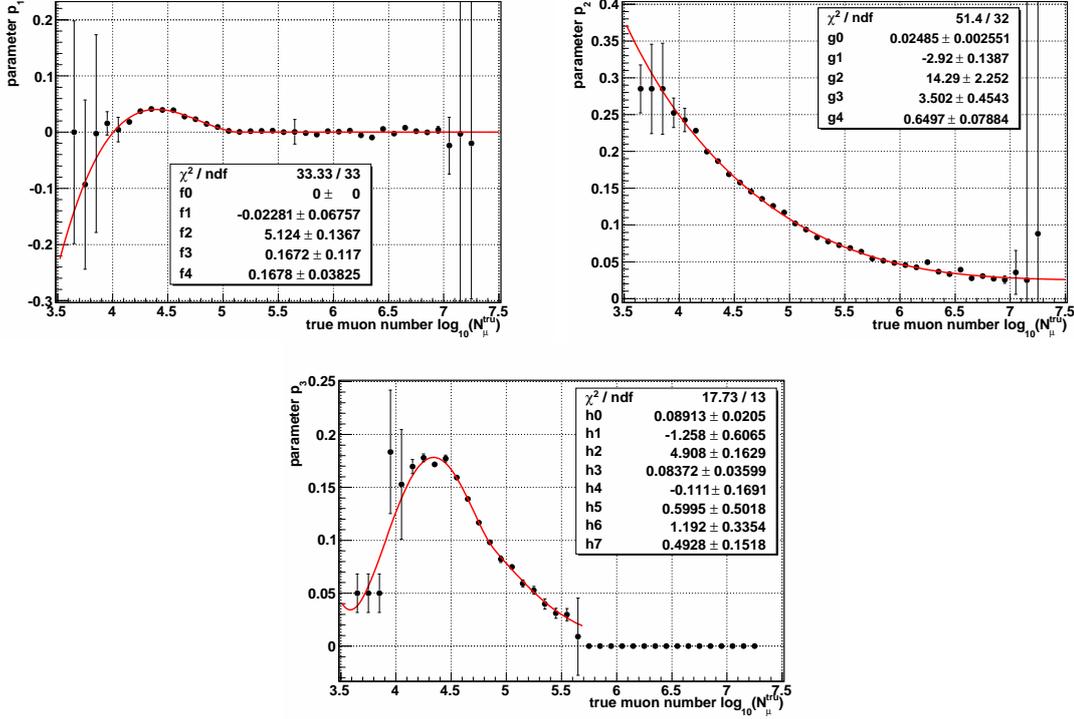


Figure 5.28: Interpolation (based on the parametrizations given by Eq.(5.22)) of the true muon number dependence of the parameters  $p_1$ ,  $p_2$  and  $p_3$  used in the parametrization of the muon reconstruction uncertainty (cf. Eq.(5.20) and Eq.(5.21)).

converges to zero. Hence, in that case the distribution will be all over fitted by the “right” part, a simple Gaussian function, what is in agreement with Eq.(5.21). Analogous to the remarks made in the previous section for the charged particle reconstruction accuracy, the parameter  $p_3$  can again be understood as a weight assigned to the exponential part. In the limit of  $p_3 \rightarrow 0$ , the parameters  $p_1$  and  $p_2$  represent the mean value and the standard deviation of the used Gaussian function, and  $p_2$  by this the statistical reconstruction uncertainties, which are the primary object of this section. As can be derived from Fig. 5.21, bottom panel, as well as from Fig. 5.28, top right panel, in case of full efficiency ( $\log_{10} N_{\mu}^{\text{tru}} \geq 5.0$ ) the resolution improves from initially 0.1 to finally 0.025 in units of the logarithmic muon number. The muon bin width for the final analysis is set to 0.07 in logarithmic scale for the muon number, taking into account the afore mentioned resolution development, as well as the results from preliminary tests trying out different binnings. Finally, the parameter  $p_0$  corresponds to the arbitrary normalization of the distributions. The two slices shown in Fig. 5.27, top left panel respectively the top right one, are fitted by the functions according to Eq.(5.20) respectively Eq.(5.21). The regarding reduced chi-squares, in case of all used true muon number intervals, are depicted at the bottom. All derived fit parameters are listed in Appendix F.

The dependence of the fit parameters  $p_1$ ,  $p_2$  and  $p_3$  (jointly used in Eqs.(5.20) and Eq.(5.21)) on the true muon number is interpolated in order to get a functional relationship, which can be used in the computation of the response function. Again, a

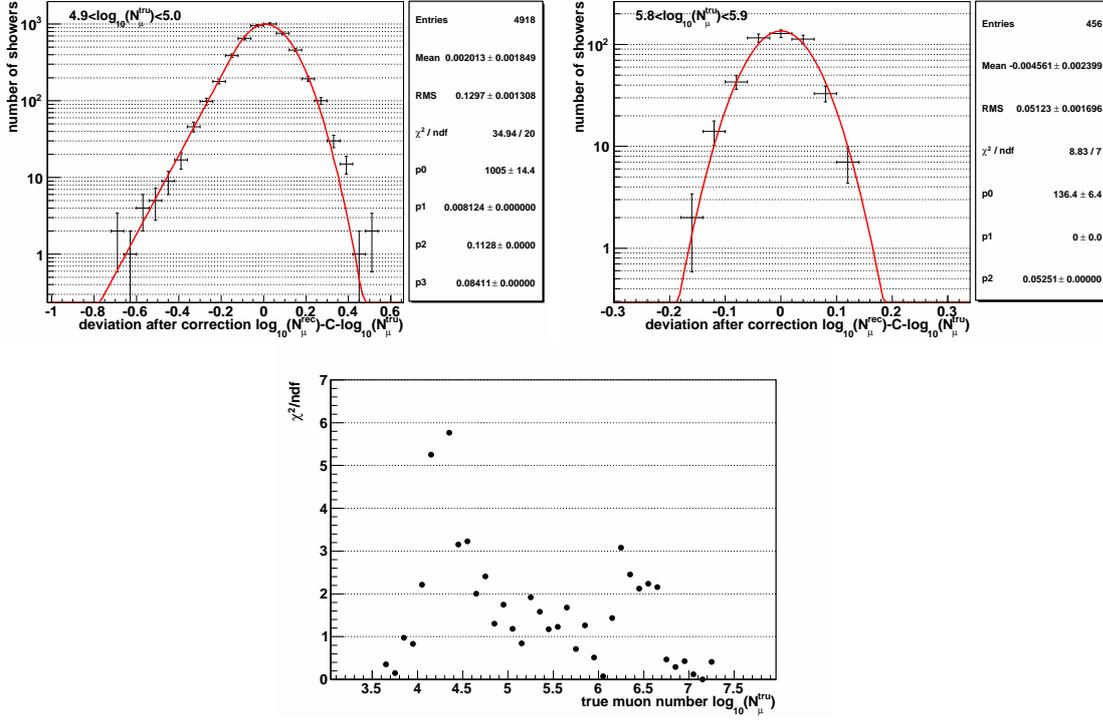


Figure 5.29: Slices  $4.9 < \log_{10}N_{\mu}^{\text{tru}} < 5.0$  (top, left) and  $5.8 < \log_{10}N_{\mu}^{\text{tru}} < 5.9$  (top, right) of the distribution of the muon reconstruction uncertainty after correction (depicted in Fig. 5.26). The distributions are fitted using Eqs.(5.20) (left panel) and (5.21) (right panel), while the parameters are fixed according to the interpolation based on Eq.(5.22). At the bottom, the reduced chi-squares of all fits, with parameters fixed according to the interpolation, are illustrated as a function of the true muon number.

phenomenological approach will be pursued:

$$\begin{aligned}
 p_1(x) &= \begin{cases} f_0 + f_1(x - f_2) + f_3(x - f_2)^2 + f_4(x - f_2)^3 & \text{if } x \leq f_2, \\ f_0 & \text{else,} \end{cases} \\
 p_2(x) &= g_0 + g_2 \times x^{g_1} \times (0.5 + 0.5 \times \text{erf}(-g_4x + g_3)), \\
 p_3(x) &= \begin{cases} h_0 + h_4(x - h_2) + h_5(x - h_2)^2 + h_6(x - h_2)^3 + h_7(x - h_2)^4 & \text{if } x < h_2, \\ h_0 + h_3 \times \text{erf}(h_1(x - h_2)) & \text{else,} \end{cases} \\
 &\text{with } x = \log_{10}N_{\mu}^{\text{tru}}.
 \end{aligned} \tag{5.22}$$

The fits are illustrated in Fig. 5.28. Additionally, the derived values of the used parameters  $f_i$ ,  $g_i$  and  $h_i$  are listed in Appendix F. In Fig. 5.29, the fits (using Eqs.(5.20) and (5.21)) to the reconstruction uncertainty distributions are shown, now with parameters fixed according to the interpolations given by Eq.(5.22). Again, the reduced chi-squares are shown (bottom), and its values are listed in Table F.5 in Appendix F.

Comparing the results shown in Fig. 5.29 to those in Fig. 5.27, one can conclude that the additional interpolation does not worsen the quality of the fits. Hence, the parametrizations given by Eqs.(5.20) and (5.21), with parameters interpolated according to Eq.(5.22), will allow to account for the statistical muon reconstruction uncertainties

whilst the computation of the response function.

Together with the parametrization of the charged particle number reconstruction uncertainty as well as with the bias correction functions from Section 5.2.3, a functional description of the reconstruction properties term  $r_n$  of Eq.(4.6) is available now. As the other terms of Eq.(4.6), namely the parametrization of the intrinsic shower fluctuations  $s_n$  (see Section 5.1) and the one of the experiment's efficiency (obtained in Section 5.2.2), are already known, by now all contributors to the response function are determined, such that it can be computed in the following.

## 5.3 Computation and Test of the Response Matrix

In the previous sections, the contributors to the kernel function  $p_n$ , introduced in Eq.(4.6), have been determined: the probabilities describing the intrinsic shower fluctuations  $s_n$ , the experiments efficiency  $\varepsilon_n$ , as well as the reconstruction properties  $r_n$ . In this section, the resulting response matrix will be computed, then its condition number will be discussed and, afterwards, some consistency checks will be performed in order to ensure its reliability.

### 5.3.1 Computation of the response matrix

In Section 4.2, the equations describing the convolution of the sought-after primary energy spectra of individual primaries are deduced. Equation (4.4) represents the folding of the differential energy fluxes  $dJ_n/d\log_{10}E$  of the primary cosmic ray nuclei  $n$  having an energy  $E$  into the measured number of showers  $N_i$  contributing to the cell  $i$ , and thus to the content of this specific charged particle and muon number bin  $(\log_{10}(N_{\text{ch}}), \log_{10}(N_{\mu}))_i$  of the measured distribution (depicted in Fig. 4.8). The term  $p_n$  represents the kernel function. For a detailed description of all variables, confer Section 4.2. As mentioned there, neither the intrinsic shower fluctuations nor the efficiencies or the reconstruction properties vary significantly within the used zenith angle range  $0^\circ \leq \theta \leq 18^\circ$ , such that its influence on all three probabilities can be neglected. Hence, the integration over the zenith angle can be carried out, such that Eq.(4.4) yields:

$$N_i = \Omega_{\text{eff}} A_f T_m \sum_{n=1}^{N_{\text{nucl}}} \int_{-\infty}^{+\infty} \frac{dJ_n}{d\log_{10}E} p_n d\log_{10}E \quad , \quad (5.23)$$

with  $p_n = p_n((\log_{10}N_{\text{ch}}, \log_{10}N_{\mu})_i | \log_{10}E)$  .

The new variable  $\Omega_{\text{eff}}$  represents the effective solid angle. While the left side of the equation is already discretized since the measurement is histogramized and filled into discrete cells of the two-dimensional shower size plane (Fig. 4.8), the sought-after differential flux (contributing to the right side) is still a continuous variable. Due to effects of limited resolution, it will not be possible to reconstruct the energy spectra with arbitrary fine bin width, such that also the solution part has to be discretized using appropriate bin widths<sup>25</sup>  $\Delta\log_{10}E$ . Thereby it is assumed that the measured shower size plane has  $N_m$  cells and the solution, the differential energy fluxes of individual primaries, will be

<sup>25</sup>In preliminary tests was found that a bin width of  $\Delta\log_{10}(E/eV) = 0.14$  allows reliable and stable results.

reconstructed for  $N_s$  energy intervals:

$$N_i = \Omega_{\text{eff}} A_f T_m \sum_{n=1}^{N_{\text{nucl}}} \sum_{j=1}^{N_s} \int_{\log_{10} E_j}^{\log_{10} E_j + \Delta \log_{10} E} \frac{dJ_n}{d \log_{10} E} p_n d \log_{10} E \quad , \quad (5.24)$$

with  $p_n = p_n((\log_{10} N_{\text{ch}}, \log_{10} N_{\mu})_i | \log_{10} E)$  ,  $i \in [1, N_m]$  .

The right side can mathematically be expanded:

$$N_i = \sum_{n=1}^{N_{\text{nucl}}} \sum_{j=1}^{N_s} \frac{\int_{\log_{10} E_j}^{\log_{10} E_j + \Delta \log_{10} E} \frac{dJ_n}{d \log_{10} E} p_n d \log_{10} E}{\underbrace{\int_{\log_{10} E_j}^{\log_{10} E_j + \Delta \log_{10} E} \frac{dJ_n}{d \log_{10} E} d \log_{10} E}_{:=R_{ij}^n}} \times \quad (5.25)$$

$$\times \underbrace{\Omega_{\text{eff}} A_f T_m \int_{\log_{10} E_j}^{\log_{10} E_j + \Delta \log_{10} E} \frac{dJ_n}{d \log_{10} E} d \log_{10} E}_{:=x_j^n} \quad ,$$

with  $p_n = p_n((\log_{10} N_{\text{ch}}, \log_{10} N_{\mu})_i | \log_{10} E)$  ,  $i \in [1, N_m]$  .

The term  $R_{ij}^n$  is the  $ij$  element of the response matrix, belonging to the primary  $n$ , while  $x_j^n$  is the number of entries in the  $j$ th energy bin of the sought-after solution<sup>26</sup> for that primary  $n$ . More precisely, the entry  $R_{ij}^n$  represents the probability that a particle  $n$ , having an energy belonging to an energy bin  $j$  of the solution vector, will finally be reconstructed with a specific charged particle and muon number  $(\log_{10}(N_{\text{ch}}), \log_{10}(N_{\mu}))_i$  corresponding to the entry number  $i$  of the measurement vector<sup>27</sup>.

For purposes of simplification, Eq.(5.25) will in the following be written as a matrix equation:

$$\vec{Y} = \mathbf{R} \vec{X} \quad , \quad \text{with } \vec{Y} \in \mathbb{R}^{N_m} \quad , \quad \vec{X} \in \mathbb{R}^{N_{\text{nucl}} N_s} \quad , \quad \mathbf{R} \in \mathbb{R}^{N_m \times N_{\text{nucl}} N_s} \quad , \quad (5.26)$$

<sup>26</sup>In order to obtain a differential energy spectrum, which is comparable to those of other experiments, at the very end the ‘‘preliminary’’ solution  $x_j^n$  has to be divided by the exposure  $\Omega_{\text{eff}} A_f T_m$  and the energy bin width.

<sup>27</sup>Vector, since the cells of the two-dimensional shower size plane were numbered continuously, such that their content can be written into a one-dimensional vector.

what, more precisely, corresponds to:

$$\underbrace{\begin{pmatrix} N_1 \\ \vdots \\ N_{N_m} \end{pmatrix}}_{\vec{Y}} = \underbrace{\left( \begin{pmatrix} R_{11}^1 & \dots & R_{1N_s}^1 \\ \vdots & \ddots & \vdots \\ R_{N_m 1}^1 & \dots & R_{N_m N_s}^1 \end{pmatrix} \dots \begin{pmatrix} R_{11}^{N_{\text{nucl}}} & \dots & R_{1N_s}^{N_{\text{nucl}}} \\ \vdots & \ddots & \vdots \\ R_{N_m 1}^{N_{\text{nucl}}} & \dots & R_{N_m N_s}^{N_{\text{nucl}}} \end{pmatrix} \right)}_{\mathbf{R}} \times \underbrace{\begin{pmatrix} x_1^1 \\ \vdots \\ x_{N_s}^1 \\ \vdots \\ x_1^{N_{\text{nucl}}} \\ \vdots \\ x_{N_s}^{N_{\text{nucl}}} \end{pmatrix}}_{\vec{X}}. \quad (5.27)$$

The in total  $N_{\text{nucl}}$  solution vectors (each of dimension  $N_s$ ), namely the individual energy spectra of different primaries, are consecutively recorded in one common vector  $\vec{X}$ . The contents of the  $N_m$  cells of the two-dimensional shower size plane are also listed in a vector  $\vec{Y}$ , as already mentioned above. Finally, the  $N_{\text{nucl}}$  individual response matrices for the specific primaries are considered as submatrices of a common matrix  $\mathbf{R}$ .

As can be deduced from Eq.(5.25), the response matrix also depends on the solution itself, hence on the sought-after differential energy fluxes  $dJ_n/d\log_{10}E$ . Since the solution is not *a priori* known, it can only be estimated. In the limit of infinitesimal small energy bin widths, the flux can be treated as constant inside the specific energy intervals. By this, the flux can be taken in front of the integral, such that it can be cancelled out from the response matrix term. While, due to the limited resolution, only energy bin widths at around  $\Delta\log_{10}(E/eV) = 0.14$  are possible, whilst the computation of the response matrix these intervals were again divided into 14 sub-intervals, each corresponding to a bin width of  $\Delta\log_{10}(E/eV) = 0.01$ . It was found<sup>28</sup> that these widths are small enough to sufficiently decouple the correlation between the matrix elements and the flux. Considering that the exact value of the flux is not important for the computation of the response matrix, it was fixed according to an  $E^{-3}$  energy spectrum. This value was chosen taking into account the result about the all-particle energy spectrum obtained by our collaboration [33], which states a spectral index of  $\approx -3.0$  for energies between  $10^{16}$  eV to  $10^{17}$  eV, and of  $\approx -3.2$  above energies of  $10^{17}$  eV.

By this, the means and the knowledge necessary for the computation of the response matrix are acquired, such that the task can be completed now. The next steps will be to check the conditioning of the derived matrix.

### 5.3.2 Conditioning of the response matrix

The response matrix  $\mathbf{R}$  is computed based on the continuous parametrization functions determined in Sections 5.1 and 5.2. Since double precision floating-point numbers are used inside the applied C++ scripts, the calculated probabilities cover some ten decades and range from just below 1 down to the precision limit of the used floating-point numbers<sup>29</sup>.

<sup>28</sup>This was tested by assuming either a  $E^{-1}$ ,  $E^{-2}$ ,  $E^{-3}$  or  $E^{-4}$  energy spectrum when computing the response matrix. Based on these four different matrices, a test data sample was unfolded. It was found that there is no significant difference between the results, such that the exact knowledge of the spectral value is not necessary.

<sup>29</sup>Till they are rounded to zero below a certain value, when the corresponding floating-point number runs out of the precision.

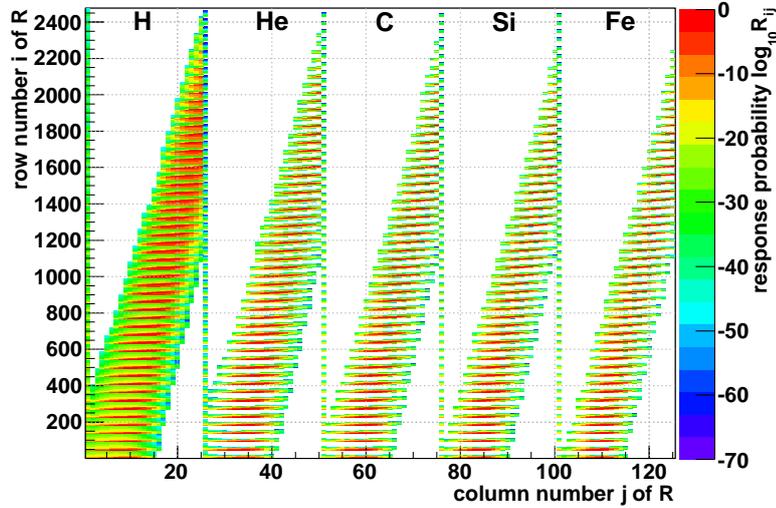


Figure 5.30: The entries (which are probability densities) of the response matrix  $\mathbf{R}$  defined in Eq.(5.27), visualized by a coloured histogram.

Furthermore, the intrinsic shower fluctuations and the reconstruction uncertainties both result in a remarkable overlap of the probability distributions of different primaries (cf. Fig. 5.37). The more similar the atomic mass numbers of the particles are, the stronger the effect is. Due to the limitation given by the precision of the calculation and the chosen bin widths, nearby primaries could induce to some degree coinciding submatrices, on which the final response matrix bases.

In Fig. 5.30, the entries<sup>30</sup> of the response matrix  $\mathbf{R}$ , defined in Eq.(5.27) and computed exemplarily using five primaries<sup>31</sup>, are visualized. White cells represent zero probability, or values that are below the precision of the floating-point numbers, and hence regarded to be zero. The abscissa represents the column numbers. The used energy range is divided into 25 bins. Thus, the solution vector  $\vec{X}$  has 25 entries for each of the 5 primaries, hence in total 125 ones resulting in the 125 columns of  $\mathbf{R}$ . The measured shower size distribution  $\log_{10}(N_{\text{ch}})$  vs.  $\log_{10}(N_{\mu})$  is divided in 45 bins for the muon and 55 bins for the charged particle number, resulting in 2475 cells. The ordinate reflects these 2475 entries of the measurement vector  $\vec{Y}$ . As defined in Section 4.2, the shower size plane is numbered from the left to the right, beginning from the lower left cell. This causes the slice-like structures in the probabilities shown in Fig. 5.30. The vertical dashed lines that are observable at all five columns, each corresponding to the lowest energy bin for the respective primary, are originating presumably from the limited precision whilst the computation of the response matrix. In case of these lowest energies (at around  $10^{15}$  eV), the probability that an air shower exhibits charged particle respectively muon numbers  $\log_{10}(N_{\text{ch}}) \geq 6.0$  and  $\log_{10}(N_{\mu}) \geq 5.0$  is very small (cf. Fig. 5.14, top panels). This small probability will be distributed over the cells of the  $\log_{10}(N_{\text{ch}})$ – $\log_{10}(N_{\mu})$  plane, such that the probability belonging to an individual cell will even be smaller. Whilst the computation of the response matrix (cf. Eq.(5.25)), these small probabilities, in combination

<sup>30</sup>Which are probability densities.

<sup>31</sup>Protons (representing a hydrogen nucleus), as well as helium, carbon (representing the carbon-nitrogen-oxygen group), silicon (representing the intermediate elemental group), and iron (representing the heavy component of cosmic rays) nuclei.

with the limited precision of the floating-point numbers, can yield artificial probability values unequal zero, finally resulting possibly in artefacts like e.g. the mentioned vertical structures<sup>32</sup>.

One has to conclude that the response matrix consists of many almost identical columns and rows. Furthermore, there is a large number of cells that are exhibiting very small probabilities and are surrounding the areas with higher probability values. Hence, the response matrix is ill-conditioned and to a certain extent singular<sup>33</sup>. Equation (5.26) is overdetermined, since  $\mathbf{R}$  is not a square matrix. An exact solution via inversion will not be possible. However, deconvolution algorithms are an appropriate means to solve such equations. Nevertheless, also these methods demand a minimum level of stability to provide reliable solutions. The stability of such an matrix equation is usually characterized by the *condition number*  $\kappa$ , in this case by the condition number  $\kappa(\mathbf{R})$  of the response matrix. This number reflects the accuracy, i.e. the statistical significance of the solution, that can be expected when performing a simple matrix inversion to solve the equation.

In case of a invertible matrix, the condition number can be calculated rather simple. The response matrix  $\mathbf{R}$  cannot be inverted; however, one can calculate a “pseudoinverse” by means of a *singular value decomposition* (SVD). A SVD transforms a matrix to a product of a diagonal matrix and orthonormal matrices, and is often used to investigate ill-conditioning. More details about the investigation or solution of ill-conditioned matrix equations can be found in [79, 53, 112, 113, 176, 85, 139, 147, 156, 167]. The response matrix  $\mathbf{R}$  is factorized based on the SVD following the instructions summarized in [79] and [80, p. 69 et seqq.], what yields:

$$\mathbf{R} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad (5.28)$$

where  $\mathbf{U} \in \mathbb{R}^{N_m \times N_m}$  and  $\mathbf{V} \in \mathbb{R}^{N_{\text{nucl}}N_s \times N_{\text{nucl}}N_s}$  are orthogonal matrices, whereas  $\mathbf{\Sigma} \in \mathbb{R}^{N_m \times N_{\text{nucl}}N_s}$  is a rectangular diagonal matrix:

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \text{ with } \mathbf{D} = \mathbf{diag}(\sigma_1, \sigma_2, \dots, \sigma_r), \text{ and } \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0. \quad (5.29)$$

The nonnegative diagonal entries  $\sigma_i$  are called the *singular values* of  $\mathbf{R}$ . The rank of  $\mathbf{R}$  is given by  $r$ .

The condition number  $\kappa(\mathbf{R})$  is given by (cf. [53]):

$$\kappa(\mathbf{R}) = \|\mathbf{R}\| \|\mathbf{R}^{-1}\|. \quad (5.30)$$

Since  $\mathbf{R}$  is not invertible,  $\kappa(\mathbf{R})$  can only be calculated based on the pseudoinverse. Furthermore, the condition number depends on the used norm<sup>34</sup>. From Eq.(5.28) one can deduce that

$$\mathbf{R}v_i = \sigma_i u_i, \quad \mathbf{R}^T u_i = \sigma_i v_i, \quad (5.31)$$

<sup>32</sup>Nevertheless, the probabilities given by the vertical structures are very small, and hence are compatible with zero. It has been observed that these lines disappear when the probabilities at small energies are on trial increased, what strengthens the assumption that especially these structures are caused by the very small probabilities for the lowest energies. The applied conditioning will remove these structures (cf. Fig. 5.31), such that they are unproblematic for the final analysis.

<sup>33</sup>More precisely, a matrix equation is called *singular* if the condition number is infinite, but only *ill-conditioned* if this number is relatively too large.

<sup>34</sup>E.g. the  $l_1$ ,  $l_2$  or the  $l_\infty$  norms, from which the Euclidean norm  $l_2$  is usually used in analyses comparable to that in this work.

and since  $\|\mathbf{R}\|_2 = \sigma_1$  and  $\|\mathbf{R}^{-1}\|_2 = \sigma_r^{-1}$  (cf. [80, p. 71]), one can write:

$$\kappa_2(\mathbf{R}) = \|\mathbf{R}\|_2 \|\mathbf{R}^{-1}\|_2 = \frac{\sigma_1}{\sigma_r}. \quad (5.32)$$

The index “2” indicates the use of the Euclidean norm  $l_2$ . Hence, the condition number can be calculated by the ratio of the largest to the smallest singular value.

Even though using double precision floating-point numbers in this work, not any matrix  $\mathbf{R}$  with arbitrary small singular values can be treated. This problem is discussed in more detail by Osborne [143]. A good rule of thumb for an indicator for the accuracy of the matrix equation solution is the logarithm of the condition number [171, p. 267]. If  $n_{\max}$  is the largest integral value not greater than the logarithm of the condition number, then the accuracy of the solution is  $15 - n_{\max}$  digits for a double precision arithmetic. The content of further digits would be statistically insignificant. Some preliminary tests have shown that based on the chosen bin widths the solution contains maximally numbers with 6 digits<sup>35</sup>. Following the rule of thumb, the condition number should not be larger than  $10^9$ . Since this is only a short and rough estimate, just to be on the safe side, in this analysis a condition number not larger than  $10^7$  is demanded. Although also matrices that fulfil this requirement are still to a certain degree ill-conditioned, it was found that the precision, respectively the stability is then accurate enough to solve the matrix equation by applying deconvolution algorithms, whereas a simple inversion method would still fail.

In general, two paths can be pursued to improve the stability of the matrix equation (given by Eq.(5.26)). The first one would be to reduce the number  $N_{\text{nucl}}$  of primary particles used whilst the unfolding. The second one would be to apply cuts to the response matrix, such that all probabilities below a certain threshold are neglected, i.e. assumed to be compatible with zero. With all examined constellations of numbers of primary particles and of possible cuts to the probabilities in  $\mathbf{R}$ , generated realistic toy data sets<sup>36</sup> are unfolded in order to verify whether or not a reliable solution can be achieved.

The quality of the solution will be judged by a comparison between the unfolded solution and the *a priori* known true test spectra, which were the input to these test runs. This can easily be done by eye. Furthermore, the unfolded solution can be forward folded, to get the two-dimensional shower size distribution belonging to the solution. This plane can be compared to the initial test data set by means of a Kolmogorov-Smirnov test. As a third criterion, the condition number  $\kappa_2(\mathbf{R})$  is computed and should not be larger than  $10^7$ , as explained above.

It was found that it is not possible to regard more than five primary particles. The comprehensive tests have furthermore shown that the solution is more reliable, if five instead of only four primaries are used for the unfolding. This result, obtained based on simulated test spectra, was confirmed by an unfolding of the shower size plane measured with KASCADE-Grande. The comparison between the measured distribution and the one obtained by a forward folding of the solution revealed a Kolmogorov-Smirnov probability of 0.85 to 0.88 (depending on which primaries are used) in case of four, and 0.97 in case of five considered particles. In case of only three primaries, the probabilities are even lower. This indicates that, based on the used high energy interaction model QGSJET-II-02, the solution is only compatible to the measurement if there are at least five particles considered whilst the deconvolution. Hence, in the following and in the main analysis,

<sup>35</sup>I.e. shower numbers up to  $\approx 10^6$ .

<sup>36</sup>The mechanism of testing the deconvolution procedure by means of test spectra will be explained in detail in the Section 6.3.

five primaries will be considered: protons, as well as helium, carbon, silicon, and iron nuclei.

The entries of the diagonal matrix  $\mathbf{D}$  defined in Eq.(5.29), if the response matrix  $\mathbf{R}$  bases on five primaries, are given in Appendix G, Eq.(G.1). The condition number<sup>37</sup> is  $\kappa_2(\mathbf{R}) = 0.54/(1.17 \times 10^{-13}) = 4.6 \times 10^{12}$ . Hence,  $\mathbf{R}$  is very ill-conditioned. Nevertheless, an unfolding based on this matrix is in general also possible in most cases, as intensive trial runs with toy spectra have shown. However, there were a few test data samples with some larger excesses, which are actually compatible with the normal statistical fluctuations and by this could also appear in the measured data set. This results in large oscillations in the unfolded solution. Hence, one would leave it to chance, whether or not the deconvolution of the measured data will have reliable results – depending on whether or not the measured shower size plane is such a “problematic” data set. Since this is not acceptable, in order to solve this problem, as aforementioned, cuts are applied to the response matrix. Tests have shown that these cuts do not deteriorate the results of the deconvolution of the already excess-free trial samples, but improve the solutions in case of the few “problematic” ones.

The suitable cuts on the response matrix  $\mathbf{R}$  have been derived based on detailed trial runs, where toy data samples have been unfolded and compared by eye to the true solution, as well as by the Kolmogorov-Smirnov test, as described afore. Furthermore, for every examined response matrix, the condition number is computed and was demanded to be below the afore-stated limit of  $10^7$ . As a result, the contents of all cells in  $\mathbf{R}$  that contain probability densities below  $p_{\min} = 10^{-4}$  will be set to zero in the final analysis. The entries of the diagonal matrix  $\mathbf{D}$  of the resulting conditioned response matrix are given in Appendix G, Eq.(G.2). The condition number is  $\kappa_2(\mathbf{R}) = 0.54/(2.27 \times 10^{-7}) = 2.4 \times 10^6$  now, and thus below the demanded limit of  $10^7$ . It was found that the exact value of the cut is only of minor importance. In case of the “unproblematic” test samples, it is anyway not necessary, but at least does not deteriorate the unfolding result. However, in case of the “problematic” test spectra, the results are significantly better when applying this cut, while its exact value  $p_{\min}$  is not important as long as it is in the range  $10^{-3}$  to  $10^{-5}$ . However, one should not cut away probabilities larger than  $10^{-3}$ , since one then affects probability ranges that are necessary to account for the intrinsic shower fluctuations, and hence are physically important. Although the fluctuations from lower to higher energy bins get less likely with increasing distance between the bins, such that the probabilities are very small from the mathematical point of view, they can have a huge impact on the final energy spectrum due to steeply decreasing flux of cosmic rays. Setting the cut value  $p_{\min}$  to probabilities larger than  $10^{-3}$  would hence induce non negligible systematic distortions.

Whilst the unfolding, for the energy spectra of individual particles only those energy bins are considered for that there is a non-zero probability that this particular primary with this specific energy contributes to at least one cell of the measured  $\log_{10}(N_{\text{ch}})$ – $\log_{10}(N_{\mu})$  shower size plane. Due to the cut on  $\mathbf{R}$ , a non-zero probability means a probability larger than  $p_{\min} = 10^{-4}$ . Consequently, the solution vector  $\vec{X}$  derived with the conditioned response matrix will no longer have necessarily all 25 entries that a solution vector derived with an unconditioned one would have. Analogous to Fig. 5.30, where the unconditioned response matrix is shown, in Fig. 5.31 the entries of the conditioned

---

<sup>37</sup>In case of four primaries it is  $\kappa_2(\mathbf{R}) \approx 5 \times 10^{10}$  to  $\kappa_2(\mathbf{R}) \approx 1 \times 10^{11}$  (depending on the chosen primaries); and thus a response matrix, only considering four primary particles, would be ill-conditioned, too, such that a reduction of the number of primaries would not solve this problem.

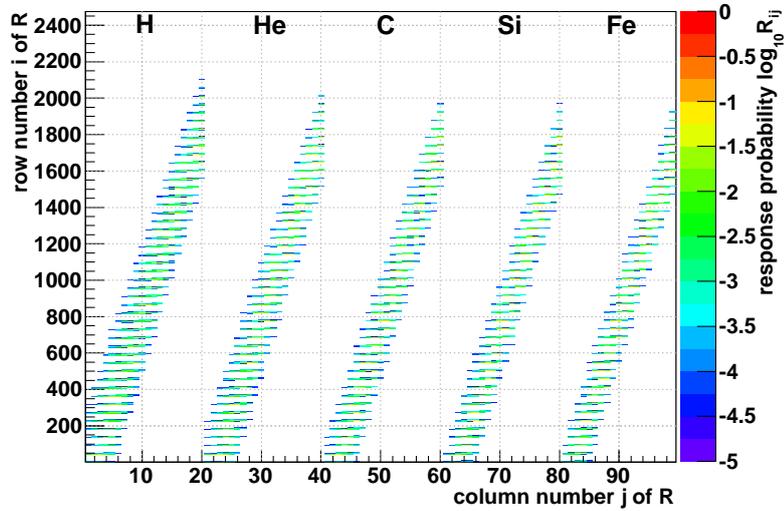


Figure 5.31: The entries (which are probability densities) of the response matrix  $\mathbf{R}$  defined in Eq.(5.27), visualized by a coloured histogram, after improving the condition number by the cut  $p_{\min} = 10^{-4}$  (see text).

one, computed applying the cut  $p_{\min} = 10^{-4}$ , are visualized. The unphysical vertical structures at the beginning of each particle’s energy interval are gone. Furthermore, the contributing probability distributions are less overlapping. As mentioned above, the individual particles contribute with less than 25 energy bins to the solution vector  $\vec{X}$ , such that the number of columns of the conditioned response matrix is smaller than in case of the unconditioned one.

Hence, if the conditioned response matrix is used, the matrix equation (given by Eq.(5.26)) has a sufficient level of stability, and can therefore be solved by means of unfolding methods. Although the reliability of the conditioned response matrix has already extensively been tested whilst deconvolutions of simulated toy data sets, a last additional check of consistency will be performed in the next section.

### 5.3.3 Consistency checks for the response matrix

The reliability of the conditioned<sup>38</sup> response matrix, introduced in Section 5.3.2, has already extensively been tested by means of deconvolutions of simulated toy data sets. However, in the following, some last checks of consistency will be performed. Initially, a cross-check with the Monte Carlo simulation set introduced in Section 5.2.1 will be accomplished, thereafter a first comparison with the measured shower size distribution follows.

As outlined in Section 5.2.1, there is a large simulation set available, in the following referred to as “direct Monte Carlo” set, which covers for one thing the air shower development, for another thing the entire measurement and reconstruction procedure. This set was simulated assuming a continuous energy spectrum with a spectral index of  $-2$ , and for five individual primaries<sup>39</sup> with equal abundance. Using this direct Monte

<sup>38</sup>The distinctive description “conditioned” will be left out in the future, since the conditioned response matrix will be the one used in the final analysis, while the unconditioned one will never be used again.

<sup>39</sup>Protons, as well as helium, carbon, silicon, and iron nuclei.

Carlo set, one can compute the resulting two-dimensional  $\log_{10}(N_{\text{ch}})$ – $\log_{10}(N_{\text{p}})$  shower size distribution, which is depicted in Fig. 5.32, left panel. All shown events pass the standard quality cuts (cf. Section 3.3) that will be used in the final analysis. In this computation, the derived shower size planes of all five primaries are combined to the shown resulting one. In addition, one can compute the same shower size distribution based on the response matrix, i.e. based on the parametrizations (again with equal abundance of the primaries). This is achieved by dividing the individual five energy spectra, which underlie the direct Monte Carlo simulations, into bins. Based on this, and regarding the response matrix as probability distribution, one can determine the two-dimensional shower size distribution by means of a random generator. For example, if there are  $N_i$  entries in the first energy bin of the direct Monte Carlo proton spectrum, one simply has to call the random generator, which regards that part of the response matrix as probability distribution that belongs to protons and this particular energy bin,  $N_i$  times. In every step, a  $(\log_{10}(N_{\text{p}}), \log_{10}(N_{\text{ch}}))$  pair will be returned. Repeating the same for every energy bin and all primaries, and combining all results in a common shower size distribution, referred to as “parametrization result”, yields the plane shown in Fig. 5.32, right panel. In general, there seems to be a good agreement between the results of the direct Monte Carlo simulations and the ones based on the parametrization. However, one has to take into account that the direct simulations are the same ones that have been used to parametrize the reconstruction properties in Section 5.2, and thus there is somehow a correlation between them and the response matrix. Nevertheless, for the parametrization of the shower fluctuations, different simulation sets are used. Furthermore, the computed response matrix was conditioned. Considering all this, the comparison between the direct Monte Carlo simulations and the parametrizations given by the response matrix is nonetheless an appropriate additional consistency check. In Fig. 5.33, the results of some projections applied to the distributions of Fig. 5.32 are illustrated. Also these one-dimensional distributions conform very well inside the statistical uncertainties.

The projections shown in Fig. 5.34 do not depend on the combined two-dimensional shower size distributions of all individual primaries. Instead, the distributions of them are considered separately now, more precisely exemplarily in case of protons or iron nuclei. Furthermore, the one-dimensional distributions based on the response matrix are now calculated immediately<sup>40</sup> using the parametrizations for protons, respectively iron nuclei. In general, also in case of single primaries there seems to be a good agreement between the direct Monte Carlo simulations and the parametrizations. Furthermore, it gets obvious that the utilization of parametrizations to derive the entries in the response matrix was necessary. The alternative that would be to derive the entries of the response matrix immediately based on the direct Monte Carlo simulations would cause problems due to the limited statistics. As it can be seen in Fig. 5.34, bottom, in case of the direct simulations the results suffer from statistical uncertainties. Furthermore, the possibility to analyse the influences of the individual probability functions, contributing to the response matrix, would get lost. The direct simulations only would allow computing the final probability that a primary with a certain energy will contribute to a specific cell of the shower size distribution, while all intermediate stages would be hidden in a kind of black box. Hence, this would also complicate the implementation of another high energy interaction model, what means in case of the parametrization, as used in this work, a simple exchange of the functions describing the shower fluctuations.

As a last consistency check, the parametrizations, used in the response matrix, are

<sup>40</sup>This time not using a random generator to compute a two-dimensional shower size distribution.

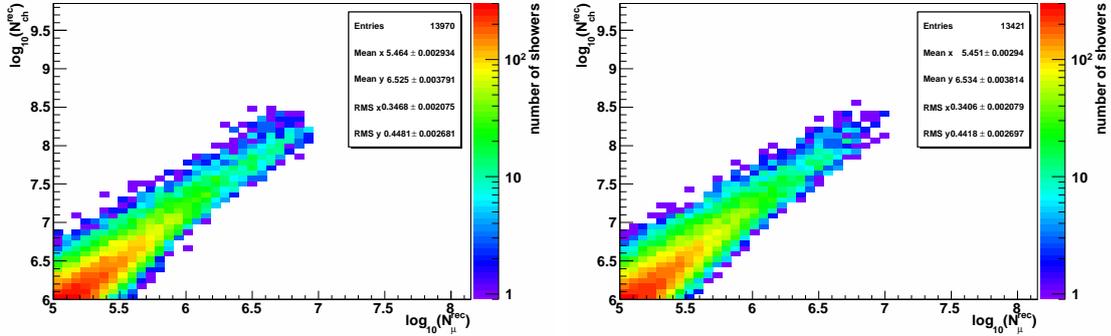


Figure 5.32: A comparison between the shower size distribution based on direct Monte Carlo simulations (left) or on the parametrizations included in the response matrix (right).

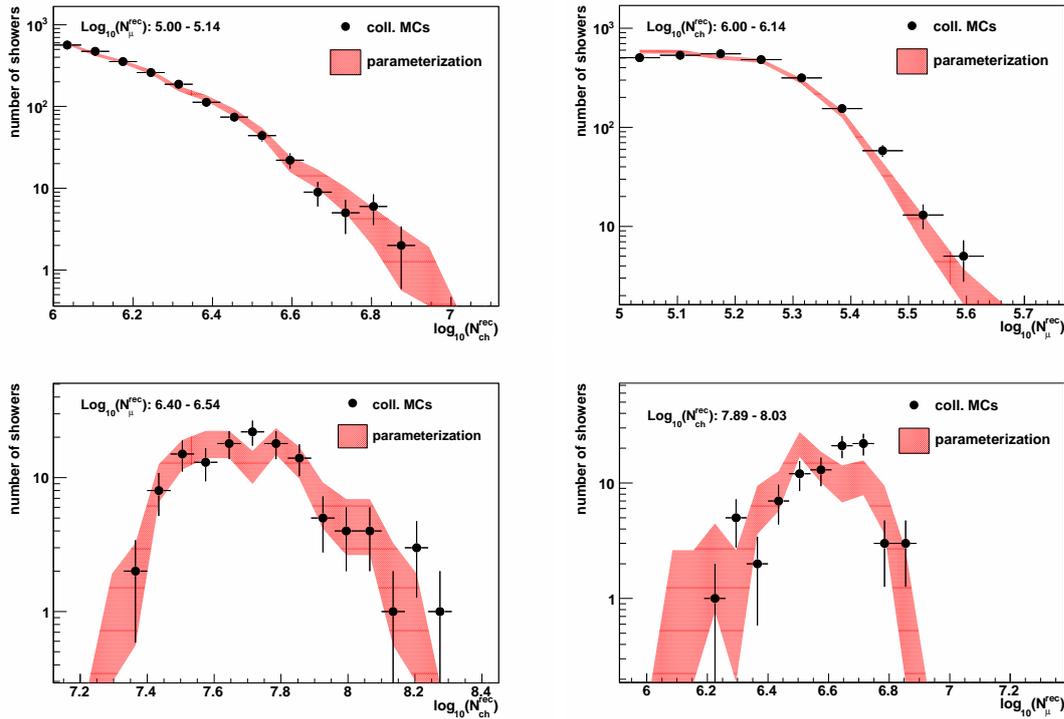


Figure 5.33: A comparison between the projections applied to the shower size distribution based on direct Monte Carlo simulations (Fig. 5.32, left panel, here labelled “coll. MCs”) and to the one based on the parametrizations (Fig. 5.32, right panel, here labelled “parameterization” and represented by their statistical error band): On the left for different  $\log_{10}(N_{\mu})$  intervals, on the right for  $\log_{10}(N_{ch})$ -intervals. The upper plots correspond to lower primary energies, and those at the bottom to higher ones.

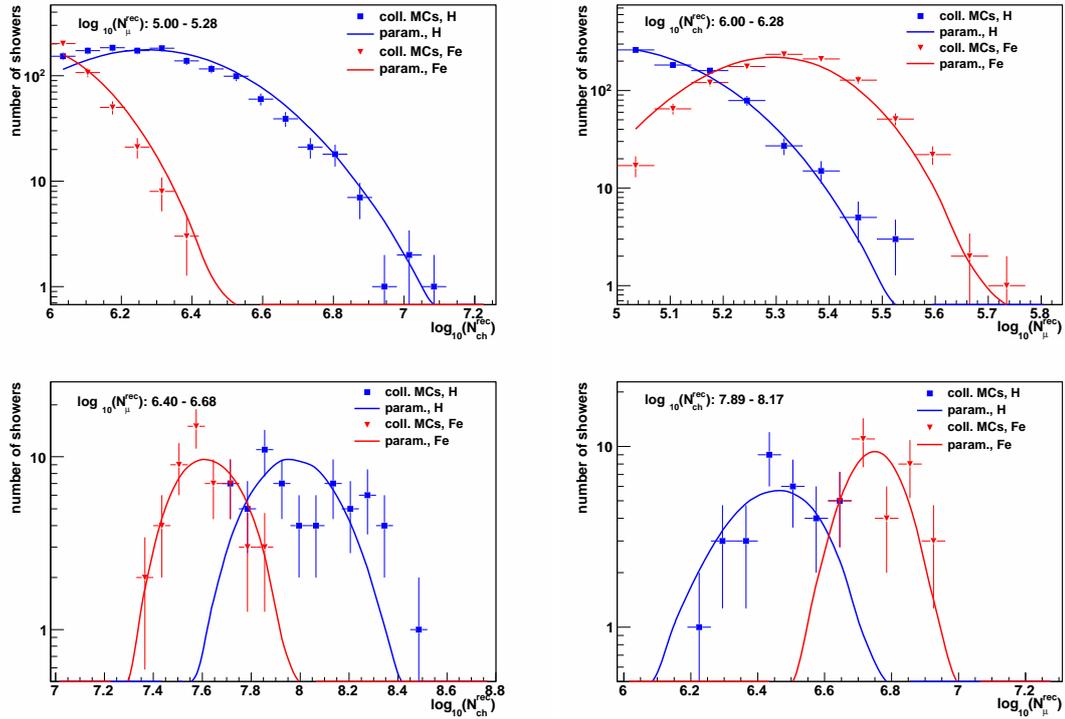


Figure 5.34: A comparison between the projections applied to the shower size distributions based on direct Monte Carlo simulations considering either pure protons or iron nuclei (labelled “coll. MCs”) and the one-dimensional distributions based on the parametrizations for protons or iron nuclei (labelled “param.”). On the left for different  $\log_{10}(N_{\mu})$  intervals, on the right for  $\log_{10}(N_{\text{ch}})$ -intervals. The upper plots corresponds to lower primary energies, and those at the bottom to higher ones.

compared to the shower size distribution measured with KASCADE-Grande. In Fig. 5.35, some projections applied to the measured shower size distribution (depicted in Fig. 4.8) are shown. In addition, the distributions computed based on the parametrizations are shown in case of protons and iron nuclei. They are normalized<sup>41</sup> such that their tails overlap with those of the measured distribution. Regarding the limited statistics, one cannot find indications that the parametrizations are not able to describe the measured distributions. On the contrary, the parametrizations are able to describe the measurement quite well. However, with these two primaries it is not possible to cover the middle part of the measured distribution, i.e. also particles with intermediate atomic mass are needed for an entire description. Furthermore, in order to describe the tails of the measured distribution reliably, the distribution based on the parametrizations for iron primaries is “higher” than that for protons. Hence, iron nuclei have to be more abundant than protons. This appears to be even more pronounced in case of higher energies, where iron nuclei have to dominate significantly over protons in order to describe the tails of the measured distribution reliably. Although this is only a rough cross-check, it gives already slight indications that the cosmic ray composition is dominated by a heavy component at the observed energies and regarding the interaction model QGSJET-II-02.

This is confirmed by Fig. 5.36, where the measured shower size distribution (grey

<sup>41</sup>This is done by hand, since this should only be a rough cross-check.

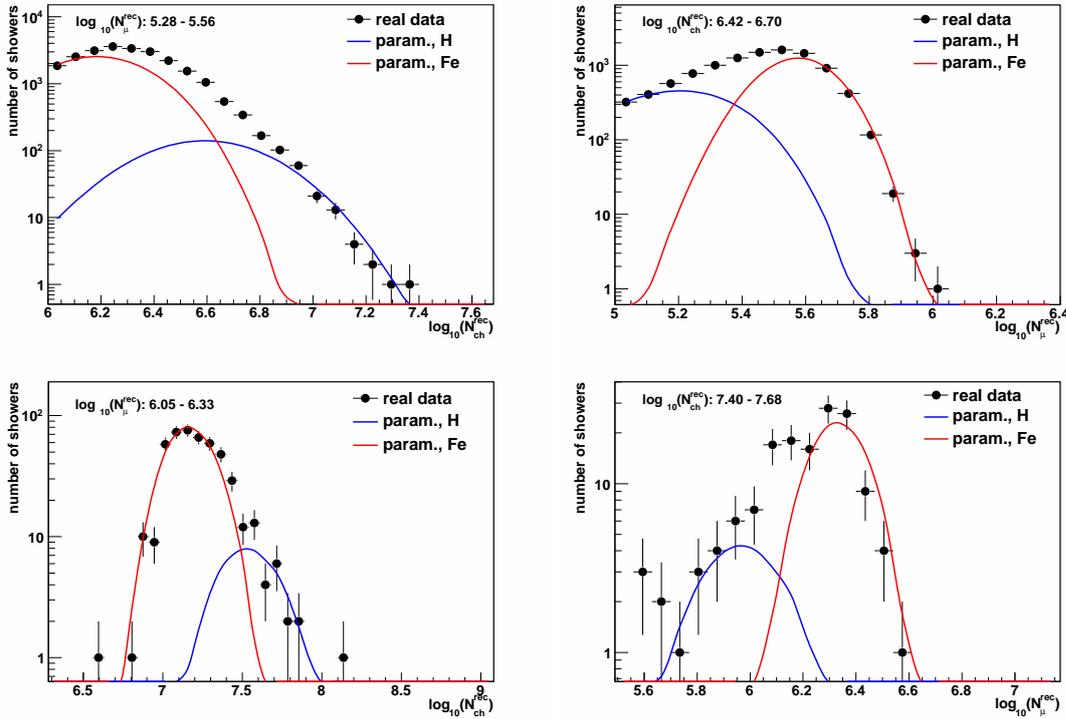


Figure 5.35: A comparison between the projections applied to the measured shower size distribution (Fig. 4.8, here labelled “real data”) and the one-dimensional distributions based on the parametrizations for protons or iron nuclei (labelled “param.”). On the left for different  $\log_{10}(N_{\mu})$  intervals, on the right for  $\log_{10}(N_{\text{ch}})$ -intervals. The upper plots corresponds to lower primary energies, and those at the bottom to higher ones.

isoareas) and the most probable  $\log_{10}(N_{\mu})$ – $\log_{10}(N_{\text{ch}})$  values according to the parametrizations (markers) are illustrated for different particles and primary energies. All over, the most probable values for the heavy nuclei (silicon, iron) are located in the central region of the measured distribution, where it has many entries, while the most probable values for lighter primaries (with same primary energy) are located at regions with smaller number of entries. Hence, this indicates that there are more heavy than light nuclei, i.e. that the composition is dominated by the heavy component.

However, one must not only consider these maximal probabilities, but also the widths of the probability distributions. In Fig. 5.37, again the measured shower size plane is shown, now in comparison to some isolines representing the cells  $(\log_{10}(N_{\text{ch}}), \log_{10}(N_{\mu}))_i$  of the data plane with constant probability (from the inner<sup>42</sup> to the outermost isoline: 0.1, 0.05 and  $10^{-4}$  probability density). Although only three primaries are illustrated, one can see that there is a large overlap between the individual distributions, such that a simple consideration of the maxima will not result in a reliable solution. Hence, the entire probability distributions have to be taken into account and used whilst the deconvolution procedure. Furthermore, the isolines, which correspond to the  $\log_{10}(N_{\mu})$ – $\log_{10}(N_{\text{ch}})$  combinations with a probability value of  $10^{-4}$ , represent the smallest probability value,

<sup>42</sup>In case of smaller energies, the widths of the probability distributions are as large that there are no individual probabilities larger than 0.1 or even 0.05, such that the inner isolines are missing in these cases.

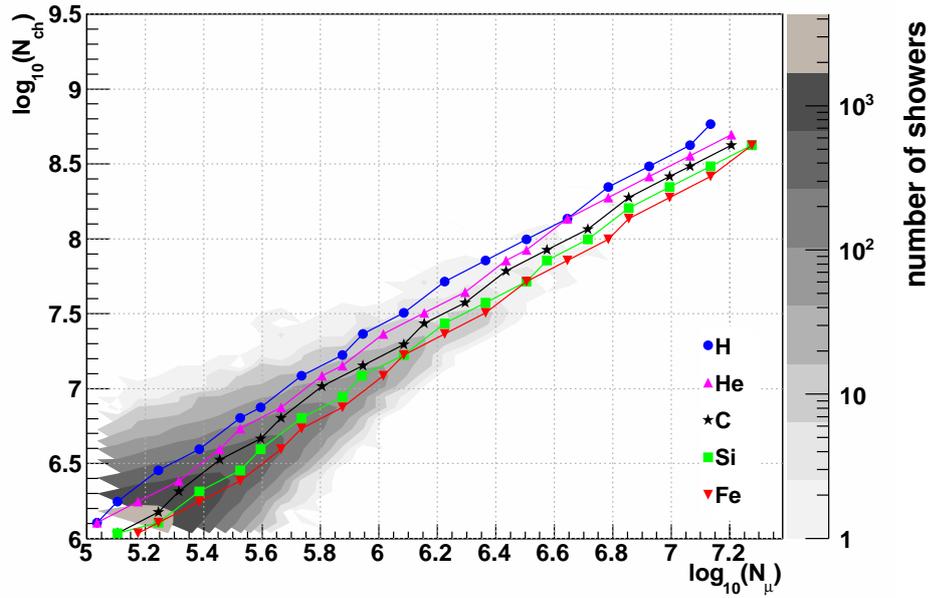


Figure 5.36: A comparison between the measured shower size distribution (grey isolines with filled areas in between) and the most probable  $\log_{10}(N_{\mu})$ – $\log_{10}(N_{\text{ch}})$  combinations according to the parametrizations, illustrated for different particles. The corresponding primary energies are out of the  $\log_{10}(E/\text{GeV})$  interval 6.98 to 9.5, which is divided into 18 bins of width 0.14 (the first bin, which covers the energy bin  $\log_{10}(E/\text{GeV})=6.98$  to 7.12, corresponds to the first markers shown in the bottom left corner). To guide the eyes, the markers are connected with straight lines.

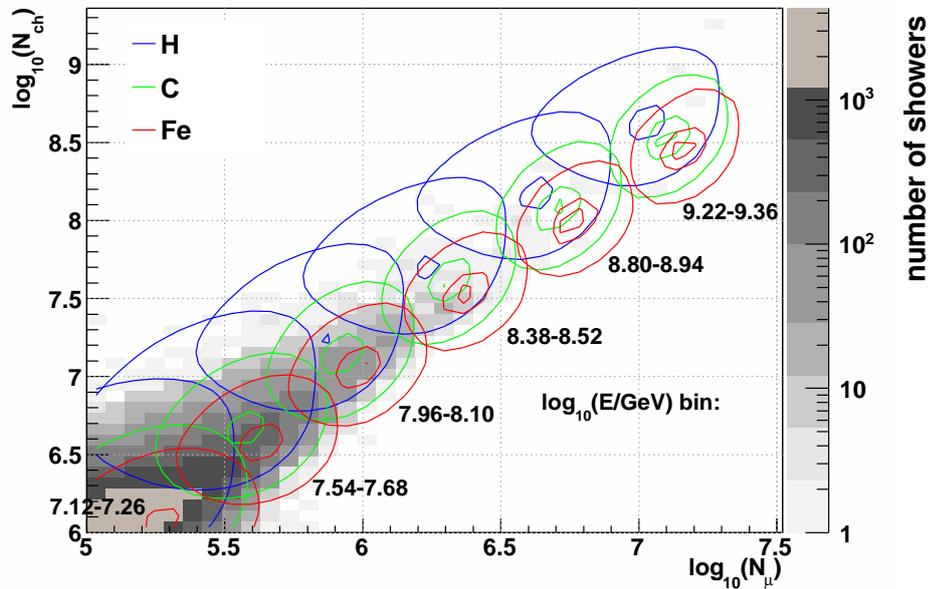


Figure 5.37: A comparison between the measured shower size distribution (grey histogram) and some isolines with  $\log_{10}(N_{\mu})$ – $\log_{10}(N_{\text{ch}})$  combinations of constant probability according to the parametrizations. This is illustrated exemplarily for protons, carbon as well as for iron, and in case of six energy bins (labelled below each isoline set). Each isoline set corresponds, from the inner to the outermost line, to 10%, 5% and 0.01% probability.

just considered in the response matrix after its conditioning (cf. Section 5.3.2). As can be seen, these outer isolines cover almost all measured data, hence the minimal probability is not set too large.

Summarizing the insights gained in this section, the response matrix, in the conditioned version according to the instructions given in Section 5.3.2, seems to represent the experiment's response to air showers, induced by an individual particle with specific primary energy, sufficiently well. The parametrizations conform with direct Monte Carlo simulations. Furthermore, the measured shower size distribution can be described based on the response matrix. Hence, together with the tests already performed in the previous section, one can conclude that the applied conditioning does not deteriorate the quality of the solution of the matrix equation Eq.(5.26). This enables to solve this equation by means of deconvolution methods, what will be the object of the next chapter.

## Chapter 6

# Unfolding Methods and Monte Carlo Tests

The main goal of this work is to solve the matrix equation (introduced in Eq.(5.26))

$$\vec{Y} = \mathbf{R}\vec{X}, \quad \text{with } \vec{Y} \in \mathbb{R}^{N_m}, \vec{X} \in \mathbb{R}^{N_{\text{nucl}}N_s}, \mathbf{R} \in \mathbb{R}^{N_m \times N_{\text{nucl}}N_s}, \quad (6.1)$$

which describes the convolution of the individual energy spectra of different primary particles, represented by  $\vec{X}$ , to the measured  $\log_{10}(N_{\text{ch}})$ – $\log_{10}(N_{\mu})$  shower size distribution, represented by  $\vec{Y}$ . There is a huge number of methods available to find a solution, however, with strongly varying quality. The general ideas or approaches that are pursued in the following base mainly on the information given in [38, 100, 54, 55, 41, 11, 39, 40]. Especially Jansson [100] gives a comprehensive overview over the multitudinous and widely spread topic of deconvolution.

As emphasized many times in the last chapter, direct methods, like for example a solution based on a simple matrix inversion, will not yield appropriate results, due to the ill-conditioning of the response matrix. Hence, these methods will only be outlined rather shortly. Afterwards, the focus will immediately be put on the deconvolution procedures<sup>1</sup>.

In the following, based on an example where a simple matrix inversion fails (Section 6.1), the usage of more complex deconvolution methods will be motivated, of which different algorithms and techniques will subsequently be discussed (Section 6.2). Afterwards, the implementation of that algorithms will be explained, and their quality will be tested based on toy data samples (Section 6.3).

### 6.1 Why Unfolding Algorithms?

In Section 4.2, it was already indicated, why a simple solution of the matrix equation (6.1) based on inversion fails in many cases: While a convolution is correlated to a smoothing of the distributions, since integral equations in general tend to damp especially higher frequencies, the deconvolution behaves conversely and amplifies fluctuations<sup>2</sup> present in the measured data sample (for further details confer [83, p. 7 et seqq.]). Hence, the solution by inversion, i.e.

$$\vec{X} = \mathbf{R}^{-1}\vec{Y}, \quad (6.2)$$

---

<sup>1</sup>Some authors also count the simple matrix inversion to the unfolding procedures, what is from the basic principle in fact true. However, in this work with unfolding methods only those are meant that use iterative algorithms or apply some other regularization techniques.

<sup>2</sup>And the KASCADE-Grande data sample will, beyond question, suffer from fluctuations due to the limited measurement time.

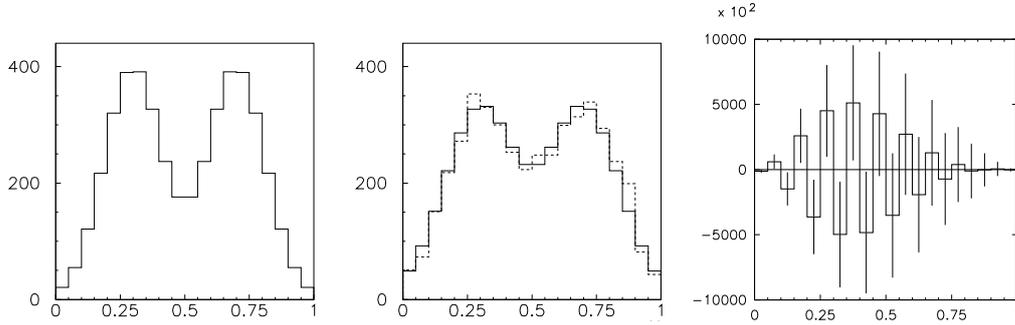


Figure 6.1: Left: A true histogram  $\vec{X}$ . Middle: The histogram based on a measurement (dashed lines, corresponding to  $\vec{Y}_{\text{meas}}$ ) as well as the one expected based on the matrix equation (6.1) (solid lines, corresponding to  $\vec{Y}$ ). Right: The estimated solution ( $\vec{X}_{\text{est}}$ ) based on a matrix inversion given by Eq.(6.3). All pictures are taken from Cowan [55].

or, more realistically, assuming that the entries of the actually measured<sup>3</sup> data vector  $\vec{Y}_{\text{meas}}$  are independent and following a Poisson distribution:

$$\vec{X}_{\text{est}} = \mathbf{R}^{-1} \vec{Y}_{\text{meas}} \quad , \quad (6.3)$$

will not yield appropriate results<sup>4</sup>.

This is illustrated exemplarily in Fig. 6.1, taken from Cowan [55]. On the left panel, an example spectrum<sup>5</sup>, what is assumed to be the true one  $\vec{X}$ , is depicted. Based on Eq.(6.1), the spectrum  $\vec{Y}$  that is expected to be observed can be computed by means of a forward folding (middle panel, solid lines). However, typically the statistics of the experiment is not infinite, such that the instead measured data sample  $\vec{Y}_{\text{meas}}$  will suffer from fluctuations, often underlying a Poisson distribution. This actually measured distribution is represented by the dashed lines, middle panel. Using this distribution as basis for the matrix inversion, given by Eq.(6.3), yields the estimated solution  $\vec{X}_{\text{est}}$  shown in the right panel. The solution oscillates, has huge variances and negative correlations between the adjoining bins. Although the solution is unbiased, since for the expectation value of  $\vec{X}_{\text{est}}$  follows:

$$\langle \vec{X}_{\text{est}} \rangle = \mathbf{R}^{-1} \langle \vec{Y}_{\text{meas}} \rangle = \vec{X} \quad , \quad (6.4)$$

it is useless in sense of a physical analysis due to the large statistical oscillations.

Nevertheless, from all unbiased estimators for the solution, the one based on the matrix inversion has at least the smallest possible variance, although it is already huge (cf. [55]). Hence, the only possibility to reduce the statistical uncertainty will be to allow a small systematic bias in the final solution. This won't yield an exact solution, but an appropriate one. The goal is not to regard any single fluctuation in the data set and to derive an exact solution for the actual energy spectra present whilst the limited time of measurements, but rather to find a best estimate for the true energy spectrum that is expected in case of infinite measurement time, and thus for infinite statistics. Hence, in the following, a good solution will be synonymous to an ‘‘appropriate’’ one, but not

<sup>3</sup>‘‘Measured’’ vector  $\vec{Y}_{\text{meas}}$  as counterpart to the ‘‘true’’ data vector  $\vec{Y}$ , which are only equal in case of infinite measurement statistics. However,  $\vec{Y}_{\text{meas}}$  is a good maximum likelihood estimator for  $\vec{Y}$ .

<sup>4</sup>Apart from the problem that  $\mathbf{R}$  in this work would not be invertible at all due to the ill-conditioning.

<sup>5</sup>This is only an arbitrary example, as well as the considered experiment and its response matrix.

to an “exact” one. Suitable methods that allow deriving such optimal solutions will be introduced now.

## 6.2 Unfolding Algorithms

Since the estimators for the solution of the matrix equation (6.1) that one would derive based on a simple matrix inversion can exhibit extremely large oscillations, one has to use alternative solution techniques that reduce these statistical uncertainties, but in turn do not yield unbiased results however. There are two classes of frequently used methods available<sup>6</sup>, which allow to solve, to a certain extent, this problem of amplifying the statistical fluctuations: unfolding methods based on regularization techniques or on iterative algorithms. Both categories will be introduced in the following.

### 6.2.1 Regularized unfolding

The principal idea underlying the regularized unfolding methods bases mainly on the investigations of Levenberg [124] and Marquardt [132]. In the following, the theoretical background will be introduced. Afterwards, a concrete implementation of these ideas will be discussed based on the method of “maximum entropy”.

#### General idea of regularized unfolding

The oscillating estimated solution that was computed in Section 6.1 by means of inversion given by Eq.(6.3) corresponds to a minimum value from a chi-square minimization<sup>7</sup> applied to Eq.(6.1):

$$\chi^2(\vec{X}) = \sum_{i=1}^{N_m} \left( \frac{y_i - \sum_{j=1}^{N_{\text{nucl}} N_s} R_{ij} x_j}{\sigma(y_i)} \right)^2 . \quad (6.5)$$

Thereby,  $\sigma(y_i)$  represents the statistical uncertainty of the data element  $y_i$ . As the obtained solution can suffer from large oscillations, a regularization has to be deployed. This can be achieved by limiting the solution space:

$$\chi^2(\vec{X}) \leq \chi_{\text{min}}^2 + \Delta\chi^2 . \quad (6.6)$$

This means to find a compromise between a systematic bias and possible statistical oscillations. However, there should still be a certain agreement between the derived solution and the measurement: i.e. the measured vector  $\vec{Y}$  should to a certain degree conform with the result  $\vec{R}\vec{X}$  obtained by a forward folding of the solution  $\vec{X}$ .

In addition to the limitation of the solution space to acceptable values, the solution should also exhibit a certain amount of smoothness. The idea will be, to select the

<sup>6</sup>A solution based on a forward folding is also often performed in comparable analyses. But, thereby, the shape of the solution that will be forward folded has to be known already *a priori*. Therefore, one cannot use this method to examine unknown structures in the solution. Furthermore, in this work the solution is not known sufficiently. Hence, the solution by forward folding will not be considered in the following.

<sup>7</sup>Respectively to a maximization of a log-likelihood function, which will be neglected in the following, since everything is analogue to the chi-square method.

smoothest solution out of the set of acceptable ones. For this purpose, a measure of smoothness has to be defined. This is done by means of a *regularization function*  $S(\vec{X})$ , while a *regularization parameter*  $\tau$  defines the strength of the regularization:

$$\chi^2(\vec{X}) = \sum_{i=1}^{N_m} \left( \frac{y_i - \sum_{j=1}^{N_{\text{nucl}}N_s} R_{ij}x_j}{\sigma(y_i)} \right)^2 + \tau S(\vec{X}) \quad . \quad (6.7)$$

If  $\tau = 0$ , the solution will oscillate as strong as the one derived by inversion does. The larger  $\tau$  is, the smoother the solution will get and the more the structures in the measured data distribution will be ignored.

There are many different approaches to realize a regularization and to implement the ideas into an unfolding algorithm. A very popular one, the *Tikhonov regularization* [172, 173, 174, 175], was developed by Tychonoff, who tried to solve ill-conditioned equations in a mathematically precise way. However, in the actual analysis this method is not applicable since it allows negative solutions, which are unphysical in terms of the sought-after fluxes. Furthermore, in case of small number of entries, the algorithm tends to reconstruct zero or small negative values. There are extensions and modifications available, which shall guarantee positiveness; however, for this analysis there are more stable methods usable. Another standard method based on regularization uses the principle of maximum entropy and will be outlined in the following.

### Principle of maximum entropy

The maximum entropy method is a regularized unfolding, which is often used in image reconstruction procedures and hence also intensively refined by the technical industry, like that of the digital cameras. In science, it is widely used, e.g. for image processing in astronomy [140]. The technique combines a least-squares method with the principle of maximum entropy. Its regularization function bases on the so-called entropy of information [160]:

$$H(\vec{X}) = - \sum_{j=1}^{N_{\text{nucl}}N_s} x_j \log_e(x_j) \quad . \quad (6.8)$$

The entropy can serve as measure for the smoothness of the solution. Hence, the regularization function of Eq.(6.7) can be written as:

$$S(\vec{X}) = -H(\vec{X}) = \sum_{j=1}^{N_{\text{nucl}}N_s} x_j \log_e(x_j) \quad . \quad (6.9)$$

The case of maximum entropy correlates to a minimal regularization function. This maximum entropy means the smoothest solution. In case of an uniformly distributed solution, the entropy would adopt the largest possible value. The regularization uses implicitly the positiveness of  $\vec{X}$ , what is an advantage over the aforementioned Tikhonov regularization, since by this only positive solutions are possible.

The maximum entropy method was modified by Schmelling [159], who applied the principle of *reduced cross-entropy*. Thereby, a best guess<sup>8</sup>  $\vec{X}_{\text{ref}}$  for the expected solution

<sup>8</sup>What is called *prior* in Bayes' theorem.

has to be stated. It is demanded that the solution  $\vec{X}$  does not deviate significantly from the reference vector  $\vec{X}_{\text{ref}}$ . Based on this, the regularization function can be written as follows:

$$S(\vec{X}) = -K(\vec{X}, \vec{X}_{\text{ref}}) = \sum_{j=1}^{N_{\text{nucl}}N_s} x_j \log_e \left( \frac{x_j}{x_{\text{ref},j}} \right) , \quad (6.10)$$

where  $K(\vec{X}, \vec{X}_{\text{ref}})$  is the so-called cross-entropy or Shannon-Jaynes entropy [102]. Using the reduced cross-entropy as regularization function in Eq.(6.7), yields:

$$\chi^2(\vec{X}) = \sum_{i=1}^{N_m} \left( \frac{y_i - \sum_{j=1}^{N_{\text{nucl}}N_s} R_{ij}x_j}{\sigma(y_i)} \right)^2 + \tau \sum_{j=1}^{N_{\text{nucl}}N_s} x_j \log_e \left( \frac{x_j}{x_{\text{ref},j}} \right) . \quad (6.11)$$

In this case, the regularization can be understood as follows: The bias of the estimated solution will be small, if the true solution is equal, or at least close to the reference distribution.

The resulting minimization functional is non-linear and of high complexity, what complicates the solution procedure. Whilst the cross-check unfoldings of the KASCADE-Grande data, the functional was solved based on the ROOT [45] class `TMinuit`, which implements the MINUIT package [98]. Thereby, the reference function was chosen to be as realistic as possible. For this purpose, the result obtained by our collaboration [33] was considered, which states a spectral index of the differential energy spectrum of  $\approx -3.0$  for energies below  $\log_{10}(E/\text{GeV}) = 8.0$ , and of  $\approx -3.2$  above. That is why the reference distribution, which describes the reference vector  $\vec{X}_{\text{ref}}$ , is assumed to follow a  $E^{-2.1}$  power law (what corresponds to a differential energy spectrum with flux index  $-3.1$ ). Furthermore, in the reconstructed integral energy spectrum there are roughly  $3 \times 10^6$  entries in the energy bin  $\log_{10}(E/\text{GeV}) = 5.93$  to  $6.07$ . Hence, the reference distribution is assumed to contain  $3 \times 10^6$  showers at an energy of  $\log_{10}(E/\text{GeV}) \approx 6.0$ . Using for simplification the logarithm of the energy and the assumptions made afore, as reference function  $x_{\text{ref}}$ , which contributes to the reference vector  $\vec{X}_{\text{ref}}$ , the following equation will be used:

$$\begin{aligned} x_{\text{ref}}(\log_{10}(E/\text{GeV})) &= 3 \times 10^6 \times \left( 10^{\log_{10}(E/\text{GeV})} \times 10^{-6.0} \right)^{-2.1} \\ &= 3 \times 10^6 \times 10^{-2.1(\log_{10}(E/\text{GeV})-6.0)} . \end{aligned} \quad (6.12)$$

Due to the steeply falling spectrum of cosmic rays and the resulting huge solution space, the following transformation will be used:

$$\begin{aligned} \tilde{x}_j &= x_j/x_{\text{ref},j} , \\ \tilde{R}_{ij} &= R_{ij} \times x_{\text{ref},j} . \end{aligned} \quad (6.13)$$

Since

$$\tilde{R}_{ij}\tilde{x}_j = R_{ij} \times x_{\text{ref},j} \times x_j/x_{\text{ref},j} = R_{ij}x_j , \quad (6.14)$$

the transformation mathematically does not change the left part of Eq.(6.11), i.e. the standard chi-square term. However, the regularization term will base on the new variables now. Thereby, the elements of the new reference vector will be  $\tilde{x}_{\text{ref},j} = 1$ . Furthermore,

since  $\vec{X}$  bases on a normalized probability distribution, also  $\vec{\tilde{X}}$  has to be used normalized in the entropy term. The new functional that has to be minimized in the final analysis is given by:

$$\chi^2(\vec{\tilde{X}}) = \sum_{i=1}^{N_m} \left( \frac{y_i - \sum_{j=1}^{N_{\text{nuc}}N_s} \tilde{R}_{ij} \tilde{x}_j}{\sigma(y_i)} \right)^2 + \tau \sum_{j=1}^{N_{\text{nuc}}N_s} \frac{\tilde{x}_j}{\sum_{k=1}^{N_{\text{nuc}}N_s} \tilde{x}_k} \log_e \left( \frac{\tilde{x}_j}{\sum_{k=1}^{N_{\text{nuc}}N_s} \tilde{x}_k} \right) . \quad (6.15)$$

The regularization parameter  $\tau$  will be chosen based on the *weighted mean squared error*, which will be described in Section 6.3.4. In case of the KASCADE-Grande data, values around  $\tau = 700$  to 900 permit reliable unfolding results.

It was found that in general the deconvolution based on the least-squares method in combination with the principle of reduced cross-entropy yields reliable results in many cases. Nevertheless, the iterative unfolding algorithms result in steadily comparable, or even better results, as ascertained by means of the unfolding of test spectra. The weakness of the entropy based method is the large number of variables<sup>9</sup> that have to be considered whilst the minimization procedure. Furthermore, the method works best in cases of rather underdetermined equation systems, for which reason it is often applied in image reconstruction analyses. However, in this work the convolution is strongly overdetermined. As a conclusion, the method of maximum entropy will only be used for the purpose of cross-checks. The slightly better iterative methods will be discussed in the following section.

### 6.2.2 Iterative unfolding

Albeit already the regularized unfolding method of maximum entropy, which bases on the least-squares method in combination with the principle of reduced cross-entropy, yields reliable results in many cases, an alternative approach will be discussed now: the iterative deconvolution algorithms. Among these, especially the algorithm developed by Gold as well as the Bayesian unfolding are extensively applied in situations comparable to that in this work. While the first one is a refinement of the Van Cittert algorithm and is based on the minimization of a chi-square functional, the second technique uses the Bayes' theorem. All iterative methods have in common that they try to approximate successively the solution by recursion. As explained in Section 6.1, the exact and unbiased solution of the matrix equation can suffer from huge statistical uncertainties such that one should rather search an "appropriate" solution instead of an exact one. The basic idea of the iterative methods is to stop the iteration at a certain depth, where the solution is as unbiased as possible, while the statistical fluctuations are still small enough. Strictly speaking, by this the iterative procedures deploy a kind of regularization, too. In the following, the algorithm of Van Cittert and its refinement by Gold as well as the method based on the Bayes' theorem will be introduced.

---

<sup>9</sup>At around 125, since the solution vector  $\vec{X}$  will have, maximally, 25 (energy bins) times 5 (primaries) entries.

### Van Cittert's algorithm

The Van Cittert algorithm is the most famous iterative deconvolution technique and the first published linear iterative unfolding method [185]. It has comprehensively been discussed by Jansson [100], while further details about the application can be found in [47, 166].

The iteration specification of the Van Cittert algorithm is given by:

$$x_i^{k+1} = x_i^k + \alpha \left( y_i - \sum_{j=1}^{N_{\text{nucl}}N_s} R_{ij}x_j^k \right), \quad i, j = 1, \dots, N_{\text{nucl}}N_s \quad . \quad (6.16)$$

The idea bases on a comparison between the measurement vector  $\vec{Y}$  and the data distribution predicted by the current estimated solution  $\vec{X}^k$ , computed by the forward folding  $R_{ij}x_j^k$ . The closer the estimated solution is to the exact one, the smaller the residual will be, i.e. the better the estimated solution will describe the measured data. The estimated solution of step  $k$  will be corrected by that residual, what yields the new estimation of step  $k + 1$ . The convergence can be steered by the parameter  $\alpha$ , the so-called *relaxation factor*, which was introduced by Jansson [101].

As can be seen, the dimensions of the data vector  $\vec{Y}$  and the one of the solution  $\vec{X}$  have to be equal. Since this is not the case in the current analysis, one has to apply the following transformations:

$$\mathbf{R} \rightarrow \mathbf{R}^T \mathbf{R} \quad \text{and} \quad \vec{Y} \rightarrow \mathbf{R}^T \vec{Y} \quad . \quad (6.17)$$

To account for the statistical errors  $\sigma(y_i)$  of the measured data, they can be considered via the error matrix  $C_{ij} = \delta_{ij} \sigma(y_i)$  in the modified response matrix  $\tilde{\mathbf{R}}$  and the modified measurement vector  $\tilde{\vec{Y}}$ :

$$\tilde{\mathbf{R}} = \mathbf{C} \mathbf{R} \quad \text{and} \quad \tilde{\vec{Y}} = \mathbf{C} \vec{Y} \quad . \quad (6.18)$$

The speed of convergence can be steered by the relaxation factor  $\alpha$ . For a save convergence, this parameter has to be  $0 < \alpha < 2/\lambda_{\text{max}}$ , with  $\lambda_{\text{max}}$  being the largest eigenvalue of the response matrix [191].

The disadvantages of Van Cittert's method are discussed by Jansson [100]. It is a general problem of most of the linear deconvolution methods that they cannot guarantee positiveness of the solution. However, negative fluxes would not physically make sense in the current analysis. This problem can be solved by utilization of nonlinear algorithms, which ensure that no channel of the solution can have negative values. In 1964, Gold [77] developed a nonlinear unfolding technique, which applies multiplicative corrections, instead of the subtraction in case of Van Cittert. This algorithm, which ensures positiveness of the solution, will be outlined in the next section.

### Gold's algorithm

The nonlinear iterative unfolding technique developed by Gold [77] in 1964 was regarded, over many years, to be unrelated to the algorithm of Van Cittert. Thirty years later, Xu et al. [191] showed that Gold's algorithm is simply a special form of the Van Cittert technique with a variable relaxation function, which does not only vary with the channel number, but also with the iteration index. In comparison to the Van Cittert algorithm,

Gold's one demands stricter conditions for convergence. It is only applicable, if all principal minors of the response matrix and all its eigenvalues are positive and real. The following deliberations base on the ones performed in [77].

In a first assumption, the response matrix  $\mathbf{R}$  is considered to be quadratic<sup>10</sup>, i.e.  $\mathbf{R} \in \mathbb{R}^{n \times n}$ . It is not singular and has only positive entries, while the diagonal elements additionally have to be non-zero. The two latter claims are satisfied in the current analysis, since only non-negative probabilities contribute to  $\mathbf{R}$ , while the non-singularity condition is sufficiently fulfilled, since the response matrix is ill-conditioned, but not completely singular. The data vector  $\vec{Y}$  has to consist of positive values only, what can simply be achieved by only regarding those cells of the measured two-dimensional shower size distribution that have non-zero entries. Last but not least, the solution vector  $\vec{X}$  has to be non-negative, what is the case since the cosmic ray flux does not adopt negative values. Under all these assumptions, stated by Gold [77], the problem can be transferred to a kind of eigenvalue problem. There exists a diagonal matrix  $\mathbf{D}$  with unique elements such that:

$$\vec{X} = \mathbf{D}\vec{Y}, \quad \text{with } D_{ii} = x_i/y_i, \quad i = 1, 2, \dots, n \quad . \quad (6.19)$$

Note that only in the trivial case, where the response matrix  $\mathbf{R}$  itself is diagonal,  $\mathbf{D}$  can be identified with the inverse of  $\mathbf{R}$ . If the diagonal matrix  $\mathbf{D}$  is known, one can simply compute  $\vec{X}$ , and hence the sought-after energy spectra. While Van Cittert's algorithm successively approximates the solution  $\vec{X}$ , in Gold's ansatz the elements of  $\mathbf{D}$  will be iteratively approximated in lieu thereof (cf. [77, p. 12 et seqq.] for a detailed derivation):

$$D_{ii}^{k+1} = \frac{D_{ii}^k y_i}{\sum_{j=1}^n R_{ij} D_{jj}^k y_j}, \quad i, j = 1, 2, \dots, n \quad . \quad (6.20)$$

However, by multiplication with the measurement  $\vec{Y}$  and using Eq.(6.19), one can transform this equation into a recursion relation between successive approximations for the solution  $\vec{X}$ :

$$x_i^{k+1} = \frac{x_i^k y_i}{\sum_{j=1}^n R_{ij} x_j^k} = x_i^k \frac{y_i}{y_i^k}, \quad i, j = 1, 2, \dots, n \quad . \quad (6.21)$$

Thereby, the solution  $\vec{X}$  can be approximated without using the inverse of  $\mathbf{R}$ . While in Van Cittert's algorithm, given by Eq.(6.16), the difference  $y_i - y_i^k$  between the measurement vector  $\vec{Y}$  and the data vector  $\vec{Y}^k$  that is predicted by the actual solution  $\vec{X}^k$  is regarded as indicator for the quality of the approximation, in Gold's approach their ratio  $y_i/y_i^k$  is considered. Therefore, it is sometimes named Gold's ratio method. For example, if the element  $y_i^k$  of the estimated data vector  $\vec{Y}^k$ , which is derived by a forward folding of the actual solution  $\vec{X}^k$ , is smaller than that one in the measured vector  $\vec{Y}$ , the ratio will be larger than one. By this, the corresponding entry in the estimated solution will be increased from iteration step  $k$  to  $k + 1$ , such that the subsequent solution estimate will conform better with the measured data sample.

<sup>10</sup>The non-quadratic case will be discussed later.

Furthermore, it is obvious that all successive estimates  $\vec{X}^k$  are positive vectors<sup>11</sup>, since the data vector consists of positive values only, as demanded afore. By this, the initially claimed condition that the solution vector has to be non-negative is accomplished. Gold [77, p. 21 et seqq.] proves that, if the conditions<sup>12</sup> claimed at the beginning of this subsection are fulfilled, the iteration given by Eq.(6.21) will converge to the exact solution, as long as it lies in the non-negative subspace.

In case of non-symmetric response matrices  $\mathbf{R} \in \mathbb{R}^{m \times n}$  with  $m > n$ , i.e. the matrix equation (6.1) is overdetermined, the method has to be modified. Gold [77, p. 22] states that if  $\mathbf{R}$  does not fulfil the required conditions,  $\mathbf{R}^T \mathbf{R}$  however does and is positive definite<sup>13</sup>. The modified matrix equation follows by multiplication of Eq.(6.1) by  $\mathbf{R}^T$ :

$$\underbrace{\mathbf{R}^T \vec{Y}}_{\vec{Y}_{\text{mod}}} = \underbrace{\mathbf{R}^T \mathbf{R}}_{\mathbf{R}_{\text{mod}}} \vec{X} \quad . \quad (6.22)$$

Since  $\vec{Y}$  is positive, also  $\vec{Y}_{\text{mod}} = \mathbf{R}^T \vec{Y}$  will be. Hence,  $\vec{X}$  will meet the requirement of a positive solution. In general, this modified equation fulfils all conditions necessary for a convergence of Gold's algorithm to the exact solution.

Gold and Bennett [78] discuss in detail how random statistical errors possessed by the measured data sample will influence the solution of the convolution equation. The regarding uncertainties depend strongly on the statistics of the data sample, and are not *a priori* considered in the response matrix<sup>14</sup>. Practically, these statistical errors of the data set are often regarded additionally in an extended response matrix  $\tilde{\mathbf{R}}$  by means of an error matrix  $\mathbf{C}$ :

$$\tilde{\mathbf{R}} = \mathbf{C} \mathbf{R} \quad , \quad \tilde{\vec{Y}} = \mathbf{C} \vec{Y} \quad , \quad \text{with } \mathbf{C} \in \mathbb{R}^{m \times m} \quad , \quad \text{and } C_{ij} = \frac{\delta_{ij}}{\sigma_i} \quad . \quad (6.23)$$

The statistical errors  $\sigma_i$  of the data vector element  $y_i$  are usually assumed to be Poissonian ones<sup>15</sup>. Applying this error extension to the modified response matrix  $\mathbf{R}_{\text{mod}}$  and the data vector  $\vec{Y}_{\text{mod}}$  from Eq.(6.22), yields:

$$\tilde{\mathbf{R}}_{\text{mod}} = (\mathbf{C} \mathbf{R})^T (\mathbf{C} \mathbf{R}) \quad , \quad \tilde{\vec{Y}}_{\text{mod}} = (\mathbf{C} \mathbf{R})^T \mathbf{C} \vec{Y} \quad , \quad \text{with } \mathbf{C} \in \mathbb{R}^{m \times m} \quad , \quad \text{and } C_{ij} = \frac{\delta_{ij}}{\sigma_i} \quad . \quad (6.24)$$

Using these modifications, Eq.(6.22) has to be written as:

$$\tilde{\vec{Y}}_{\text{mod}} = \tilde{\mathbf{R}}_{\text{mod}} \vec{X} \iff \mathbf{R}^T \mathbf{C} \mathbf{C} \vec{Y} = \mathbf{R}^T \mathbf{C} \mathbf{C} \mathbf{R} \vec{X} \quad . \quad (6.25)$$

Since  $\mathbf{C}$  is a diagonal matrix<sup>16</sup>,  $\mathbf{R}$  and  $\tilde{\mathbf{R}}$  have the same rank. The new response matrix  $\tilde{\mathbf{R}}_{\text{mod}}$  is still positive definite. Finally, Gold's iteration method, given in Eq.(6.21), can

<sup>11</sup>If the initial vector  $\vec{X}^0$  is also positive, what can, however, be fulfilled, since this vector has to be set manually at the very beginning. In this context, a short remark: Gold [77, p. 27 et seqq.] has shown that the convergence of the iteration is independent of the choice of  $\vec{X}^0$  (as long as it is positive). Only the rate of convergence will be influenced. Practically, as starting point  $\vec{X}^0$  a positive vector that is preferably close to the expected solution should be chosen.

<sup>12</sup>Primarily, that the response matrix has only positive principal minors and all its eigenvalues are real and positive, which especially prevail with matrices that are positive definite.

<sup>13</sup>Although not mentioned by Gold, this statement bases on an application of the ideas going along with the Cholesky decomposition [80, p. 143 et seqq.].

<sup>14</sup>In which, due to the application of parametrizations, effects of the limited Monte Carlo statistics, which correlate in one sense to the uncertainty in the data sample, are cancelled out to a certain extent.

<sup>15</sup>In this analysis,  $\sigma_i = \sqrt{y_i}$  will be assumed.

<sup>16</sup>Which may have positive elements only, what is, however, fulfilled in this analysis due to the specific choice of Poissonian errors.

be written as:

$$x_i^{k+1} = \frac{x_i^k \tilde{y}_{\text{mod},i}}{\sum_{j=1}^n \tilde{R}_{\text{mod},ij} x_j^k}, \quad i, j = 1, 2, \dots, n \quad . \quad (6.26)$$

This recurrence formula will be used in the final unfolding analysis of this work.

Whilst his reasoning, Gold [77, p. 23] introduced the positive definite quadratic form:

$$F(\vec{X}^k) = (\vec{X}^k - \vec{X})^T \mathbf{R} (\vec{X}^k - \vec{X}) \quad . \quad (6.27)$$

Mathematically, this functional, which is also named error function, is the norm of the difference vector  $\vec{X}^k - \vec{X}$  between the estimated  $\vec{X}^k$  and the exact solution  $\vec{X}$ , while  $\mathbf{R}$  is the underlying metric respectively. The convergence of Eq.(6.26) is equivalent to the convergence  $F \xrightarrow{k \rightarrow \infty} 0$ . Using Eq.(6.26) and after a short calculation, this functional can be transferred to the following one:

$$F(\vec{X}^k) = \sum_i \frac{\left( \sum_{j=1}^n R_{ij} x_j^k - y_i \right)^2}{\sigma(y_i)^2} = \chi_k^2 \xrightarrow{k \rightarrow \infty} 0 \quad . \quad (6.28)$$

Therefore, the convergence of Gold's algorithm is equivalent to a chi-square minimization, which converge to the exact solution of the matrix equation. However, as explained in Section 6.1, this exact and unbiased solution that one would achieve in case of a large number of iteration steps ( $k \rightarrow \infty$ ) is not wanted, since it can suffer from large oscillations. Hence, one will have to stop Gold's iteration at an appropriate depth, such that the estimated solution will be biased, but, as a *quid pro quo*, will exhibit significantly smaller fluctuations than the unbiased exact solution. The optimal stopping criterion will be discussed in Section 6.3.4.

As a conclusion, Gold's algorithm, in the form given by Eq.(6.26), is an easy to use and robust algorithm. Furthermore, no defining constants have to be specified. Only the first guess for the solution, based on which the iteration will start, has to be chosen at the very beginning. Thereby, not the convergence itself, but only its rate will depend on the initial value. This starting point should be set as close as possible to the expected true solution, while the accuracy of its definition is uncritical however. Furthermore, the introduced method is very fast and hence an appropriate mean to deconvolute also two-dimensional problems. In comparison to Van Cittert's algorithm, Gold's one guarantees the positiveness of the solution implicitly and by this physically meaningful results. In addition, it is less sensitive to ill-conditioning in the response matrix. Nevertheless, also in case of his rather stable algorithm, Gold [77, p. 36 et seqq.] concluded that, in case of strongly ill-conditioned response matrices, "the rate of convergence may be too low to afford a proper definition of the appropriate solution". Among other things, he suggested to apply smoothing techniques or to limit the domain of the solution. Both ideas are in general already considered in this work. The conditioning of the response matrix was improved in Section 5.3.2 by application of specific cuts on the domain of the probabilities, whereas the use of parametrizations whilst the computation of the response matrix introduced a kind of smoothing. In summary, the course for a successful implementation of Gold's unfolding method is set. Preliminary trials based on test spectra have shown that reliable solutions are derivable, such that Gold's method will be the primary one used in this work. Nevertheless, beside the already introduced regularized deconvolution

technique based on the principle of maximum entropy (Section 6.2.1), a second comparative method will be used in this analysis in order to verify the results derived with Gold's algorithm. For this purpose, an iterative method that applies Bayes' theorem will be used, which is also widely used in comparable analyses and will be outlined now.

### Bayesian unfolding

Algorithms based on the Bayes' theorem are often applied solution techniques for convolutions equivalent to that given by the matrix equation (6.1). While in Bayes' native approach the *a priori* knowledge of the initial distribution is a basic prerequisite, the deduced deconvolution methods overcome this limitation by iterative procedures. The considerations performed in the following base mainly on [56, 155, 130, 96, 137, 136, 109, 110, 111].

If a solution  $\vec{X}$  of Eq.(6.1) exists, what is the case as already shown in the sections afore, its elements  $x_i$  can be written as a linear combination based on the elements  $y_j$  of the measurement vector  $\vec{Y}$ :

$$P(x_i) = \sum_j P_{ij}P(y_j) = \sum_j P(x_i|y_j)P(y_j), \quad \text{or } P(\vec{X}) = \mathbf{P}P(\vec{Y}) \quad . \quad (6.29)$$

Thereby,  $\mathbf{P}$  is an appropriate transfer matrix, which contains the conditional probability of a true event at  $x_i$ , under the condition of a measured event at  $y_j$ . As in case of Gold's algorithm, where the transfer matrix  $\mathbf{D}$  (cf. Eq.(6.19)) is not the inverse of the response matrix  $\mathbf{R}$ , also  $\mathbf{P}$  is not! Since  $\mathbf{P}$  contains only non-negative values by construction, again the non-negativeness of the solution vector  $\vec{X}$  is ensured. While in Gold's algorithm  $\mathbf{D}$  has to be estimated to solve the matrix equation for  $\vec{X}$ , in the actual Bayesian approach the conditional probability  $\mathbf{P}$  will finally have to be approximated successively. This probability can be reformulated based on Bayes' theorem:

$$P(x_i|y_j) \stackrel{\text{Bayes' theorem}}{=} \frac{P(y_j|x_i)P(x_i)}{P(y_j)} = \frac{P(y_j|x_i)P(x_i)}{\sum_l P(y_j|x_l)P(x_l)} \quad . \quad (6.30)$$

This expression for  $\mathbf{P}$  can now be substituted into Eq.(6.29):

$$P(x_i) = \sum_j \frac{P(y_j|x_i)P(x_i)P(y_j)}{\sum_l P(y_j|x_l)P(x_l)} \quad . \quad (6.31)$$

However, this equation cannot be solved immediately since the prior  $P(\vec{X})$ , which is the sought-after solution itself, is not known, but contributes also to the right-hand side of the expression. As mentioned in the introductory words to this subsection, the problem due to the unknown prior can be solved by means of iterative procedures. Thereby, Bayes' postulate, occasionally also referred to as *principle of equidistribution of ignorance*, will be used, which assumes that, if there exists no further knowledge, the prior  $P(\vec{X})$  can be approximated by a uniform probability distribution<sup>17</sup>  $P(\vec{X}^0) = 1/I$ , where  $I$  is the dimension of the vector  $\vec{X}$ . Taking  $P(\vec{X}^0)$  as first guess, the exact solution  $P(\vec{X})$  can

<sup>17</sup>Alternatively, one can also use a physically motivated starting value, or a first guess based on the results of other unfolding methods (Gold, maximum entropy, etc.).

now be computed iteratively:

$$P(x_i^{k+1}) = P(x_i^k) \sum_j \frac{P(y_j|x_i)P(y_j)}{\sum_l P(y_j|x_l)P(x_l^k)} . \quad (6.32)$$

The probabilities  $P(x_i)$  and  $P(y_j)$  can be expressed as follows:

$$P(x_i) = \frac{x_i}{\sum_{l=1}^L x_l} \quad \text{and} \quad P(y_{j'}) = \frac{y_{j'}}{\sum_{j=1}^J y_j} , \quad (6.33)$$

where  $L$  and  $J$  are the dimensions of  $\vec{X}$  and  $\vec{Y}$  respectively. Furthermore, the transfer probability  $P(y_j|x_i)$  is nothing less than the response matrix:

$$P(y_j|x_i) = R_{ji} . \quad (6.34)$$

Hence, Eq.(6.32) can be replaced by the following recurrence formula:

$$x_i^{k+1} = x_i^k \sum_j \frac{P(y_j|x_i)y_j}{\sum_l P(y_j|x_l)x_l^k} = x_i^k \sum_j \frac{R_{ji}y_j}{\sum_l R_{jl}x_l^k} . \quad (6.35)$$

The summation over  $j$  can be understood as a regularization or, more precisely, as a correction factor affecting  $x_i^k$  analogue to that in Van Cittert's or Gold's approach. Usually, the denominator tends to be the dominant initiator of the correction. If e.g. the estimated solution  $x_i^k$  is too small, the recurrences of  $x_i^k$  in the denominator would usually enlarge the value of  $x_i^k$  that is in front of the summation, and a new and typically larger solution  $x_i^{k+1}$  will be approximated.

Because the Bayesian approach rests on the concept of probability, one has to take into account that full efficiency of the experiment will not occur in all cases. Especially for air showers with smaller primary energies, the probability to induce an entry in the measured two-dimensional shower size spectrum is possibly smaller than 1. Hence, in case of the Bayesian unfolding the efficiency  $\varepsilon$  has to be considered explicitly in the used algorithm. Since the response matrix contains all probabilities, the corresponding efficiencies can be computed by:

$$\varepsilon_i = \sum_j R_{ji} . \quad (6.36)$$

Based on this expression, the algorithm that will finally be used in this work can be written as:

$$x_i^{k+1} = \frac{x_i^k}{\sum_j R_{ji}} \sum_j \frac{R_{ji}y_j}{\sum_l R_{jl}x_l^k} . \quad (6.37)$$

The convergence of the Bayesian unfolding method was topic of many publications. For example, Kennett et al. [109] have investigated the convergent properties in detail and investigated the relationship between the depth of the iteration and the achieved resolution. Summarizing the results of all contributions, the convergence is not in every case sure, contrary to Gold's algorithm, which guarantees this property *a priori*. Furthermore, the implementation is more difficult, and the determination of the right stopping criterion is not as simple as in case of Gold's approach. However, the Bayesian deconvolution has performed rather stable and offered good results, as trial unfoldings based on

test spectra have shown. As stated in the previous subsection, Gold’s algorithm will be the first choice for this work, while, nevertheless, the main results will be cross-checked by means of the Bayesian method, beside the third approach based on the principle of maximum entropy.

## 6.3 Implementation and Quality of the Unfolding Algorithms

In Section 6.2, different deconvolution techniques were introduced, which can be used to solve the matrix equation

$$\vec{Y} = \mathbf{R}\vec{X}, \quad \text{with } \vec{Y} \in \mathbb{R}^{N_m}, \vec{X} \in \mathbb{R}^{N_{\text{nuc1}}N_s}, \mathbf{R} \in \mathbb{R}^{N_m \times N_{\text{nuc1}}N_s}, \quad (6.38)$$

for  $\vec{X}$  that contains the information about the primary energy spectra of individual mass groups. These methods have been chosen attentively out of the large number of unfolding procedures that can be used to solve such matrix equations. The preselected<sup>18</sup> algorithms will now be tested by means of Monte Carlo methods. Firstly, the meaning of the term “iteration depth”, which is related to the number of iteration steps and will be used several times in the following, will be defined precisely. Then, appropriate toy data samples will be generated (Section 6.3.2) that can be used as basis for trial unfoldings, which serve to investigate the propagation of errors (Section 6.3.3), to find appropriate stopping criterions for the iterative unfolding methods (Section 6.3.4), and to analyse the quality and robustness of the algorithms (Section 6.3.5).

### 6.3.1 Definition of the “iteration depth”

In case of the iterative unfolding procedures, a kind of regularization of the reconstruction process can be applied by stopping the recursion after a specific number of iteration steps, before it converges to the exact, but probably oscillating solution. The number of iterations is rather a mathematical size, and less physically motivated. Hence, an alternative quantity will be introduced now that represents the number of iteration steps in a more meaningful way. As stated in Section 6.2.2, Gold’s algorithm is equivalent to a chi-square minimization (cf. Eq.(6.28)). Hence, the reduced chi-square value<sup>19</sup>

$$\chi_k^2 = \frac{1}{M} \sum_{i=1}^M \frac{\left( \sum_{j=1}^N R_{ij} x_j^k - y_i \right)^2}{\sigma(y_i)^2} := \frac{1}{M} \sum_{i=1}^M \chi_{k,i}^2. \quad (6.39)$$

is a good measure, how accurately the actual solution estimate  $\vec{X}^k$ , that is derived after  $k$  iteration steps, describes the measured data sample  $\vec{Y}$ . Thereby,  $M$  and  $N$  are the dimensions of  $\vec{Y}$  and  $\vec{X}^k$  respectively. The used statistical errors  $\sigma(y_i)$  of the data sample

<sup>18</sup>As already mentioned in Section 6.2, Gold’s method turned out to be the most reliable one for this analysis. Hence, in the following, mainly results for this algorithm will be shown, whereas, however, all considerations have been applied to the other techniques, too.

<sup>19</sup>However, since some of the entries of the data vector  $\vec{Y}$  are small numbers, the chi-square test have to be interpreted with caution. As a replacement for the physically less meaningful number of iterations it is suitable, while, in the final analysis, a precise comparison between the unfolded solution and the measurement will additionally be performed by means of a Kolmogorov-Smirnov test [117, 163].

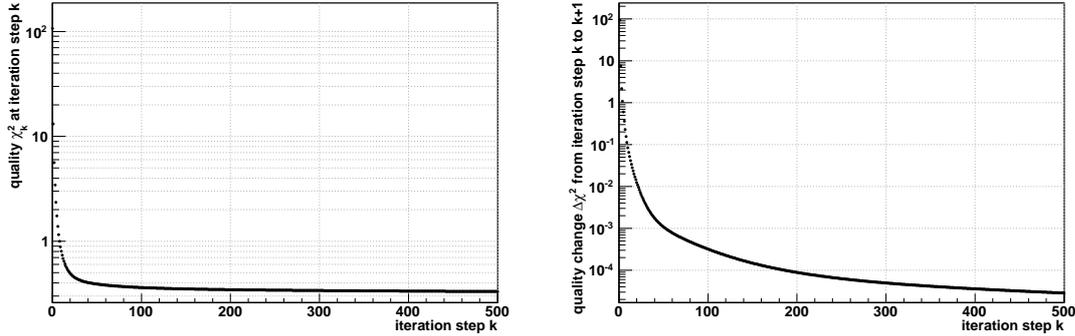


Figure 6.2: The quality  $\chi_k^2$  (given by Eq.(6.39)) of the description of the measured data by the unfolded solution (left panel), as well as the quality improvement  $\Delta\chi^2$  (given by Eq.(6.40)) from iteration step  $k$  to  $k+1$  (right panel), both as a function of the number or iterations. The presented results base on an unfolding of a realistic toy dataset that will be introduced in Section 6.3.2.

$\vec{Y}$  are assumed to be Poissonian ones, and hence are set to  $\sigma(y_i) = \sqrt{y_i}$ . Alternatively, instead of the quality measure  $\chi_k^2$  itself, the degree of its improvement

$$\Delta\chi^2 = \chi_k^2 - \chi_{k+1}^2 \quad (6.40)$$

from iteration step  $k$  to  $k+1$  can be considered. It is more sensitive to changes in the number of iterations, and hence a better replacement for the iteration number, while still possessing a physical meaning.

In Fig. 6.2, both  $\chi_k^2$  and  $\Delta\chi^2$  are illustrated as a function of the iteration number  $k$ . The results base on an unfolding of the realistic toy data sample that will be introduced in Section 6.3.2. It is obvious, that a larger number of iterations will yield a smaller  $\chi_k^2$  value, what corresponds to a better reproduction of the data sample by the estimated solution. While the reduced  $\chi_k^2$  values level out at around 0.3 to 0.4 after some tens of iterations, its improvement  $\Delta\chi^2$  from one step to another decreases significantly. For the toy dataset, which was exemplarily used for this comparison, an optimal solution will be found for  $\Delta\chi^2$  values at around  $3 \times 10^{-4}$ , as will be shown later on (cf. Section 6.3.4). This correlates roughly to 100 iteration steps, while the corresponding  $\chi_k^2$  is round about 0.35. Henceforth, the change  $\Delta\chi^2$  will be used as reference value instead of the number of iterations. The term “iteration depth”, which will be used frequently in the further analysis, will be used synonymously<sup>20</sup> for  $\Delta\chi^2$ .

While, in case of the aforementioned example, the estimated solution already seems to conform with the data quite accurately after round about 100 iteration steps, further recursions, however, would again improve the description. In this context, it should be pointed out, that the defined chi-square value describes the quality of the reproduction of the measured data by the solution, but not the quality of the solution itself. This is an important difference, since, as mentioned in Section 6.1, not an exact solution is desired that reproduces every single fluctuation of the data sample, but, instead, an “appropriate” one. In order to judge about the quality of the solution itself, a better indicator will be defined (cf. Section 6.3.4), which allows to find an iteration depth that yields an optimal

<sup>20</sup>This is reasonable, since  $\Delta\chi^2$  corresponds indeed to a certain number of iterations, and hence to the “depth” of the recursion

primary	$\gamma_1$	$\gamma_2$	$E_k/\text{PeV}$	relative abundance $\alpha$ , ( $\sum = 1$ )
H	-2.62	-4.76	4.47	0.1956
He	-2.62	-4.63	8.94	0.4058
C	-2.62	-4.37	26.82	0.1550
Si	-2.62	-3.84	62.58	0.0812
Fe	-2.62	-3.05	116.22	0.1624

Table 6.1: The parameters of the spectra of individual cosmic ray mass groups measured with KASCADE [9] and computed in [36]. These parameters are used as basis for, to a certain extent realistic, trial spectra.

solution (in this example the iteration depth  $\Delta\chi^2 = 3 \times 10^{-4}$ ). But before that, the focus will be put on the generation of toy datasets that serve for further investigations.

### 6.3.2 Generation of toy datasets

Monte Carlo methods are used in order to test the unfolding procedure, to derive its uncertainties etc. Thereby, arbitrary energy spectra of cosmic rays can be assumed. Based on those, toy datasets can be generated, which can be unfolded by means of the respective unfolding algorithms. As the assumed “true” spectra are known *a priori*, the unfolded solution can be compared to it. By this, the quality of the unfolding procedure can be judged.

As starting point, any spectrum could be assumed. However, since it is accepted widely that the spectrum of cosmic rays follows a power law, in this work only trial spectra are considered that are distributed according to a single or a double power law. For the mathematical description of the differential spectrum  $dJ(E)/dE$  of a certain cosmic ray mass group, a formula, inspired by that used by Hörandel [95] for the polygonato<sup>21</sup> model, is used:

$$\frac{dJ(E)}{dE} = J_0 E^{\gamma_1} \left( 1 + \left( \frac{E}{E_k} \right)^\varepsilon \right)^{(\gamma_2 - \gamma_1)/\varepsilon}, \quad (6.41)$$

where the absolute flux  $J_0$  serves as normalization, while the power law can be a broken one with a change of index from  $\gamma_1$  to  $\gamma_2$  at the transition energy  $E_k$ . The index “k” alludes to the word “knee”, since the structures that are expected in the observed energy range of cosmic rays are called the knees of the cosmic ray spectrum, caused by a steepening of the spectra of individual mass groups. However, vice versa, also spectra that get harder (“ankle”-like structure) or those with no change in index can be regarded. The smoothness of the transition between the two power laws is characterized by  $\varepsilon$ , whereby  $\varepsilon = 1$  corresponds to a smooth change over about one decade of energy, while  $\varepsilon = 4$  means a faster change within already 1/5 of a decade [95].

To get trial spectra that are close to reality<sup>22</sup>, the results of Bindig [36] are taken into account, who fitted the spectra of individual mass groups measured with the KASCADE experiment [9]. The results are summarized in Table 6.1. The parameter of smoothness can be set to  $\varepsilon = 3$ , while its exact value is not important for the tests. The absolute normalization  $J_0$  that belongs to the respective mass group is not important, either,

<sup>21</sup>Greek notation for *many knees*.

<sup>22</sup>The trial spectra should be more or less realistic, while the exact properties are not of importance, since they are used only to test the unfolding algorithms, but are not used in the final analysis of the measured data.

since Eq.(6.41) will serve only as a probability distribution used as basis for a random generator, and will thereby be normalized to 1 anyway.

Lastly, the generation of toy data samples is sketched in Fig. 6.3. The procedure can be divided into three basic steps:

1. Arbitrary energy spectra for the five used mass groups have to be stated, in this example the realistic ones based on a broken power law according to Eq.(6.41) with parameters as summarized in Table 6.1. These initial spectra will be considered as the “true ones” in the following.
2. Since the measurement time of KASCADE-Grande is strongly limited to a couple of years only, even perfectly reconstructed energy spectra would only represent a sample of the true<sup>23</sup> energy spectra. This sample will suffer from statistical fluctuations. Hence, based on the assumed smooth energy spectra, a new set will be generated by means of a random generator<sup>24</sup> that considers the initial power law functions as probability distributions. By this, more realistic energy spectra, which include statistical fluctuations, are attained. Remark: in the example of the realistic test spectra, for a specific primary the random generator will be run  $7.86 \times 10^6 \times \alpha$  times, where  $\alpha$  is the relative abundance of this primary as given in Table 6.1. The factor  $7.86 \times 10^6$  was defined such that the toy dataset that will be generated in step 3 has roughly the same number of entries<sup>25</sup> as the shower size plane actually measured with KASCADE-Grande and used in this analysis. Thus, the generated toy dataset will suffer from comparable statistical uncertainties like the measured one.
3. Now, the toy dataset can be generated. This is again done by means of the random generator. Thereby, the response matrix is considered as probability distribution, since it contains the information with which probability an air shower that was induced by a primary having a certain energy will contribute to a specific cell of the measured two-dimensional shower size plane. The number of runs of the random generator for a specific energy bin and primary particle is given by the number of entries in the associated energy bin of the particle’s trial spectrum generated in step 2.

The artificial toy dataset can now be used for further investigations, and, thereby, e.g. can be unfolded. As the initial “true” energy spectra are known *a priori*, the estimated solutions can be compared to them in order to review the quality of different unfolding techniques. Furthermore, the error propagation through the algorithms can be examined, what will be the topic of the next section.

---

<sup>23</sup>Which can only be determined with unlimited measurement time, under the condition that they are constant over that time and their variations have only a statistical origin.

<sup>24</sup>In all further steps, as random generator the Mersenne and Twister pseudo-random number generator, developed by Matsumoto and Nishimura [135], is used in form of the implementation in the ROOT [45] class `TRandom3`. It is fast and of high quality.

<sup>25</sup>Actually, in the toy data sample there are a little bit less events than in case of the measurement.

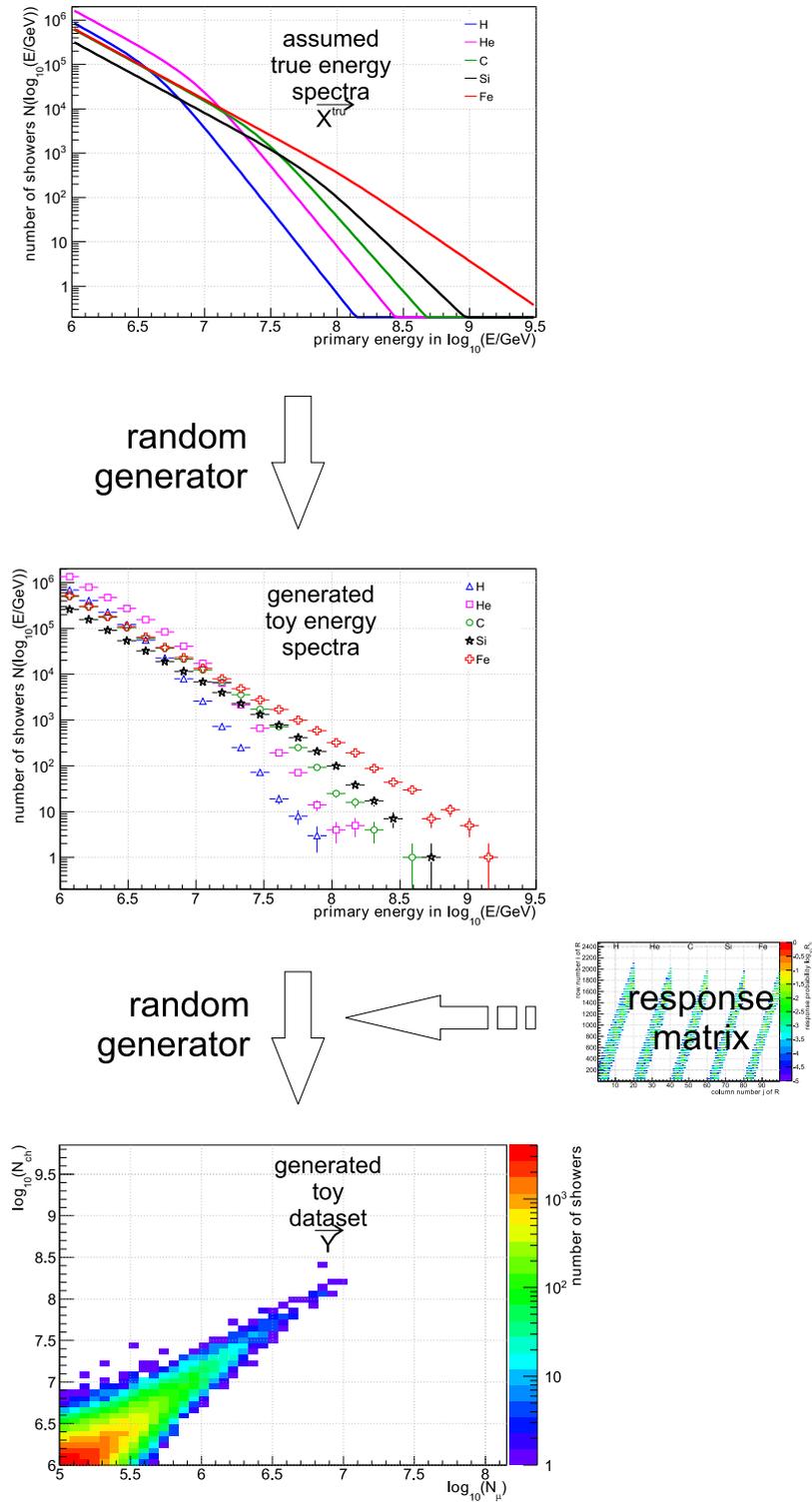


Figure 6.3: Sketch of the generation of toy datasets (see text for further details).

### 6.3.3 Error propagation

In this section, the uncertainties imposed to the unfolded solution are examined. Thereby, all considerations are performed exemplarily for the “realistic” toy spectrum introduced in the previous section. However, everything was repeated for diverse spectra, as well as for other<sup>26</sup> unfolding procedures than that based on Gold’s algorithm. Firstly, an overview over all possible sources of uncertainties is given. Thereafter, appropriate methods to account for these uncertainties are elaborated. Finally, a first comparison of the impacts of different sources of uncertainties on the unfolded solution based on the shower size distribution measured with KASCADE-Grande will be presented.

#### Sources of uncertainties

In this analysis, four different sources of uncertainties have to be considered and will now be discussed:

1. *Statistical uncertainties due to the limited measurement time:* Due to the limited exposure, the measured data sample will suffer from unpreventable statistical uncertainties, which are expected to be Poisson distributed. These uncertainties will be propagated through the applied unfolding algorithm and are usually amplified thereby. In case of the iterative deconvolution techniques, this amplification can be controlled by the number of iterations, or, in case of the regularized methods, by the regularization parameter respectively.
2. *Systematic bias induced by the unfolding method:* based on the convergence properties of the iterative algorithms, small numbers of iteration steps will for one thing reduce the amplification of the statistical uncertainties of the data sample, for another thing will result in a solution that is deviating from the exact one. In case of the regularized techniques it is similar, since the regularization damps oscillations, but, conversely, results in a biased solution.
3. *Systematic uncertainties due to the limited Monte Carlo statistics:* Due to limited computing time, only Monte Carlo simulation sets with limited statistics can be generated. The uncertainties in the simulated distributions used in Sections 5.1 and 5.2 to parametrize the air shower development and the reconstruction properties result in an uncertainty in the computed response matrix. Hence, the response matrix can be systematically wrong. Furthermore, the conditioning (cf. Section 5.3.2) that was applied to the matrix could have systematic impacts, which are, however, small enough to be neglected. Since the response matrix is used in the deconvolution procedure, all this induces finally a systematic uncertainty in the estimated solution.
4. *Systematic uncertainties due to the systematic uncertainty in the Monte Carlo simulations:* The Monte Carlo simulations used to compute the response matrix base on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68]. As indicated by the word “model”, they are theoretical constructs, even though experimental experiences are considered. It is not guaranteed that they describe the truth sufficiently well. d’Enterria et al. [59] compared the first Large Hadron Collider (LHC) data with the predictions of various Monte Carlo models, including e.g.

---

<sup>26</sup>Since everything would be analogue to the way of proceeding pursued for Gold’s algorithm, the other methods are not treated separately.

the models QGSJET 01, QGSJET-II, SIBYLL 2.1, and EPOS 1.99. They stated that none of the investigated models can describe consistently all characteristics of the LHC measurements. Nevertheless, whilst the final analysis of this work, it will be shown that the used model QGSJET-II-02 yields results, that agree with the data measured with KASCADE-Grande, and hence it can be expected that the result is not quite far off the truth. Coming back to the response matrix, a possible deficient description of the contributing physical processes would result in systematic errors in the response matrix, and hence to a systematic wrong deconvolution result. However, this is a common problem of all analyses that need to utilize simulations. This kind of uncertainties cannot be computed within this analysis. Nevertheless, as mentioned just afore, the quality of the description of the measured data by the solution can be investigated to have a rough impression whether the used interaction models are rather wrong, or, possibly, proper candidates. Hence, all results in this work can only be given under the assumption that the applied models QGSJET-II-02 and FLUKA 2002.4 are the right ones, such that the additional model uncertainty can be neglected in the following. From both, the low energy interaction model, here FLUKA, has less influence on the final result, as already the analyses based on the KASCADE measurements have proved [12]. An impression about the model dependence of the solution can only be obtained by repeating the whole analysis using different high energy interaction models. Whilst the main analysis of this work, a first but roughly done cross-check based on the high energy interaction model EPOS 1.99 [188] will be performed (Section 7.2.3). Precise analyses based on that or alternative interaction models are carried out actually by our collaboration and will be presented in the near future, but cannot be shown in this work right now.

In the following, adequate measures are adopted in order to account for the mentioned sources of uncertainties and to consider them in the final results.

### “True” statistical uncertainty and “true” bias

The “true” statistical uncertainty, caused by the limited exposure, as well as the “true” bias, caused by the unfolding algorithm itself, will be computed now. The determination of these “true”<sup>27</sup> uncertainties is only possible in case of trial energy spectra and toy datasets, since there the “true” solution is known and can be compared to the estimated one. In case of the measurement, however, one can only try to estimate the uncertainties, what will be shown thereafter. The principle of the computation of the “true” uncertainties is sketched in Fig. 6.4 and can be divided into several steps:

1. The first step is the generation of a toy dataset, as already explained in Section 6.3.2. However, based on the assumed “true” energy spectra  $\vec{X}^{\text{tru}}$ , instead of one  $N$  toy energy spectra are generated<sup>28</sup>. Again, based on these spectra, for each set  $i$  a toy dataset  $\vec{Y}_i$  will be generated.
2. Every toy dataset  $\vec{Y}_i$  is unfolded based on Gold’s algorithm with the same defined iteration depth, which yields the estimated solutions  $\vec{X}_i^{\text{est}}$  belonging to the respective set  $i$  and the chosen iteration depth.

<sup>27</sup>Strictly speaking, these are not the really true uncertainties and hence are written with quotation marks, since the true uncertainties can only be derived with infinite simulation statistics. However, the computed “true” uncertainties are reliable estimates.

<sup>28</sup>In this work,  $N = 50$  sets are used.



The “true” statistical uncertainties are given by the fluctuations of the  $N$  estimated solutions  $\vec{X}_i^{\text{est}}$ . More precisely, the variance of the distribution of the estimated solutions  $x_{k,i}^{\text{est}}$  around their mean value  $\overline{x_k^{\text{est}}}$  for an individual energy bin  $k$  will be used as measure for the absolute statistical uncertainty  $\sigma_k$  of the flux at that energy bin:

$$\sigma_k^2 = \frac{1}{N} \sum_{i=1}^N \left( x_{k,i}^{\text{est}} - \overline{x_k^{\text{est}}} \right)^2, \quad \text{with} \quad \overline{x_k^{\text{est}}} = \frac{1}{N} \sum_{i=1}^N x_{k,i}^{\text{est}}. \quad (6.42)$$

The absolute bias  $b_k$  of the flux at an individual energy bin  $k$  is given by the mean deviation of the estimated solutions  $x_{k,i}^{\text{est}}$  from the true one  $x_k^{\text{tru}}$  for that specific energy bin  $k$ :

$$b_k = \frac{1}{N} \sum_{i=1}^N \left( x_{k,i}^{\text{est}} - x_k^{\text{tru}} \right). \quad (6.43)$$

Whilst the aforementioned computations based on toy datasets, the whole deconvolution procedure is considered as a kind of “black box”. Thereby, the exact knowledge about the propagation process of the errors is not necessary. However, as already emphasized, in case of the KASCADE-Grande measurements the “true” energy spectra are not known, such that the unfolded ones cannot be compared to them. For this purpose, the error computation has to be modified, what will be shown hereinafter. Firstly, the estimation of the statistical uncertainty will be explained, and, thereafter, the estimation of the bias.

### Estimation of the statistical uncertainty

Since in case of the measurement the “true” energy spectra are not known, the methods explained afore to compute the “true” statistical uncertainties are not immediately applicable. However, by means of a frequentist approach, the uncertainties can be estimated. The basic idea is sketched in Fig. 6.5, and can be divided into two steps:

1. Based on the measured data sample  $\vec{Y}$ ,  $N$  toy datasets  $\vec{Y}_i$  are generated. This is done by means of a random generator, that regards the measured sample  $\vec{Y}$  as probability distribution. The generated sets will differ from the measured one in the context of Poisson statistics only.
2. The generated toy datasets  $\vec{Y}_i$  are now unfolded, what yields the estimated solutions  $\vec{X}_i^{\text{est}}$ .

Analogue to the “true” absolute statistical uncertainty, the estimated one  $\sigma_k^{\text{est}}$  can now be determined by the variance of the distribution of the estimated solutions  $x_{k,i}^{\text{est}}$  around their mean value  $\overline{x_k^{\text{est}}}$  for an individual energy bin  $k$ :

$$(\sigma_k^{\text{est}})^2 = \frac{1}{N} \sum_{i=1}^N \left( x_{k,i}^{\text{est}} - \overline{x_k^{\text{est}}} \right)^2, \quad \text{with} \quad \overline{x_k^{\text{est}}} = \frac{1}{N} \sum_{i=1}^N x_{k,i}^{\text{est}}. \quad (6.44)$$

In Fig. 6.6, a comparison between the “true” (computed using Eq.(6.42)) and the estimated (computed using Eq.(6.44)) statistical uncertainties, based on the realistic toy dataset introduced in Section 6.3.2, is shown. The data samples have been unfolded with either a smaller iteration depth  $\Delta\chi^2 = 3 \times 10^{-3}$  (top panels), or with a significantly larger number of iteration steps corresponding to a  $\Delta\chi^2 = 1 \times 10^{-8}$  (bottom panels), of

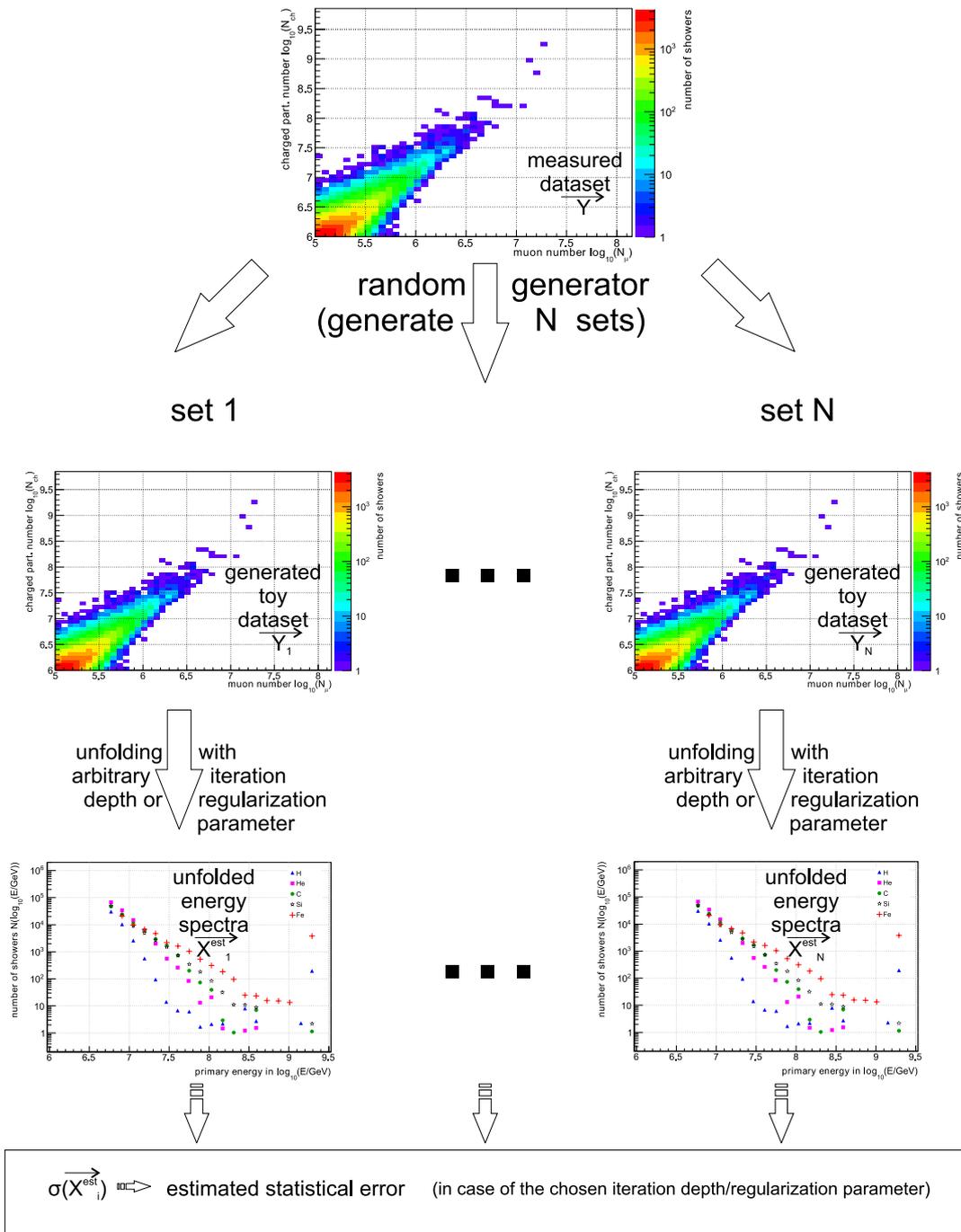


Figure 6.5: Sketch of the computation of the estimated statistical uncertainties (see text for further details).

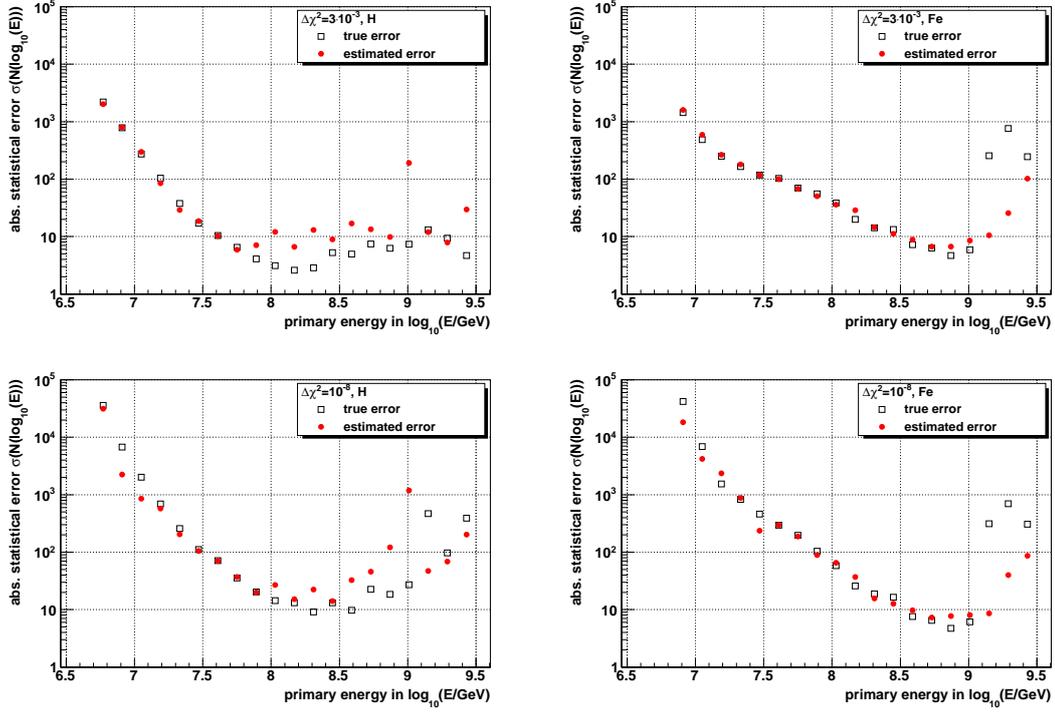


Figure 6.6: Comparison between the “true” (Eq.(6.42)) and the estimated (Eq.(6.44)) absolute statistical uncertainties of the flux based on the realistic toy dataset introduced in Section 6.3.2, which was unfolded with either a smaller iteration depth  $\Delta\chi^2 = 3 \times 10^{-3}$  (top panels), or with a significantly larger one  $\Delta\chi^2 = 1 \times 10^{-8}$  (bottom panels). This is exemplarily shown in case of primary protons (left) or iron nuclei (right).

which the latter iteration yields a solution closer to the exact one and by this results in larger statistical fluctuations. This is exemplarily shown in case of primary protons (left) or iron nuclei (right). Overall, there is a good agreement; hence, the estimation of the statistical uncertainties performs reliable. If there are deviations, they are mostly based on an overestimation of the statistical uncertainties, what will result in slightly too large error bars of the flux at that energy bins. Nevertheless, this appears more acceptable than a presentation with error bars that are too small. However, most of the significant deviations are in energy ranges for that the spectrum of the specific primary will have one-digit entry numbers only, such that both the estimated and the “true” statistical uncertainties are large enough to tag this range as precarious in equal measure. The exact uncertainty values are less important there. Above energies of  $\log_{10}(E/\text{GeV}) \approx 8.5$ , *de facto* no energy spectra will be shown in the final result of this work<sup>29</sup>, such that this range is less important anyhow.

<sup>29</sup>Due to the huge overall uncertainties there.

### Estimation of the bias

In case of the measurement, the bias can be estimated only, since, naturally, the “true” spectra are not known such that the unfolded estimates cannot be compared to them. The estimation of the bias is again performed based on a frequentist approach. The basic idea bases on the principle of the *bootstrap methods* and is illustrated in Fig. 6.7. The procedure can be divided into the following steps:

1. The “true” bias was computed by a comparison between some estimated solutions and the initially assumed true elemental energy spectra. Since in case of the measurement the true energy spectra are sought-after, and hence not known, the situation is more difficult. An appropriate compromise that was found to solve this problem is to guess the “true” solution by means of an unfolding of the measured data sample  $\vec{Y}$  for a specific iteration depth. For purpose of the estimation of the bias it is not necessary that the guessed solution is right. This will only be important in the final analysis, and a method to find the best guess will be introduced in Section 6.3.4. Therefore, currently, the properness of the guessed solution is not of interest, and hence the used iteration depth can be arbitrary without loss of generality. The unfolded solution estimate will be considered, for lack of a better knowledge, to be the true one in the following steps. However, it will be named  $\vec{X}^{\text{temp}}$  to point out, that it is rather a template for the further computations and not really the definitely true solution.
2. Based on the solution template  $\vec{X}^{\text{temp}}$ ,  $N$  toy datasets  $\vec{Y}_i$  can be generated by means of a random generator and based on the response matrix as probability distribution. Hence, the following steps will be equivalent to that performed whilst the computation of the “true” bias.
3. All toy datasets  $\vec{Y}_i$  are unfolded, what yields the estimated solutions  $\vec{X}_i^{\text{est}}$ . Thereby, for a correct estimation of the bias, the same iteration depth has to be used as whilst the computation of the template solution in step 1.

Analogue to the “true” absolute bias, the estimated one  $b_k^{\text{est}}$  at an energy bin  $k$  is given by the mean deviation of the estimated solutions  $x_{k,i}^{\text{est}}$  from the template  $x_k^{\text{temp}}$ :

$$b_k^{\text{est}} = \frac{1}{N} \sum_{i=1}^N \left( x_{k,i}^{\text{est}} - x_k^{\text{temp}} \right) . \quad (6.45)$$

The basic idea behind this technique is that the bias is only an intrinsic one induced by the unfolding algorithm itself. Hence, if the algorithm is biased<sup>30</sup>, whilst the unfolding of the toy datasets that were generated based on the template a certain bias will be introduced. This bias is a specific one for the unfolding of that data sample<sup>31</sup> under the chosen iteration depth. In the final analysis, the task will be to find a template that is close to the true<sup>32</sup> solution, and then to estimate the associated bias by the outlined technique.

In Fig. 6.8, a comparison between the “true” (Eq.(6.43)) and the estimated (Eq.(6.45)) absolute bias, based on the realistic toy dataset introduced in Section 6.3.2, is shown<sup>33</sup>.

<sup>30</sup>What is the case for the used ones, as explained in Section 6.2.

<sup>31</sup>Or comparable datasets that are compatible within statistics.

<sup>32</sup>In Section 6.3.4, a method will be introduced that allows to judge whether or not the template is close to the truth.

<sup>33</sup>Since the bias varies over some orders of magnitudes in the observed energy range, a logarithmic scale was used such that only the absolute values  $|b^{\text{est}}|$  and  $|b|$  can be shown.

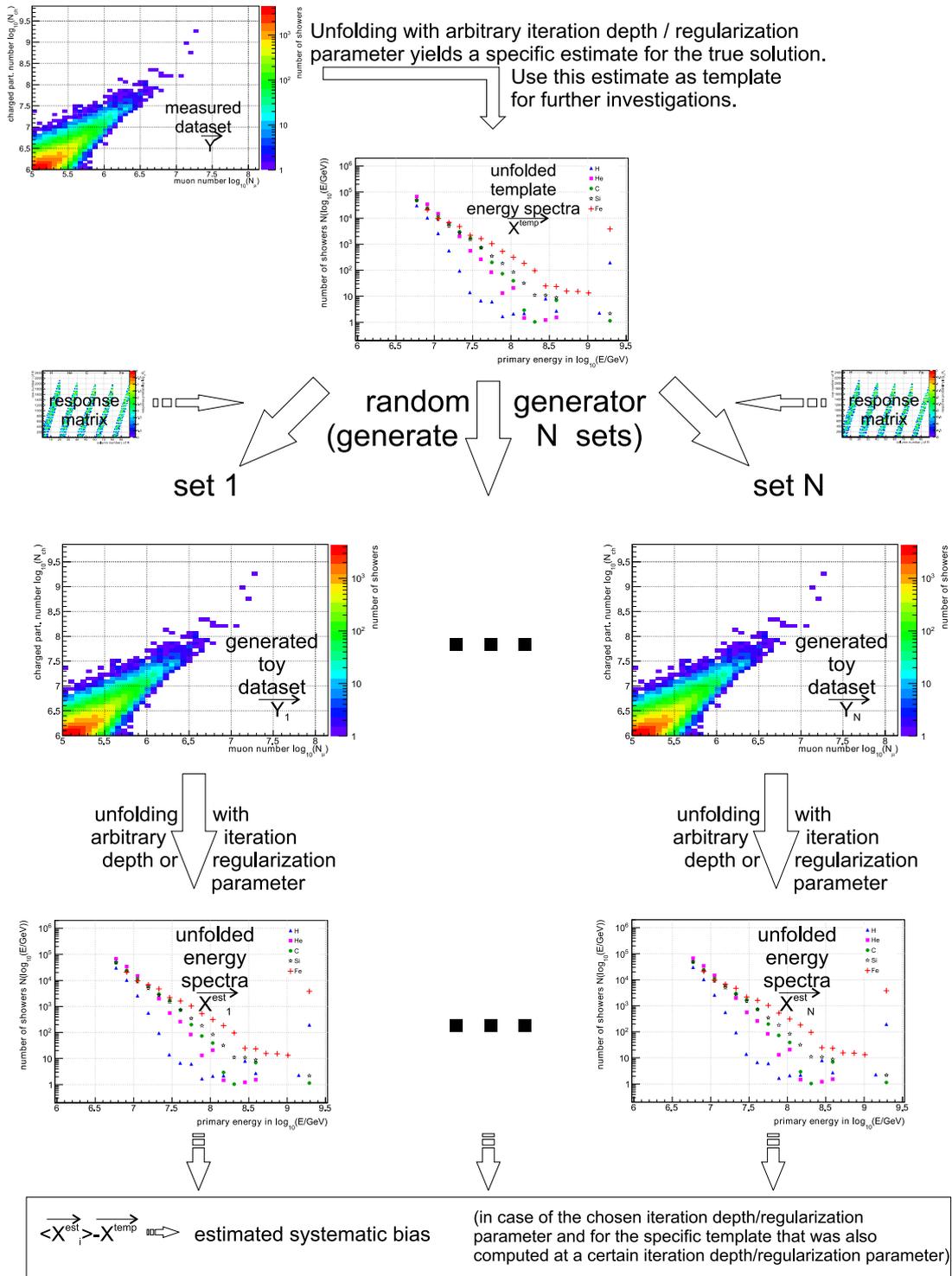


Figure 6.7: Sketch of the computation of the estimated bias (see text for further details).

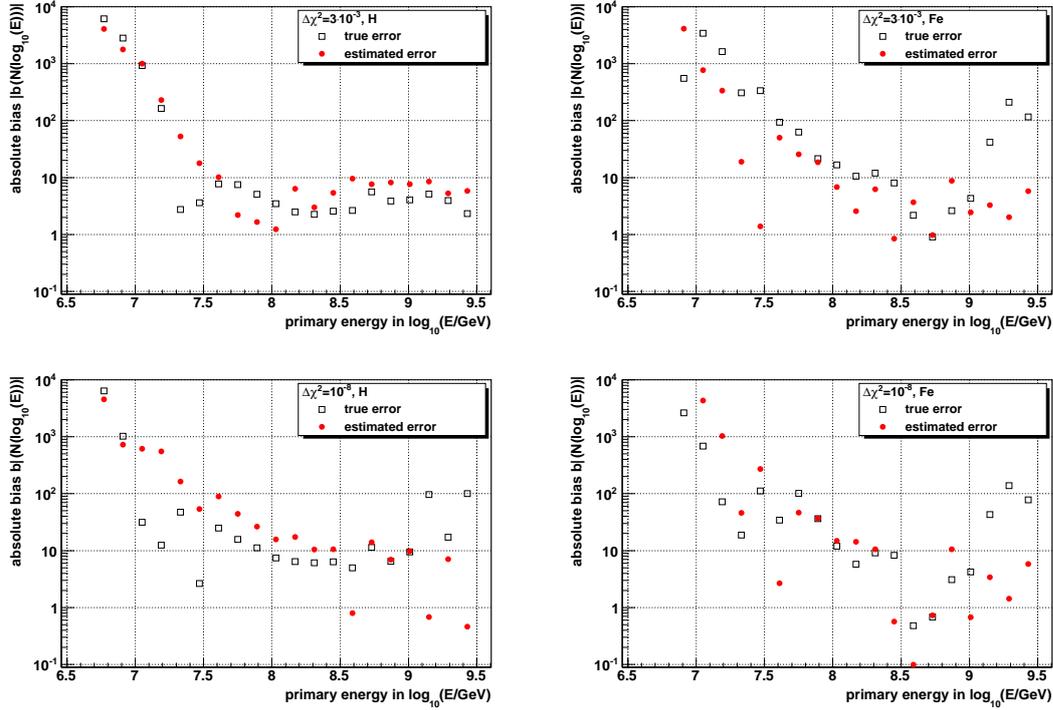


Figure 6.8: Comparison between the “true” (Eq.(6.43)) and the estimated (Eq.(6.45)) absolute bias of the flux based on the realistic toy dataset introduced in Section 6.3.2, which was unfolded with either a smaller iteration depth  $\Delta\chi^2 = 3 \times 10^{-3}$  (top panels), or with a significantly larger one  $\Delta\chi^2 = 1 \times 10^{-8}$  (bottom panels). This is exemplarily shown in case of primary proton (left) or iron nuclei (right).

Both the template solution  $\vec{X}^{\text{temp}}$  as well as the estimated solutions have been derived by an unfolding with the same iteration depth: either both with  $\Delta\chi^2 = 3 \times 10^{-3}$  (top panels) or  $\Delta\chi^2 = 1 \times 10^{-8}$  (bottom panels). This is exemplarily shown in case of primary protons (left) or iron nuclei (right). Tendentially, the estimated and the “true” bias agree well. The orders of magnitude as well as the course are roughly the same. However, looking at individual energy bins, in several cases one can observe larger deviations between the “true” and the estimated bias. In a few cases, also the signs of the values are different<sup>34</sup>, what happens seldom however. While the estimation of the statistical uncertainties performs as well that it can be done for each energy bin separately, this will not be possible for the bias. Instead, a conservative approach will be pursued: the limits of the domain of the estimated bias will be determined under the demand that the “true” bias will be covered and, as possible, that the bias will not be underestimated. These limits will be used as measure for the systematic bias of the used unfolding algorithm.

For this purpose, it is more useful to regard the relative bias instead of the absolute one. In Fig. 6.9, the relative bias, i.e. the relation between the number of entries in the estimated solution vector for a certain energy to that in the true solution, is depicted as a function of the number of entries in the estimated solution vector, i.e. as a function of the

<sup>34</sup>Not visible here, since only the absolute values are shown. Confer Fig. 6.9 to get a better impression.

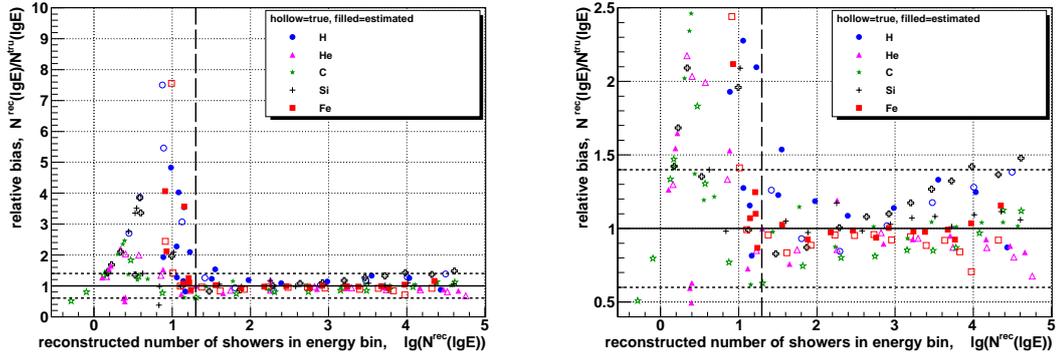


Figure 6.9: Comparison between the “true” and the estimated relative bias for different primaries and based on the realistic toy dataset introduced in Section 6.3.2, which was unfolded with the optimal iteration depth of  $\Delta\chi^2 = 3 \times 10^{-4}$  (cf. Section 6.3.4). On the right panel, the range of the axis of ordinates is scaled-down for reasons of a better distinguishability.

number of showers<sup>35</sup>  $\log_{10}(N^{\text{rec}})$ . The result bases on the realistic toy dataset introduced in Section 6.3.2, which was unfolded with the optimal iteration depth of  $\Delta\chi^2 = 3 \times 10^{-4}$  (cf. Section 6.3.4). Both the “true” and the estimated relative bias are shown.

Above values of  $\log_{10}(N^{\text{rec}}) = 1.3$  (indicated by the vertical dashed line), what corresponds to around  $N = 20$  air shower events and hence to 20 entries in the related energy spectrum, the relative systematic bias can be bracketed overall as done in the figure by the two dashed horizontal lines. On the right panel, the axis of ordinates is scaled-down for a better distinguishability. Thereby, only the estimated relative biases are taken into account, since the “true” ones would not be available in case of the real measurement. Tendentially, the “true” and the estimated biases conform to a certain extent. However, it will be more reliable to consider, conservatively, the value derived by the bracketing as common bias for all primaries. By this, for this example a common relative bias of 40% can be stated. This bracketing can also be done in case of the measured data sample, since it covers not only the estimated relative bias, but also the “true” one in almost all cases, as extensive tests based on diverse trial spectra have shown. Only in a few examples, the bias was underestimated slightly (maximal 5%), but only at one or two shower numbers  $\log_{10}(N^{\text{rec}})$ , what is still an acceptable result and does not give any indication that the applied bias estimation will fail in case of real data.

For smaller numbers of entries in the solution vector, in this example in case of less than 20 air shower events, the relative bias gets huge. This is mainly caused by the positive definiteness of Gold’s algorithm, that results in non-zero flux entries, often quite far away from zero, even when the true flux is zero. This results in overestimations of the flux, and hence in the discernible large relative bias. Furthermore, the estimated bias agrees only slightly with the “true” one, while both are subject to bin-by-bin fluctuations. Some tests have shown that it is more appropriate to assume a fixed absolute bias, instead of a relative one, below that problematic threshold. In this example, based on the estimated relative bias, an absolute one of  $b = 40$  was defined, if  $\log_{10}(N^{\text{rec}}) < 1.3$ . With decreasing

<sup>35</sup>The relative bias in the number of showers is shown in dependence on the number of showers itself instead on the energy. This simplifies the reversion to an absolute bias that will be assigned to the fluxes in terms of error bands in the final depiction.

number of events, the relative bias will get larger in case of this assumption. For example, at  $\log_{10}(N^{\text{rec}}) = 1$ , what corresponds to  $N = 10$  events,  $b = 40$  would mean that there could also be 50 events, what is equivalent to a relative bias of 400%. This would already tag the corresponding ranges of the spectrum as not trustable. Hence, the knowledge of the exact value of the bias is less important in those cases.

Extensive tests with diverse trial spectra have shown, that the common relative bias estimated based on the bracketing, in combination with the defined fixed absolute bias below a certain number of events, yields reliable results in all observed cases. Hence, this method yields an acceptable estimate for the relative bias over the whole observed energy range, and will finally be used in case of measured data, too. Again, it should be emphasized that both the estimation of the statistical uncertainty as well as of the bias can be applied in the same manner for all used unfolding techniques, and hence are not restricted to Gold's algorithm.

### Systematic uncertainties in the response matrix

One key part of the unfolding analysis is the response matrix, which contains the probabilities that specific charged particle and muon number combinations are measured by KASCADE-Grande, on condition that the air shower, in which the particles were produced, was induced initially by a certain cosmic ray primary with a specific energy. The probability densities that contribute to the response matrix have been determined based on Monte Carlo simulations (cf. Section 5.1 and 5.2). Due to the limited Monte Carlo statistics, the response matrix computation will suffer from uncertainties. Furthermore, the response matrix was optimized in order to avoid a strong ill-conditioning (cf. Section 5.3.2). All this will affect the solution gained by an unfolding.

The last-mentioned source of uncertainties was intensively analysed by means of diverse trial spectra. It was found that the applied conditioning is as weak that it cannot introduce a significant systematic bias to the unfolded solution. One has to keep in mind, that the used cuts on the probability values of the response matrix are covering only charged particle and muon number combinations that are already very unlikely. Hence, and since the measured dataset has strongly limited statistics (at around 78 000 used events only), those combinations will less likely occur. However, in case of some problematic trial spectra, the conditioning could help to damp huge unphysical oscillations in the solution, while in the unproblematic cases that can be unfolded without these statistical distortions the conditioning will not have any effect, neither a positive nor a negative one. In all observed cases where the conditioned response matrix was used, the statistical uncertainties of the solution could be estimated reliably. Since the conditioning affects mainly the propagation of the statistical uncertainties of the limited data sample through the unfolding algorithm, and hence influences the statistical uncertainties<sup>36</sup> of the solution while the systematic ones remain unaffected, the conditioning does not have to be considered as possible systematic error source further.

Now, the systematic uncertainty in the solution caused by the limited Monte Carlo statistics will be investigated. The methods introduced so far and sketched in Fig. 6.7 allow, regarding the systematic uncertainties, only to investigate the intrinsic bias caused by the unfolding method itself, but not the effects of systematic uncertainties in the response matrix. The latter ones would already affect the generation of toy data samples, such that in the following unfolding procedure this influence would be compensated to a

<sup>36</sup>That are already considered by the afore-explained estimation of the statistical uncertainties.

certain degree, since the same wrong response matrix would be used again. Consequently, the systematic uncertainties will instead be analysed by another conservative approach.

But, firstly, we shall investigate which of the contributors<sup>37</sup> to the response matrix, i.e. the intrinsic shower fluctuations or the reconstruction resolution of KASCADE-Grande, have to be considered in this context. If e.g. the resolution is as bad that the reconstructed shower sizes have relative deviations from the true values in the order of some hundred per cent, a preceding variation within e.g. some ten per cent caused by the intrinsic shower fluctuations will get lost in the following strong smearing by the reconstruction uncertainties. Hence, in the context of this example, the uncertainty in the parametrization of the shower fluctuations would have less influence on the solution. In Fig. 6.10, a comparison between the intrinsic shower fluctuations and the reconstruction resolution of KASCADE-Grande is accomplished (at the top panel for the charged particle number, at the bottom one for the muon number). The shower fluctuations as a function of the respective mean true shower size are obtained, for each set of mono-energetic shower simulations separately, by plotting the RMS<sup>38</sup> of the distributions of charged particles respectively muons as a function of their mean values<sup>39</sup>. The depiction of the resolution of the charged particle respectively muon reconstruction bases on the RMS<sup>40</sup> of the deviations between the true and the reconstructed<sup>41</sup> shower sizes. The abscissa, i.e. the mean true value<sup>42</sup> of the respective shower size, is the same as in case of the shower fluctuations. Whilst these comparisons, the actually asymmetric distributions are considered to be symmetric ones. However, this simplification will affect the shown results only slightly.

In case of the charged particles, the intrinsic fluctuations whilst the development of the air showers that are induced by protons are by far the largest uncertainties. In general, the intrinsic shower fluctuations of heavier primaries with atomic mass  $A$  are smaller, since they can be considered as a superposition of  $A$  hydrogen nuclei, and by this the fluctuations roughly<sup>43</sup> decrease by  $\sqrt{A}$ , since the primary energy is split in  $A$  independent cascades (cf. Section 2.4). This is confirmed by the shown result, where the intrinsic fluctuations of iron induced showers are significantly smaller than that of proton induced ones. While in case of protons the uncertainty caused by the reconstruction resolution is considerably smaller than that induced by the shower fluctuations, in case of iron nuclei both sources are with less diverse impacts. Above the threshold of full efficiency at around  $\log_{10}(N_{\text{ch}}^{\text{tru}}) \approx 6.0$ , the shower fluctuations and the experiment's resolution in case of iron nuclei as well as the resolution in case of protons are of comparable magnitude. Hence, especially in case of heavier primaries, the resolution is a crucial source of uncertainties. Consequently, in case of the charged particle number the resolution will be taken into account additionally to the intrinsic shower fluctuations in the following considerations.

The situation is similar in case of the muon number. Again, the intrinsic shower fluctuations are smaller for heavier primaries. Above the threshold of full efficiency at around

<sup>37</sup>The efficiency is contributing, too, but with less influence since only charged particle and muon number combinations that belong to 100% efficiency are considered in the final analysis. The small uncertainty in the parametrization of the efficiencies is negligible in comparison to the other discussed sources.

<sup>38</sup>See exemplarily Fig. 5.6 respectively 5.12, where the RMS is given in the statistics box.

<sup>39</sup>The mean is also given in the statistics box of the mentioned example Fig. 5.6 respectively 5.12.

<sup>40</sup>Can be seen exemplarily in the statistics boxes of Fig. 5.25 respectively 5.29.

<sup>41</sup>That are already corrected for a bias by  $C$  that is given by Eq.(5.15) respectively (5.16).

<sup>42</sup>The mean is given by the middle of the actually chosen interval of the true sizes, that is e.g. given in the caption of the mentioned example Fig. 5.25 respectively 5.29.

<sup>43</sup>In more precise computations, the intrinsic shower fluctuations for heavy primaries are decreasing a little bit less than  $\sqrt{A}$ ; however, they are still significantly smaller than that in case of hydrogen nuclei.

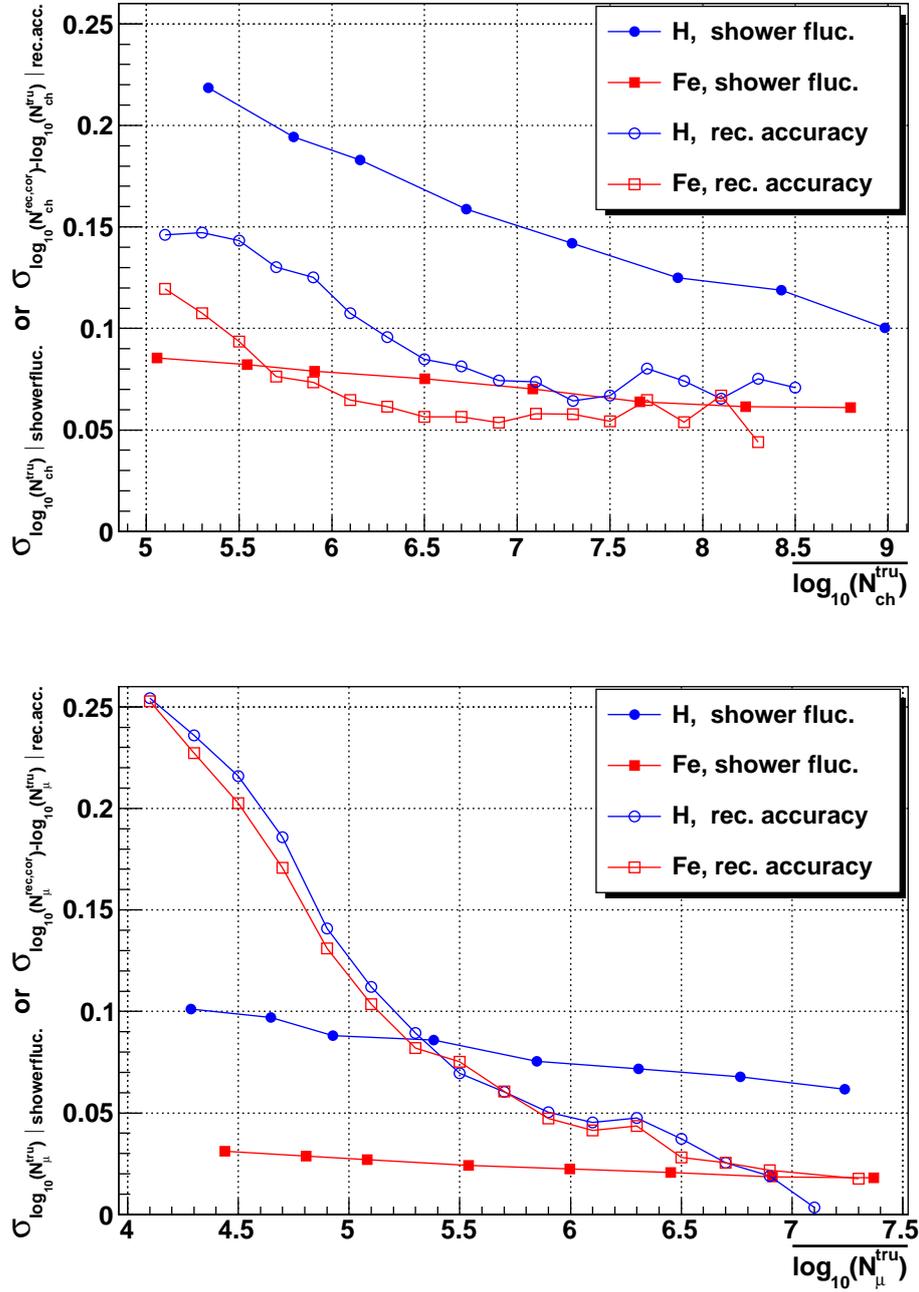


Figure 6.10: Comparison between the intrinsic shower fluctuations and the reconstruction resolution of KASCADE-Grande, on the top panel for the charged particles, on the bottom one for muons (exemplarily for proton and iron primaries). To guide the eye, the markers are connected with straight lines.

$\log_{10}(N_{\mu}^{\text{tru}}) \approx 5.0$ , the fluctuations in the air showers induced by protons are dominant. However, below that limit, the muon reconstruction causes the largest uncertainties. Nevertheless, the resolution gets rapidly better in case of full efficiency. Hence, in case of the muon number all four sources will be considered, too.

As already mentioned, the distributions used in Section 5.1 and 5.2 to parametrize the contributors to the response matrix suffer from statistical uncertainties, since they base on Monte Carlo simulations with limited statistics. Due to this, the derived parametrizations will suffer from these uncertainties, too. The impacts on the unfolded solution are investigated by the following procedure:

1. Whilst fitting the distributions in Section 5.1 and 5.2, the fit routine takes into account also their statistical uncertainties. Hence, to all fit parameters appropriate uncertainties are given, which reflect the uncertainty caused by the limited Monte Carlo statistics. In a very conservative approach, all fit parameters are assumed to be uncorrelated and to follow a Gaussian probability density function. The fit parameter is treated as the mean value of a Gaussian function, while the parameter's error is regarded as the standard deviation. For every parameter an individual Gaussian function is defined.
2. By regarding the Gaussian functions that have been defined in step 1 as probability distribution for the respective parameters, and using a random generator, for every parameter a new one can be computed, which conforms with the initial one in terms of the statistical uncertainties.
3. As done in Section 5.1 and 5.2, the energy dependence of the parameters is interpolated based on the new parameter set.
4. All afore-listed steps are repeated for each of the error sources, i.e. for the parametrization of the intrinsic shower fluctuations as well as of the reconstruction accuracy (both in case of the muon and charged particle number).
5. Now, a new set of parametrizations for all four types of distributions is available, based on which a new response matrix is computed. The new response matrix will differ from the usual one only within statistical uncertainties.

The whole chain is repeated 50 times in order to get a sufficiently large set of response matrices, which all are only differing within statistical uncertainties. The data sample can now be unfolded with each of the 50 response matrices, resulting in 50 solutions. The bin-wise variance of all solutions can be interpreted as the systematic uncertainty arising from the uncertainty in the response matrix, i.e., more precisely, from the limited Monte Carlo statistics.

It was found that the aforementioned technique varies the parametrizations only slightly. Particularly, in the middle part of the distributions the Monte Carlo statistics is rather good, such that the fit is already confined to a certain extent and will reveal only small parameter uncertainties. Furthermore, the tails of the distributions will be affected only marginally by these modifications. From the mathematical point of view, these tails are absolutely unimportant for the fit procedure, as their content is negligible small in comparison to that of the middle part of the distributions. The tails of the parametrizations can be modified without changing the reduced chi-square value of the fit. However, from the physical perspective they are of high importance. In particular, the right tail that describes the fluctuations in direction to higher energies can have an

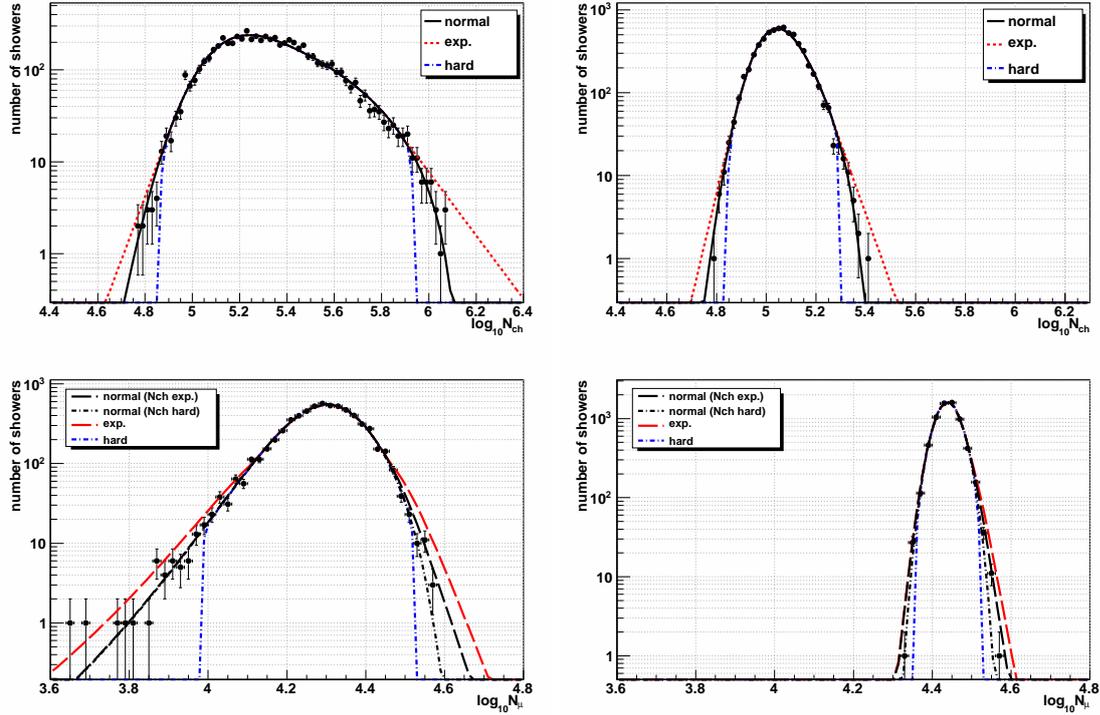


Figure 6.11: Different parametrizations for the intrinsic shower fluctuations of charged particles (top) and muons (bottom). Exemplarily, the distributions for proton (left) or iron (right) induced showers are shown, both for a primary energy of 2 PeV.

important impact on the unfolded solution due to the steeply falling flux of cosmic rays. These bin-to-bin migration effects can affect the unfolded solution significantly. That is why the influence of the tails will now be investigated additionally.

In order to examine the impacts of the uncertainties in the distributions' tails on the final solution, a conservative approach is pursued. The tails of all four different types of distributions that have been used to parametrize the intrinsic shower fluctuations and the experiment's resolution, for the number of charged particles as well as of muons, are parametrized additionally based on different approaches. Thereby, all used fit functions describe the distributions equally well, i.e. all modifications yields parametrizations that are compatible within the statistical uncertainties.

In Fig 6.11, different parametrizations for the intrinsic shower fluctuations of charged particles (top) and muons (bottom) are depicted. Exemplarily, the distributions for proton (left) or iron (right) induced showers are shown, both for a primary energy of 2 PeV.

On the top panels, the black lines (“normal”) represent the standard parametrization that were already shown in Fig. 5.6. Taking into account the large statistical uncertainties of the tails, the standard parametrizations can be bracketed by two extreme cases: for one thing by a faster decreasing fit function (“hard” cutoff), for another thing by a less steeper one (“exp.”, i.e. exponential decrease). Both additional parametrizations are chosen on the one hand to cover almost all possible scenarios, but on the other hand such that the resulting fit quality does not get worse. All three approaches are consistent in the context of the statistical uncertainties in the Monte Carlo simulations. For all three

cases, the parametrization bases mainly on that given by Eq.(5.3). Only the tails have been modified. The hard cutoff was realized by assuming zero probability if the number of showers falls below 1/20 of the number at the distribution's maximum. Thereby, Eq.(5.3) will be replaced by:

$$p_{\text{hard}} \left( \log_{10} N_{\text{ch}}^{\text{true}} \mid \log_{10} E \right) = \begin{cases} 0 & \text{if } \log_{10} N_{\text{ch}}^{\text{true}} \leq p_6 , \\ \text{Eq.(5.3) unchanged} & \text{if } p_6 < \log_{10} N_{\text{ch}}^{\text{true}} < p_7 , \\ 0 & \text{if } \log_{10} N_{\text{ch}}^{\text{true}} \geq p_7 , \end{cases} \quad (6.46)$$

where  $p_6$  and  $p_7$  are the charged particle numbers at that the distribution falls below 1/20 of its maximum. In case of the exponential decrease approach, the continuous differential transition to the exponential function starts when the distribution falls below 1/10 of its maximum. More precisely, Eq.(5.3) will be replaced by:

$$p_{\text{exp.}} \left( \log_{10} N_{\text{ch}}^{\text{true}} \mid \log_{10} E \right) = \begin{cases} p_{10} \times \exp \left( (p_6 - \log_{10} N_{\text{ch}}^{\text{true}}) p_8 \right) & \text{if } \log_{10} N_{\text{ch}}^{\text{true}} \leq p_6 , \\ \text{Eq.(5.3) unchanged} & \text{if } p_6 < \log_{10} N_{\text{ch}}^{\text{true}} < p_7 , \\ p_{10} \times \exp \left( (p_7 - \log_{10} N_{\text{ch}}^{\text{true}}) p_9 \right) & \text{if } \log_{10} N_{\text{ch}}^{\text{true}} \geq p_7 . \end{cases} \quad (6.47)$$

Thereby, the value  $p_{10}$  is 1/10 of the distribution's maximum and ensures continuity, while  $p_6$  and  $p_7$  correspond to the charged particle numbers where the distribution falls below that defined threshold value, and hence mark the transition region. The parameters  $p_8$  and  $p_9$  are the derivatives<sup>44</sup> of the original function at the transition points, and hence ensure the differentiability there.

On the bottom panels, the two black lines given in each plot correspond to the standard parametrization of the muon number distribution already shown in Fig. 5.12. Since the charged particle and the muon numbers are correlated sizes, this was considered in Section 5.1 whilst the determination of the standard parametrizations, which finally include this correlation. Hence, using the modified description of the charged particle shower fluctuations, also the correlated  $\log_{10} N_{\text{ch}} - \log_{10} N_{\mu}$  distribution would be affected, and by this also the parametrization for the muon number distribution given by Eq.(5.11). The two black lines represent the by now unchanged parametrization according to Eq.(5.11) under the assumption of the hard decrease ("Nch hard") of the charged particle distribution, or of an exponential one ("Nch exp.") respectively.

Additionally, two extreme parametrizations for the description of the muon shower fluctuations are shown: a fast cutoff ("hard"), and an exponential decrease ("exp"). Again, all parametrizations are chosen such that the quality of the description, which is represented by the reduced chi-square value, does not get worse. The modification of Eq.(5.11) was done somewhat different to that applied to Eq.(5.3) by means of the new Eqs.(6.46) or (6.47). While the modification with the hard cutoff can equally be applied to the muon distribution<sup>45</sup>, the modification considering an exponential decrease have to be implemented based on another approach. Otherwise, if the description of the tails would simply be replaced by an exponential term as a function of the muon number, the correlation between the two shower sizes would get lost, since the new functional description of the tails would depend only on the muon number. In order to keep the correlation, the quadratic function, given by Eq.(5.8) and used for the parametrization of the muon

<sup>44</sup>More precisely, the derivatives divided by  $p_{10}$ .

<sup>45</sup>And since the implementation of the hard cutoff is done analogue to that in case of the charged particles (cf. Eq.(6.46)), it is trivial and will not be shown again.

number distribution by Eq.(5.11), was modified in such a way that the desired effect can be achieved. For the modification of the left tail, the old quadratic function is replaced by the new one:

$$\begin{aligned} \log_{10}N_0 \left( \log_{10}N_{\mu,t}^{\text{true}} \right) &= b_0 + b_1 \times \log_{10}N_{\mu,t}^{\text{true}} - b_2 \times \left( \log_{10}N_{\mu,t}^{\text{true}} \right)^2 \\ &\quad - (b_3 - \log_{10}N_{\mu,t}^{\text{true}}) \times 0.3 \times \frac{\left( \log_{10}N_{\mu,t}^{\text{true}} \right)^2}{16} \end{aligned} \quad (6.48)$$

if  $\log_{10}N_{\mu,t}^{\text{true}} \leq b_3$ . The right tail is modified by the replacement:

$$\begin{aligned} \log_{10}N_0 \left( \log_{10}N_{\mu,t}^{\text{true}} \right) &= b_0 + b_1 \times \log_{10}N_{\mu,t}^{\text{true}} - b_2 \times \left( \log_{10}N_{\mu,t}^{\text{true}} \right)^2 \\ &\quad - (b_4 - \log_{10}N_{\mu,t}^{\text{true}}) \times 0.4 \times \frac{\left( \log_{10}N_{\mu,t}^{\text{true}} \right)^2}{16} \end{aligned} \quad (6.49)$$

if  $\log_{10}N_{\mu,t}^{\text{true}} \geq b_4$ . The parameters  $b_3$  and  $b_4$  correspond to the muon numbers where the distribution falls below a defined threshold value, and hence mark the transition region. Thereby, the hard cutoff is assumed to start at 1/40 of the distributions maximum, and the exponential decrease at 1/5 respectively. The values 0.3 respectively 0.4 cause the less faster decrease of the tails of the final parametrization of the muon number fluctuations. More precisely, by this modification the final parametrizations will not base on a true exponential description of the tails. The standard tail will only be deformed such that it decreases slower. Hence, the description ‘‘exponential decrease’’ is strictly speaking not hundred per cent correct here<sup>46</sup>. The term  $\left( \log_{10}N_{\mu,t}^{\text{true}} \right)^2 / 16$  causes a variation of the strength of the modification in dependence on the muon number, i.e. in principle in dependence on the primary energy.

Now, we shall focus on the reconstruction resolution. In Fig 6.12, different parametrizations for the reconstruction accuracies of the charged particles (left) and muons (right) are depicted, exemplarily for the slices  $6.6 < \log_{10}N_{\text{ch}}^{\text{tru}} < 6.7$  and  $4.9 < \log_{10}N_{\mu}^{\text{tru}} < 5.0$  respectively. The black lines (‘‘normal’’) represent the standard parametrization that are already shown in Fig. 5.25 and 5.29 respectively. Taking into account the large statistical uncertainties of the tails, the standard parametrizations can again be bracketed by two extreme cases: for one thing by a faster decreasing fit function (‘‘hard’’ cutoff), for another thing by a less steeper one (‘‘exp.’’, i.e. exponential decrease). The hard cutoff was realized by assuming zero probability if the number of showers falls below 1/25 of the number at the distribution’s maximum (both for the charged particles and for the muons). For the exponential decrease, the threshold values are 1/4 of the distribution’s maximum in case of the charged particles, and 1/3 in case of the muons. While the hard cutoff is again realized as done for the shower fluctuations, the exponential decrease is implemented somewhat different. In case of the distribution of the charged particle reconstruction accuracy (cf. Eq.(5.17)), the left part of the parametrization was given by a Gaussian function, and can hence be modified analogue to Eq.(6.47), whereas the right part bases already on an exponential function. In order to additionally reduce the velocity of that exponential decrease of the right tail, its strength is reduced by substituting  $p_3$  of Eq.(5.17) by  $p_3 \rightarrow 1.4 \times p_3$ . Vice versa, in case of the distribution of the

<sup>46</sup>Nevertheless, the applied modification yields the desired effect, as can be seen from Fig. 6.11, bottom panel, such that the description ‘‘exponential decrease’’ will be used synonymously for the in truth only slower decreasing tails in comparison to the standard parametrization.

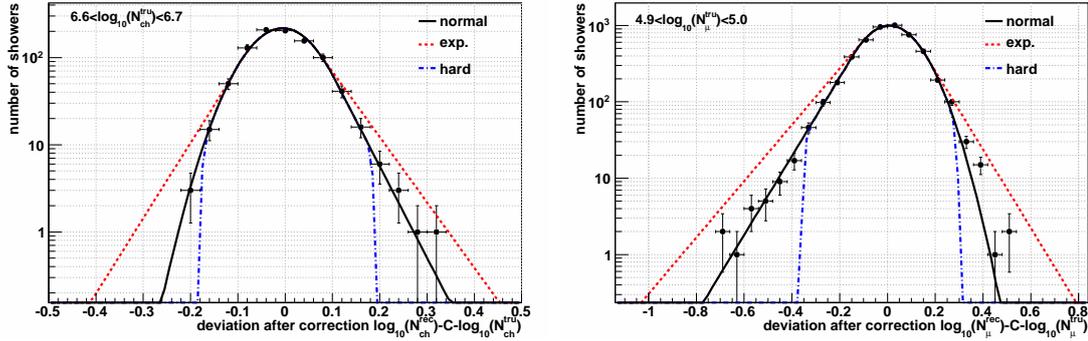


Figure 6.12: Different parametrizations for the reconstruction accuracies of charged particles (left) and muons (right), depicted exemplarily for the slices  $6.6 < \log_{10} N_{\text{ch}}^{\text{tru}} < 6.7$  and  $4.9 < \log_{10} N_{\mu}^{\text{tru}} < 5.0$  respectively.

muon reconstruction accuracy (cf. Eq.(5.20)) the left part of the distribution follows an exponential law, while the right one is given by a Gaussian function. The latter one can be modified analogously to Eq.(6.47), while the exponential decrease can again be attenuated by applying the aforementioned substitution, now to Eq.(5.20).

Under consideration of the two additional parametrization approaches, i.e. either the assumption of a hard cutoff or the weaker exponential decrease for all distributions, two new response matrices can be computed. Based on these matrices, the dataset can be unfolded again. The “standard” solution  $\vec{X}$  is bracketed by the two “extreme” estimates  $\vec{X}^{\text{hard}}$  and  $\vec{X}^{\text{exp.}}$ . The range that is spanned by the “extreme” solutions can be considered as a conservative estimate for the systematic uncertainty caused by the ambiguity of the distributions’ tails due to the limited Monte Carlo statistics. In the following, the range of this systematic uncertainty that affects the bin  $k$  of the estimated solution  $\vec{X}$  will be considered to be given by  $\pm |x_k^{\text{exp.}} - x_k^{\text{hard}}|/2$ .

The absolute value of the estimated range  $|x_k^{\text{exp.}} - x_k^{\text{hard}}|/2$  of systematic uncertainty in the unfolded solution  $\vec{X}$  in dependence on the primary energy is depicted in Fig. 6.13, exemplarily for protons (top panel) or iron nuclei (bottom panel). By this, the influence of the tail’s uncertainties can be judged. The result is based on an unfolding of the realistic toy data sample, which was introduced in Section 6.3.2. While the dataset was generated by means of the standard response matrix, it was unfolded with different sets of response matrices in that either some or all new “extreme” parametrizations have been considered. Shown are four different cases: While in the first one the response matrix used in the unfolding was computed based on the new parametrizations of the charged particle shower fluctuations only, in the other cases in addition to that also the other parametrizations are changed successively, i.e. the fluctuations in the muon number, the parametrizations of the resolution for the charged particles, and, finally, all four new parametrizations together. In case of protons, already a large systematic effect is induced by a modification of the shower fluctuations of the charged particles, while the changes applied to the other three types of parametrizations are with less additional influence. This was already expected, since the results shown in Fig. 6.10 predicted that in case of protons the charged particle shower fluctuations are the dominant uncertainties. For iron nuclei, the circumstances are different. Again, the systematic uncertainty induced by the one in the charged particle shower fluctuations is dominating again, while the shower fluctuations in

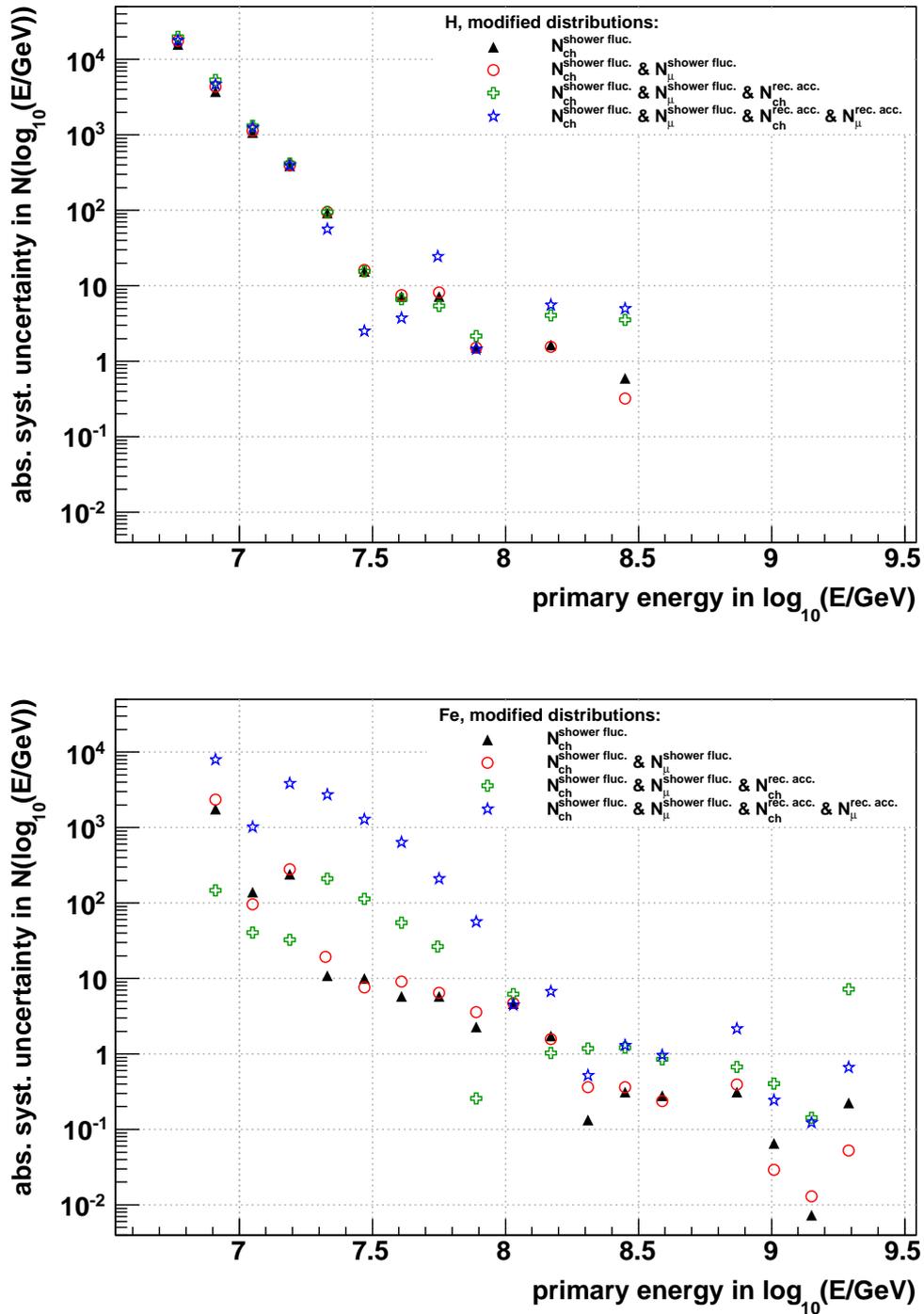


Figure 6.13: The range of the systematic uncertainty in the unfolded solution caused by the ambiguity of the distributions' tails due to the limited Monte Carlo statistics as a function of the primary energy (top: protons, bottom: iron nuclei). Shown are different estimates that are gained by modification of either only some or of all four parametrizations that are contributing to the response matrix (see legend and text for details).

the muon number are less important (again already predicted by Fig. 6.10). However, in case of iron nuclei the shower fluctuations are in general smaller than in case of protons, while the reconstruction resolution is comparable to that of protons. Hence, in case of heavier nuclei, the resolution gets more and more important. That is why in case of iron nuclei also the modifications of the parametrizations of the reconstruction accuracies will have significant impacts on the final solution. This can be ascertained from Fig. 6.13, where, especially close to the threshold of full efficiency at around  $\log_{10}(E/\text{GeV}) \approx 7.0$ , the additional modification of the parametrizations of the resolution induces further systematic uncertainties. To summarize, the preliminary considerations that based on Fig. 6.10 and motivated the decision to take into account all four parametrizations as sources of significant systematic uncertainties, have finally been confirmed.

### Impacts of the uncertainties in case of real data

Since, appropriate means are now available to estimate all sources of uncertainties, the impacts of them on the unfolded solution that bases on the shower size distribution measured with KASCADE-Grande can now be examined. While the calculation of the uncertainties for the unfolding of that real dataset will be performed later on, at this point the results are already anticipated, since the comparison fits contextually well in this section. Furthermore, the uncertainty of the response matrix is more meaningful in case of the unfolding of the measured shower sizes distribution than in case of trial spectra, which are already generated using the perhaps systematically “wrong” response matrix. The results shown here are derived by an unfolding with Gold’s algorithm under an iteration depth of  $\Delta\chi^2 = 2 \times 10^{-3}$ , which will later on be found to be the best one in case of the real data.

In Fig. 6.14, a comparison between all uncertainties the solution suffers from is presented (top panel for protons, bottom panel for iron nuclei). Since real data are used, the statistical uncertainty due to the limited measurement time as well as the bias caused by Gold’s algorithm could only be estimated. While the former was computed according to Eq.(6.44), the latter one bases on the relative systematic bias bracketed analogue to Fig. 6.9 and reconverted to an absolute value, i.e. to units of the number of showers. The overall uncertainties in the parametrizations that are used in the response matrix and that are caused by the limited Monte Carlo statistics are computed by the afore-explained approach that bases on the randomized generation of new parametrizations, which are all compatible within statistics. The influences of the uncertainties of the distributions’ tails are again analysed additionally by means of the two extreme approaches assuming either a hard cutoff or an exponential decrease of all four parametrizations that contribute to the response matrix.

While at the lower energies the bias as well as the systematic uncertainty caused by the ambiguity in the tails of the distributions dominate, at higher energies the statistical uncertainty due to the limited measurement time gets important, too. At the highest observed energies the number of air showers contributing to a specific energy bin gets as small that the tails of the distributions are covered only rarely, such that their uncertainties get less important. The systematic uncertainty that was caused by the overall statistical uncertainty in the distributions is, compared to the other sources, without significance over the whole energy range. That means that the Monte Carlo statistics is in general already good. Considering the limited data statistics, an increase of the Monte Carlo statistics does not seem to be needed. Nevertheless, the tails of the simulated distributions are very ambiguous. From the statistical point of view, they are unimportant.

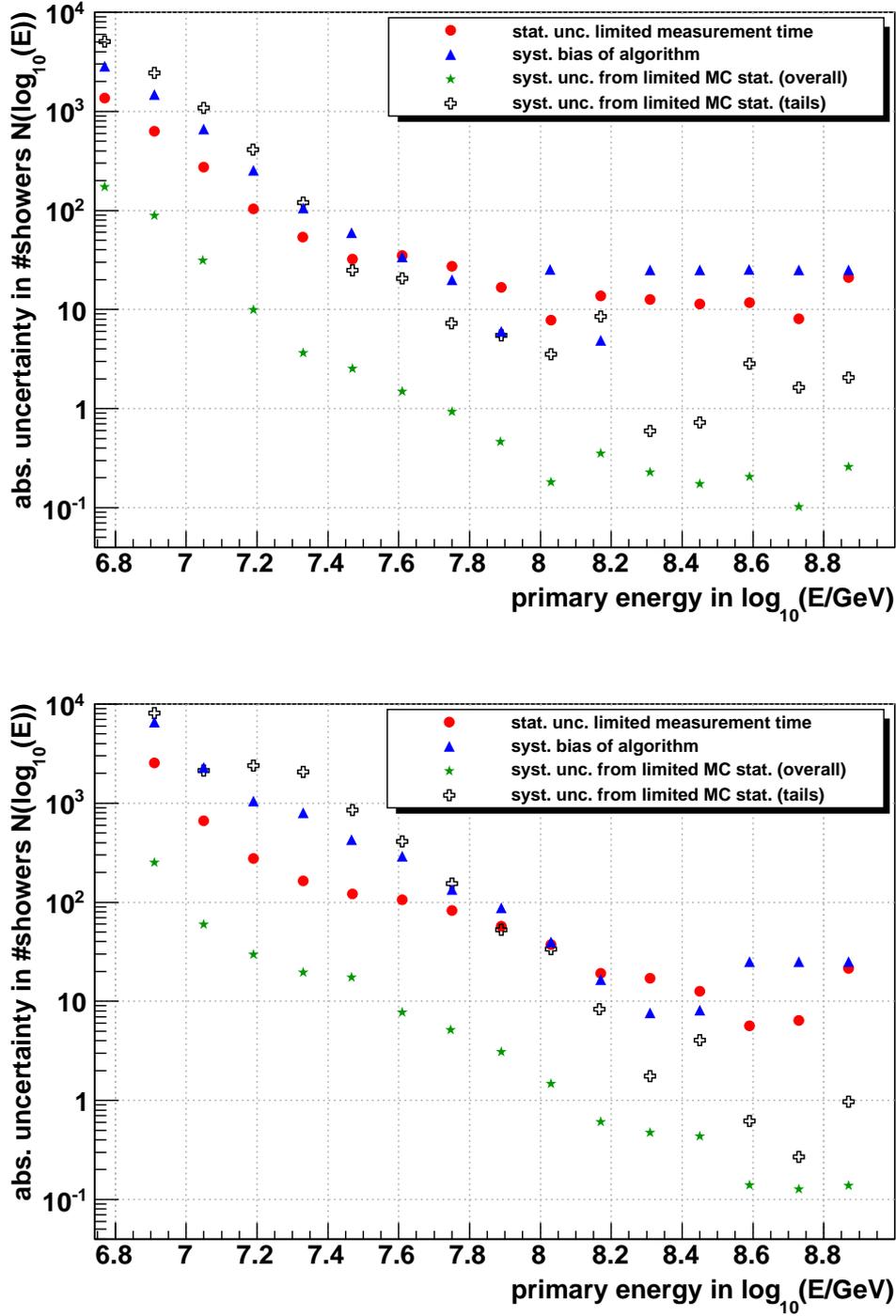


Figure 6.14: Comparison between all uncertainties the solution suffers from (top: protons, bottom: iron nuclei). The results base on an unfolding of the measured shower size distribution. Since real data are used, the statistical uncertainty due to the limited measurement time as well as the bias caused by Gold's algorithm are both estimated only (see text for further details).

From the physical one, as explained afore, this uncertainty can have huge impacts on the final result. In order to simulate distributions with significantly better confined tails, an enormous number of air showers have to be generated, which would mean a very long computing time. However, the effects of the tail's uncertainties are not significantly larger than the systematic bias induced by the unfolding algorithm itself, such that it was decided not to generate a larger simulation set.

Summing up the insights gained in this section, all possible sources of uncertainties are known and can be estimated with high precision. All results can be applied without limitation to the case of real data. In the final presentation of the energy spectra for elemental groups of cosmic rays, the statistical uncertainties caused by the limited measurement time will be represented by error bars, while all three sources of systematic uncertainties, i.e. the bias caused by the unfolding algorithm as well as both uncertainties induced by the limited Monte Carlo statistics, are visualized by a common systematic error band. Hence, the presented results will agree with the truth within the given uncertainties, under the premise that the used high energy interaction model QGSJET-II-02 describes the air shower development correctly.

#### 6.3.4 Stopping criterion for Gold's unfolding algorithm

The iterative unfolding algorithms approximate the exact solution  $\vec{X}$  of the matrix equation  $\vec{Y} = \mathbf{R}\vec{X}$  successively. As explained, the more iteration steps are performed, the less<sup>47</sup> biased the estimated solution will be, but the more possible fluctuations in the data set will be amplified. It will be the goal to find an "optimal" iteration depth that yields an "appropriate" solution. The question to answer is: what characterizes such an optimal solution?

##### The weighted mean squared error

In Section 6.3.1, a chi-square test (cf. Eq.(6.39)) was introduced in order to check the quality of the description of the measured data by the unfolded solution. However, it was emphasized that by this test the quality of the solution itself can only be judged peripherally. For sure, the unfolded solution should be able, to a certain extent, to reproduce the measured data sample. Nevertheless, the solution is not wanted to reproduce any random fluctuation in the data sample.

Instead, an "appropriate" solution will be characterized by an optimal balance between a remaining bias and the unavoidable statistical uncertainties. Thereby, only the statistical uncertainties of the data sample that are propagated through the unfolding algorithm, as well as the bias of the algorithm itself are of interest. The additional uncertainties that are caused by the systematic uncertainties in the response matrix are meaningless for the question about an optimal iteration depth. Cowan [54] suggests as a measure of the goodness of the unfolded solution the so-called *Mean Squared Error*:

$$MSE = \frac{1}{N} \sum_{k=1}^N (\sigma_k^2 + b_k^2) \quad . \quad (6.50)$$

Thereby, for all<sup>48</sup>  $N$  energy bins  $k$  of the estimated solution vector  $\vec{X}^{\text{est}}$  the true statistical

<sup>47</sup>Since Gold's algorithm is designed to converge to the exact solution.

<sup>48</sup>Note: the vector  $\vec{X}^{\text{est}}$  contains, by construction, the estimated solutions of all five regarded primary particles, i.e. five energy spectra, such that the sum in Eq.(6.50) is carried out over all considered energies for all used primaries. Thus, the *MSE* includes also an averaging over all nuclei.

uncertainty  $\sigma_k$  (given by Eq.(6.42)) and the true bias  $b_k$  (given by Eq.(6.43)) in the current bin are considered. An optimal solution estimate  $\vec{X}^{\text{est}}$  is achieved when the *MSE* is minimal, since in this case the best compromise between a remaining bias and statistical oscillations is realized.

However, since the contents  $x_k^{\text{est}}$  of the bins of the solution vector can have different statistical accuracies, in a correct approach the uncertainties  $\sigma_k$  and  $b_k$  of individual bins should have varying strong influences on the *MSE*. Assuming, in a simplified argumentation, Poissonian uncertainties in the bin contents of the solution  $\vec{X}^{\text{est}}$ , the uncertainty in bin  $k$  of the solution is given by  $\sqrt{x_k^{\text{est}}}$ . Using this, the *Weighted Mean Squared Error* can be defined:

$$WMSE = \frac{1}{N} \sum_{k=1}^N \frac{\sigma_k^2 + b_k^2}{x_k^{\text{est}}} . \quad (6.51)$$

In the further analysis, the iteration depth that is used to unfold the optimal solution with the smallest *WMSE* possible will be called the “optimal iteration depth”.

In [55], Cowan introduces another approach to find the optimal solution. He observed the ratio between the bias  $b$  and the accuracy  $\sigma_b^2$  of its determination itself by means of the *Relative Variance of the Bias*:

$$RVB = \frac{1}{N} \sum_{k=1}^N \frac{b_k^2}{\sigma_{b,k}^2} . \quad (6.52)$$

In Eq.(6.43), the bias  $b$  is computed as the mean deviation between the true solution and 50 estimates that base on toy datasets generated by means of a random generator. The fluctuation of these deviations can be considered as the uncertainty  $\sigma_b^2$  in the bias estimation itself. The larger the iteration depth is, the smaller the bias, but the larger the uncertainty in its determination will be. An optimal situation would be, if  $RVB = 1$ . In this case, the bias and its uncertainty are of comparable size. A deeper iteration would be unsuitable, since on the one hand the bias would get smaller, while on the other hand its uncertainty will increase such that the determination of the bias will fail.

In Fig. 6.15, on the top panels, the weighted mean variance

$$\text{weighted mean variance} = \frac{1}{N} \sum_{k=1}^N \frac{\sigma_k^2}{x_k^{\text{est}}} , \quad (6.53)$$

as well as the weighted mean squared bias

$$\text{weighted mean squared bias} = \frac{1}{N} \sum_{k=1}^N \frac{b_k^2}{x_k^{\text{est}}} \quad (6.54)$$

are illustrated in dependence on the iteration depth, which both contribute<sup>49</sup> to the *WMSE* given by Eq.(6.51) and shown in the middle plot. Additionally, the alternative approach that bases on the *RVB*, given by Eq.(6.52), is depicted on the bottom panel. All results are exemplarily shown in case of the realistic trial spectra that is introduced in Section 6.3.2 and unfolded with the respective iteration depths. As already predicted, the

<sup>49</sup>The weighted mean variance can be understood as the “left” term of Eq.(6.51), while the weighted mean squared bias gives the “right” one.

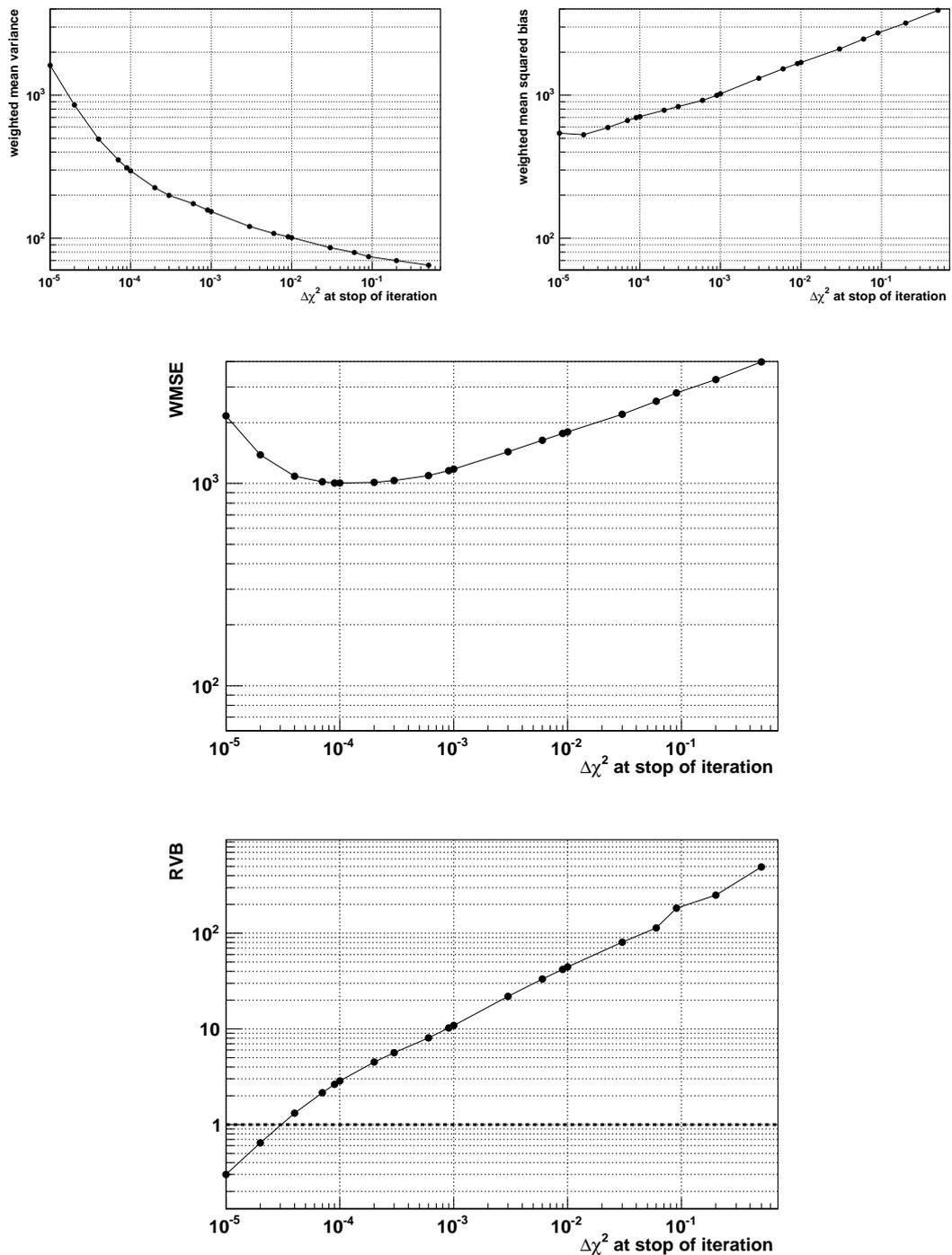


Figure 6.15: On the top panels, the weighted mean variance (cf. Eq.(6.53)) and the weighted mean squared bias (cf. Eq.(6.54)) are illustrated in dependence on the iteration depth, which both contribute to the *WMSE* (cf. Eq.(6.51)) shown in the middle plot, again as a function of the iteration depth. Additionally, the alternative approach that bases on the *RVB* (cf. Eq.(6.52)) is depicted on the bottom panel. All results are exemplarily shown in case of the realistic trial spectra that is introduced in Section 6.3.2. To guide the eyes, all markers are connected by straight lines.

larger<sup>50</sup> the depth of the iteration is, the larger the statistical uncertainties will get, but the smaller the bias will be. The resulting  $WMSE$  has a minimum at around  $\Delta\chi^2 \approx 10^{-4}$ . An optimal balance between the statistical uncertainties and the remaining bias is given in case of that iteration depth. The  $RVB$  is 1 at around  $\Delta\chi^2 \approx 3 \times 10^{-5}$ , what conforms with an iteration depth for that the  $WMSE$  is still minimal in good approximation, as the range of the minimum is relatively broad. Also the  $RVB$  implies that more iteration steps are useless, since the uncertainty in the bias estimate will be larger than the bias itself.

Also in case of other trial spectra, both approaches yield compatible optimal iteration depths, such that they can be used optionally. In a further cross-check based on diverse toy data samples and performed by eye, the unfolded solutions have been compared to the true ones in order to review the quality of the prediction of the optimal iteration depth by both methods. Thereby, both approaches performed reliably in all cases. The next question to be answered is how that can be applied to real data, where the “true” uncertainties are unknown and hence can only be estimated.

### The estimated weighted mean squared error

For the determination of the optimal iteration depth for the unfolding of the shower size distribution measured with KASCADE-Grande, basically the same approach as in case of the toy spectra can be used. The only difference will be that now the estimated uncertainties, i.e. the estimated statistical uncertainty  $\sigma_k^{\text{est}}$  given by Eq.(6.44) and the estimated bias  $b_k^{\text{est}}$  given by Eq.(6.45), have to be regarded. The estimated  $WMSE^{\text{est}}$  is given by:

$$WMSE^{\text{est}} = \frac{1}{N} \sum_{k=1}^N \frac{(\sigma_k^{\text{est}})^2 + (b_k^{\text{est}})^2}{x_k^{\text{est}}} \quad , \quad (6.55)$$

while the estimated  $RVB^{\text{est}}$  follows to:

$$RVB^{\text{est}} = \frac{1}{N} \sum_{k=1}^N \frac{(b_k^{\text{est}})^2}{(\sigma_{b,k}^{\text{est}})^2} \quad . \quad (6.56)$$

However, since for the estimation of the systematic bias templates are used that have been reconstructed by an unfolding under specific iteration depths and are assumed to represent the “true” solution in the subsequent steps, the answer which of the templates is really close to the truth has still to be given. The template that is as close to the truth as possible can be considered as the best estimate in the final analysis. As a first indication, the toy dataset that will be generated based on the template solution should be, within statistics, compatible to the initial shower size distribution that was measured with KASCADE-Grande. This criterion will be considered in the final unfolding analysis. However, as already explained, the quality of the data description is something different than the quality of the solution itself. Hence, a further criterion has to be found.

The following chain of thoughts will illustrate how to find the best solution template: Each template is gained by an unfolding of the measured shower size distribution under a specific iteration depth. Whilst the estimation process of the systematic bias, a certain number of toy datasets are generated based on the template by means of a random generator and regarding the response matrix as probability distribution. Apart from that

<sup>50</sup>Note: A large depth of iteration corresponds to a large number of iterations, and hence finally to a small value of  $\Delta\chi^2$ . Thus, the utilization of a large number of iterations means to follow the abscissas from the right to the left.

a random generator is used, this generation is an exact one in the sense that there is no algorithm used that could cause any bias. The generated toy dataset is now unfolded with diverse iteration depths. The unfolding results can be compared to the template solution, i.e. to the solution that is temporarily assumed to be the “true” one. Hence, for the toy datasets an optimal iteration depth can be determined by means of the minimum in the *WMSE*. In case of using the same iteration depth for the unfolding of the toy datasets as already used for the computation of the template, the best agreement between the template and the estimated solutions should be given. This can be understood when taking into account that the generation of the toy datasets is primarily an exact and unbiased procedure, such that the unfolding of all toy sets is only a repetition of the initial computation of the template, i.e. of the unfolding of the measured shower size distribution. Since it is only a repetition, the best iteration depth should be the same as in case of the initial unfolding. However, if the *WMSE* predicts another optimal iteration depth than that used in the computation of the template, something goes wrong. There are two possible scenarios: The optimal iteration depth for the unfolding of the toy datasets is smaller<sup>51</sup> than the depth used for the template computation, or vice versa. The first case means that already in the initial unfolding of the measured distribution too many iteration steps have been performed. That means that the bias in the template is reasonable, but the statistical oscillations can get problematic. In the second case, the iteration whilst the computation of the template was stopped too early. The template will suffer from a too large bias. Consequently, the template is most probable close to the truth, if the toy datasets have to be unfolded with the same<sup>52</sup> number of iteration steps as already used in the initial template computation. By this, additionally to the comparison of the toy datasets with the measured data sample by means of a chi-square test, the comparison between the iteration depth of the template and the optimal depth predicted by the *WMSE* for the toy datasets will offer a criterion which of all templates is most probable close to the “truth”. In this analysis, this template solution will finally be considered as the best estimate for the elemental energy spectra of cosmic rays.

The application of the afore introduced criterion will now be shown exemplarily in case of the realistic trial spectrum, which is introduced in Section 6.3.2 and was also already used for the figures that depict the true *WMSE* (Fig. 6.15). In Fig. 6.16, a comparison between that true and the estimated *WMSE* or *RVB* is shown for diverse templates. Each template, computed under a specific iteration depth (see legend), was used to generate toy datasets that have been unfolded with diverse iteration depths. For each depth, a  $WMSE^{\text{est}}$  can be determined. For each template, these estimated  $WMSE^{\text{est}}$  are shown as a function of the iteration depth used to unfold the respective toy datasets. The  $RVB^{\text{est}}$  is also shown for cross-check purposes. On the two bottom panels, the axis has been rescaled to the important ranges. The vertical lines mark, on the x-axis, the iteration depths that have been used whilst the computation of the respective template. The afore-introduced criterion, i.e. that the best iteration depth used to unfold the toy datasets should be equal or slightly less deeper than that used whilst the computation of the respective template, is fulfilled in case of the template characterized by the iteration depth  $\Delta\chi^2 = 3 \times 10^{-4}$ . The same reasoning used for the  $WMSE^{\text{est}}$  can also be applied to the  $RVB^{\text{est}}$ . Also the iteration depth that is tagged to be the optimal one by the criterion  $RVB^{\text{est}} = 1$  should correspond to the depth initially used to compute the respective template. This is fulfilled for the template unfolded with an iteration depth  $\Delta\chi^2 = 3 \times 10^{-4}$ . Hence, both

<sup>51</sup>The iteration is not as deep, i.e. less steps are necessary.

<sup>52</sup>Or slightly smaller.

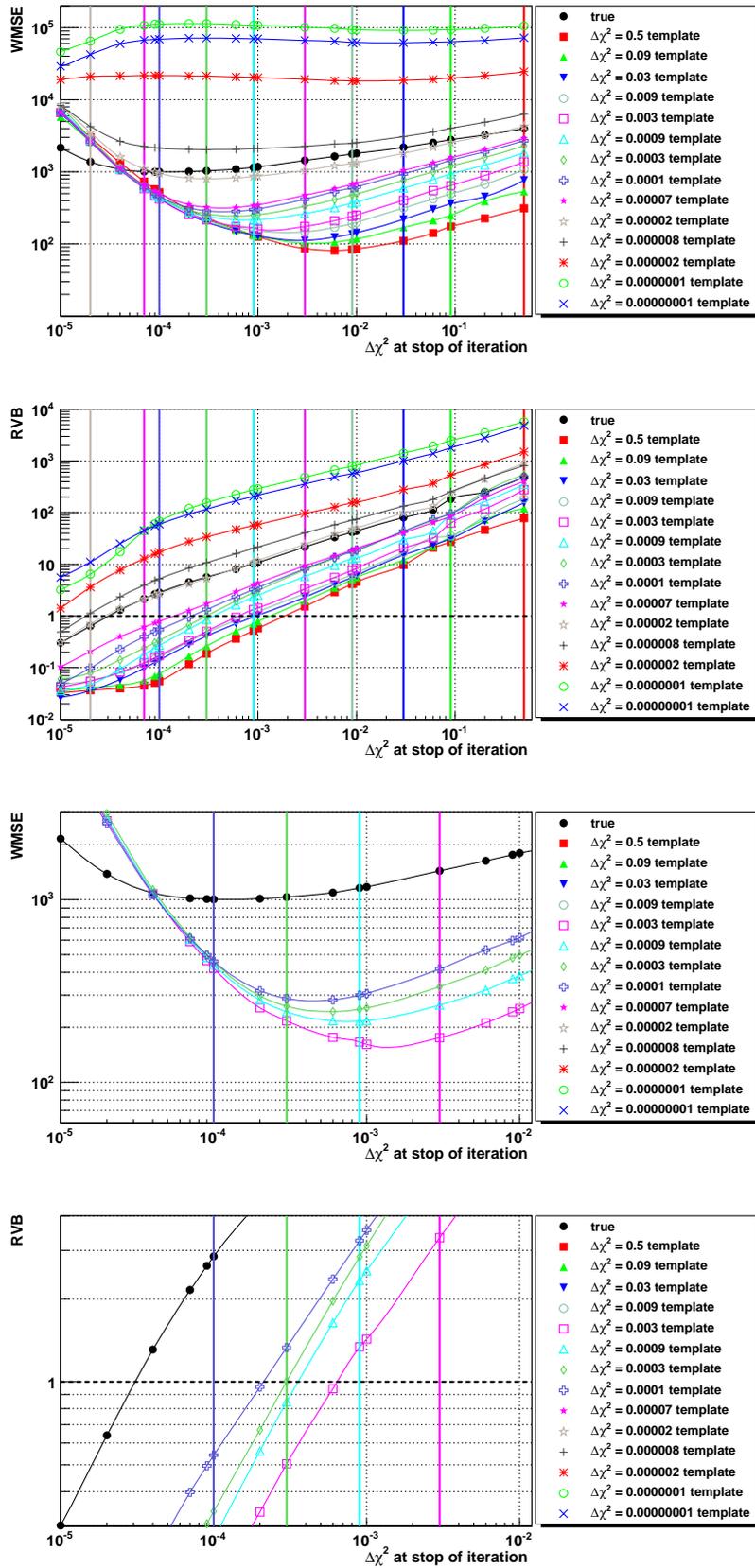


Figure 6.16: Comparison between the true and the estimated  $WMSE$  or  $RVB$  for different templates. On the bottom panels, the important axis ranges are zoomed in. The results base on the realistic trial spectrum that is introduced in Section 6.3.2.

approaches yield the same optimal iteration depth, and agree<sup>53</sup> with the ones gained by the true  $WMSE$  ( $\Delta\chi^2 \approx 10^{-4}$ ) or the true  $RVB$  ( $\Delta\chi^2 \approx 3 \times 10^{-5}$ ). Consequently, an optimal iteration depth can also be determined if the true spectra are unknown, as it is the case for the real measurement. As a result and based on the  $WMSE^{\text{est}}$ , for the used realistic trial spectra an optimal iteration depth of  $\Delta\chi^2 = 3 \times 10^{-4}$  is selected.

As a conclusion, the determination of the optimal iteration depth based on the  $WMSE^{\text{est}}$  works reliable. This was also tested by means of diverse alternative trial spectra. The  $RVB^{\text{est}}$  confirms the results. Hence, also in case of the real data distribution measured with KASCADE-Grande, an appropriate means to determine the optimal iteration depth is available now. In the next section, the quality of the unfolding procedure will be tested by means of diverse exemplary trial spectra.

### 6.3.5 Quality of Gold's unfolding algorithm

The reliability of Gold's unfolding algorithm was examined by means of a vast number of toy spectra. A large variety of constellations have been tested that cover physically meaningful spectra, but also very abstract ones that are expected to be present in nature less likely. In almost all observed cases, Gold's unfolding algorithm yielded solutions that were compatible with the true initial spectra within the estimated uncertainties. Only in a very few examples, some deviations between the true and the unfolded spectra were found that were not covered by the estimated uncertainties. However, experiences have been gained that allow to judge, e.g. based on the  $WMSE$ , whether or not such problems have to be expected. Showing all these intensive tests would go beyond the scope of this thesis. Hence, only some exemplary trial unfoldings will be presented in the following.

#### Realistic toy spectra

Firstly, the unfolding procedure shall be tested by means of a toy spectrum that is close to the cosmic ray spectrum expected to be present in nature. For this purpose, the realistic<sup>54</sup> toy dataset, introduced in Section 6.3.2 and based on Eq.(6.41), with parameters given in Table 6.1, was used. As determined by the estimated  $WMSE$  (cf. Section 6.3.4), the optimal iteration depth for the unfolding of that toy dataset is  $\Delta\chi^2 = 3 \times 10^{-4}$ .

In Fig. 6.17, the solution to the toy dataset estimated by means of Gold's unfolding algorithm and based on the aforementioned iteration depth is shown for all five considered primaries, representing different mass groups, as well as for the all-particle spectrum, which is the bin-wise sum of all five individual spectra. Additionally shown are, with hollow markers, the initially assumed true energy spectra. For a better distinguishability, the lighter (top panel) and heavier (bottom panel) mass groups are depicted in individual diagrams. The error bars represent the estimated statistical uncertainties caused by the limited statistics of the data sample, while the error bands mark the maximal range of the systematic uncertainties induced by the unfolding algorithm itself<sup>55</sup>. All uncertainties are estimated as elaborated in Section 6.3.3. The spectra are scaled by the potential law  $(E/\text{GeV})^{1.62}$ , which is the reciprocal of the one assumed in the initial toy spectra to be present before the knee (cf. Table 6.1, where the indices of the differential toy energy

<sup>53</sup>Considering the broad minimum of the  $WMSE$ .

<sup>54</sup>Or at least expected to be close to reality.

<sup>55</sup>Maximal 40% relative bias if  $\log_{10}(N^{\text{rec}}) > 1.3$  (corresponding to around  $N > 20$  air shower events), or, else, an absolute one of  $b = 40$ , as explained in Section 6.3.3 based on Fig. 6.9.

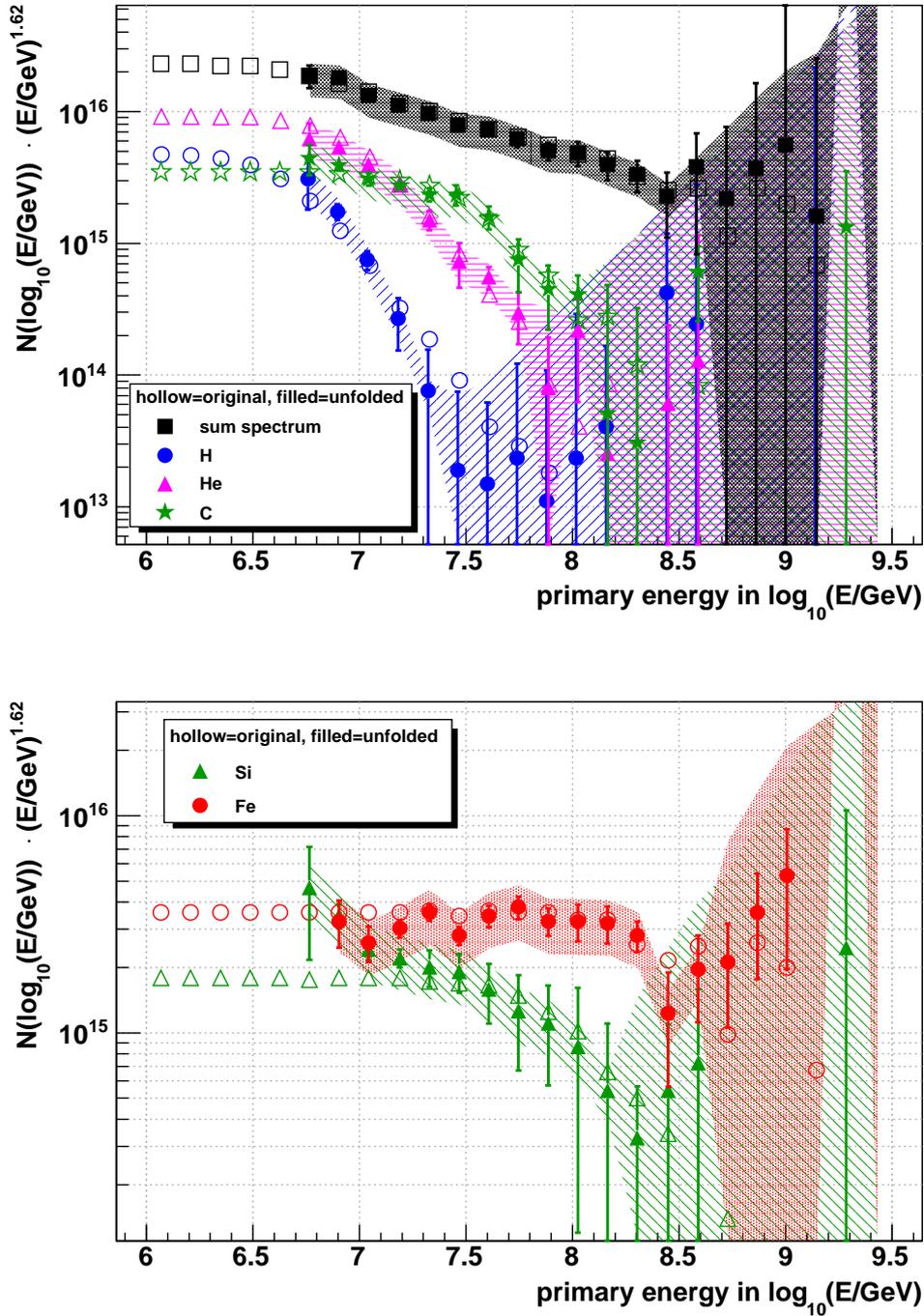


Figure 6.17: True (hollow markers) and unfolded (filled markers, Gold unfolding with iteration depth  $\Delta\chi^2 = 3 \times 10^{-4}$ ) energy spectra based on the realistic toy dataset that was introduced in Section 6.3.2. On the top panel for the all-particle spectrum and the lighter mass groups, represented by protons, helium, as well as by carbon, and on the bottom panel for the heavier mass groups represented by silicon and iron. The error bars represent the statistical uncertainties, while the error bands mark the maximal range of systematic uncertainty. For a better distinguishability, the true spectra are shown without error bars, albeit they suffer from statistical uncertainties, too.

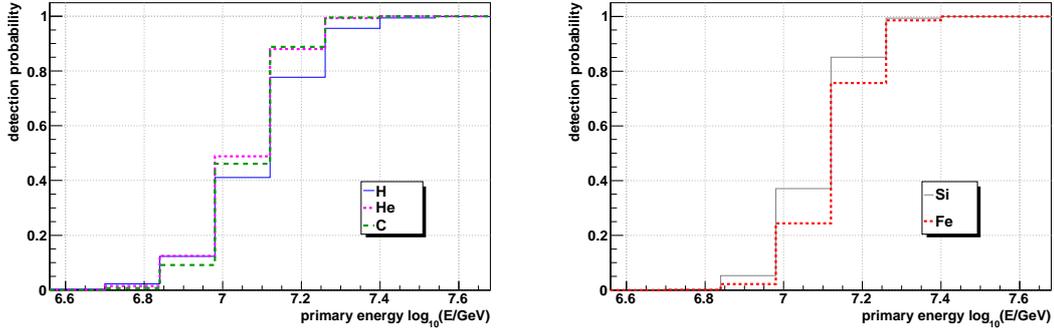


Figure 6.18: The detection probability, which incorporates the trigger and reconstruction efficiency (cf. Section 5.2.2) as well as the cut  $\log_{10}(N_{\text{ch}}) \geq 6.0$  and  $\log_{10}(N_{\text{p}}) \geq 5.0$  that showers have to pass in order to get considered in the  $\log_{10}(N_{\text{ch}})$ – $\log_{10}(N_{\text{p}})$  shower size plane.

spectra are summarized). By this, the knee structures are discernible more easily in the depiction.

In general, there is a good agreement between the true and the reconstructed solution, and most of the deviations conform with either the statistical uncertainties or with the systematic bias, or at least with a combination of both. Especially the all-particle spectrum can be unfolded with only small uncertainties. At higher energies, the spectra suffer from very large deviations or fluctuations. However, these ranges are tagged successfully to be insignificant by either the error bars or the error bands, or by both. One has to keep in mind that at these energies one has to deal with bins with actually only one-figure, or even zero content. Due to the positive definiteness of Gold’s unfolding algorithm, these contents are artificially enlarged. These overestimation effects are intensified visually by the diagrammatic representation due to the scaling of the spectra by  $(E/\text{GeV})^{1.62}$ .

As explained in Section 5.3.2, the conditioning applied to the response matrix will affect the number of reconstructed energy bins, since whilst the unfolding of the energy spectra of individual particles only those energy bins are considered for that there is a non-zero probability that this particular primary with this specific energy contributes to at least one cell of the considered  $\log_{10}(N_{\text{ch}})$ – $\log_{10}(N_{\text{p}})$  shower size plane, with  $\log_{10}(N_{\text{ch}}) \geq 6.0$  and  $\log_{10}(N_{\text{p}}) \geq 5.0$ . That is why not all 25 energy bins are covered by the solution, but only the last 19 or 20 bins.

While the unfolded spectra of the lighter mass groups often agree with the true ones already in the first considered bins, in case of the spectra of silicon and iron nuclei the first bins are often affected by larger deviations, which are not covered by the given uncertainties. This underestimation of the uncertainties for silicon and iron nuclei next to the threshold of full detection efficiency was observed in almost all performed trial unfoldings. The mentioned detection efficiency includes on the one hand the trigger and reconstruction efficiency (cf. Section 5.2.2), but in this specific analysis also the probability that an individual air shower contributes to at least one cell of the defined  $\log_{10}(N_{\text{ch}})$ – $\log_{10}(N_{\text{p}})$  shower size plane. Hence, additionally, all used showers have to pass the cuts  $\log_{10}(N_{\text{ch}}) \geq 6.0$  and  $\log_{10}(N_{\text{p}}) \geq 5.0$  that define the lower limits of the considered plane. In Fig. 6.18, that detection probability is illustrated in dependence on the primary energy. It was found that the unfolded spectra are only reliable for energies for that the respective primary has at least a detection probability of around 40% to

50%. In case of protons, helium, and carbon, this minimum probability is given for energies beyond  $\log_{10}(E/\text{GeV}) \approx 7.0$ , and for silicon and iron at higher energies at around  $\log_{10}(E/\text{GeV}) \approx 7.1$  to  $7.2$ . Already in case of the pure trigger and reconstruction efficiency (cf. Fig. 5.19, bottom left panel), for the heavier mass groups the efficiency curves were shifted to higher energies with respect to those of the lighter ones.

The reason for the dependence of the quality of the unfolded solution on the detection probability bases on differences in the convergence rates. Whilst the unfolding process, the quality improvement from one to another iteration step is judged by means of a chi-square test between the initial data sample and the forward folded solution, i.e. the reconstructed data sample (cf. Section 6.3.1). In case of low detection probabilities, a change in the solution will have less impacts on the reconstructed data sample than in case of full detection efficiency. Hence, for lower energies, going along with smaller detection probabilities, more iteration steps are necessary than for energy ranges with full detection efficiency. This imbalance between the convergence rates for the solution at smaller and higher energies is the reason for the bias visible at the first energy bins of the spectra of silicon and iron nuclei. In principle, if one would only take into account the smaller energies, the minimum in the *WMSE* would be located more at deeper iteration depths than in case of considering only higher energies whilst the *WMSE* computation. Hence, simply enlarging the number of iteration steps would for one thing remove the bias at the first energy bins, but for another thing would result in arising oscillations in the solution at higher energies. Since only the first energy bins<sup>56</sup> are affected by that bias, in the final analysis the afore-gained results will be considered such that the unfolded spectra are only shown for energies beyond the aforementioned minimum detection probability limits  $\log_{10}(E/\text{GeV}) \approx 7.0$  for the lighter mass groups, and  $\log_{10}(E/\text{GeV}) \approx 7.1$  to  $7.2$  for silicon and iron. Adapting these limits also for the results shown in Fig. 6.17, i.e. one would have to remove the first two markers of the unfolded energy spectra for protons, helium, and carbon, as well as the first three respectively four markers in case of iron respectively silicon, would improve the quality of the shown results significantly. Extensive tests have shown that all aforementioned insights can also be transferred to other toy spectra, such that, without loss of generality, everything will also be applicable for the unfolding of the dataset measured with KASCADE-Grande. Hence, a high quality solution, which will only deviate from the “true” one within the given uncertainties, is guaranteed.

### Improvement of the unfolding results

Although already all conceivable measures are adopted to get an optimal estimate for the true energy spectra, it will now be focused briefly on the question what else could be done to improve the results further. One main problem is the reconstruction resolution of the KASCADE-Grande experiment. While in the former KASCADE experiment the unavoidable shower fluctuations were dominating significantly over the reconstruction uncertainties, in KASCADE-Grande both sources are of more equivalent weight. Especially the muon reconstruction suffers from the limited detector coverage. In principle, the resolution can only be improved significantly by comprehensive and costly extensions of the experimental setup. Another problem is the limited measurement time. In Fig. 6.19, the same toy spectra are shown as already used for Fig. 6.17, but with round about a

<sup>56</sup>These are rather unimportant for this thesis, since one of the main goals is to examine whether or not the heavy component, i.e. especially the iron-like one, shows a knee-like structure at energies at around  $E = 10^{17}$  eV, what would be far enough from the first problematic energy bins.

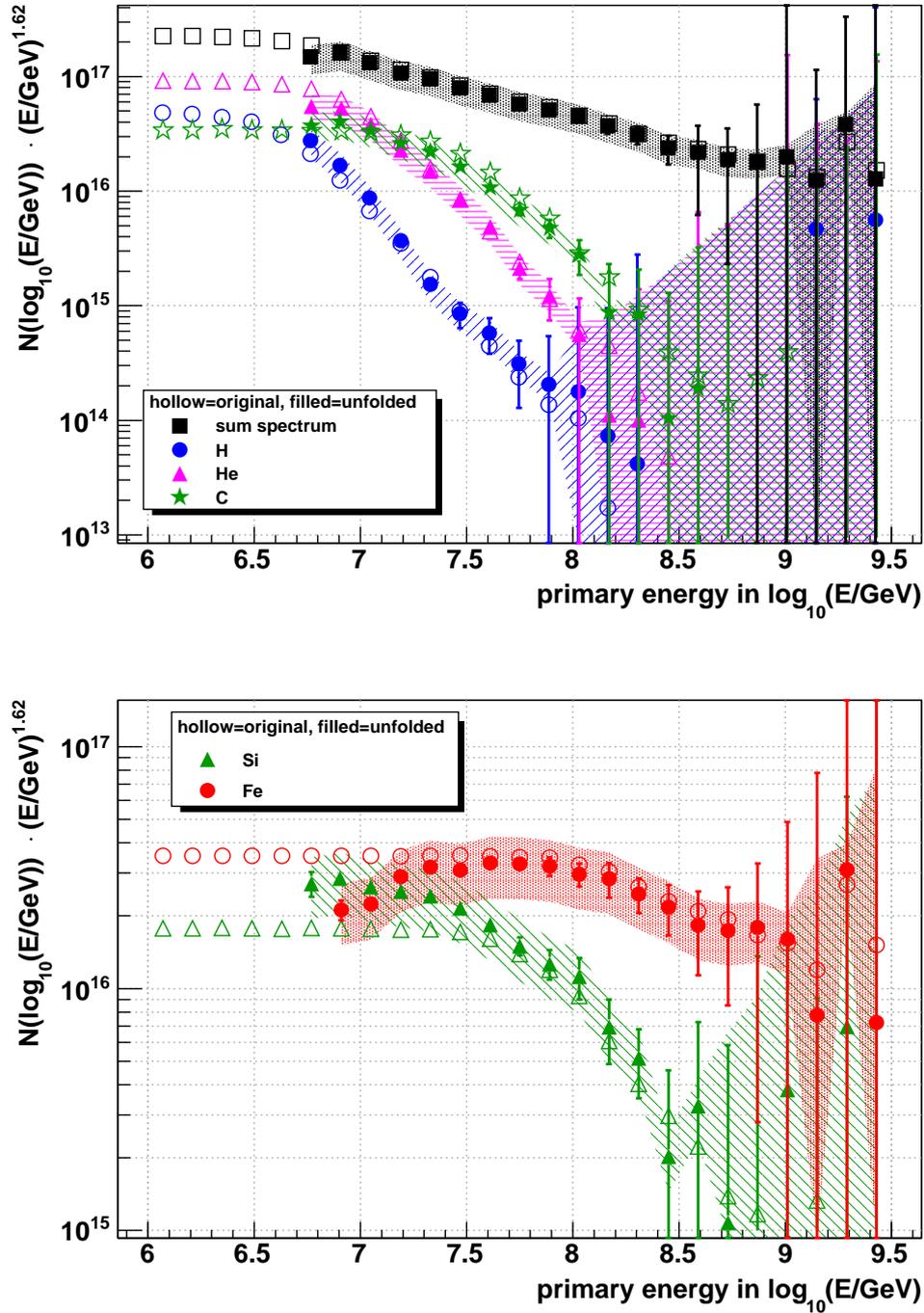


Figure 6.19: True (hollow markers) and unfolded (filled markers, iteration depth  $\Delta\chi^2 = 3 \times 10^{-4}$ ) energy spectra based on the realistic toy dataset that was introduced in Section 6.3.2, but now with a statistics enhanced by a factor of 10. On the top panel for the all-particle spectrum and the lighter mass groups, represented by protons, helium, as well as by carbon, and on the bottom panel for the heavier mass groups represented by silicon and iron. The error bars represent the statistical uncertainties, while the error bands mark the maximal range of systematic uncertainty. For a better distinguishability, the true spectra are shown without error bars, albeit they suffer from statistical uncertainties, too.

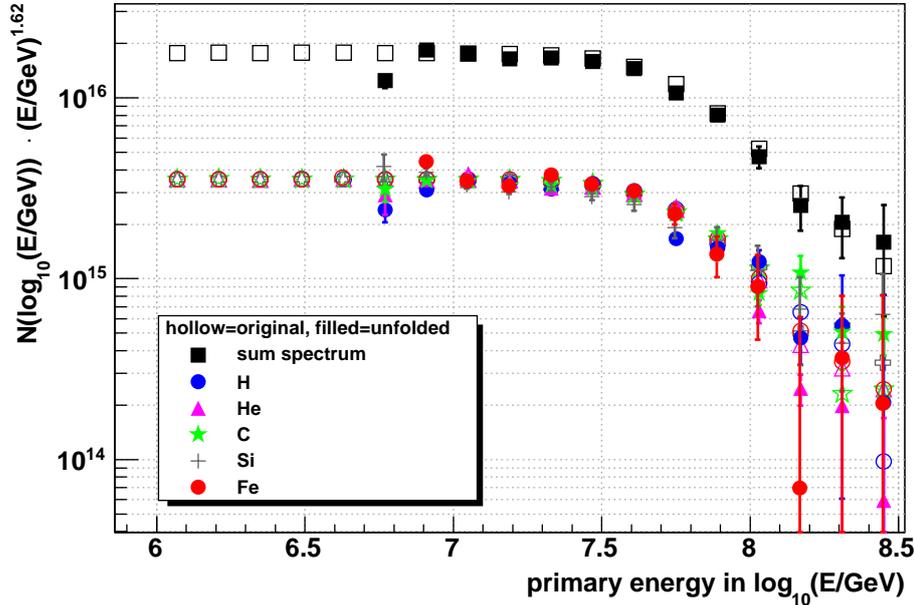


Figure 6.20: An exemplary result for an unfolding of a toy dataset that bases on “true” elemental energy spectra with equal indices, knee positions, and normalizations for all five primaries. For a better distinguishability, only the statistical uncertainties of the unfolded solutions are shown.

factor 10 more statistics. Except for the initial bias affecting again the energy ranges below full detection efficiency, the enhanced statistics yields a significantly better unfolding result. The spectra are overall smoother. Furthermore, the unfolded spectra are reliable up to energies half (lighter mass groups) a decade, or even a complete one (heavier mass groups) more than in case of the realistic data sample. Since the data set used in this work was collected during  $\sim 6.5$  years, a factor 10 more in statistics would require an additional operation for approximately 60 years, which is unrealistic. Alternatively, the observation area could be enlarged significantly, what would again cause additional costs.

As a conclusion, significant improvements of the estimation of the elemental energy spectra are principally achievable at KASCADE-Grande, but only under rather unrealizable efforts.

### Equidistribution

Many trial unfoldings have been performed in order to check whether or not the applied unfolding methods are, maybe, favouring certain primaries due to their different response matrices. None of the tests has revealed indications for such a favour. In Fig. 6.20, exemplarily the result of such a trial unfolding is shown. The true spectra base on a double power law with a knee at  $6.3 \times 10^{16}$  eV, and indices of the differential power law of -2.62 before, and -4.76 after the knee. The primaries have equal abundances. Except for the well-known bias at the first energy bins, there is no indication that the applied algorithm of Gold favours a certain mass group. Within the statistical uncertainties, the unfolded spectra are still overlapping.

Consequently, if the unfolding of the measured shower size spectra will yield that

an individual mass group is dominating over the others, this is not due to a systematic distortion caused by the applied algorithm, but certainly a true physical result<sup>57</sup>.

### Problematic spectra

In Fig. 6.21, exemplarily, for one of the rare “problematic” toy spectra sets the unfolding result is depicted. The toy energy spectra are characterized by very steep falling carbon and silicon spectra in comparison to that of the other primaries. Especially the silicon to iron ratio decreases quickly. Hence, the unfolding algorithm has to deal with neighbouring mass groups with considerably different abundances, and hence possibly different convergence rates to an exact solution. While the solutions agree with the true spectra in general, there are a few excesses discernible, which are not covered by the given uncertainties, e.g. in the spectrum of silicon at around  $\log_{10}(E/\text{GeV}) \approx 8$ . Since these excesses are not tagged to be insignificant by respective large uncertainties, the risk arises that they could be interpreted as real physical structures.

However, it was found that in most cases those excesses that are not covered by the given uncertainties are only affecting one single bin, such that one would anyway not dare to pay too much attention to this possible “structure”. Furthermore, the *WMSE* often gives indications whether or not such spectra are “problematic” ones. In Fig. 6.22, the *WMSE* in case of the problematic toy dataset used in Fig. 6.21 is depicted. Additionally shown are the *WMSE* of individual primaries that finally contribute to the common *WMSE*. The *WMSE* of the individual primaries have strongly different minima. While e.g. protons and carbon need smaller iteration depths, silicon or iron need significantly more iteration steps. A common optimal iteration depth is not achievable. Hence, if one selects an iteration depth that is as deep that all elemental spectra can be unfolded with a small bias only, some of the spectra of primaries with smaller necessary iteration number can already suffer from huge oscillations. For example, in case of the carbon spectrum, at around  $\log_{10}(E/\text{GeV}) \approx 8$ , one can observe a large jump from around  $N \times (E/\text{GeV})^{1.62} \approx 10^{15}$  to below  $N \times (E/\text{GeV})^{1.62} \approx 10^{13}$ . As a consequence and due to the correlation between the unfolded spectra, the spectrum of silicon is pulled up there, although in case of that primary the iteration depth is optimal, such that oscillations actually should not occur.

In such cases, where the correlated spectra of individual primaries underlie different convergence rates whilst the unfolding process, the influence of these excesses can to a certain degree be weakened by combining some individual spectra to a representative one. This is illustrated exemplarily in Fig. 6.23, where the individual spectra of helium, carbon, and silicon nuclei, shown in Fig. 6.21, are summed up bin-wise to a representative one. Now, at least within the combination of the statistical uncertainty and the bias, the unfolded spectra agree better with the true ones. Nevertheless, one knows that also this result has to be interpreted with caution, since the used datasets are already tagged to be “problematic” ones by the individual *WMSE*.

Based on intensive trial unfoldings, on the one hand technical means to tag problematic datasets have been elaborated, but on the other hand also experiences have been gained that allow to judge the reliability of the derived solution.

<sup>57</sup>But only in the framework of the used interaction models, here QGSJET-II-02, as already mentioned several times.

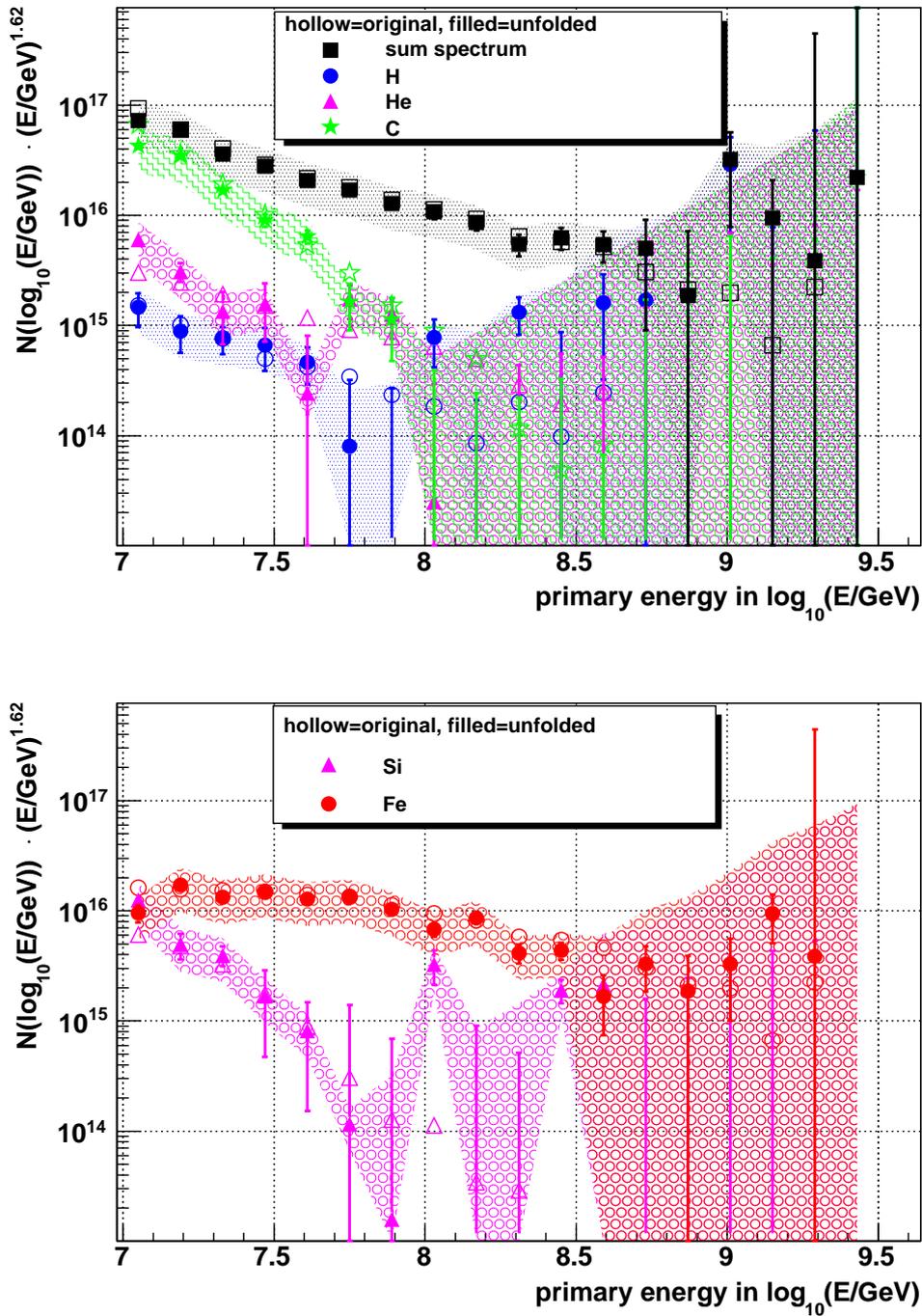


Figure 6.21: True (hollow markers) and unfolded (filled markers, iteration depth  $\Delta\chi^2 = 4 \times 10^{-6}$ ) energy spectra based on an arbitrary chosen “problematic” toy dataset. On the top panel for the all-particle spectrum and the lighter mass groups, represented by protons, helium, as well as by carbon, and on the bottom panel for the heavier mass groups represented by silicon and iron. The error bars represent the statistical uncertainties, while the error bands mark the maximal range of systematic uncertainty. For a better distinguishability, the true spectra are shown without error bars, albeit they suffer from statistical uncertainties, too.

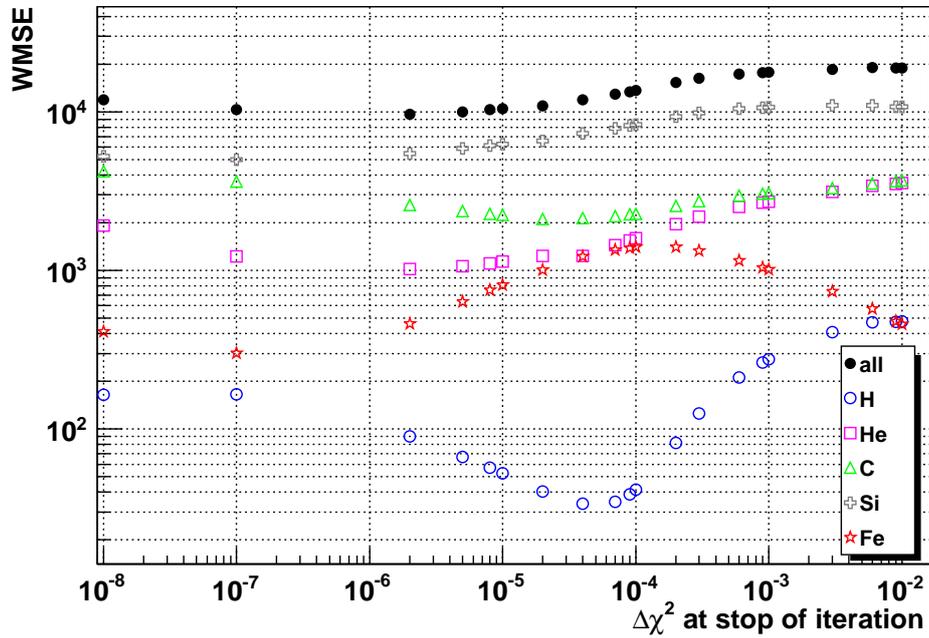


Figure 6.22: The  $WMSE$  (“all”) in case of the problematic toy dataset used in Fig. 6.21. Additionally shown are the  $WMSE$  of individual primaries finally contributing to the common  $WMSE$  (“all”).

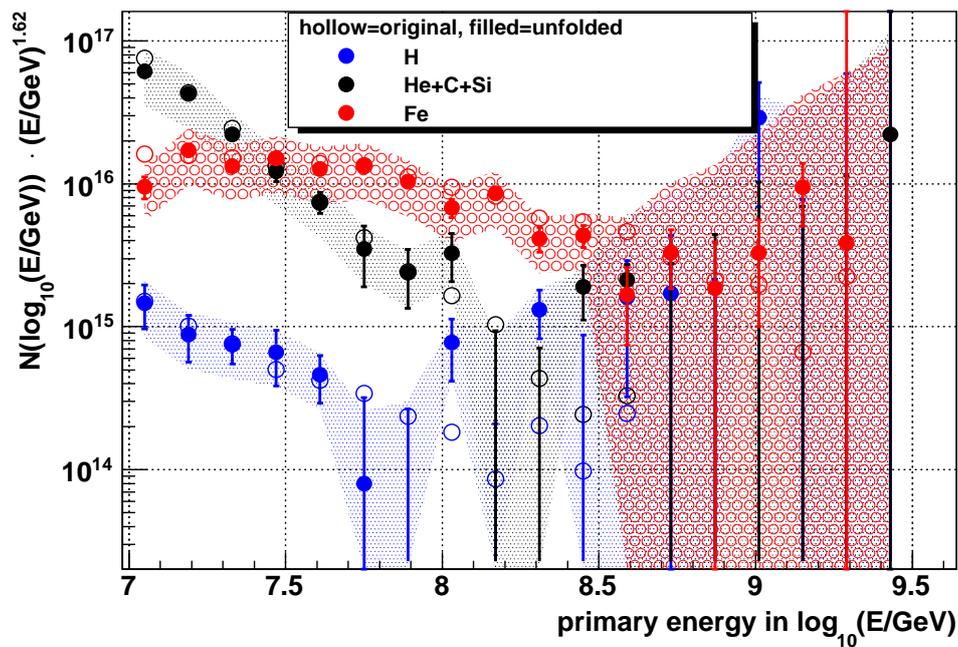


Figure 6.23: The true and unfolded spectra already shown in Fig. 6.21. But now the spectra of helium, carbon, and silicon are summed up bin-wisely to a representative one.

### Comparison between different unfolding algorithms

Finally, a brief comparison between the unfolding results gained by the iterative unfolding methods applying Gold's algorithm or Bayes' theorem, and by the regularized unfolding based on the principle of maximum entropy will be presented. Again, the realistic toy dataset, introduced in Section 6.3.2, is unfolded. In Fig. 6.24, a comparison between the unfolding results of all three algorithms is illustrated. As also seen in other trial unfoldings, the unfolding techniques applying Gold's algorithm or Bayes' theorem yield comparable results, with sometimes slight advantages for Gold's one. The method based on the principle of maximum entropy sometimes does not agree at all with the results of the two aforementioned techniques within the given uncertainties, while in general the unfolded solutions are of less quality. The all-particle spectrum can only be unfolded with larger uncertainties, too. As a conclusion, for the main unfolding analysis of the dataset measured with KASCADE-Grande Gold's algorithm will be used, while the two other ones will only be applied for cross-check purposes.

Summarizing the content of this chapter, appropriate methods for unfolding the shower size spectrum measured with KASCADE-Grande have been elaborated. The uncertainties that have to be taken into account are well known. Furthermore, the reliability of the whole unfolding procedure was tested intensively. Hence, all means that are needed to successfully apply these techniques to the measurement are available now, what will be the objective of the next chapter.

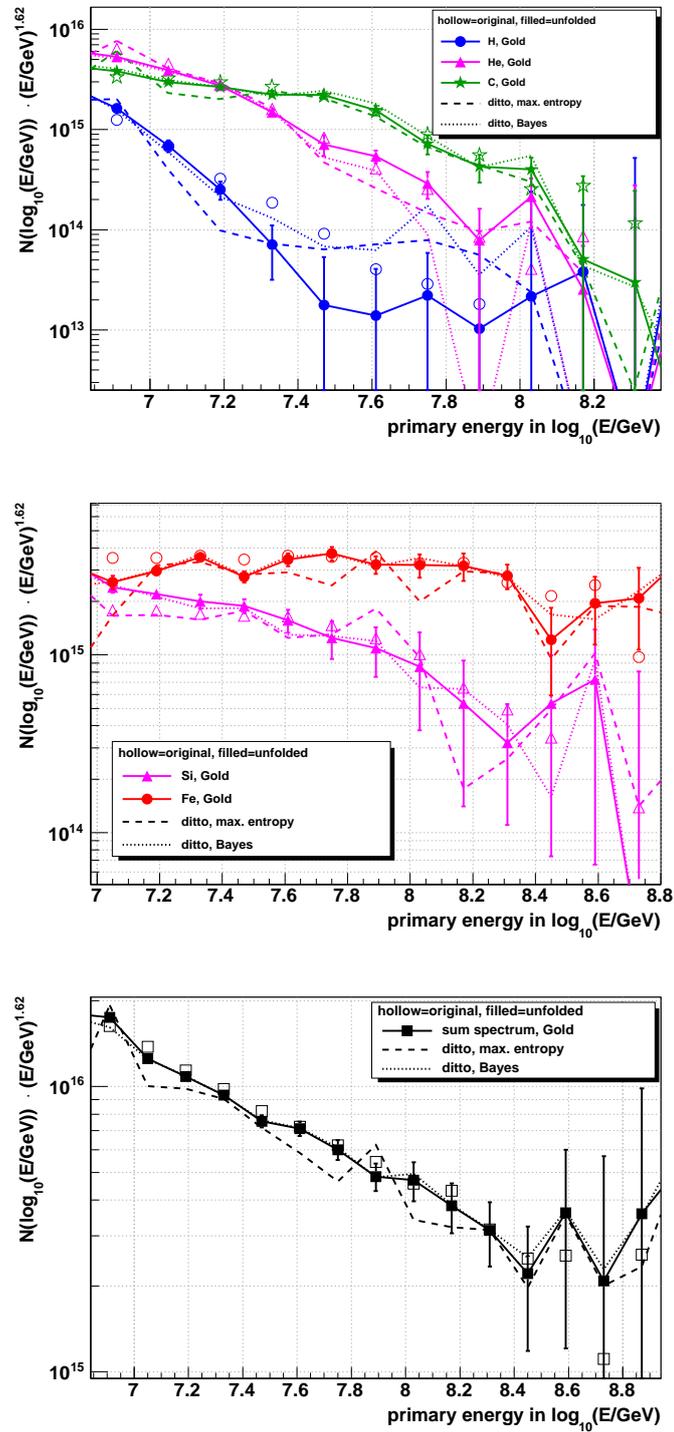


Figure 6.24: The true and unfolded spectra in case of the realistic toy dataset. But now, additionally, the unfolding results (dashed lines) gained with the algorithms based on Bayes' theorem or the principle of maximum entropy are shown. For a better distinguishability, only the statistical uncertainties (error bars) in case of Gold's algorithm are given.



## Chapter 7

# Unfolding the Energy Spectra for Elemental Groups of Cosmic Rays

This work's objective is to determine the composition and the elemental energy spectra of cosmic rays outside the Earth's atmosphere. In the previous chapters, appropriate techniques to derive these energy spectra by means of an unfolding of the two-dimensional  $\log_{10}(N_{\text{ch}})$ – $\log_{10}(N_{\text{p}})$  shower size distribution (cf. Fig. 4.8), measured with KASCADE-Grande, have been elaborated. Suitable deconvolution algorithms have been introduced, from which that of Gold revealed to be the most robust one. A method to estimate the uncertainties of the reconstructed solution was developed, and the reliability of the unfolding techniques has been tested extensively. Accordingly, the unfolding of the measured shower size spectrum is well prepared now and will be presented in the following.

After some preparatory work in Section 7.1, where things like the optimal iteration depth and the uncertainties are determined, in Section 7.2, the measured dataset will be unfolded, and the quality of the solution will be judged. Thereafter, the significance of the structures observed in some spectra will be evaluated (cf. Section 7.3). Finally, the compatibility between the results obtained in this thesis and those obtained by other analyses will be examined (see Section 7.4).

### 7.1 Some Preparations for the Unfolding of the Measured Data

The first step in the unfolding analysis of the measured dataset will be to determine the optimal iteration depth for Gold's iterative algorithm (Section 7.1.1), and to compute the uncertainties imposed to the final solutions (Section 7.1.2).

#### 7.1.1 The optimal iteration depth

The procedure that was used in Section 6.3.4 to determine the optimal iteration depth for the toy datasets can also be applied to the measurement without limitation, and will hence not be explained again. In Fig. 7.1, top panel, the *WMSE*, estimated according to Eq.(6.55), is shown for different templates generated by means of a preliminary unfolding of the measured shower size distribution. In the middle diagram, the important axis ranges are zoomed in, whereby uninteresting curves are omitted for a better distinguishability. Following the line of reasoning given in Section 6.3.4, the template that was unfolded with an iteration depth of  $\Delta\chi^2 = 2 \times 10^{-3}$  is the one expected to be the closest to the optimal solution of the convolution problem. In case of this template, the

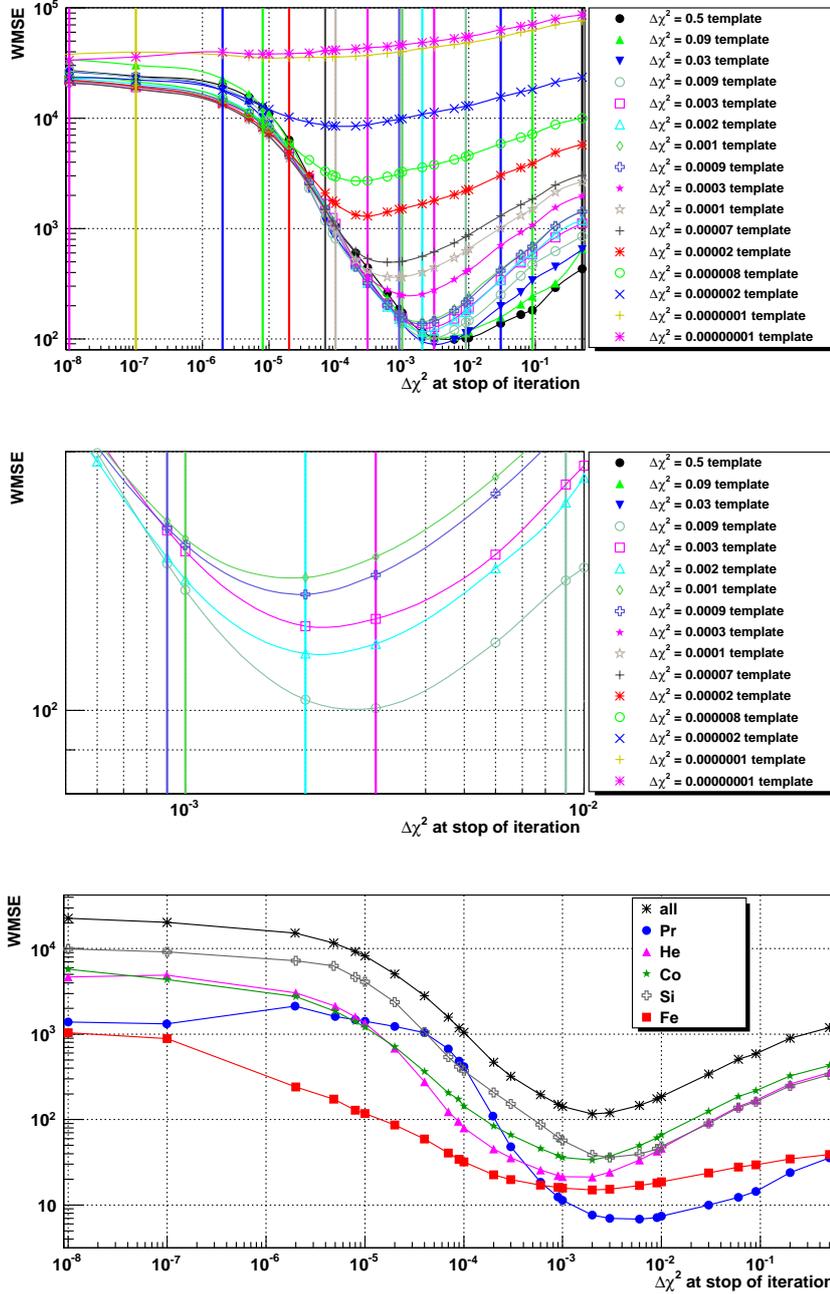


Figure 7.1: The estimated  $WMSE$  (cf. Eq.(6.55)) for different templates generated by an unfolding of the measured shower size distribution under different iteration depths (top panel). On the middle panel, the important axis ranges are zoomed in, whereby uninteresting curves are omitted. On the bottom panel is shown, exemplarily in case of the  $\Delta\chi^2 = 2 \times 10^{-3}$  template, how the individual  $WMSE$  of the primaries contribute to the sum  $WMSE$  (“all”) that is shown in the two top panels.

estimated  $WMSE$  has a minimum at around  $\Delta\chi^2 = 2.1 \times 10^{-3}$ , what agrees<sup>1</sup> with the depth  $\Delta\chi^2 = 2 \times 10^{-3}$  that was used while generating the template.

Hence, the measured dataset will finally be unfolded under this iteration depth of  $\Delta\chi^2 = 2 \times 10^{-3}$ . In Fig. 7.1, bottom panel, for the template generated with this  $\Delta\chi^2$  value is illustrated how the individual  $WMSE$  of certain primaries contribute to the sum  $WMSE$  (“all”). While the absolute values of the specific  $WMSE$  are slightly different at around  $\Delta\chi^2 = 2 \times 10^{-3}$ , what is unproblematic however, the positions of their minima are sufficiently compatible. Hence, there is no indication that whilst the unfolding process the spectra of different primaries are subjected to different convergence rates. Considering the comprehensive experiences gained by trial unfoldings of toy datasets, partly summarized in Section 6.3.5, the measured dataset does not have to be categorized to be a “problematic” one. On the contrary: A very good-natured solution can be expected in case of the KASCADE-Grande measurement. Since now the optimal iteration depth is found, the uncertainties imposed to the final solution can be computed.

### 7.1.2 Error propagation

The elemental energy spectra derived by an unfolding of the shower size distribution measured with KASCADE-Grande will suffer from four different sources of uncertainties, as described at the beginning of Section 6.3.3. The statistical uncertainties due to the limited measurement time as well as the bias caused by the unfolding algorithm itself can be estimated according to Eqs.(6.44) and (6.45). The systematic uncertainty due to the uncertainty in the response matrix, caused by the limited Monte Carlo statistics, can be estimated by the two additional “extreme” response matrices introduced in Section 6.3.3, which base on either slower, or faster decreasing tails of the used parametrizations in comparison to the “standard” case. As also explained in the aforementioned section, the fourth source of uncertainty, i.e. the interaction models used for the simulations, cannot be analysed in detail in this work. Hence, all results have to be interpreted under the premise that especially the used model QGSJET-II-02 describes the physics of hadronic interactions at these energies with a high level of reliability.

The computation of the statistical uncertainties as well as of the systematic uncertainties caused by the limited Monte Carlo statistics is straightforward, such that the result will immediately be shown in the final energy spectra in terms of error bars and error bands, without presenting further calculations. The estimation of the bias induced by the unfolding algorithm itself needs an intermediate step however. As shown in Section 6.3.3, this bias can overall be estimated correctly, while in some individual exceptional energy bins the estimation could fail however. It was stated that a bin-wise computation, as done for the other sources of uncertainties, will not be possible in case of the bias. Instead, it is more convenient to conservatively estimate the maximal range of the relative bias.

In Fig. 7.2, the estimated relative bias for individual primaries as a function of the number of entries in the estimated solution vector<sup>2</sup> is depicted. Above values of  $\log_{10}(N^{\text{rec}}) = 1.4$  (indicated by the vertical dashed line), what corresponds to around  $N = 25$  air shower events and hence to 25 entries in the related energy spectrum, the relative systematic bias can be bracketed overall as done in the figure by the two dashed horizontal lines. On the right panel, the axis of ordinates is scaled-down for a

<sup>1</sup>Reminder: The minimum in the estimated  $WMSE$  belonging to a certain template should be at the same, or a slightly larger  $\Delta\chi^2$  value than that used to generate the template. This was the criterion for which of the templates is expected to be the correct one (cf. Section 6.3.4).

<sup>2</sup>I.e. as a function of the number of showers  $\log_{10}(N^{\text{rec}})$ .

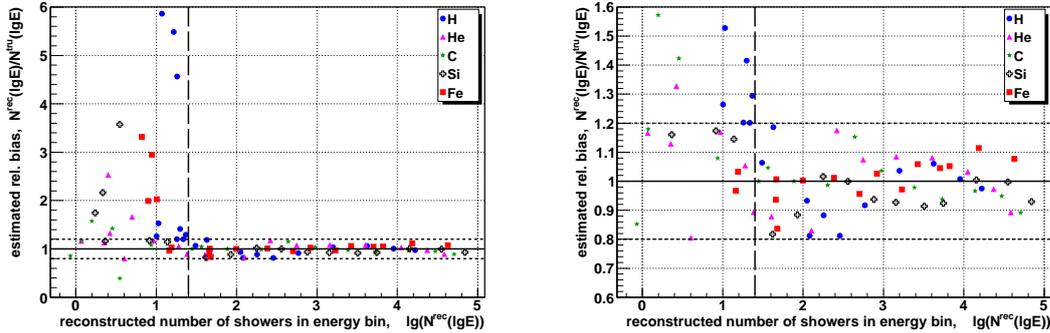


Figure 7.2: The estimated relative bias for different primaries in case of the measured dataset, which was unfolded with the optimal iteration depth of  $\Delta\chi^2 = 2 \times 10^{-3}$  (cf. Section 7.1.1). On the right panel, the range of the axis of ordinates is scaled-down for reasons of a better distinguishability.

better distinguishability. By this, a common relative bias of 20% can be estimated for  $\log_{10}(N^{\text{rec}}) > 1.4$ .

For smaller numbers of entries in the solution vector, i.e. in case of less than 25 air shower events, the relative bias gets huge. This is mainly caused by the positive definiteness of Gold's algorithm, what results in a reconstruction of always non-zero flux entries, often quite far away from zero, even when the true flux is zero. This results in an overestimation of the flux, and hence in the discernible large relative bias. As shown in Section 6.3.3, it is more appropriate to assume a fixed absolute bias, instead of a relative one, below that threshold. Based on the aforementioned deliberations and the depicted estimated relative bias, an absolute one of  $b = 25$  can be defined, if  $\log_{10}(N^{\text{rec}}) \leq 1.4$ . For the final depictions of the unfolded elemental energy spectra, the relative bias of 20% can simply be converted to respective absolute values for the specific energy bins. Then, the computed absolute bias will be shown together with the systematic uncertainty that was caused by the uncertainty in the response matrix in terms of a common resulting error band.

As the optimal iteration depth is known, and the expected uncertainties are discussed, the measured shower size distribution can be unfolded now.

## 7.2 Unfolding of the KASCADE-Grande Dataset

In this section, the key result of this thesis will be presented: The energy spectra for elemental groups of cosmic rays. The shower size distribution measured with KASCADE-Grande is unfolded by means of Gold's algorithm with an iteration depth of  $\Delta\chi^2 = 2 \times 10^{-3}$ . For the sake of completeness, it is emphasized that the results are obtained based on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68].

Firstly, the reconstructed fluxes will be presented (Section 7.2.1). Thereafter, the consistency and the quality of the obtained spectra will be judged (Section 7.2.2). Finally, some preliminary results based on the high energy interaction model EPOS 1.99 [188] will be discussed (Section 7.2.3).

### 7.2.1 Energy spectra for elemental groups of cosmic rays

Fig. 7.3 comprises the unfolded energy spectra for elemental groups of cosmic rays, represented by protons, helium, and carbon nuclei (top panel) as well as by silicon and iron nuclei (bottom panel). The all-particle spectrum is also shown. The differential flux values of the unfolded energy spectra and their uncertainties are listed in Appendix H, Table H.1 to H.6.

The error bars represent the statistical uncertainties caused by the limited measurement time. The error bands mark the maximal range of systematic uncertainties, considering both the bias caused by Gold's algorithm as well as the uncertainty caused by the one in the parametrizations used for the response matrix. Thereby, for energy bins with more than  $N = 25$  contributing air shower events, a constant relative bias of 20% (cf. Section 7.1.2) is assumed. The uncertainty caused by the parametrizations depends more strongly on the primary energy: while at lower energies the shower fluctuations are larger, and hence the uncertainty in their description is more important, at higher ones the tails of the distributions are additionally only rarely covered by events due to the small statistics, such that the parametrization uncertainties are even less important there. This was already illustrated by Fig. 6.14. The uncertainties due to the parametrizations (caused by the limited Monte Carlo statistics) are dominating at energy ranges up to  $\log_{10}(E/\text{GeV}) \approx 7.4$  for lighter mass groups, or up to  $\log_{10}(E/\text{GeV}) \approx 7.8$  for heavier ones. That is why in Fig. 7.3 the error bands get larger at lower<sup>3</sup> energies.

The demand for a minimum detection probability (cf. Section 6.3.5) of around 40% to 50% determines which is the smallest shown energy bin for a specific primary. The all-particle spectrum is the bin-wise sum of all five individual spectra; hence, the effect of the correlated biases at the very first energy bins for individual primaries is compensated to a certain degree, such that the all-particle spectrum can be shown up to slightly smaller energies. However, as will be discussed in Section 7.4.1 whilst a comparison of the all-particle spectra based on KASCADE and KASCADE-Grande, there are indications that the first shown all-particle flux value (at  $\log_{10}(E/\text{GeV}) \approx 6.9$ ) is overestimated to a certain extent. Hence, in most of the further depictions, this first data point is omitted. The right axis limit for the shown spectra at around  $\log_{10}(E/\text{GeV}) \approx 8.6$  is determined by the all-particle as well as the iron spectrum: above this energy, these unfolded spectra are unreliable, as tagged by the large error bars and error bands there. The spectra of the lighter mass groups are subject to larger uncertainties already at smaller energies due to their comparatively smaller fluxes. The energy ranges tagged by these large uncertainties should be treated with extreme caution, since no physically meaningful interpretation is possible there.

The abundance of the respective primaries seems to increase successively from protons, over helium, carbon, silicon up to iron nuclei. Hence, the heavier cosmic ray mass groups seem to be the most abundant ones in the observed energy range. This is confirmed by the depiction of the mean logarithmic atomic mass in Fig. 7.4. The markers represent the mean logarithmic mass derived based on the elemental spectra shown in Fig. 7.3. Up to energies of  $\log_{10}(E/\text{GeV}) \approx 8.0$  the composition seems to get steadily heavier. However, at this energy, there seems to be a change in composition in direction back to lighter elements.

---

<sup>3</sup>And also larger at the highest energies, since there not a constant relative bias, but a constant absolute one is assumed if the number of showers contributing to a specific energy bin is less than  $N = 25$ . The unfolded solution is poor at these energies. But, by means of the large assumed absolute bias, these ranges are successfully tagged to be insignificant.

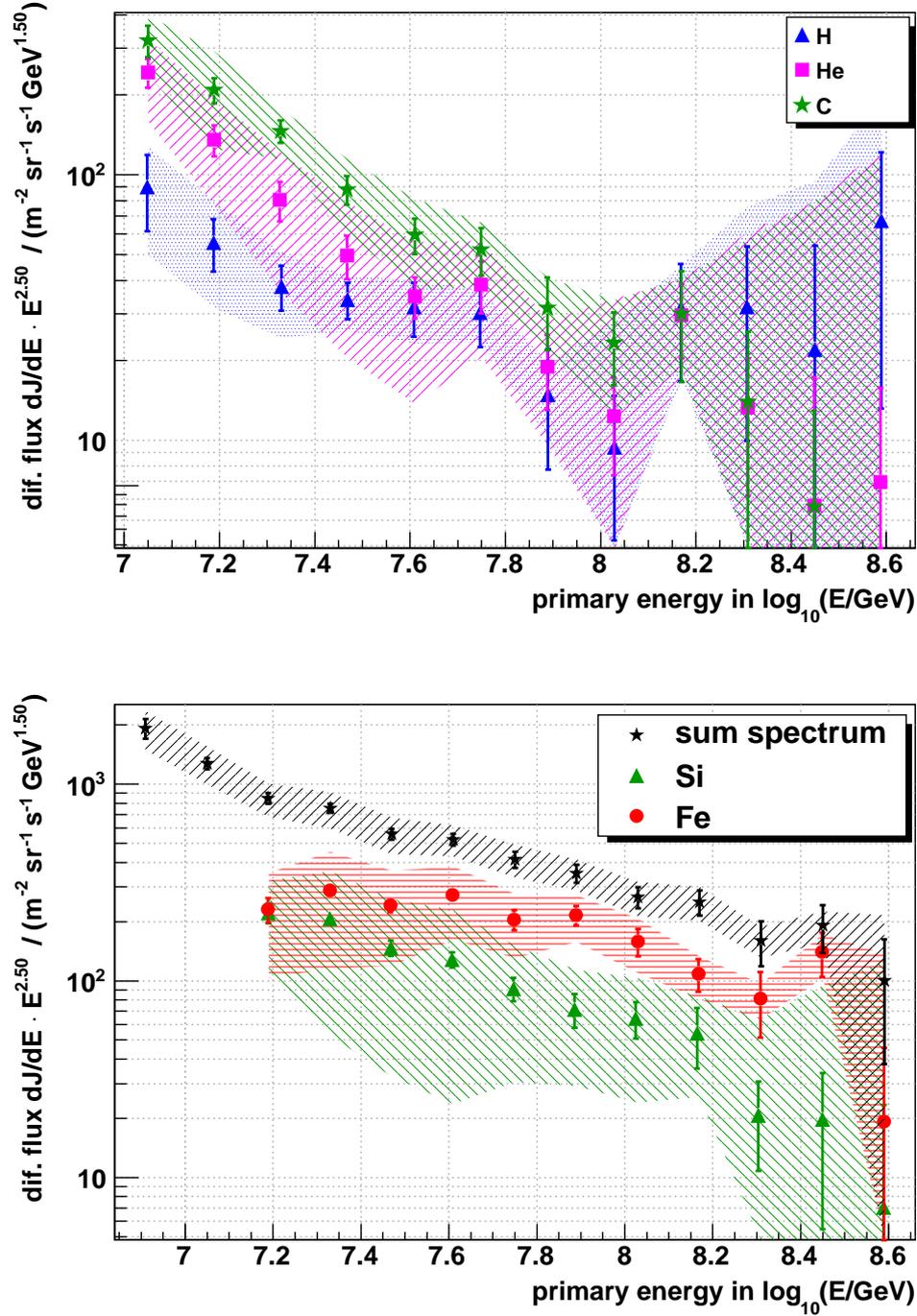


Figure 7.3: The unfolded energy spectra for elemental groups of cosmic rays, represented by protons, helium, and carbon nuclei (top panel) as well as by silicon and iron nuclei (bottom panel). The all-particle spectrum that is the sum of all five individual spectra is also shown. The error bars represent the statistical uncertainties, while the error bands mark the maximal range of systematic uncertainties. The result bases on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68].

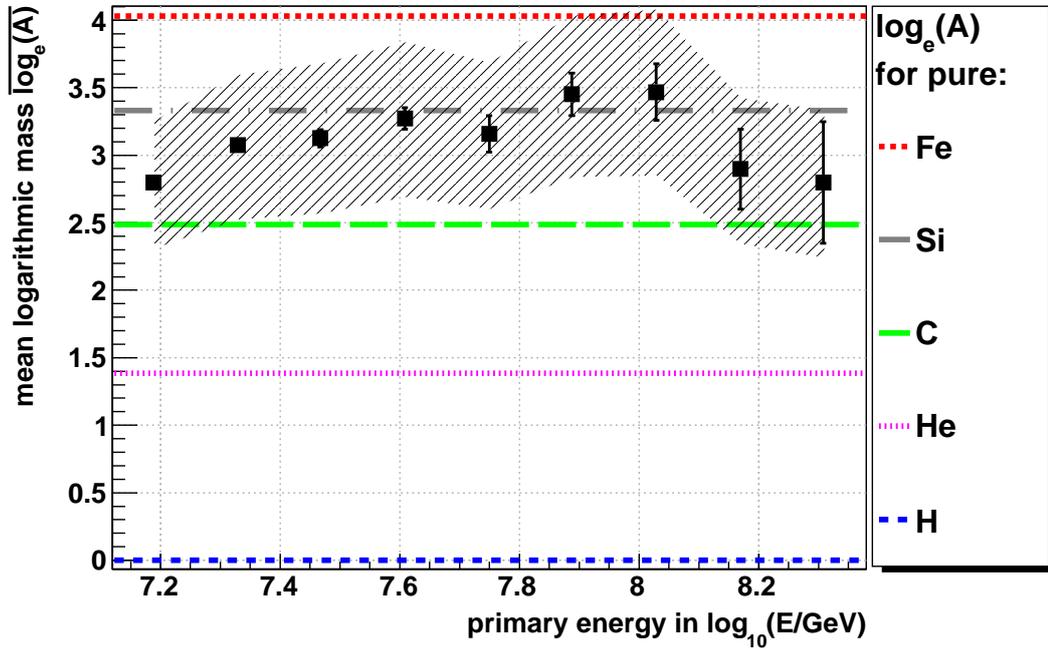


Figure 7.4: The development of the mean logarithmic mass (black markers) that was derived based on the elemental spectra shown in Fig. 7.3 in dependence on the primary energy. The error bars represent the statistical uncertainties, the error band the maximal range of the systematic ones. The horizontal dashed lines represent the logarithmic atomic masses of specific nuclei (see legend).

Investigating the results depicted in Fig. 7.3 further shows that the spectra of protons, as well as of helium, carbon, and silicon nuclei are rather featureless, if one only takes into account the energy ranges that are not tagged to be meaningless by large uncertainties. But, the spectra of protons as well as of helium nuclei seem to get harder above  $\log_{10}(E/\text{GeV}) > 7.4$  respectively  $\log_{10}(E/\text{GeV}) > 7.6$ . However, these parts of the spectra have to be interpreted with caution since the statistics is poor there, as reflected by the large error bars. As mentioned in Section 6.3.5, Gold’s algorithm tends to pull up the spectra at ranges of low statistics due to its positive definiteness. Hence, although the obtained results give slight indications for a hardening especially in the spectrum of protons, this cannot be stated with absolute certainty. From the statistical point of view, none of the spectra of protons, helium, carbon, or silicon, as well as the all-particle spectrum shows a significant structure.

Things are different in case of iron nuclei. The flux of this most abundant primary can be reconstructed over a wide energy range with only small uncertainties. There is a steepening observable in the spectrum at around  $\log_{10}(E/\text{GeV}) \approx 7.9$  to  $\log_{10}(E/\text{GeV}) \approx 8.0$ . The significance of this “knee-like” structure will be judged in Section 7.3.1. The relative increase of the flux of iron nuclei at around  $\log_{10}(E/\text{GeV}) \approx 8.45$  has to be interpreted with caution. For one thing, the statistical uncertainties are already large there, for another thing, the immediately succeeding energy bins are already tagged to be insignificant by large systematic uncertainties. Hence, beside the possibility that this increase at that single energy bin is simply a statistical effect, it could also be caused by a

systematic artificial effect of the algorithm; e.g. due to the positive definiteness of Gold's algorithm, since the number of contributing air showers to that energy bin is  $N = 49$ , and hence already very small. As shown in Fig. 7.5, this relative flux increase is primarily visible in case of the two iterative unfolding methods (Gold, Bayes' theorem), while the spectrum of iron nuclei derived by the regularized unfolding based on the principle of maximum entropy is rather structureless at this energy range. Hence, this could be an indication that this increase is an impact of the used algorithm. However, if the flux of iron nuclei really suffers from such a systematic distortion at this energy, the performed estimation of the bias would have failed to a certain extent at this energy bin. The number of entries that one would expect if this data point would follow the trend of the predecessors is round about  $N \approx 20$ . To explain the deviation by a systematic effect, the relative bias would be rather 40% instead of the assumed 20% at this specific energy bin. Such a single underestimation is possible however, as discussed in Section 6.3.3, since the bracketing of the "estimated" relative biases in order to get an estimate for the maximal range of uncertainty works reliably in many cases, but could indeed fail to cover a few single "true" relative biases. This was especially observed in case of smaller statistics. As a conclusion, it cannot be decided whether or not the increase of the flux of iron nuclei at around  $\log_{10}(E/\text{GeV}) \approx 8.45$  is a real physical feature, even though there are more indications that it is not. Hence, in the following analyses, this data point as well as all subsequent ones will be neglected.

To summarize, the preliminary expectation of Section 5.3.3 (cf. Fig. 5.35 and 5.36) that the composition of cosmic rays is dominated by the heavy component is confirmed now. While the spectra of protons, as well as of helium, carbon, and silicon nuclei are featureless within the scope of uncertainties, the flux of iron nuclei seems to be characterized by a steepening at around  $\log_{10}(E/\text{GeV}) \approx 7.9$  to  $\log_{10}(E/\text{GeV}) \approx 8.0$ . However, concerning protons, it might be possible that its spectrum recovers above  $\log_{10}(E/\text{GeV}) > 7.4$ . Before the significance of the knee-like structure in the spectrum of iron nuclei will be analysed in more detail, or comparisons with other results are performed, first the quality of the solution will be examined in the next section.

## 7.2.2 Consistency and quality of the result

In order to review the reliability of the solution, in a first step, the spectra unfolded with Gold's algorithm can be compared to those derived with the other introduced unfolding algorithms. By this, a potential systematic impact of the applied technique can be ruled out. However, since the true energy spectra are unknown, it is hard to judge about the quality of the derived solution. The only expedient will be to compare the results with the measured shower size distribution by means of a forward folding, to get at least an impression about the quality of the data description, even though this is not exactly the same as the quality of the solution itself.

### Cross-check with other unfolding algorithms

For cross-check purposes, the shower size distribution measured with KASCADE-Grande is unfolded with the two additional methods introduced also in Section 6.2: The iterative algorithm using the Bayes' theorem, as well as the regularized unfolding based on the principle of maximum entropy. the optimal iteration depth ( $\Delta\chi^2 = 4 \times 10^{-3}$ ) respectively the appropriate regularization parameter ( $\tau = 900$ ) have been determined based on the *WMSE*, as already done in case of Gold's technique.

Fig. 7.5 comprises the energy spectra for elemental groups of cosmic rays, represented by protons, helium, and carbon nuclei (top panel) as well as by silicon and iron nuclei (bottom panel). The all-particle spectrum is also shown. The markers (connected by straight lines to guide the eyes) are the solution derived by Gold's unfolding method, while the dashed lines represent the spectra unfolded by the techniques based on Bayes' theorem or the principle of maximum entropy (see legend for details). For a better distinguishability, the error bars represent the statistical uncertainties of the solution derived by Gold's algorithm only, while the systematic error bands are omitted entirely<sup>4</sup>. All results base again on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68].

In case of the all-particle spectrum, all three results agree perfectly. The solutions for the spectra of individual particles derived by the two iterative procedures conform very well, too. Only the method based on the principle of maximum entropy yields spectra farther away from those of the two iterative algorithms. As stated in Section 6.3.5, the quality of this regularized technique was already found to be poorer in case of trial unfoldings based on toy datasets, such that this result is not surprising. Nevertheless, at least the courses of the fluxes agree tendentially. Furthermore, in all three unfolded spectra of iron nuclei, the steepening at around  $\log_{10}(E/\text{GeV}) \approx 7.9$  to  $\log_{10}(E/\text{GeV}) \approx 8.0$  is likewise observable, such that there is no indication that this knee-like feature is an artificial effect of the applied unfolding algorithm itself.

As a conclusion, the solution based on Gold's algorithm has been confirmed by means of two additional unfolding methods. Especially the knee-like structure in the spectrum of iron nuclei is observable independently from the applied unfolding technique.

### Quality of data description

Since in case of the analysis of the measured shower size distribution the true solution is unknown, the quality of the solution itself cannot be reviewed immediately. In Section 6.3.1, a chi-square test (cf. Eq.(6.39)) was introduced in order to check the quality of the description of the measured data by the unfolded solution. It was emphasized that by this test the quality of the solution itself can only be judged peripherally. For sure, the unfolded solution should be able to reproduce the measured data sample to a certain extent. However, the solution is not wanted to reproduce any random fluctuation in the data sample. Nevertheless, to check the quality of the description of the measured data by the unfolded solution, i.e. that the solution agrees in principle with the measured dataset, is the only expedient.

The chi-square test (cf. Eq.(6.39)) between the forward folded solution (Fig. 7.6, top right panel) and the measured shower size distribution (Fig. 7.6, top left panel) yields a reduced chi-square value of 0.5 (chi-square probability of agreement is 100%). An additional Kolmogorov-Smirnov test results in a probability of 97% for a compatibility of both distributions. Hence, in principle, the unfolded solution agrees with the actually measured dataset. Overall, the quality of the data description is high.

However, it has to be checked whether the small difference between both distributions is caused by statistical excesses in the measured dataset, which are neglected successfully by Gold's algorithm, or if there are indications for systematic problems, like e.g. a deficit description of the physics of hadronic interactions at these energies by the used interaction models. In this context, one has to call to mind that a forward folding introduces a smoothing (cf. explanations in Section 4.2), such that the forward folded solution will

<sup>4</sup>Anyway, the uncertainties are only slightly different for the applied unfolding techniques.

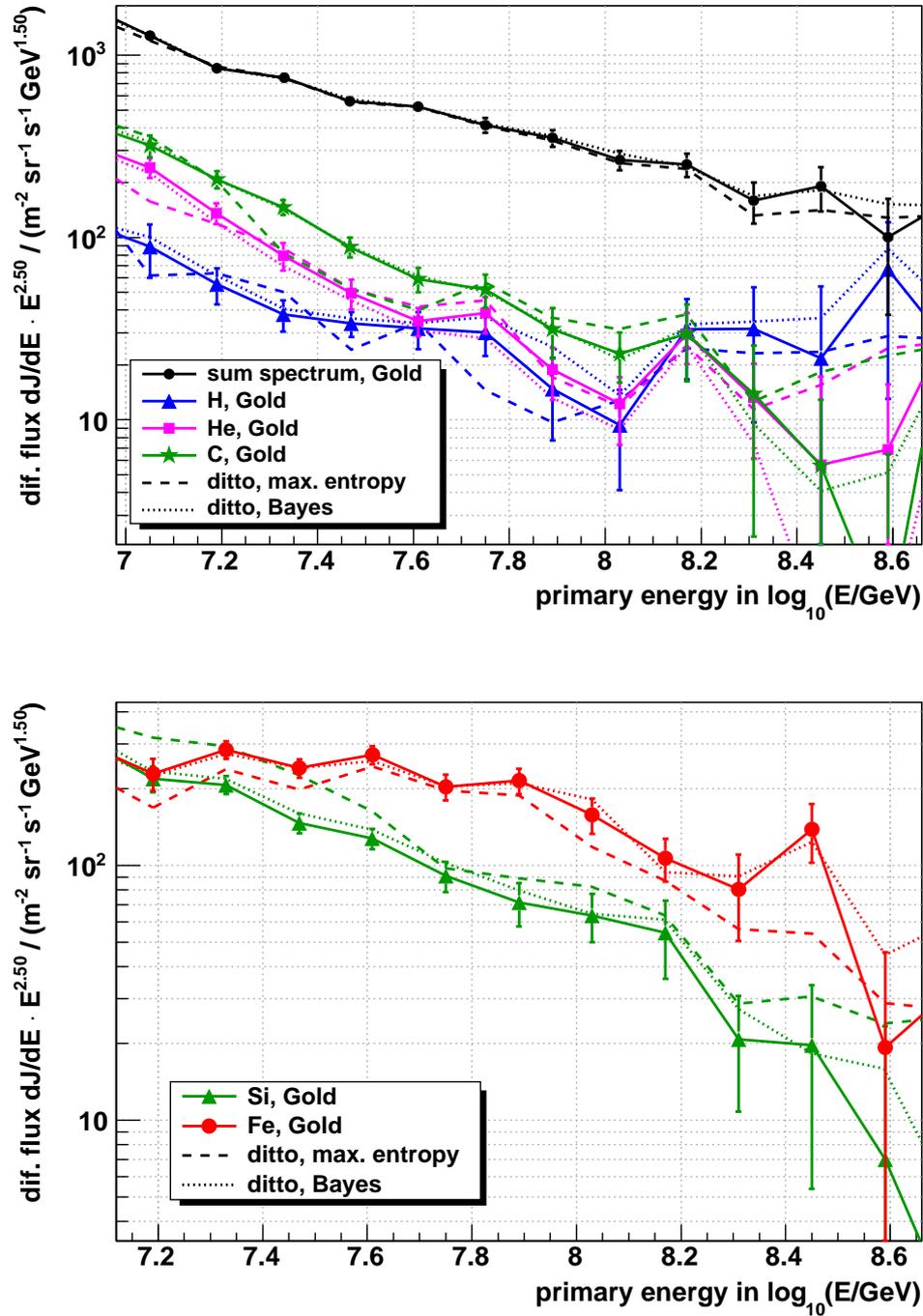


Figure 7.5: The energy spectra for elemental groups of cosmic rays, represented by protons, helium, and carbon nuclei (top panel) as well as by silicon and iron nuclei (bottom panel). The all-particle spectrum is also shown. The markers (connected by straight lines to guide the eyes) are the solution derived by Gold's unfolding method, while the dashed lines represent the spectra unfolded by the techniques based on Bayes' theorem or the principle of maximum entropy (see legend for details). The error bars represent the statistical uncertainties of the solution derived by Gold's algorithm. All results base on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68].

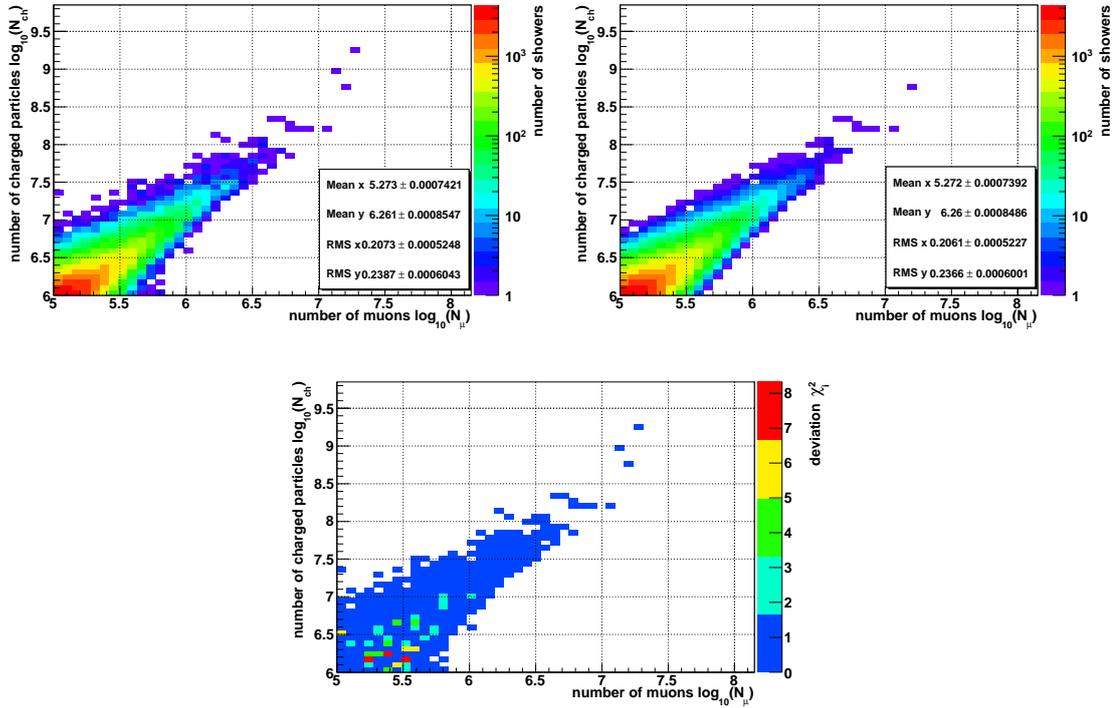


Figure 7.6: The shower size distribution measured with KASCADE-Grande (top left) and the one derived by a forward folding of the unfolding solution (top right). On the bottom panel, the distribution of the  $\chi_i^2$ -deviations (that are the  $M$  summands  $\chi_i^2$  of Eq.(6.39) that contribute to the chi-square value, where  $M$  is the dimension of the data vector) between these two shower size planes is depicted.

naturally be smoother than the measured data sample. The forward folded solution reflects rather a data sample with high statistics<sup>5</sup>, and hence with less fluctuations than in case of the KASCADE-Grande dataset measured in a few years only, and thus with very limited statistics. More precisely, it has to be checked in the following if the forward folded solution agrees with the measured dataset in the framework of statistics, albeit such an agreement is already highly indicated by the large chi-square and Kolmogorov-Smirnov probabilities.

For this purpose, the distribution of the  $\chi_i^2$ -deviations (that are the  $M$  summands  $\chi_i^2$  of Eq.(6.39) that contribute to the chi-square value, where  $M$  is the dimension of the data vector) between the two shower size planes is examined more in-depth (cf. Fig. 7.6, bottom panel). Overall, there seems to be a good agreement between the forward folded solution and the measurement ( $\chi_i^2$  values at around 1 or less). Only a few cells exhibit larger  $\chi_i^2$  values. This is confirmed by the distributions comprised in Fig. 7.7. There, different one-dimensional slices of the two two-dimensional distributions depicted in Fig. 7.6 are shown: for different fixed charged particle number (left panels) respectively muon number (right panels) intervals<sup>6</sup>, as denoted in the legends. Additionally

<sup>5</sup>What would be the ideal case: An experiment with unlimited statistics, which does not suffer from fluctuations, such that the “true” energy spectra can be reconstructed without the need of strong regularization.

<sup>6</sup>In this exemplary depictions, only the interesting intervals are covered that are corresponding to smaller or intermediate energies, as only there large deviations are discernible (cf. Fig. 7.6, bottom

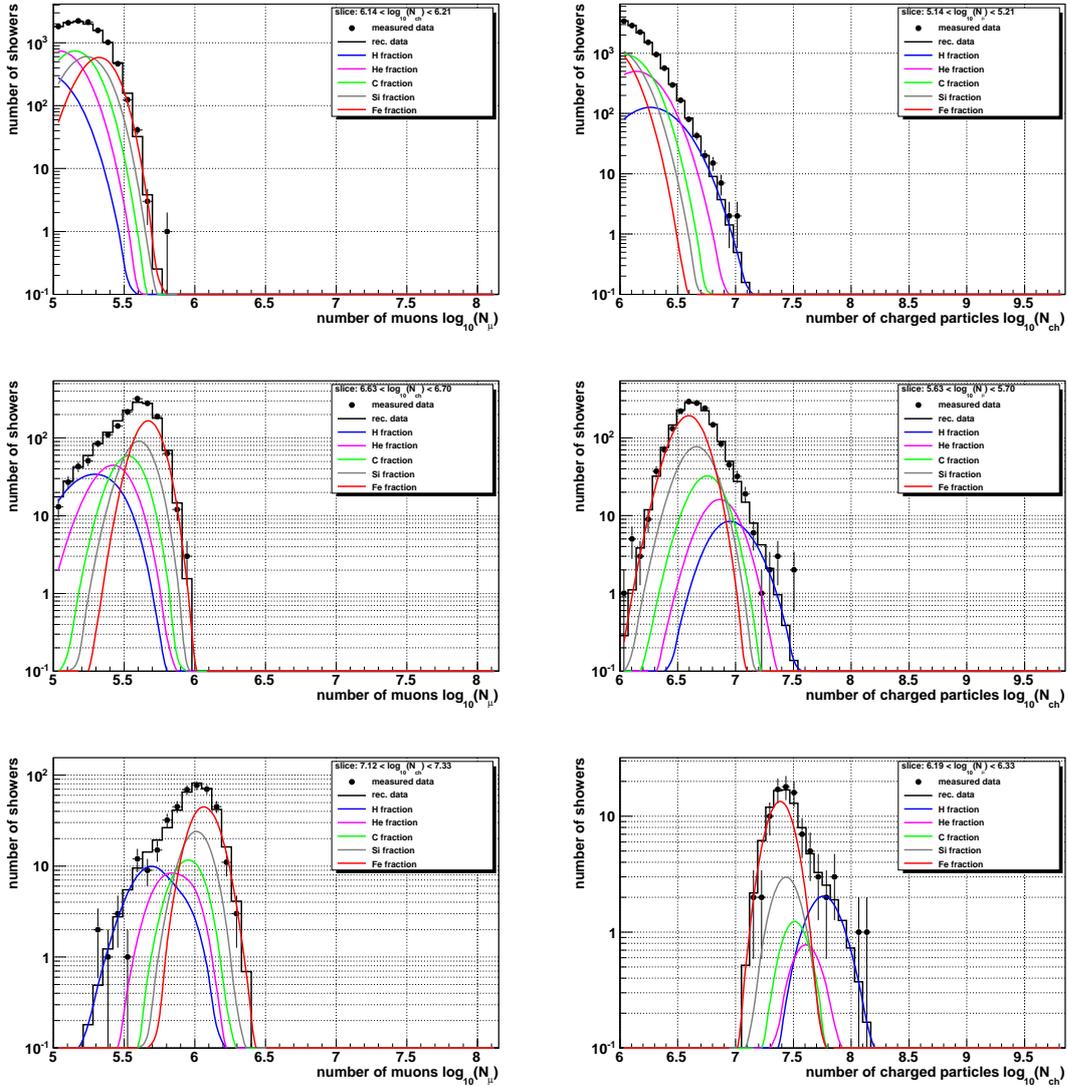


Figure 7.7: Comparison between the measured dataset (“measured data”) and the reconstructed one (“rec. data”) that was derived by a forward folding of the unfolded solution. Shown are different slices of the two-dimensional distributions depicted in Fig. 7.6: for different fixed charged particle number (left panels) respectively muon number (right panels) intervals, as denoted in the legends. Additionally shown are the contributions of the individual primaries to the reconstructed dataset.

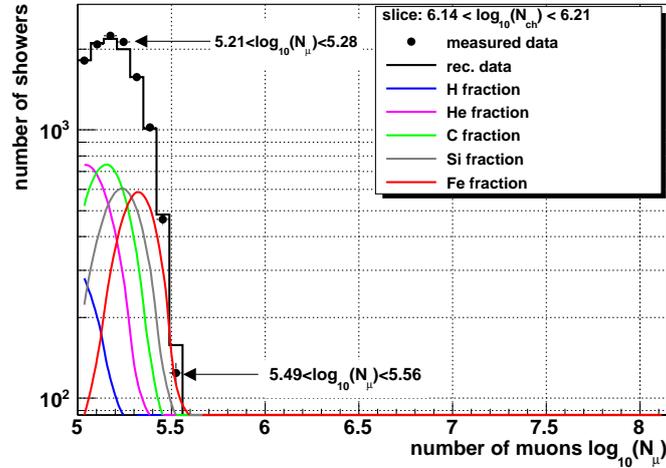


Figure 7.8: The distribution already shown in Fig. 7.7, top left panel, now zoomed in in order to better point out the deviations between the measured dataset and the reconstructed one.

shown are the contributions of the individual primaries to the reconstructed dataset. In general, there is a good agreement between the measurement and the forward folded solution. Furthermore, especially from the right panels, one can deduce that a dominating heavy component is needed in order to describe sufficiently the left tail of the measured distribution. However, also lighter components are still needed at these energy ranges in order to describe the right tail.

Exemplarily, the excesses observable in Fig. 7.6 (bottom panel) in the interval  $6.14 < \log_{10}(N_{\text{ch}}) < 6.21$  at the bins  $5.21 < \log_{10}(N_{\mu}) < 5.28$  or  $5.49 < \log_{10}(N_{\mu}) < 5.56$  will now be examined. For this purpose, the range of the y-axis of the distribution shown in Fig. 7.7, top left panel, is reduced in order to better point out the respective deviations. The result is depicted in Fig. 7.8. The two muon number bins of this slice where these larger excesses are observable in the chi-square distribution are marked. In general, the measured distribution is described reliably. Taking into account that the description is of high quality for almost all cells of the shower size plane, these single larger deviations are less likely of physical origin, but rather compatible with random statistical excesses due to the limited measurement time. Due to the smoothing effect of the forward folding, the reconstructed dataset does not exhibit such excesses.

As a conclusion, the unfolded energy spectra for elemental groups of cosmic rays enable a very good description of the measured shower size distribution. The few statistical excesses in the data sample are ignored successfully by the unfolding procedure. This results in very small isolated deficits in the description of the measured data by the unfolded solution, however, guarantees an optimal non-oscillating solution what is the primary goal of the analysis. There are no indications that the used interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68] fail to describe the physics of hadronic interactions at these energies reliably<sup>7</sup>.

panel).

<sup>7</sup>What, however, does not necessarily mean that the used models are 100% correct.

### 7.2.3 Preliminary result based on EPOS 1.99

It was emphasized in Section 6.3.3 that the unfolded solution will suffer from four sources of uncertainties. While the uncertainties caused by the limited measurement time, the limited Monte Carlo statistics, as well as the bias of the unfolding algorithm can be estimated reliably, the uncertainty that originates from a possibly wrong description of the physical processes in the air shower development by the used interaction models cannot be estimated immediately, since the truth is not known. The reliability of the error estimation cannot be reviewed as done in case of the other sources by the comparison between “true” and “estimated” uncertainties. However, the influence of different models on the final result can be examined in order to get a first impression about the model dependence; but, without knowing which of the models is the right one, or at least close to truth. Only accelerator experiments enable an independent test of the hadronic interaction processes considered in the different interaction models. Nevertheless, the comparison between the measured dataset and the one predicted by the solution (forward folding) can at least give first indications whether or not the used models seem to describe the physical processes in a right way.

In order to get a rough impression about the model dependence of the unfolded solution, the unfolding procedure is repeated in a very preliminary analysis based on the high energy interaction model EPOS 1.99 [188]. After some words about the preparatory steps for the EPOS 1.99 analysis, the results will be presented and compared to the QGSJET-II-02 solution. Deficits in the data description by EPOS 1.99 will be discussed, and preliminary suggestions for a possible improvement of that model will be given.

### Preparations

Since the low energy interaction model (here FLUKA 2002.4) is less important for the final result<sup>8</sup>, only the influence of the high energy interaction models (in the main analysis QGSJET-II-02 [144, 145]) has to be examined. A change of that model means that the whole unfolding analysis<sup>9</sup> has to be repeated again based on the new model. Since this is very time consuming, only a rather quick cross-check based on the high energy interaction model EPOS 1.99 [188] has been performed for this thesis. This model is widely used in actual analyses, such that it was selected for the first cross-check. In principle the same steps as already done in case of the main analysis based on QGSJET-II-02 have been performed. However, in case of the new model EPOS 1.99 only very quickly, while in case of the standard model of this thesis every step was reviewed very carefully. Hence, all EPOS based results shown in this section are only preliminary results! Accurate and final analyses based on different high energy interaction models are actually in work and will be presented by our collaboration soon.

Since things like detector efficiency and resolution do not depend<sup>10</sup> significantly on the model used to simulate the air shower, only the part of the response matrix that contains the parametrization of the air shower development has to be re-parametrized using EPOS. The used EPOS 1.99 simulation set is comparable to the QGSJET-II-02 set

<sup>8</sup>As already the analyses based on KASCADE measurements have proved [12].

<sup>9</sup>Beginning from the parametrizations of the air shower development, the computation of the response matrix and its conditioning, test of the reliability of the applied unfolding algorithms in case of the current model, determination of the optimal iteration depth, computation of uncertainties etc.

<sup>10</sup>Or at least should not depend on the used shower development simulation code, what has to be checked in case of a final EPOS analysis. However, in first preliminary checks, this appeared to be confirmed.

(cf. Section 5.1.1) used for the main analysis of this work. The intrinsic shower fluctuations are parametrized for EPOS analogous to those for QGSJET-II-02 in Section 5.1. The remaining parametrizations for the efficiency and the reconstruction uncertainties are inherited from the QGSJET-II-02 response matrix. Finally, a new response matrix, which bases on EPOS 1.99 and FLUKA 2002.4, can be computed and used for an unfolding analysis.

A first comparison between the preliminary parametrizations based on the EPOS 1.99 response matrix and the measured dataset is depicted in Fig. 7.9 and 7.10. In the first mentioned figure, the measured shower size distribution (grey isoareas) and the most probable  $\log_{10}(N_{\mu}^{\text{rec}}) - \log_{10}(N_{\text{ch}}^{\text{rec}})$  values according to the response matrix (markers) that bases on EPOS 1.99 and FLUKA 2002.4 are illustrated for different particles and primary energies. While in case of QGSJET-II-02 (see Fig. 5.36) the most probable values for the heavy nuclei (silicon, iron) are located in the central region of the measured distribution (where it has many entries), in case of EPOS 1.99 the most probable values for the intermediate mass group (represented by carbon nuclei) are matching this central region. That gives first indications that in case of EPOS 1.99 the reconstructed composition of cosmic rays would be lighter than in case of QGSJET-II-02.

In Fig. 7.10, again the measured shower size plane is shown, but now in comparison to some isolines representing the cells  $(\log_{10}(N_{\text{ch}}^{\text{rec}}), \log_{10}(N_{\mu}^{\text{rec}}))_i$  of the data plane with constant probability (from the inner<sup>11</sup> to the outermost isoline: 0.1, 0.05 and  $10^{-4}$  probability density). One can see that the widths of the distributions are of the same order as those in case of QGSJET-II-02, such that again a simple consideration of the maxima will not result in a reliable solution. Hence, also in case of EPOS 1.99, the entire probability distributions have to be taken into account and used whilst the deconvolution procedure.

### Unfolded energy spectra based on EPOS 1.99

The measured  $\log_{10}(N_{\mu}^{\text{rec}}) - \log_{10}(N_{\text{ch}}^{\text{rec}})$  shower size distribution is unfolded by means of Gold's algorithm and using the response matrix based on the interaction models EPOS 1.99 [188] and FLUKA 2002.4 [24, 67, 68]. The optimal iteration depth<sup>12</sup> for this purpose was found to be  $\Delta\chi^2 = 2 \times 10^{-4}$  (determined based on the *WMSE*).

Fig. 7.11 comprises the unfolded energy spectra for elemental groups of cosmic rays, represented by protons, helium, and carbon nuclei (top panel) as well as by silicon and iron nuclei (bottom panel). The all-particle spectrum that is the bin-wise sum of all five individual spectra is also shown. The error bars represent the statistical uncertainties caused by the limited measurement time. The systematic uncertainties<sup>13</sup> are not yet shown in this preliminary result.

Comparing this preliminary result to that derived by means of the interaction models QGSJET-II-02 and FLUKA 2002.4 (cf. Fig. 7.3) reveals some significant differences between both solutions. The spectra of the lighter mass groups, represented by protons

<sup>11</sup>In case of smaller energies, the widths of the probability distributions are as large that there are no individual probabilities larger than 0.1 or even 0.05, such that the inner isolines are missing in these cases.

<sup>12</sup>The determined optimal iteration depth differs from that used in the unfolding based on the QGSJET-II-02 model. It will be discussed later on that EPOS 1.99 does not need iron and silicon nuclei to describe the measurement. Hence, the unfolding using the EPOS model is primarily based on only three primaries, instead of five. This could be one reason, why in case of EPOS 1.99 Gold's algorithm has different convergence rates than in case of QGSJET-II-02.

<sup>13</sup>The bias of the unfolding algorithm is again at around 20% to 30%, while the systematic uncertainties due to that in the response matrix, caused by the limited Monte Carlo statistics, are not yet computed.

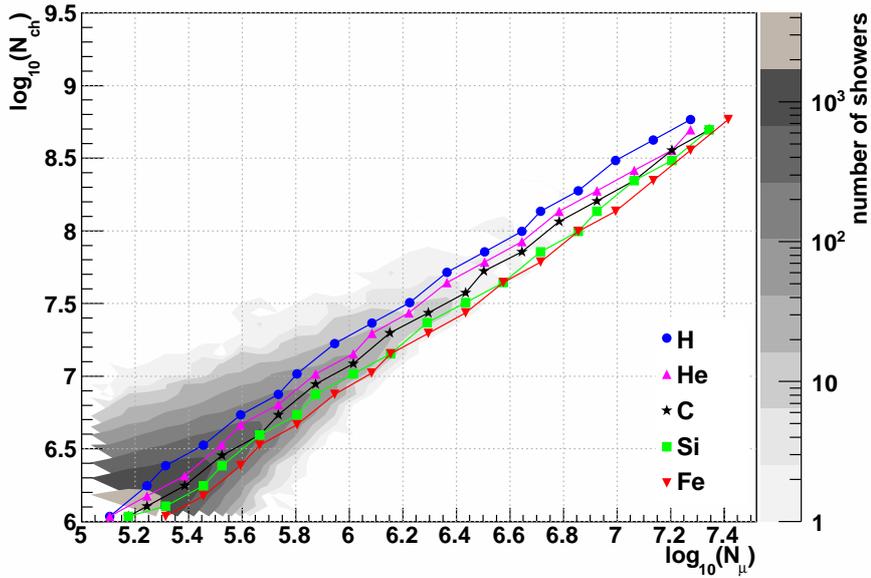


Figure 7.9: A comparison between the measured shower size distribution (grey isolines with filled areas in between) and the most probable  $\log_{10}(N_{\mu}^{rec})-\log_{10}(N_{ch}^{rec})$  combinations according to the preliminary parametrizations, illustrated for different particles and based on EPOS 1.99 and FLUKA 2002.4. The corresponding primary energies are out of the  $\log_{10}(E/\text{GeV})$  interval 6.98 to 9.5, which is divided into 18 bins of width 0.14 (the first bin, which covers the energy bin  $\log_{10}(E/\text{GeV})=6.98$  to 7.12, corresponds to the first markers shown in the bottom left corner). To guide the eyes, the markers are connected with straight lines.

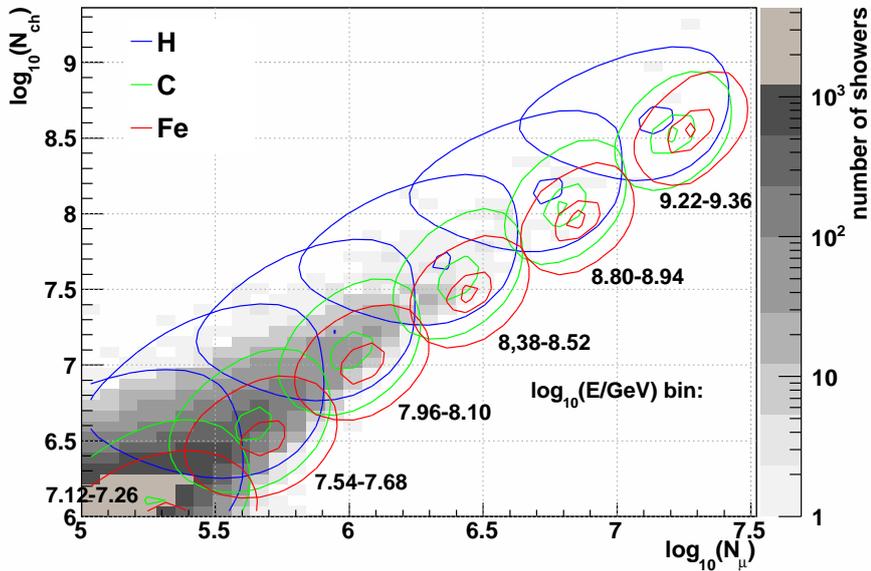


Figure 7.10: A comparison between the measured shower size distribution (grey histogram) and some isolines with  $\log_{10}(N_{\mu}^{rec})-\log_{10}(N_{ch}^{rec})$  combinations of constant probability according to the preliminary parametrizations based on EPOS 1.99 and FLUKA 2002.4. This is illustrated exemplarily for protons, carbon as well as for iron nuclei, and in case of six energy bins (labelled below each isoline set). Each isoline set corresponds, from the inner to the outermost line, to 10%, 5% and 0.01% probability.

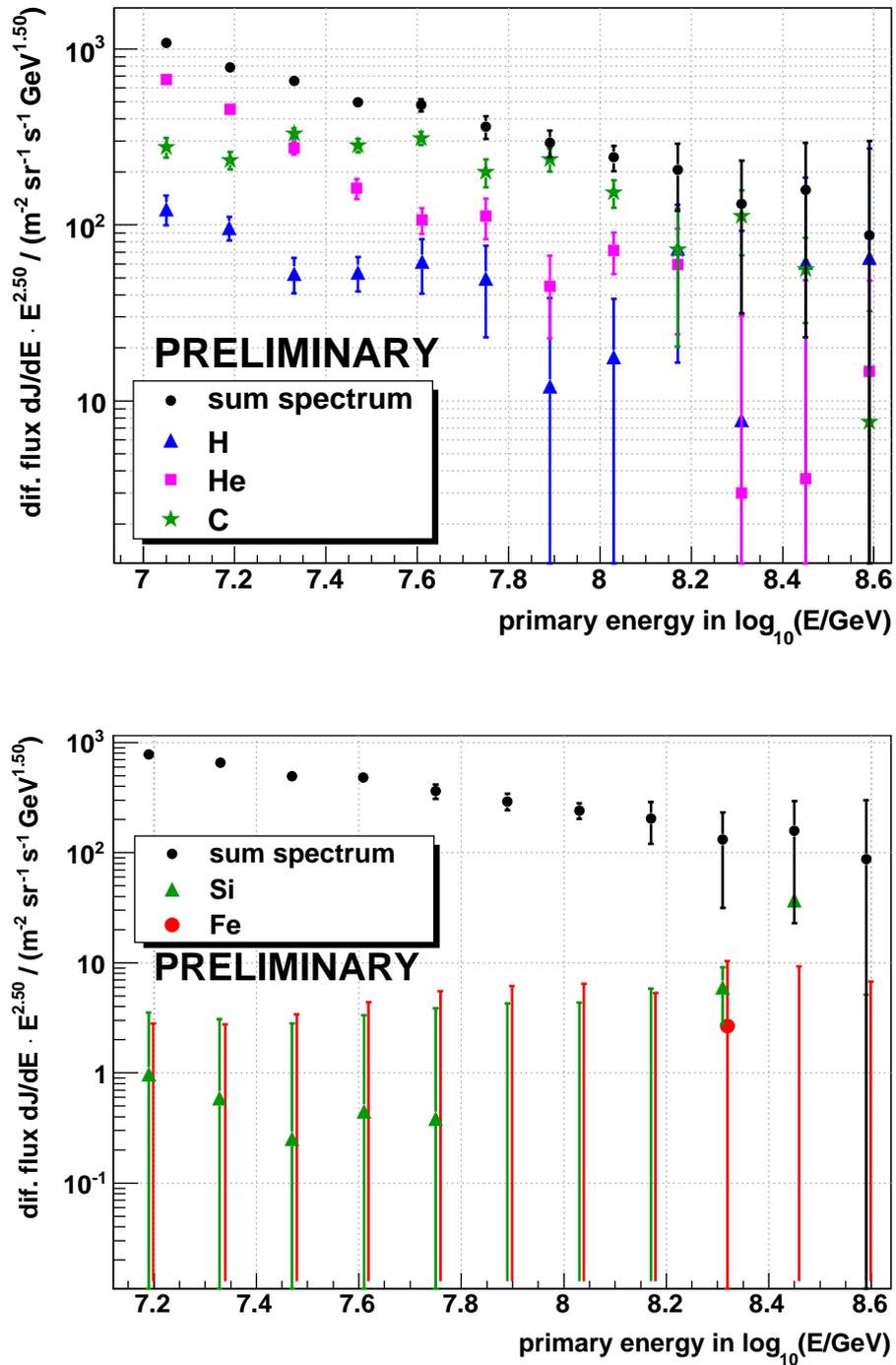


Figure 7.11: The unfolded preliminary energy spectra for elemental groups of cosmic rays, represented by protons, helium, and carbon nuclei (top panel) as well as by silicon and iron nuclei (bottom panel; the spectrum of iron nuclei is slightly shifted along the x-axis for a better discriminability). The all-particle spectrum that is the sum of all five individual spectra is also shown. The error bars represent the statistical uncertainties, while the systematic uncertainties are not yet shown. The result bases on the interaction models EPOS 1.99 [188] and FLUKA 2002.4 [24, 67, 68].

and helium nuclei, are still very similar (only the absolute normalizations, i.e. the relative abundances, are different). Especially the impression that the proton spectrum gets harder beyond  $\log_{10}(E/\text{GeV}) > 7.4$  seems to be confirmed by the EPOS result<sup>14</sup>. However, the spectrum of the intermediate mass group, represented by carbon nuclei, shows a significantly different structure in comparison to the QGSJET-II-02 result. The differential flux of carbon nuclei is characterized by a rather small power law index of around  $\gamma \approx 2.5$  up to energies of around  $\log_{10}(E/\text{GeV}) \approx 7.9$  to  $\log_{10}(E/\text{GeV}) \approx 8.0$ . Beyond this energy, the spectrum seems to get steeper. Such a knee-like feature is not observable in the spectrum of carbon nuclei derived using QGSJET-II-02. It seems that this structure present in the data is assigned to iron nuclei in case of QGSJET-II-02, and to carbon nuclei in case of EPOS 1.99. Since all EPOS 1.99 results are only preliminary and EPOS 1.99 has some deficits in the data description (as shown later), the significance of this structure has to be interpreted with caution and will not be analysed in more detail in this work. In case of the heavier mass groups, represented by silicon and iron nuclei, the results are completely different from that obtained in case of using QGSJET-II-02. There is only a vanishingly low flux of silicon nuclei observable, while iron is even less important. In truth, both fluxes are presumably compatible with zero, and are only determined to be non-zero due to the positive definiteness of Gold's algorithm. This result was already expected, as discussed in the context of Fig. 7.9. In case of EPOS 1.99, the cosmic ray composition is dominated by the intermediate or the lighter mass groups. The contents of the spectra of silicon and iron nuclei unfolded based on QGSJET-II-02 seem, in case of EPOS 1.99, to be assigned to the lighter mass groups, especially to the one represented by the spectrum of carbon nuclei. Due to the very low abundance of iron nuclei, in the context of the given measurement statistics of KASCADE-Grande no statement about a possible iron-knee can be given. This does not necessarily mean that there is not one, but only that the KASCADE-Grande experiment is not able to measure it if the high energy interaction model EPOS 1.99 describes the air shower development reliably. However, it should again be emphasized that this conclusion is preliminary.

In Fig. 7.12, a preliminary comparison between the all-particle spectra derived by an unfolding based on the high energy interaction models EPOS 1.99 or QGSJET-II-02 in combination with the low energy interaction model FLUKA 2002.4 is illustrated. For a better distinguishability, only the statistical (error bars) and systematic (error band) uncertainties in case of QGSJET-II-02 are shown. Both spectra show the same slope and structures. This is again an indication that on the one hand individual primaries have different abundances for different interaction models. But, on the other hand, specific structures in the data seem to be conserved, however, with an assignment to different primaries in dependence on the used model. Regarding the absolute normalization, the EPOS 1.99 flux is constantly shifted slightly to smaller values in comparison to the QGSJET-II-02 one. The deviation is within the given error band, i.e. within the systematic uncertainties of the QGSJET-II-02 solution. It could be caused by a systematic bias<sup>15</sup> of the algorithm itself, or by the uncertainties in the response matrices. However, it could also be a real physical property: a different mean energy assignment by the models. As one can derive from a comparison between Fig. 5.36 and 7.9, in case of EPOS 1.99, at a fixed energy, the air showers exhibit on average more muons, but less charged particles (and hence less electrons) than in case of QGSJET-II-02. That means that in case of EPOS 1.99 more energy will be assigned to muons, and less energy to electrons (if

<sup>14</sup>However, also there without statistical significance.

<sup>15</sup>That is maybe differently large in case of unfolding the data based on the EPOS 1.99 or QGSJET-II-02 response matrices.

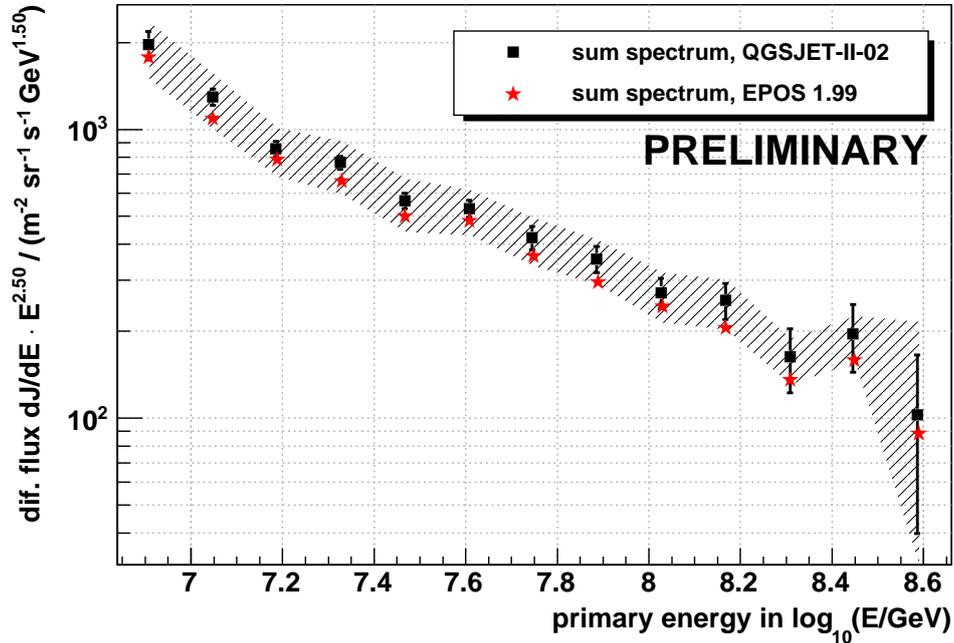


Figure 7.12: A preliminary comparison between the all-particle spectra derived by an unfolding based on the high energy interaction models EPOS 1.99 [188] or QGSJET-II-02 [144, 145] in combination with the low energy interaction model FLUKA 2002.4 [24, 67, 68]. For a better distinguishability, only the statistical (error bars) and systematic (error band) uncertainties in case of QGSJET-II-02 are shown.

the assignment per particle is not different, too). Such effects can finally influence the absolute scale of the all-particle spectra.

### Quality of data description by EPOS 1.99

As already impossible in case of QGSJET-II-02, also in case of EPOS 1.99 the quality of the unfolded solution cannot be judged immediately. The only expedient was to examine the quality of the data description by the solution and to demand an agreement within statistics (cf. Section 7.2.2). By this, a first indication is given whether or not the used interaction model seems to have serious deficits in the description of the physics of hadronic interactions at these energies.

In case of EPOS 1.99, a Kolmogorov-Smirnov probability of only 69% for an agreement between the measured dataset and the one computed by a forward folding of the unfolded solution was achievable. Even with very deep iteration depths, which yield a less biased<sup>16</sup> solution, and therefore describe the data sample better, no higher probability than 75% was achievable. Consequently, EPOS 1.99 seems to have problems to describe the shower development reliably.

To examine the deficits in more detail, in Fig. 7.13, top panel, the distribution of the  $\chi_i^2$ -deviations (that are the  $M$  summands  $\chi_i^2$  of Eq.(6.39) that contribute to the chi-square value, where  $M$  is the dimension of the data vector) between the measured dataset and

<sup>16</sup>But, a more oscillating, and hence meaningless solution.

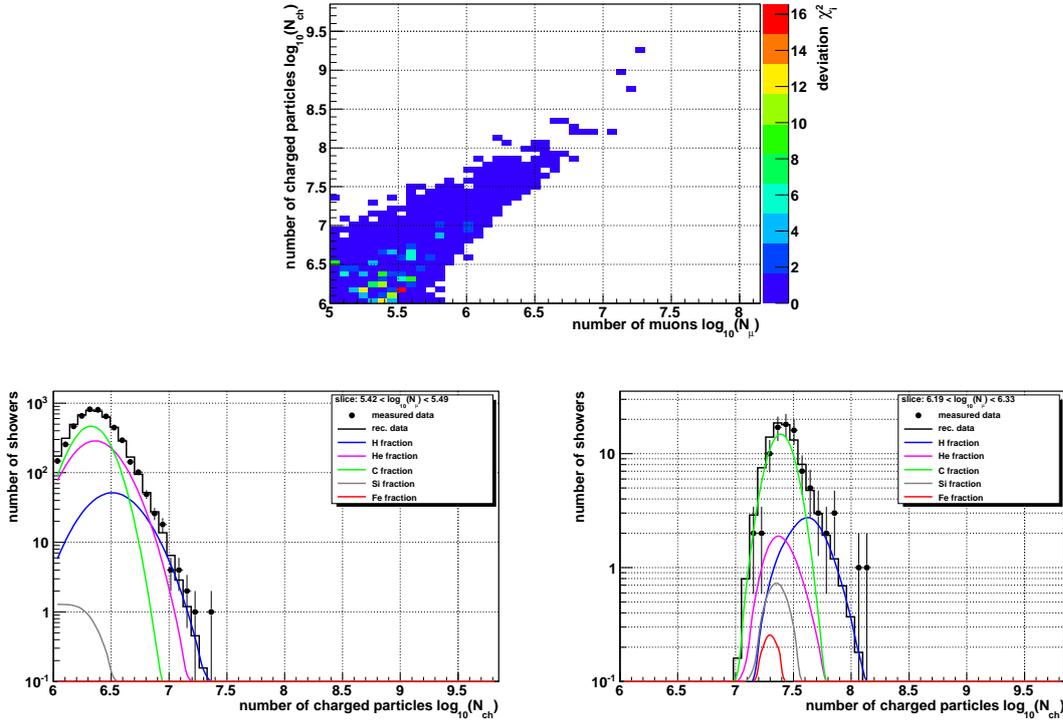


Figure 7.13: On the top panel, the distribution of the  $\chi_i^2$ -deviations (that are the  $M$  summands  $\chi_i^2$  of Eq.(6.39) that contribute to the chi-square value, where  $M$  is the dimension of the data vector) between the measured shower size plane and the one computed by a forward folding of the EPOS 1.99 solution is depicted (note the different z-axis scaling between this depiction and that in Fig. 7.6, bottom panel). On the bottom panels, a comparison between the measured dataset (“measured data”) and the reconstructed one (“rec. data”) that was derived by the forward folding of the unfolded solution is performed: Shown are different slices of the two-dimensional distributions for different fixed muon number intervals, as denoted in the legends. Additionally shown are the contributions of the individual primaries to the reconstructed dataset.

the one derived by a forward folding of the EPOS 1.99 solution is illustrated. Comparing this distribution to that for QGSJET-II-02 (see Fig. 7.6, bottom panel; note the different scaling of the z-axis!) reveals significantly larger deviations in case of EPOS 1.99. They are primarily located at lower energies and next to the heavy edge<sup>17</sup> of the distribution.

In Fig. 7.13, bottom panel, one-dimensional slices through the respective two-dimensional shower size distributions are examined, exemplarily for the two fixed muon number<sup>18</sup> intervals  $5.42 < \log_{10}(N_\mu) < 5.49$  and  $6.19 < \log_{10}(N_\mu) < 6.33$ . One can see that already the lighter and intermediate mass groups basically allow to describe the data in the main, without the need of a dominant heavy component. More precisely, in case of the shown muon number interval  $5.42 < \log_{10}(N_\mu) < 5.49$ , a heavy component (especially iron nuclei) is not allowed at all. The lighter primaries are already needed in such high

<sup>17</sup>As can be seen from Fig. 7.9, the upper edge (called “light edge”) of the distribution corresponds rather to the lighter mass groups, while the lower edge (called “heavy edge”) corresponds the heavy ones.

<sup>18</sup>Slices of fixed charged particle number intervals yield comparable results, and hence are not shown additionally.

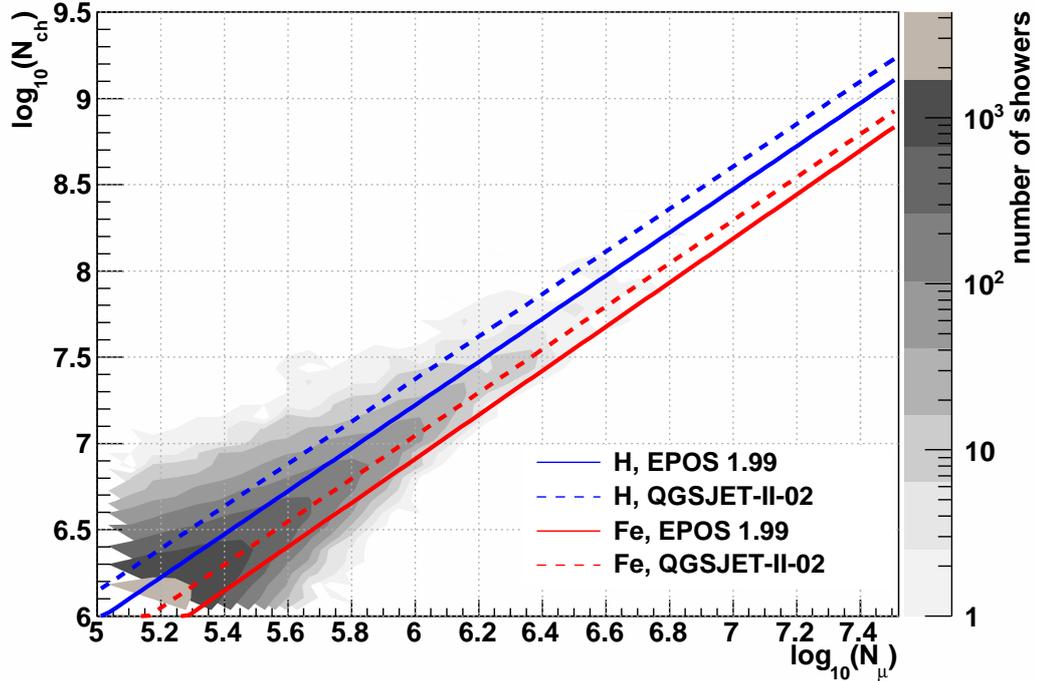


Figure 7.14: A comparison between the measured shower size distribution (grey isolines with filled areas in between) and the energy dependence of the most probable  $\log_{10}(N_{\mu}^{\text{rec}}) - \log_{10}(N_{\text{ch}}^{\text{rec}})$  combinations according to the parametrizations, illustrated by the straight lines for different particles and based on EPOS 1.99 respectively QGSJET-II-02 [144, 145] in combination with FLUKA 2002.4.

abundances in order to describe the right tail reliably that at the left tail this causes a systematic overestimation of the measured dataset by the reconstructed one. An additional heavy component would further deteriorate the description. Such an overestimation at the left tail can also be seen for the neighbouring muon number intervals, such that these deviations are not compatible with single statistical excesses (as it was the case for QGSJET-II-02), but are caused by systematic effects. The overestimation corresponds to the deviations observable for smaller energies at the heavy edge (cf. Fig. 7.13, top panel). In some other intervals<sup>19</sup>, one can observe an underestimation of the right tail of the measured data by the solution. Hence, this is an interplay between sufficiently abundant lighter or intermediate mass groups in order to describe the right tail correctly, and a sufficiently description of the left tail without overestimation. Especially at the lower energies, it is not possible to fulfil both requirements without under- or overestimating the measured distribution at the respective tails.

One could interpret this deficit such that EPOS 1.99 predicts too many muons<sup>20</sup>, or too few charged particles, respectively electrons, than the data would require. This can better be explained based on Fig. 7.14. There, a comparison between the measured shower size distribution (grey isolines with filled areas in between) and the energy dependence of the most probable  $\log_{10}(N_{\mu}^{\text{rec}}) - \log_{10}(N_{\text{ch}}^{\text{rec}})$  combinations according to the parametriza-

<sup>19</sup>In the actually depicted interval only slightly visible for the last four or six shown bins.

<sup>20</sup>What is surprising, since most of the other experiments recognize rather a muon deficit in the interaction models like QGSJET-II-02 (cf. e.g. [58, 192]), and hence it was actually the idea that larger muon numbers are needed.

tions, illustrated by the straight lines for different particles and based on EPOS 1.99 respectively QGSJET-II-02 [144, 145] and FLUKA 2002.4, is depicted. The lines of maximal probability in case of a certain primary for QGSJET-II-02 and EPOS 1.99 are not exactly parallel. Their distance seems to be larger in case of smaller energies, i.e. at energies where EPOS 1.99 seems to have some deficits. If the EPOS 1.99 lines would be shifted towards higher charged particle numbers (or to smaller muon numbers) at these energies, such that they are finally more parallel to the QGSJET-II-02 lines, one would already expect an improvement of the data description in case of EPOS 1.99. Since the maximal probability lines for the lighter and intermediate mass groups would get shifted farther away from the centre of the measured shower size plane, their abundance would be smaller, while also the heavy component would begin to be important. The large imbalance between light and heavy primaries that did not allow to describe both the right and left tails of the one-dimensional distributions (Fig. 7.13, bottom panel) equally well would be reduced to a certain degree. Another possibility to improve the data description in case of EPOS 1.99 could be to assume smaller intrinsic shower fluctuations: the distributions needed to describe the right tails and belonging primarily to the light or intermediate mass groups would be narrower, and hence would less affect the left tail. However, the aforementioned possible optimizations are rather educated guesses than assured facts. They base on simplified interpretations of the deficits of EPOS 1.99 in the data description. The physics of hadronic interactions at these energies is nontrivial and to a great extent still unknown, and can only be extrapolated tentatively from results gained by particle accelerator experiments at lower energies. Hence, the afore-suggested solutions, which furthermore base on preliminary results only, could also be wrong and should be treated with care.

### Summary of the EPOS 1.99 results

Comparing the energy spectra unfolded based on the interaction model EPOS 1.99 to those based on QGSJET-II-02 reveals that both models result in comparable all-particle spectra, which are slightly shifted against each other, presumably due to a different mean energy assignment whilst the air shower development. Another reason could be that the uncertainties in the response matrix (mainly the uncertainties in the description of the tails of the shower fluctuations) or the bias caused by the unfolding algorithm itself are of different size for different high energy interaction models.

The spectra of individual mass groups are less compatible. Especially the relative abundances are very different. That different high energy interaction models yield different relative abundances of the mass groups conforms with the result based on the KASCADE unfolding analysis [9], where the interaction models QGSJET-01 and SIBYLL yielded different relative abundances of individual primaries, while the structures in the specific spectra were less model dependent. Especially the knee-like features in the spectra were observable independently from the applied model. The latter point cannot be confirmed by the KASCADE-Grande analysis. In case of QGSJET-II-02, there is a knee-like structure visible in the flux of iron nuclei. In case of EPOS 1.99, due to the very low abundance of iron nuclei going along with a very low statistics, no statement about a structure in the spectrum of iron nuclei can be given. However, the structure presumably present in the data and assigned to iron nuclei in case of QGSJET-II-02, seems to be assigned to carbon nuclei in case of EPOS 1.99.

A comparison between the measured dataset and the one obtained by a forward folding of the EPOS 1.99 solution reveals systematic deficits of this model in the data

description. Hence, the obtained energy spectra should be interpreted with caution. Since in case of QGSJET-II-02 there are no indications for serious deficits in the data description, the QGSJET-II-02 based results are considered to be reliable at the moment, while the EPOS 1.99 based results are ignored in the following analyses in this thesis. Emphasizing that the EPOS 1.99 results base on a preliminary analysis only, it can be supposed that EPOS 1.99 seems to predict too many muons respectively too few electrons. Alternatively, the intrinsic shower fluctuations could be too large in case of EPOS 1.99. However, this are rather educated guesses than assured facts.

Detailed unfolding analyses based on other high energy interaction models than QGSJET-II-02 are currently in progress and will be presented by our collaboration soon. They will replace this preliminary EPOS 1.99 analysis presented in this work and will allow to draw final conclusions about the model dependence of the result obtained in the main analysis of this work based on QGSJET-II-02. Furthermore, the Large Hadron Collider (LHC) has started operation. It is designed to collide protons at centre-of-mass energies of up to 14 TeV, corresponding to a fixed-target energy of about  $\sim 10^{17}$  eV, and hence covering the energy range observable with KASCADE-Grande to a large extent. Hence, the results of this experiment will help to tune the parameters of the high energy interaction models in the near future.

## 7.3 Analysis of the Energy Spectra – Physical Conclusions

In the last preceding section, the energy spectra for elemental groups of cosmic rays have been unfolded. Some consistency checks (cf. Section 7.2.2) yielded a good agreement between the predictions of the model QGSJET-II-02 and the measurement, such that there is no indication so far that the derived solution could be unreliable<sup>21</sup>. The unfolded flux of iron nuclei, which is representing the heavy mass group of cosmic rays, was characterized by a steepening at a primary energy of about  $\sim 10^{17}$  eV, as observable by eye. The significance of this structure will now be analysed mathematically (Section 7.3.1). Thereafter, the position of the knee-like feature will be compared to those in the spectra of protons and helium nuclei determined with the KASCADE experiment (Section 7.3.2). This allows drawing inferences about the physical background of cosmic ray acceleration. Finally, the all-particle spectrum will be analysed and its spectral index will be determined (Section 7.3.3).

### 7.3.1 Significance of the knee-like structure in the flux of iron nuclei

The unfolded flux of iron nuclei, based on the interaction models QGSJET-II-02 as well as FLUKA 2002.4 and depicted in Fig. 7.3, is characterized by a discernible structure at a primary energy of about  $\sim 10^{17}$  eV. The significance of this knee-like feature will be judged in this section.

In Fig. 7.15, left panel, again the unfolded flux of iron nuclei is shown, but now unscaled. The last two data points of the spectrum of iron nuclei, still shown in Fig. 7.3, are omitted in the current figure, since they are precarious as discussed in Section 7.2.1. The error bars represent the statistical uncertainties, while the error bands mark the maximal range of systematic uncertainties. The flux is fitted by means of a single power

---

<sup>21</sup>While a preliminary unfolding based on EPOS 1.99 revealed deficits in the data description by that model, such that the EPOS 1.99 solution is probably unreliable.

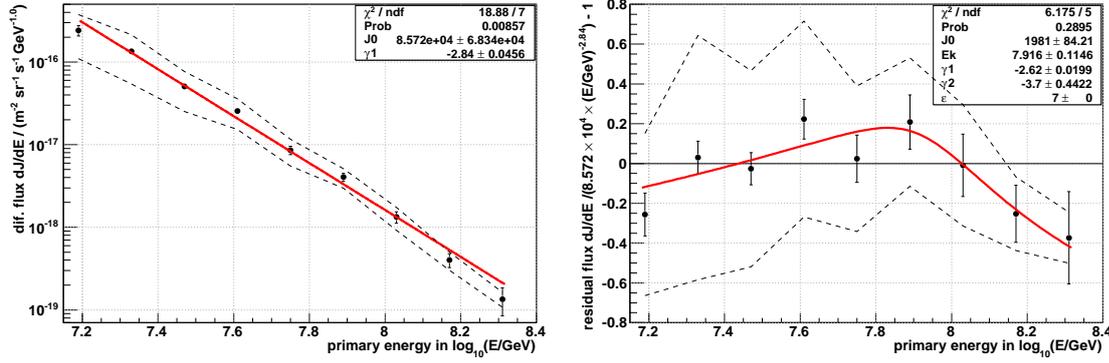


Figure 7.15: On the left panel, the unfolded flux of iron nuclei is depicted. The error bars represent the statistical uncertainties, while the error bands mark the maximal range of systematic uncertainties. Additionally, the flux is fitted by means of a single power law according to Eq.(7.1). On the right panel, the residual flux between the spectrum predicted by that power law fit and the unfolded spectrum of iron nuclei is illustrated. Additionally, the spectrum of iron nuclei is fitted by a double power law function according to Eq.(7.2).

law according to the following function:

$$\frac{dJ(E)}{dE} = J_0 E^{\gamma_1} \quad , \quad (7.1)$$

where the absolute flux  $J_0$  serves as normalization, while  $\gamma_1$  is the index of the power law. The quality of the fit is very poor however (chi-square probability of less than 1% with  $\chi^2/ndf = 18.9/7$ ). In order to better emphasize the deviations, on the right panel, the residual flux between the spectrum predicted by that single power law fit and the unfolded spectrum of iron nuclei is illustrated. Additionally, the spectrum of iron nuclei is fitted by a double power law function according to the poly-gonato function (cf. Eq.(6.41)) already introduced in Section 6.3.2:

$$\frac{dJ(E)}{dE} = J_0 E^{\gamma_1} \left( 1 + \left( \frac{E}{E_k} \right)^\varepsilon \right)^{(\gamma_2 - \gamma_1)/\varepsilon} \quad , \quad (7.2)$$

where the absolute flux  $J_0$  serves as normalization, while the power law can be a broken one with a change of index from  $\gamma_1$  to  $\gamma_2$  at the transition energy  $E_k$ . The index “k” alludes to the word “knee”, since the structure that is observed is knee-like. The smoothness of the transition between the two power laws is characterized by  $\varepsilon$ , whereby  $\varepsilon = 1$  corresponds to a smooth change over about one decade of energy, while  $\varepsilon = 4$  means a faster change within already 1/5 of a decade [95]. The latter parameter can be kept fixed to  $\varepsilon = 7$  without worsening the fit quality. The spectrum of iron nuclei can be fitted significantly better by this double power law (chi-square probability of around 30% with  $\chi^2/ndf = 6.2/5$ ). The poor fit quality in case of the single power law and the significantly better one in case of the double power law give strong indications for the presence of an “iron-knee”<sup>22</sup> at around 80 PeV. The spectral indices of the power laws are  $\gamma_1 = -2.62 \pm 0.02$  before, respectively  $\gamma_2 = -3.7 \pm 0.4$  beyond this knee. This result was confirmed by spectra of heavy primaries obtained by our collaboration in independent

<sup>22</sup>If the mass group represented in this analysis by the iron nuclei actually consists only, or at least primarily, of iron nuclei. Else, one would have to speak about the “heavy-knee” instead.

analyses, which also yield a significant knee-like steepening at about 90 PeV [13, 34] or 62 PeV [49, 50].

In this context, a short remark about the systematic uncertainties, which are shown in the depictions, but not used whilst the fit procedure, will now be given. In further tests it was found that the systematic uncertainties primarily shift the whole solution up- or downwards. This was observed for both important sources of uncertainties: the ones caused by the uncertainties in the response matrix as well as the bias induced by the unfolding algorithm itself. It is very unlikely that the systematic uncertainties cause a sharp asymmetric distortion from one bin to another. As a conclusion, the systematic uncertainties seem rather to affect the absolute overall normalizations, and thereby conserve the structures of the spectra without changing them significantly. Hence, in this rather quick analysis, the systematic uncertainties have been neglected.

Furthermore, neighbouring bins of the individual elemental energy spectra are correlated to a certain degree, among others due to internal mechanisms in the unfolding methods. Hence, in a more accurate judgement of the significance of the observed structure in the flux of iron nuclei, one would have to take into account these correlations, too. In that case, as a first guess, one would expect a slight improvement of the chi-square probability of the fits. While both the single as well as the double power law function would describe the unfolded spectrum better, also in this case the relative advantage of the double power law fit would still persist. Since, the effect of the correlations on the computed reduced chi-square values is presumably rather moderate, it is scarcely to be expected that the single power law could describe the flux of iron nuclei satisfactorily, even when the correlations are taken into account. Hence, in the framework of this quick analysis, the effect of the bin-to-bin correlations was neglected.

To summarize, if the high energy interaction model QGSJET-II-02 describes the physics of hadronic interactions at these energies with a sufficient level of reliability<sup>23</sup>, there is a strong indication for a knee-like structure at around 80 PeV caused by a change in the spectra of the heavy nuclei of cosmic rays. Furthermore, if the mass group that is represented by iron nuclei in this thesis really consists only, or at least primarily of iron nuclei, this structure can more precisely be identified as the “iron-knee”.

### 7.3.2 Comparison of knee positions

In the last section, it was shown that there is a strong indication for the presence of a knee-like structure in the energy fluxes of the heavy component of cosmic rays, represented by iron nuclei in this analysis. This was declared under the premise that the used interaction models QGSJET-II-02 and FLUKA 2002.4 are reliable. In order to draw inferences about the physical background of cosmic ray acceleration, the position of the “heavy-knee” can be compared to the positions of the knees of the lighter mass groups.

Such knee-like structures in the spectra of lighter nuclei have already been observed by the KASCADE experiment [9] based on the high energy interaction models QGSJET 01 or SIBYLL 2.1 in combination with the low energy interaction model GHEISHA 2002. For the comparison, the update [70] to this KASCADE analysis, now also based on QGSJET-II-02 and FLUKA 2002.4, is considered in the following. In Fig. 7.16, top panels, the fluxes of protons (left) and helium nuclei (right) reconstructed with KASCADE are depicted. On the bottom panel, again the flux of iron nuclei determined in this thesis and reconstructed with KASCADE-Grande is shown. All three spectra are fitted by a double power law function according to Eq.(7.2). In case of protons and iron nuclei, the

<sup>23</sup>And, up to now, none of the performed tests gave rise to the assumption that it does not.

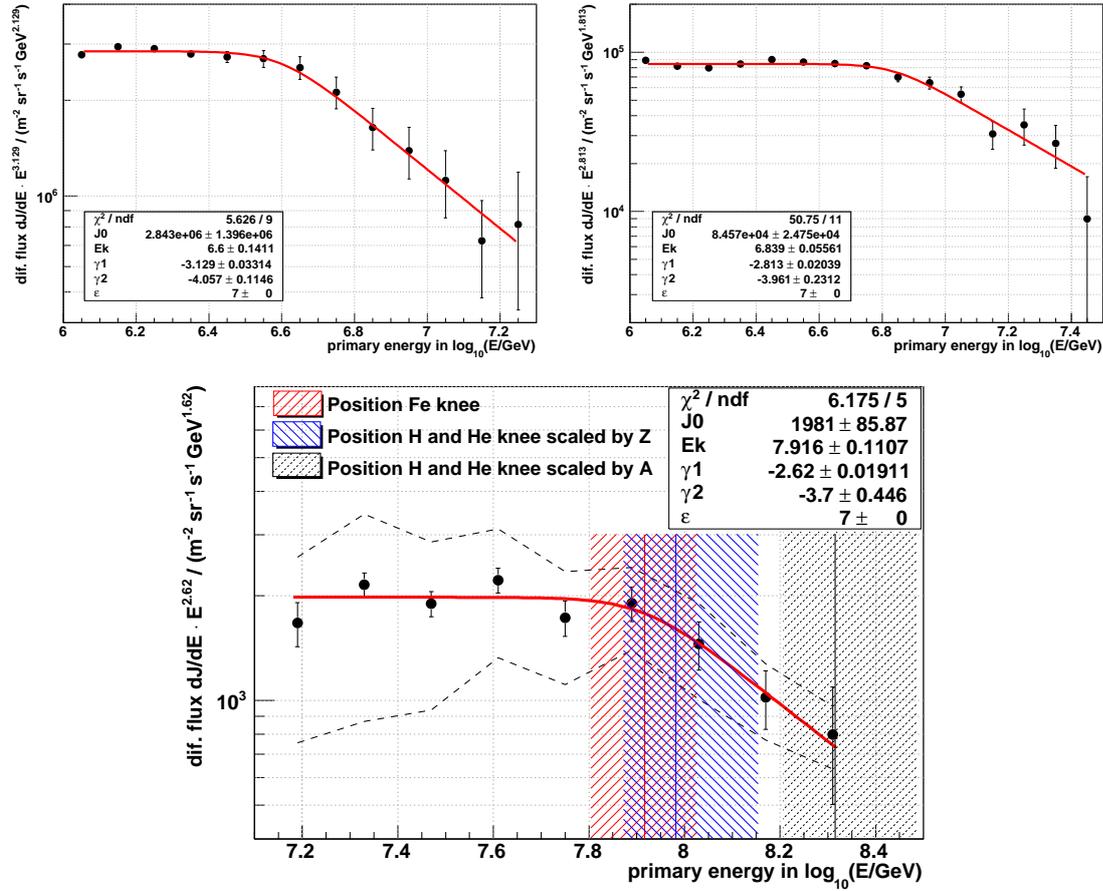


Figure 7.16: On the top panels, the fluxes of protons (left) and helium nuclei (right) are shown as they are reconstructed by the KASCADE experiment [70] based on the interaction models QGSJET-II-02 and FLUKA 2002.4. The error bars represent the statistical uncertainties. On the bottom panel, again the flux of iron nuclei determined in this thesis is shown. It is based on the aforementioned interaction models, too, but reconstructed with KASCADE-Grande (for an explanation of the vertical lines and bands, see text). All three spectra are fitted by a double power law function according to Eq.(7.2).

fit performs well, while in case of helium nuclei the reduced chi-square value indicates problems. Especially the first data points are fluctuating around the fit significantly stronger than the statistical uncertainties would allow. However, since the structure in the flux of helium nuclei at around  $\log_{10}(E/\text{GeV}) \approx 6.8$  seems to be a significant one, as one can judge by eye, and since only its position is of interest in the following, the poorness of the fit quality will be ignored.

The fits yield knee-like structures at  $\log_{10}(E/\text{GeV}) = 6.6 \pm 0.1$  in the spectrum of protons, at  $\log_{10}(E/\text{GeV}) = 6.8 \pm 0.1$  in the spectrum of helium nuclei, and at  $\log_{10}(E/\text{GeV}) = 7.9 \pm 0.1$  in the spectrum of iron nuclei. At the bottom panel, the position of the “iron-knee” and its uncertainty range are marked by the vertical red line and the shaded red band. In order to compare this knee position to those of the lighter primaries, the knee positions of protons and helium nuclei are scaled by either the factor  $Z_{\text{Fe}}/Z_{\text{H or He}}$  or  $A_{\text{Fe}}/A_{\text{H or He}}$ , where  $Z_i$  is the atomic number (the “charge”) and  $A_i$  the mass number of the respective nucleus  $i$ . In both cases, after the multiplication,

the weighted mean value (the uncertainties are the weights) of the results for protons and helium nuclei is computed. The uncertainties of the two computed mean values are estimated by conservatively assuming the uncertainties of the knee positions as the uncertainty range. The two computed mean values and its uncertainty ranges are shown additionally in the bottom panel. The vertical blue line represents the knee positions of protons and helium nuclei scaled proportional to the atomic number, while the shaded blue band marks the range of uncertainty. The grey line and band are obtained by the scaling in dependence on the atomic mass number.

One can see that under the assumption of a rigidity dependence the knee positions of heavy and light nuclei agree within the uncertainties, while a mass dependent scaling seems to be rather unlikely. In this context, it has to be emphasized again that this conclusion is only true if the used interaction models are reliable and if the mass groups represented by protons, as well as by helium and iron nuclei actually consist only, or at least primarily, of these respective primaries. While in case of protons this assumption is presumably fulfilled sufficiently (the nucleus of next following element of the periodic table, helium, is considered in the analysis, too), in case of iron nuclei things are different: The next following lighter nucleus considered in the analysis is that of silicon, what means a larger jump in the periodic table omitting several elements. Hence, in this analysis, iron has to represent also some lighter elements<sup>24</sup>. This would result in a slightly shifted knee position towards smaller energies, what could explain the small offset between the charge-scaled (with respect to pure iron nuclei) knee position of the lighter elements and that of iron nuclei. For lack of better resolution, the number of primaries that can be regarded is limited, such that no more accurate analysis is possible.

Although it is slightly off-topic, it is maybe worth to mention that the spectrum below the knee (represented by the parameter  $\gamma_1$ ) as well as beyond the knee (represented by the parameter  $\gamma_2$ ) seems to get harder with increasing mass of the primaries. However, taking into account the given uncertainties, this result is not significant. Apart from this constraint, the mass dependent hardening before the knee would agree well with the finding of the balloon-borne Cosmic Ray Energetics And Mass experiment (CREAM) [4], where, at energies below the knee, also a harder spectrum for helium nuclei in comparison to protons was observed. Based on this finding, and based on their observation of a general hardening above energies of 200 GeV/nucleon, the authors stated that their results would challenge the view that the cosmic ray spectra of individual primaries follow a simple power law below the knee. In contrast, for the behaviour above the knee, there is no simple explanation available, but it could give hints to the theory of a mixture of source populations for the galactic cosmic rays at highest energies, e.g. described in [151]. But, due to the possible non-significance of the differences between the extracted spectral indices of different primaries in this work, this point will not be discussed further, such that it will be left at that mention.

To summarize, the comparison between the knee-like structures discernible in the spectra of the light and heavy cosmic ray nuclei gives indications for a rigidity dependence of the knee position. This would encourage the cosmic ray acceleration and/or propagation models that are assuming correlations to magnetic fields. The particle physics approaches that assume connections to the mass of the nuclei seem to be less likely<sup>25</sup>. However, these findings have to be treated with caution, since they base on the one hand on the simplified assumption that the mass groups represented by the respective

<sup>24</sup>But, possibly, also heavier ones.

<sup>25</sup>If it is not a mixture of many effects, like for instance a mass dependent acceleration in combination with a charge dependent propagation.

primaries primarily consist only of these primaries, on the other hand on the supposition that the used interaction models are reliable.

### 7.3.3 Spectral indices of the all-particle spectrum

In the previous sections, the flux of iron nuclei was analysed in detail. Since the remaining elemental spectra suffer from larger uncertainties, especially due to their lower abundances, a comparable analysis like fitting these spectra with appropriate functions appears unprofitable. However, the all-particle energy flux is more stable, such that it will be investigated more in-depth.

An all-particle spectrum was already determined by our collaboration by means of a more robust method<sup>26</sup> [33], which is more native, i.e. closer to the measured data sample, and is based on a subsample with larger statistics. The residual flux between that all-particle spectrum and a single power law according to Eq.(7.1) with index  $-3.015$  is depicted in Fig. 7.17, top panel. In that analysis, the differential all-particle spectrum was fitted by different power law functions resulting in a spectral index of  $\approx -3.0$  for energies below  $\log_{10}(E/\text{GeV}) = 7.9$  to  $\log_{10}(E/\text{GeV}) = 8.0$ , and of  $\approx -3.2$  above. The significance of the change of the power law index was estimated to be at the level of 99.8%. Hence, there is a strong indication for a knee-like structure in the all-particle spectrum. The position of this structure conforms very well with the knee-like feature observed in the spectrum of iron nuclei at around  $\log_{10}(E/\text{GeV}) = 7.9$  (cf. Section 7.3.1). Hence, this indicates that the knee-like feature in the all-particle spectrum is caused by a steepening in the fluxes of the heavy cosmic ray primaries. In case of the all-particle spectrum, which was derived in this thesis by summing up all five individual elemental spectra, no significant structure can be observed<sup>27</sup> at around energies of  $\log_{10}(E/\text{GeV}) = 7.9$  to  $\log_{10}(E/\text{GeV}) = 8.0$ , as can be seen in Fig. 7.17, bottom panel. One has to keep in mind that the uncertainties of the individual elemental fluxes are propagated to this sum flux. Hence, possible structures in the all-particle flux can perhaps get lost in the larger uncertainties.

However, the all-particle flux exhibits a “concave” behaviour in the energy range from  $\log_{10}(E/\text{GeV}) = 6.9$  to  $\log_{10}(E/\text{GeV}) = 7.2$ . This “concave” structure is also observed by our collaboration based on the more robust analysis [33], which is *inter alia* based on a subsample with larger statistics, and was found to be significant. But, as will be discussed in Section 7.4.1, the first shown flux value of the all-particle spectrum determined in this work has to be interpreted with caution, since it is probably overestimated. However, ignoring this data point completely, gives still slight indications for a concave behaviour.

The all-particle flux determined in this thesis is fitted by a double power law function according to Eq.(7.2) (cf. Fig. 7.17, bottom panel). The parameter  $E_k$  corresponds to the energy at around  $\log_{10}(E/\text{GeV}) = 7.1$  to  $\log_{10}(E/\text{GeV}) = 7.2$ , where the index change takes place. In this case, it marks a “dip-position”, rather than a knee-position. The spectral index changes from  $\approx -3.8$  to  $\approx -3.1$ . This conforms well with the result given in [33], that states an index of  $\approx -3.0$  for energies beyond  $\log_{10}(E/\text{GeV}) = 7.2$  to  $\log_{10}(E/\text{GeV}) = 7.3$ , but below  $\log_{10}(E/\text{GeV}) = 7.9$  to  $\log_{10}(E/\text{GeV}) = 8.0$ .

<sup>26</sup>But, as a *quid pro quo*, do not enable to separate the elemental spectra with such a precision the unfolding technique allows.

<sup>27</sup>This was tested by means of a triple power law fit: such a fit that also contains a break in the all-particle spectrum at around  $\log_{10}(E/\text{GeV}) \approx 8.0$  does not yield better chi-square values than a double power law fit that allows a concave structure at around  $\log_{10}(E/\text{GeV}) \approx 7.2$  only, but no change in the spectrum at around  $\log_{10}(E/\text{GeV}) \approx 8.0$ .

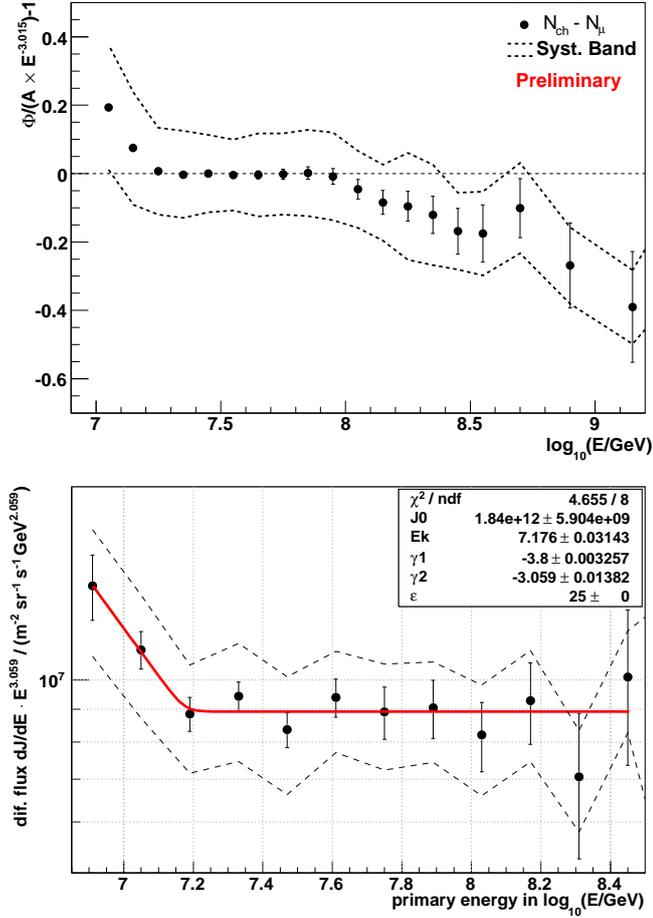


Figure 7.17: On the top panel, the residual flux between a single power law according to Eq.(7.1) with index  $-3.015$  and the all-particle spectrum derived by a more robust method by our collaboration is depicted (taken from [33]). On the bottom panel, the differential all-particle spectrum computed by summing up all five individual elemental spectra unfolded in this thesis is shown. The statistical uncertainties are given by the error bars, while the error bands mark the maximal range of systematic uncertainty. The flux is fitted by a double power law function according to Eq.(7.2).

As a conclusion, at energies at around  $\log_{10}(E/\text{GeV}) = 7.9$  to  $\log_{10}(E/\text{GeV}) = 8.0$ , where the flux of iron nuclei was characterized by a knee-like feature, in case of the computed all-particle spectrum no structure is discernible, but it could also be veiled by the large flux uncertainties. However, based on a more robust alternative analysis method [33], which is *inter alia* based on a subsample with larger statistics, our collaboration found a high significant change of the spectral index of the all-particle spectrum at this very energy range. Hence, there is a strong indication for a knee-like structure in the cosmic ray spectrum at around  $\log_{10}(E/\text{GeV}) = 7.9$  to  $\log_{10}(E/\text{GeV}) = 8.0$ , which is caused by a steepening in the spectra of heavy primaries; especially by one in the flux of iron nuclei<sup>28</sup>, as the unfolding analysis suggests. Furthermore, both the unfolding as well as the alternative method [33] yield a concave behaviour of the all-particle spectrum at the energy range from  $\log_{10}(E/\text{GeV}) = 6.9$  to  $\log_{10}(E/\text{GeV}) = 7.2$ . Viewing the

<sup>28</sup>Or at least the mass group that is represented by iron.

all-particle spectra of the GAMMA and the Tibet-III experiment, depicted in Fig. 7.22 (bottom panel), or that of the TUNKA and the Tibet experiment, depicted in Fig. 8.3, one can observe such concave structures at the same energy range there, too. However, this is a first impression only, and does not replace a comprehensive analysis!

In the following, more detailed comparisons between the energy spectra unfolded in this thesis and those based on other analysis methods or other experiments will be performed, in order to get further insights about the compatibility of the obtained results.

## 7.4 Comparison with Other Results

In the previous sections, the reliability of the elemental energy spectra unfolded based on the interaction models QGSJET-II-02 and FLUKA 2002.4 has been discussed, and the specific structures of the fluxes have been judged in the framework of statistics. It was found that based on the determined solution the measured dataset can be described well, such that there are no indications for an unreliability of the unfolded spectra. Thereafter, the significance of the distinct structures discernible in the flux of iron nuclei and in the all-particle spectrum was judged. A comparison of the derived solution with spectra reconstructed in other analyses will be the last check of consistency, and will offer insights about the compatibility of the results. Thereby, the unfolded energy fluxes will be compared to those derived with other unfolding analyses based either on KASCADE-Grande, too, or on KASCADE measurements (Section 7.4.1). Then, a comparison with KASCADE-Grande results obtained by our collaboration in analyses that are more closely related to the measured dataset will follow (Section 7.4.2). Finally, the unfolded all-particle spectrum will be compared to those obtained by other cosmic ray experiments (Section 7.4.3).

### 7.4.1 Comparison with spectra obtained by other unfolding analyses of our collaboration

Some time ago, the energy spectra for elemental groups of cosmic rays at lower energies, more precisely at the energy range of 1 PeV to some tens of PeV, have already been determined by our collaboration by means of an unfolding analysis, too, but based on KASCADE measurements [9]. Since that analysis bases on the older high energy interaction models QGSJET 01 or SIBYLL 2.1 and the low energy interaction model GHEISHA 2002, it was repeated in [70] based on the interaction models QGSJET-II-02 and FLUKA 2002.4. Additionally, in [70], the unfolding technique was also applied to KASCADE-Grande data. The aforementioned unfolding analyses base on the measured two-dimensional electron vs. muon number distribution, while that in this thesis bases on the distribution of the charged particle vs. muon number. Nevertheless, as discussed in Section 4.1, the final solution should not depend on which of the two sets will be chosen for the analysis. KASCADE and KASCADE-Grande base principally on separate setups<sup>29</sup>, such that the results of them can to a certain extent be interpreted as those of two different experiments. But, the results obtained in this thesis and in [70] by an unfolding of the KASCADE-Grande data should be more similar, since both base on the

<sup>29</sup>Indeed, KASCADE-Grande uses the former KASCADE detectors in order to measure the energy deposits of muons separately from those of electrons. However, things like core position, angle of incidence, etc., needed for the muon reconstruction, are determined based on KASCADE-Grande measurements. Hence, the remaining reconstruction procedure is independent from the KASCADE experiment.

same measurements<sup>30</sup>. In a first step, the spectra for elemental groups will be compared, and, thereafter, the all-particle spectra, which are the sums of the respective individual spectra.

### Elemental spectra

In Fig. 7.18, the energy spectra for elemental groups of cosmic rays obtained in this work (labelled with “Grande...Nch-N $\mu$ -unfolding”) are compared to those obtained by other unfolding analyses [70] that are based on the KASCADE-Grande data (labelled with “Grande...Ne-N $\mu$ -unfolding”, see top panel), too, or on KASCADE data (labelled with “KASCADE...Ne-N $\mu$ -unfolding”, see bottom panel). The error bars represent statistical uncertainties, the error bands the systematic ones (the latter ones are only shown for the results of this thesis). For a better distinguishability, the spectra of helium, carbon, and silicon nuclei are summed up to one resulting intermediate mass group (“medium”), while the spectra of protons and iron nuclei, representing the “light” and “heavy” mass groups, are still shown individually. All results base on the interaction models QGSJET-II-02 and FLUKA 2002.4. Instead of using the charged particle number as second observable beside the muon number, in case of the analyses performed in [70] the electron number was chosen. However, this difference is expected to have only small, or even negligible influences on the final result.

Comparing the two results obtained by unfolding analyses based on KASCADE-Grande data (Fig. 7.18, top panel) reveals differences in the energy spectra. While the slopes of the fluxes are very similar<sup>31</sup>, their absolute values are significantly different. However, taking into account the systematic uncertainties of both analyses, there is in the main an agreement between the results. The origin for the deviations is not known at the moment, and can only be ascertained in detailed investigations. Possible reasons could be the different iteration depths, slightly different quality cuts<sup>32</sup>, or the fact that different observable sets<sup>33</sup> have been used. Furthermore, there are slight differences in the functions used for the parametrizations of the response matrix, and the conditioning of the response matrix is treated in a different way. The interplay between all these sources could cause such deviations, but is not easy to analyse. Since the spectra coincide within the systematic uncertainties, there are no indications for systematic mistakes. Ignoring the systematic shift, especially the spectra of iron nuclei show a very similar slope; and even most individual bin-to-bin variations are identical. Both spectra are characterized likewise by a knee-like structure at around 80 PeV, such that the existence of the iron-knee/heavy-knee is confirmed by both unfolding analyses.

Comparing the spectra of this thesis with the ones obtained by the unfolding of the KASCADE dataset [70] (Fig. 7.18, bottom panel), yields similar results as in case of the afore-performed comparison. Again, there are differences in the relative abundances.

<sup>30</sup>Both analyses have been performed independently, such that the comparison of the results can also be understood as cross-check to identify possible human-caused technical mistakes in the analysis process.

<sup>31</sup>Beside the fact that the results derived in [70] seem to suffer from larger bin-to-bin fluctuations, possibly explainable by the larger number of iteration steps ( $\Delta\chi^2 = 6 \times 10^{-4}$ ), in comparison to the ones used in this thesis ( $\Delta\chi^2 = 2 \times 10^{-3}$ ).

<sup>32</sup>Especially the fiducial areas are different.

<sup>33</sup>Basically, the solution should not depend on which of the two possible observable sets is chosen. However, if the used interaction models have deficits especially in the description of one observable, different observable sets could yield differing solutions. As an example: If the electron number prediction is wrong, it will be a difference if the second observable is the pure deficient electron number, or the charged particle number, which also contains contributions of the perhaps well predicted muon number.

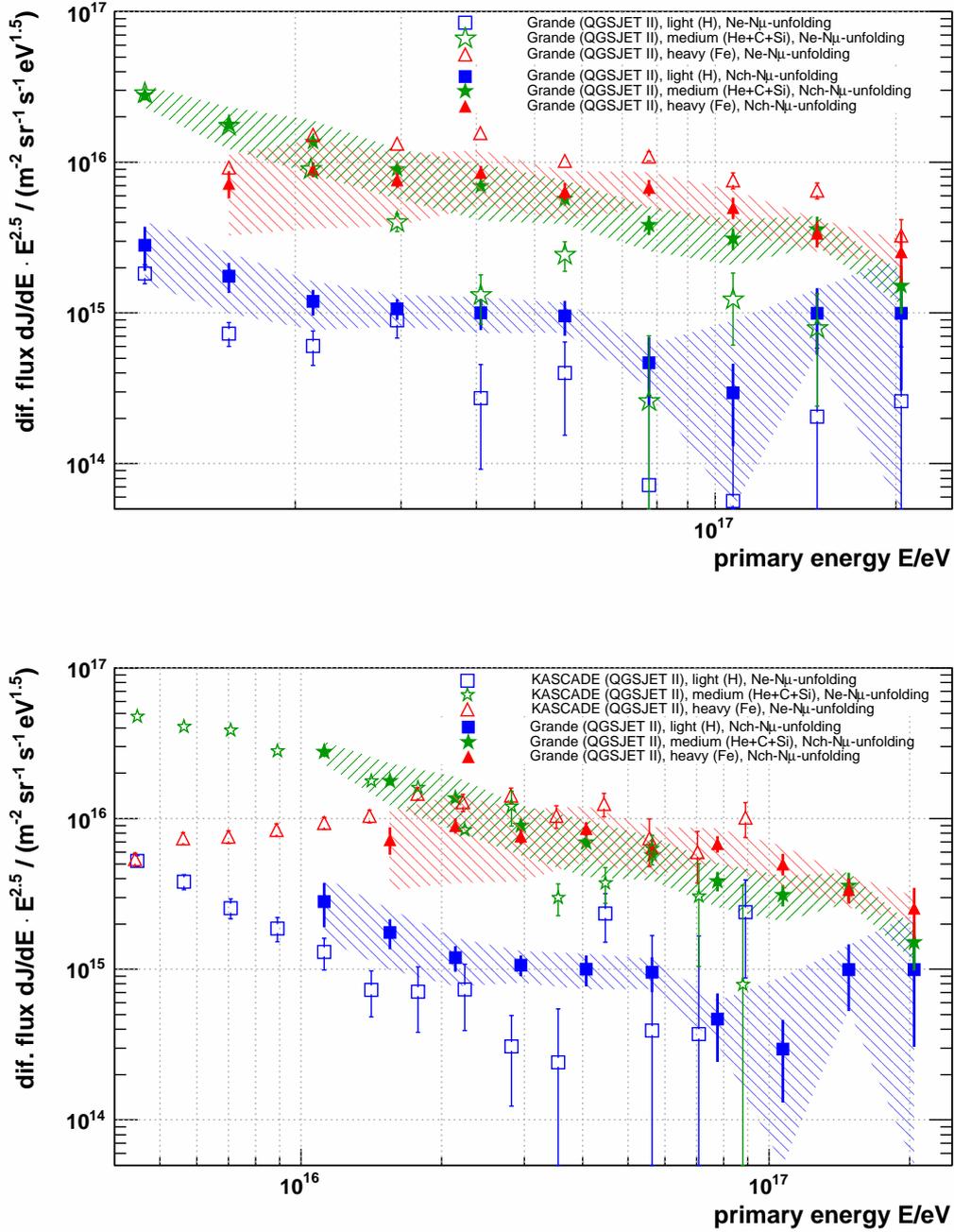


Figure 7.18: The energy spectra for elemental groups of cosmic rays obtained in this work (“Grande...Nch-Nµ-unfolding”) are compared to those obtained by other unfolding analyses [70] that are based on the KASCADE-Grande data (“Grande...Ne-Nµ-unfolding”, see top panel), too, or on KASCADE data (“KASCADE...Ne-Nµ-unfolding”, see bottom panel). The error bars represent statistical uncertainties, error bands the systematic ones (the latter ones are only shown for the results of this thesis). For a better distinguishability, the spectra of helium, carbon, and silicon nuclei are summed up to one resulting intermediate mass group (“medium”), while the spectra of protons and iron nuclei, representing the “light” and “heavy” mass groups, are still shown individually. All results base on the interaction models QGSJET-II-02 and FLUKA 2002.4.

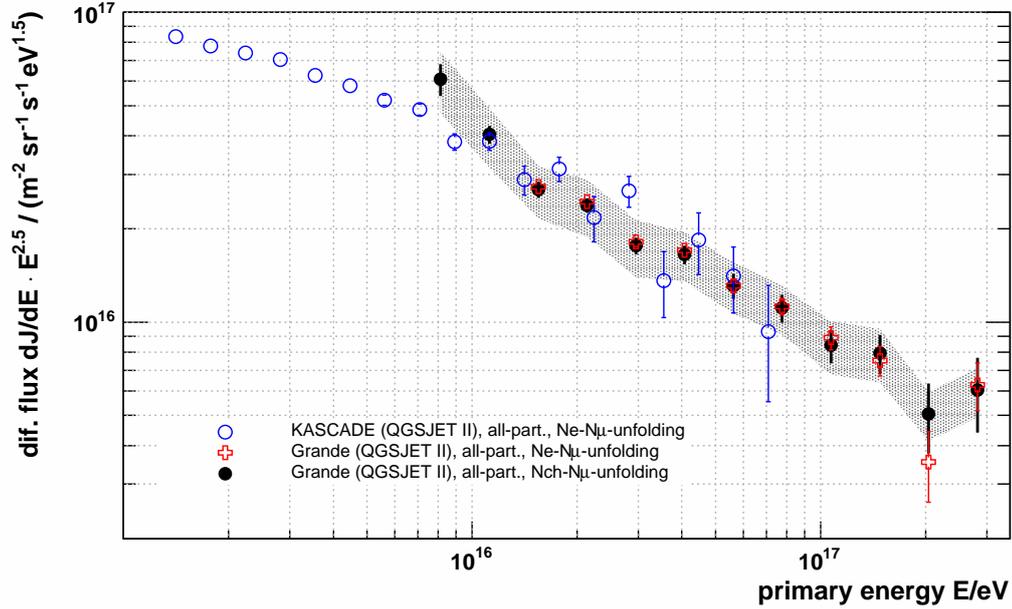


Figure 7.19: The all-particle spectrum, obtained in this work (“Grande...Nch- $N_{\mu}$ -unfolding”), is compared to those obtained by other unfolding analyses [70] that are based on the KASCADE-Grande data (“Grande...Ne- $N_{\mu}$ -unfolding”), too, or on KASCADE data (“KASCADE...Ne- $N_{\mu}$ -unfolding”). The error bars represent statistical uncertainties, error band the systematic ones (the latter one is only shown for the result of this thesis). All results base on the interaction models QGSJET-II-02 and FLUKA 2002.4.

Nevertheless, the deviation is now smaller in case of the spectra of iron nuclei, and negligible small in case of the intermediate mass groups. It is surprising that the differences are smaller than that discernible whilst the comparison between the two unfolding results both basing on KASCADE-Grande data, since now the spectra of two “different” experiments are compared. Taking into account the systematic uncertainties, both results agree perfectly. Since the KASCADE experiment suffers from very low statistics at these energies, no conclusion about a possible structure in the spectrum of iron nuclei at around 80 PeV can be drawn. However, at least the slope of the flux of iron nuclei before that energy agrees very well. Finally, the slight<sup>34</sup> hardening in the proton spectrum observed in this thesis beyond energies of about  $\sim 25$  PeV can also be seen in the proton spectrum obtained based on the KASCADE measurements<sup>35</sup>.

### All-particle spectra

After the comparison between the energy fluxes for elemental groups, now that between the all-particle spectra, which are computed by summing up all individual spectra, follows. This is illustrated in Fig. 7.19: The all-particle spectrum obtained in this work (“Grande...Nch- $N_{\mu}$ -unfolding”) is compared to those obtained by other unfolding analyses [70] that are based on the KASCADE-Grande data (“Grande...Ne- $N_{\mu}$ -unfolding”),

<sup>34</sup>Albeit statistically not significant.

<sup>35</sup>But, also there without statistical significance.

too, or on KASCADE data (“KASCADE...Ne-N<sub>Fe</sub>-unfolding”). The error bars represent statistical uncertainties, the error band the systematic ones (the latter one is only shown for the result of this thesis). All results base on the interaction models QGSJET-II-02 and FLUKA 2002.4.

While in case of the spectra of individual mass groups differences in the absolute flux values could be observed, in case of the all-particle spectra a perfect agreement between the two KASCADE-Grande results is discernible. Furthermore, the results conform with that of KASCADE, at least within the statistical uncertainties. Only at the first data point of the all-particle spectrum derived in this work (“Grande...Nch-N<sub>Fe</sub>-unfolding”), there is a larger deviation to the KASCADE result (“KASCADE...Ne-N<sub>Fe</sub>-unfolding”) observable. However, as discussed in Section 7.2.1, it was found that, below energies of  $10^{16}$  eV, the individual elemental spectra suffer from larger systematic uncertainties, which are not considered sufficiently in the error estimation. Hence, below that energy, no individual spectra are shown. The all-particle spectrum, which is the sum of all single spectra, was assumed to be more stable, and hence was also shown below  $10^{16}$  eV. But, considering the result obtained right now, the first shown flux value should be interpreted with caution<sup>36</sup>.

Nevertheless, ignoring this first data point of the all-particle spectrum obtained in this work, there is still a slight indication for the concave behaviour at the energy range from  $\log_{10}(E/\text{GeV}) = 6.9$  to  $\log_{10}(E/\text{GeV}) = 7.2$  remaining (cf. Section 7.3.3, and Fig. 7.17). Such a concave behaviour is not ascertainable in case of the KASCADE data. In this context, one has to keep in mind that, despite the comparatively good KASCADE-Grande statistics in that energy range, this structure was already determined with notable less significance<sup>37</sup> than in case of more robust analysis methods like [33]. Because of the smaller statistics in case of KASCADE at these energies, it could be possible that this concave structure is smeared out to a certain degree, what could be the reason why it is not observable there. However, a final explanation can currently not be given in this thesis.

To summarize, both the comparison of the unfolded individual energy spectra and that of the all-particle fluxes revealed a good agreement between all results, at least within the given uncertainties. The absolute values of the individual spectra are slightly different, but the heavy mass group is dominating the cosmic ray composition at energies between  $\sim 10$  PeV and  $\sim 100$  PeV in all cases. Furthermore, also most of the specific structures in the spectra are equal. Especially, the iron/heavy knee can be observed in both KASCADE-Grande unfolding results. For purpose of a further compatibility check, the energy spectra unfolded in this thesis will now be compared to ones derived by other analysis methods of our collaboration.

#### 7.4.2 Comparison with spectra obtained by other analysis methods of our collaboration

The all-particle flux as well as the spectra of individual mass groups have also been determined by our collaboration by means of more robust methods<sup>38</sup>, which are more native, i.e. closer to the measured data sample, and which are based on a subsample with

<sup>36</sup>Furthermore, this first data point of the all-particle spectrum will not be shown in the following depictions any longer, since its reliability within the given uncertainties is not guaranteed.

<sup>37</sup>In case of the more robust methods, this concave structure was found to be of high significance.

<sup>38</sup>They are more robust, but, as a *quid pro quo*, do not enable to separate the elemental spectra with such a precision the unfolding technique allows.

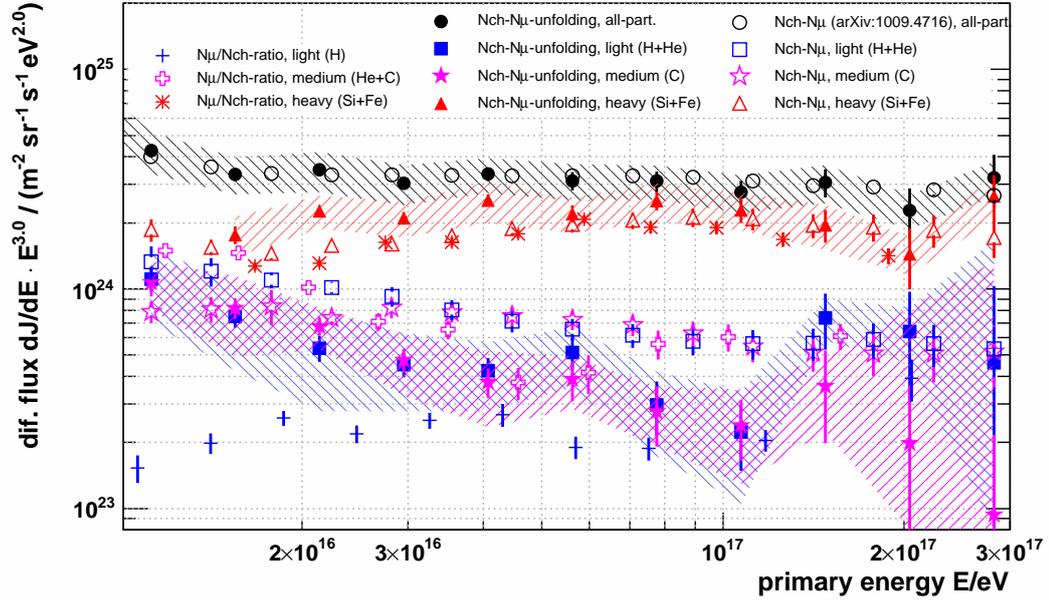


Figure 7.20: The energy spectra for elemental groups of cosmic rays, obtained in this work (“Nch-N $\mu$ -unfolding”), are compared to those obtained by other analyses of our collaboration [49, 50] (“N $\mu$ /Nch-ratio”) or [13, 34] (“Nch-N $\mu$ ”). The error bars represent statistical uncertainties, error bands the systematic ones (the latter ones are only shown for the results of this thesis). All results base on the interaction models QGSJET-II and FLUKA 2002.4. See text for further details.

larger statistics. In these alternative methods, parametrized energy estimators are used, which are rather simple, but solid, and are based immediately on the measured shower sizes. The results derived in this thesis will now be compared to those elemental energy spectra, and, thereafter, to those all-particle fluxes.

### Elemental spectra

In Fig. 7.20, the energy spectra for elemental groups of cosmic rays, obtained in this work (“Nch-N $\mu$ -unfolding”), are compared to those obtained by other analyses of our collaboration [49, 50] (“N $\mu$ /Nch-ratio”) or [13, 34] (“Nch-N $\mu$ ”). The error bars represent statistical uncertainties, error bands the systematic ones (the latter ones are only shown for the results of this thesis). Based on the specific technical aspects of the alternative methods, the spectra are divided into ones of a lighter, an intermediate, and a heavy mass group. These mass groups correspond to different primary particles<sup>39</sup> (see legends). For a better comparability, in case of the unfolding analysis results the unfolded individual spectra are summed up: protons and helium nuclei represent the “light” mass group, silicon and iron the “heavy” one. The spectrum of carbon nuclei is still shown individually and represents the intermediate (“medium”) mass group. For orientation, the all-particle

<sup>39</sup>The classification into light, intermediate, and heavy mass groups is done in a rather simple, but robust way. These mass groups are roughly correlated to the primaries labelled in the legends, but do not correspond exactly to them as in case of the unfolding analysis.

spectra derived in [33] and in this thesis are shown additionally. All results base on the interaction models QGSJET-II and FLUKA 2002.4.

In general, taking into account that e.g. H+He in case of the analysis [13, 34] is not exactly the same<sup>40</sup> as in case of the unfolding analysis, the agreement between both results is alright. The results of [49, 50] are more different, since the light mass group is represented by protons only, while the intermediate mass group is given by helium and carbon nuclei. Due to this, especially the “light” flux is below the fluxes of the two other methods. However, all methods have in common that the spectra of the heavy component exhibit a knee-like structure: in case of [49, 50] at around 62 PeV, for [13, 34] at about 90 PeV, and at about 80 PeV for the unfolding analysis. Hence, although the two alternative methods divide the primaries into mass groups only roughly, there is a nice agreement between all results.

### All-particle spectra

Fig. 7.21 comprises a comparison of the all-particle spectrum<sup>41</sup> obtained in this work (“Nch-N $\mu$ -unfolding”) with those obtained by other analyses of our collaboration [15] (“Nch-N $\mu$ ”, “Nch”, or “N $\mu$ ”). The error bars represent statistical uncertainties, error bands the systematic ones (the latter ones are only shown for the results of this thesis). All results base on the interaction models QGSJET-II and FLUKA 2002.4. Thereby, the results that are labelled “Nch” or “N $\mu$ ” and are represented by the dashed lines are all-particle spectra derived under the assumption that the cosmic rays are pure protons (“Proton”) or iron (“Iron”) nuclei; and hence, they represent rather limits for the all-particle spectrum.

The all-particle spectrum “Nch-N $\mu$ ” and that determined in this thesis agree very well within the given uncertainties. They are bracketed by the spectra determined under the assumptions of a pure proton or iron nuclei composition; but, they are always closer to the spectra based on the iron nuclei hypothesis. Furthermore, the spectra “Nch-N $\mu$ -unfolding”, “Nch-N $\mu$ ”, as well as “N $\mu$ ” (in case of the hypothesis of pure iron nuclei) are characterized by the concave behaviour below  $2 \times 10^{16}$  eV.

As a conclusion, albeit the assignment of primary nuclei to specific mass groups is technically not exactly identical, the spectra of individual mass groups derived by different analysis methods and based on different observables agree very well. In all results, the heavy mass group is dominating the cosmic ray composition and exhibits a knee-like structure at around 60 PeV to 90 PeV. The discussed all-particle spectra agree very well, too, and also predict a domination of the heavy mass group. Again, the concave structure, which is observable in the unfolding result below  $2 \times 10^{16}$  eV, is confirmed by some of the alternative analysis methods. Since now the conformity of the results obtained with different analysis methods based on the KASCADE/KASCADE-Grande experiment is ensured, finally, a comparison with spectra of other experiments follows.

<sup>40</sup>In case of the alternative methods, the particles finally representing a specific mass group are combined already whilst the analysis itself, and hence assumptions about the relative abundances have to be made (equal abundances are assumed). In case of the unfolding analysis, every primary is treated separately, such that no *a priori* assumption has to be made. Only at the very end, primaries are combined, if desired.

<sup>41</sup>As discussed in Section 7.4.1, only data points beyond  $10^{16}$  eV are shown, since for smaller energies their reliability cannot be guaranteed.

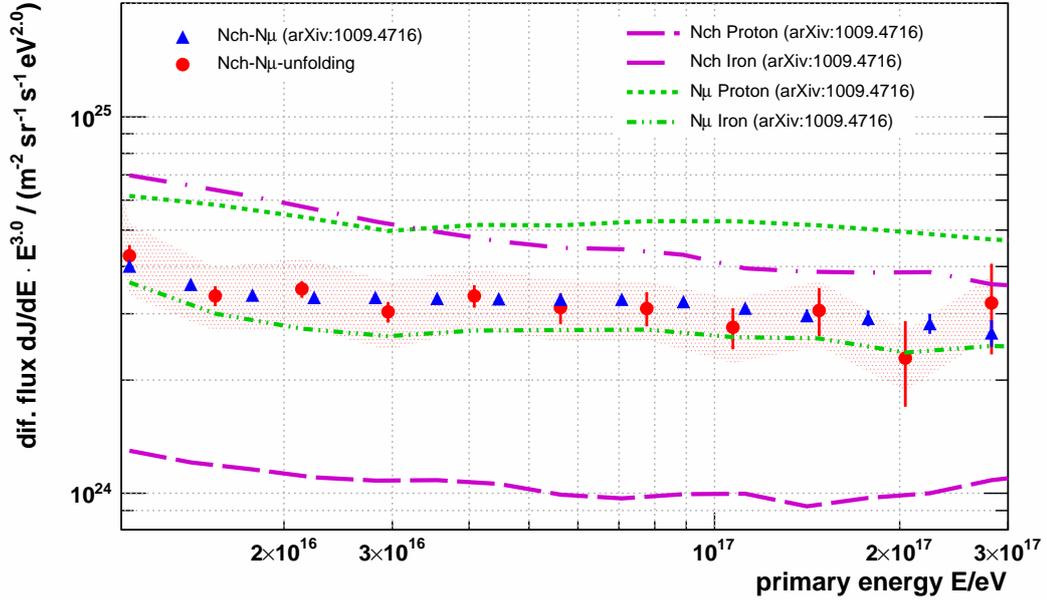


Figure 7.21: The all-particle spectrum obtained in this work (“Nch-N $\mu$ -unfolding”) is compared to those obtained by other analyses of our collaboration [15] (“Nch-N $\mu$ ”, “Nch”, or “N $\mu$ ”, see text for further details). The error bars represent statistical uncertainties, error bands the systematic ones (the latter ones are only shown for the results of this thesis). All results base on the interaction models QGSJET-II and FLUKA 2002.4.

### 7.4.3 Comparison with spectra obtained by other experiments

After, in the previous section, the unfolding results obtained in this thesis have been checked against those obtained by further analyses of our collaboration, the comparison with spectra determined by other experiments will be the last step now.

In Fig. 7.22, the all-particle spectrum obtained in this work (“Grande ... Nch-N $\mu$ -unfolding”) based on KASCADE-Grande measurements<sup>42</sup> and the one obtained in [70] (“KASCADE ... Ne-N $\mu$ -unfolding”) based on the KASCADE experiment are compared to those obtained by other experiments (see legend for references). The error bars represent statistical uncertainties, error bands the systematic ones (the latter ones are only shown for the results of this thesis). On the bottom panel, only the energy range covered by the KASCADE-Grande unfolding is shown in order to get a better distinguishability between the results of the different experiments.

In the main, the all-particle spectrum obtained in this work agrees well with that of most of the other experiments within the uncertainties. Extrapolating the trend of the all-particle flux determined with KASCADE-Grande by eye up to energies beyond  $10^{18}$  eV, yields a good agreement with the results obtained by the HiRes-II experiment or the Pierre Auger Observatory.

Also in case of the comparison depicted at the bottom panel, there is a good agreement with the results of most of the other experiments discernible: The slopes are compatible,

<sup>42</sup>As discussed in Section 7.4.1, only KASCADE-Grande data points beyond  $10^{16}$  eV are shown, since for smaller energies their reliability cannot be guaranteed.

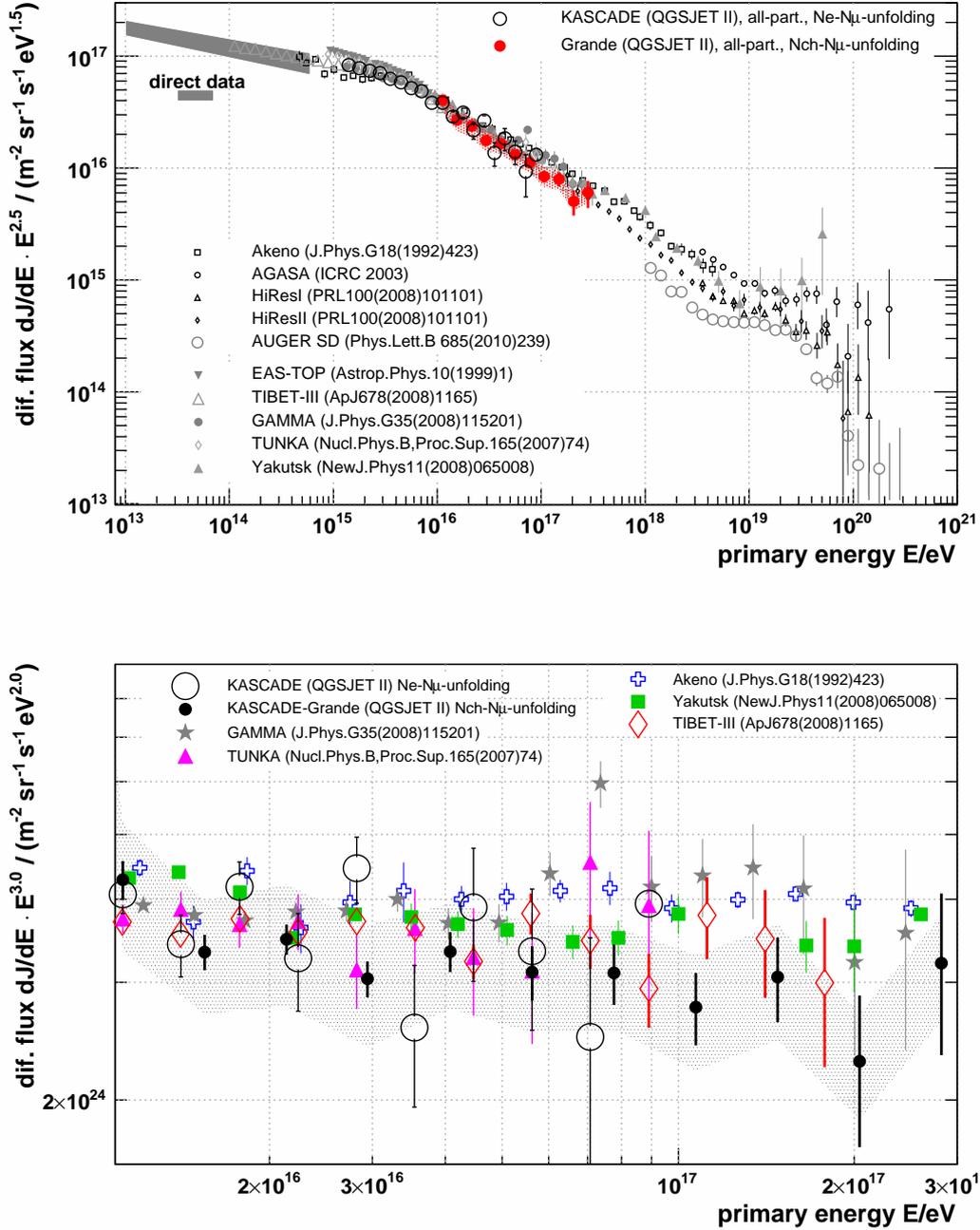


Figure 7.22: The all-particle spectrum obtained in this work (“Grande ... Nch- $N_{\mu}$ -unfolding”) based on KASCADE-Grande measurements and the one obtained in [70] (“KASCADE ... Ne- $N_{\mu}$ -unfolding”) based on the KASCADE experiment are compared to those obtained by other experiments (see legend for references). The error bars represent statistical uncertainties, error bands the systematic ones (the latter ones are only shown for the results of this thesis). On the bottom panel, only the energy range covered by the KASCADE-Grande unfolding is shown in order to get a better distinguishability between the results of different experiments.

and also the absolute values are very similar, at least within the systematic uncertainties. Especially, the spectra derived by the Tibet-III and the GAMMA experiment show a slight concave behaviour below energies of  $2 \times 10^{16}$  eV, too. However, at around  $6 \times 10^{16}$  eV to  $8 \times 10^{16}$  eV, things are different: while the result of the GAMMA experiment<sup>43</sup> exhibits a “bump”, this cannot be confirmed with the KASCADE-Grande result of this work.

To summarize, the all-particle spectrum obtained in this work by means of an unfolding analysis based on KASCADE-Grande measurements agrees very well with that of most of the other air shower experiments. Based on a guesstimate by eye, the slight concave behaviour below energies of  $2 \times 10^{16}$  eV is observable in the spectra determined by the GAMMA and the Tibet-III experiment, too. However, the “bump” at around  $6 \times 10^{16}$  eV to  $8 \times 10^{16}$  eV, seen by the GAMMA experiment, cannot be confirmed with the current KASCADE-Grande results. Extrapolating the all-particle spectrum obtained in this work by eye towards higher energies, yields a good agreement with the spectra determined by the Pierre Auger Observatory or the HiRes-II experiment.

---

<sup>43</sup>And also the TUNKA experiment, but with less significance.



## Chapter 8

# Summary, Final Discussion, and Outlook

The ultimate goal of the extension of the former KASCADE to the KASCADE-Grande experiment was to measure the energy spectra for elemental groups of cosmic rays in the energy range from 10 PeV to beyond 100 PeV. By this, the question about the existence of a so-called “second knee” in the spectrum of cosmic rays, caused by a steepening in the spectra of heavy primaries, have to be answered. Furthermore, the knowledge about the cosmic ray composition at the energies covered by KASCADE-Grande allows drawing inferences about a transition from galactic to extra-galactic cosmic rays. This goal, i.e. the determination of spectra for elemental groups of cosmic rays in that energy range, was accomplished in this work. The respective KASCADE-Grande results are of up to now unrivalled quality.

### Summary of the main steps of the analysis strategy

In order to get such high quality results, a very stable experiment is necessary, and its properties have to be understood accurately. Furthermore, a very well working air shower event reconstruction procedure is needed. The reconstruction software was optimized and checked extensively in the past time, making such unfolding analyses possible in the first place. Hence, contributing to these indispensable improvements was the preparatory work for this thesis. After at a certain stage of the optimization process a sufficient reconstruction quality of the shower sizes was realized, the main analysis was started. Since the sought-after energy spectra are folded with the probability distributions for the intrinsic shower fluctuations, as well as for the detector properties like efficiency and resolution, it was decided to apply unfolding techniques.

After preliminary considerations, the optimal set of KASCADE-Grande observables for the unfolding analysis of this work was found: the two-dimensional charged particle number vs. muon number shower size spectrum. Then, the response matrix, which inter-relates the sought-after spectra with the measured dataset, and which has to be known for the unfolding, was computed. For this purpose, shower simulations based on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68] have been used. Whilst the determination of the response matrix, smoothing respectively regularization techniques have been applied by using parametrization functions, and by conditioning the derived response matrix itself under verification by means of a *singular value decomposition*. All this ensures a stable solution. The final response matrix was checked intensively.

Different solution techniques for the present convolution problem have been discussed and tested. The unfolding procedure introduced by Gold [77] was found to be the most suitable for the current analysis. However, alternative methods based on Bayes' theorem or the principle of maximum entropy serve for cross-checks. Things like the optimal iteration depth or the appropriate regularization parameter have been determined based on the *weighted mean squared error* or the *relative variance of the bias*. A well working error estimation was realized by means of a frequentist approach.

Afterwards, the whole unfolding procedure was tested thoroughly by means of toy datasets. The applied techniques were found to guarantee a high quality solution. Problems occurred in case of very few trial spectra only; but, appropriate indicators for the identification of such problems have been elaborated.

### Main analysis and discussion of the results

Finally, the gained insights and techniques have been applied to the measured shower size spectrum. The most important results are<sup>1</sup>:

- The cosmic ray **composition is dominated by the heavy mass group** in the energy range from about 10 PeV to 100 PeV.
- The energy spectrum of the **heavy mass group** of cosmic rays, in this analysis represented by iron nuclei, exhibits a **knee-like structure at around 80 PeV**. This position corresponds to the knee-like structure observed by our collaboration [33] in the all-particle spectrum. Furthermore, this result was confirmed by spectra of heavy primaries obtained in other analyses [13] of our collaboration.
- The unfolded spectra for the other primaries (**protons, as well as helium, carbon and silicon nuclei**) suffer from low statistics. Within the given uncertainties, the fluxes exhibit **no significant structures**. There are only slight indications that the proton spectrum possibly gets harder for energies  $\log_{10}(E/\text{GeV}) \gtrsim 7.4$ , what is, from a pure statistical point of view, without significance however<sup>2</sup>.
- A comparison between the knees in the spectra of protons, of helium nuclei (both determined based on KASCADE measurements [70]), and of iron nuclei yields strong indications for a **dependence of the knee positions on the atomic number**, what would encourage the models predicting a rigidity dependent acceleration and/or propagation of cosmic rays connected with magnetic fields. This conclusion was drawn under the premise that the mass groups represented in this thesis by protons, as well as by helium or iron nuclei actually consist only, or at least primarily, of these respective primaries.
- There is a **slight concave behaviour of the all-particle spectrum at around  $\log_{10}(E/\text{GeV}) = 6.9$  to  $\log_{10}(E/\text{GeV}) = 7.2$**  observable. This “dip-like” feature (respectively hardening) is confirmed by another analysis of our collaboration [33], where this structure was found to be significant.

<sup>1</sup>Based on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68].

<sup>2</sup>However, slight indications for such a “recovering” can also be observed in an unfolding analysis based on KASCADE measurements (see Fig. 7.18, bottom panel). But also there, without statistical significance.

The quality of the unfolded energy spectra was reviewed extensively. Thereby, the quality of the description of the measured dataset by the derived solution was examined. Furthermore, the results have been counter-checked based on the alternative unfolding algorithms. There were no indications found that the used interaction models QGSJET-II-02 and FLUKA 2002.4 fail to describe the physics of hadronic interactions at these energies reliably. Accordingly, there are no signs that the derived energy spectra are unreliable. This finding could be considered to be somewhat contrasting with the conclusion drawn in [59], where the first Large Hadron Collider (LHC) data have been compared to predictions of diverse hadron-interaction Monte Carlos, like QGSJET 01, QGSJET-II, SIBYLL 2.1, and EPOS 1.99. The authors point out that “none of the models is in perfect agreement with all the hadronic observables measured at the LHC” ([59, p. 6]). However, these findings are not inconsistent with the results obtained in this work: The comparison between the unfolded solution and the measured data is indeed a means to detect rather serious and significant problems of the models, but is not as sensitive that it would allow to judge about slighter deficits that prevent only a “perfect agreement”, to use the words of the authors. Hence, that the model QGSJET-II-02 enables a good description of the measured data has to be understood as indication that it is probably not deficient at all, but does not automatically claim a “perfect” description of the physics of hadronic interactions at these energies.

The LHC is still operating at lower energies up to centre-of-mass energies at around  $7 \times 10^{12}$  eV, corresponding to a fixed-target energy of  $\sim 2.4 \times 10^{16}$  eV. Thereby, it reaches the lower energy threshold of KASCADE-Grande. But, it is designed to collide protons at centre-of-mass energies of up to  $1.4 \times 10^{13}$  eV, corresponding to a fixed-target energy of about  $\sim 10^{17}$  eV; and thus its results can prospectively help to tune the parameters of the high energy interaction models at almost the whole energy range covered by KASCADE-Grande. In this context, especially the findings of the LHC experiments TOTEM (TOTAl Elastic and diffractive cross section Measurement) and LHCf (Large Hadron Collider forward) will be of interest, as well as the measurements of the CASTOR (Centauro And Strange Object Research) calorimeter of the CMS (Compact Muon Solenoid) experiment. They are all designed to study the particles generated in the forward region<sup>3</sup> of collisions, what is the relevant kinematic range for cosmic ray air shower physics, as it dominantly influences the air shower development by transferring efficiently a large fraction of the energy of the projectile into the atmosphere. Hence, improved versions of the hadronic interaction models are expected in the near future.

In order to get a rough impression how different interaction models could influence the final result, a first, but very preliminary cross-check was performed by an unfolding based on the high energy interaction model EPOS 1.99 [188]. It was found that EPOS 1.99 seems to have deficits in the description of the measured data. Nevertheless, beside the finding that the results obtained based on EPOS 1.99 are less trustable, this countercheck allows to assess the influence of different models on the solution: As far as this rough and preliminary test allows deducing such conclusions, different models seem to affect primarily the relative abundances of different mass groups, while most of the characteristic structures in the individual spectra seem to be less model dependent<sup>4</sup>. This confirms the results obtained in [9], where the high energy interaction models QGSJET 01 and SIBYLL yielded

<sup>3</sup>While typical collider experiments investigate rather the particles emitted transversely in order to examine hard interactions with large momentum transfer and high particle multiplicities.

<sup>4</sup>Especially, whether or not the structures are there seems to be independent of the used model; nevertheless, the energies, at which the structures are observable, can vary slightly. Furthermore, the assignment of the structures to certain primaries can be different.

different relative abundances of the individual primaries, while the structures in the specific spectra were less model dependent. Based on the obtained preliminary results, a first supposition developed that EPOS 1.99 possibly predicts air showers with too many muons<sup>5</sup>, or too little charged particles.

In context of the interpretation<sup>6</sup> of the results obtained in this work, one has to ensure that the observed knee structures are actually connected to things like nuclear mass composition, acceleration procedure, or propagation effects, and are not caused by changes in the hadronic interaction mechanisms in the air shower development possibly not considered so far. Such alternative “new TeV physics” are e.g. discussed in [23]. However, d’Enterria et al. [59, p. 6] stated that their insights gained based on LHC data “give[...] a strong support to the interpretation of the results in the CR “knee” energy range ( $E_{lab} = 10^{15.5}$  eV) in terms of conventional primary spectrum and nuclear mass composition and disfavours some proposed speculative ideas that the change of the CR spectral slope could be due to a sudden change in the hadronic interaction mechanism above 2 TeV c.m. energy.” Hence, such alternative explanations for the cosmic ray “knee” observation appear to be less likely. This conclusion can be strengthened by the results obtained in this thesis, too: most of the alternative explanation attempts, e.g. the one given in [23], predict that the knee positions scale with the atomic mass for different primaries. However, in this thesis, indications for a charge dependent scaling have been found. To conclude this line of thought, such alternative explanations that need “new physics” to describe the discernible cosmic rays knees appear rather implausible. Hence, the primary supposition of this thesis, that both the light and the heavy knee exist, and are primarily connected with a maximal reachable acceleration energy of an astrophysical accelerator interrelated with magnetic fields, and are maybe additionally<sup>7</sup> connected with a modulation in dependence on the charge whilst the propagation process, is still reasonable (for further theoretical details, see Section 2.2).

The energy spectra based on QGSJET-II-02 and FLUKA 2002.4 have been compared to spectra obtained in other analyses of our collaboration. Although the results are obtained with varying levels of complexity, precision and stability, and in some cases even on different observables, no serious deviations<sup>8</sup> could be observed. Especially the main findings, i.e. the dominance of the cosmic ray composition by the heavy component and the structures observable in the iron as well as the all-particle spectrum, have been confirmed likewise with all methods.

A comparison of the unfolded elemental spectra with those of other experiments is not possible, since exactly comparable results are not published yet. However, the TUNKA collaboration has recently shown the development of the mean logarithmic atomic mass of cosmic rays in dependence on the primary energy based on TUNKA-133 measurements [29]. The presented compilation is depicted in Fig. 8.1. For purpose of comparison, the corresponding result (cf. Fig. 7.4) obtained in this thesis based on KASCADE-Grande

---

<sup>5</sup>What is surprising, since most of the other experiments recognize rather a muon deficit in the interaction models like QGSJET-II-02, and hence it was actually the idea to increase the muon number by using models like EPOS 1.99 that predict larger muon numbers (cf. e.g. [58, 192]).

<sup>6</sup>E.g. the interpretation that the observed charge dependent scaling of the knee position encourages cosmic ray acceleration and/or propagation models based on magnetic fields.

<sup>7</sup>Whereas, due to the sharp turn down of the spectra, it appears less likely that the knees are caused by a pure leakage of cosmic rays from the Galaxy (cf. [91]). Hence, if also the propagation effects contribute, it can presumably only be a combination of both: a break of the spectra at the source and a leakage whilst the propagation process (cf. [94]).

<sup>8</sup>And smaller deviations are traceable to systematic differences in the methods, like different assignments of primaries to the specific mass groups, etc.

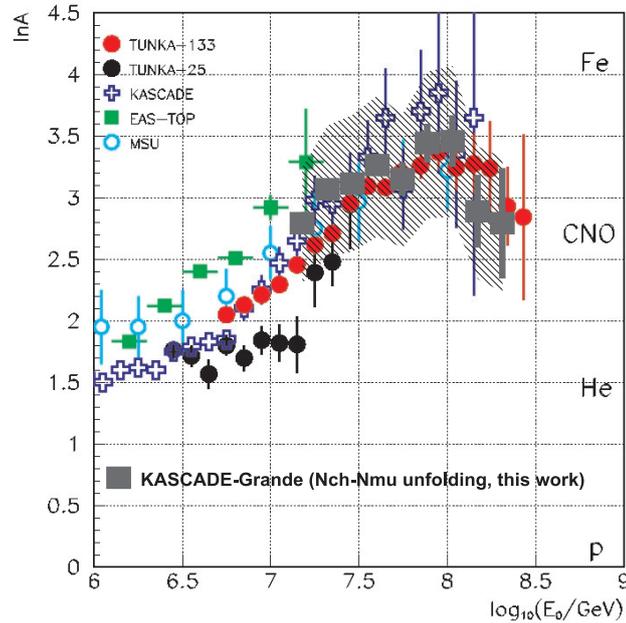


Figure 8.1: Compilation of Prosin [29] comprising the mean logarithmic atomic mass as a function of the primary energy derived by the TUNKA experiment in comparison with those of diverse other experiments (see legend). The logarithmic atomic masses of protons (“p”), helium (“He”), the carbon-nitrogen-oxygen group (“CNO”), as well as of iron (“Fe”) are indicated by the position of the respective labelling on the right side of the panel. Overlaid is the development of the mean logarithmic mass (grey, quadratic markers) derived in this work based on QGSJET-II-02 and FLUKA 2002.4 (cf. Fig. 7.4); the error bars represent the statistical uncertainties, the error band the maximal range of the systematic ones.

measurements is overlaid (grey, quadratic markers). Taking into account the uncertainties, most of the shown results agree very well with that obtained in this work. All experiments predict an increase of the ratio of the heavy to lighter cosmic ray mass groups with increasing energy. However, at energies of about  $\log_{10}(E/\text{GeV}) \approx 8.0$  the abundance of the heavy component gets again smaller. The energy where this turnround in composition takes place corresponds to that where the knee-like structure in the spectrum of iron nuclei was observed in this thesis. In this context, one has to take into account that this turnround in the mean logarithmic mass can also be caused by a recover in the spectra of the lighter mass groups. The final answer can only be given in analyses like that in this work, where the elemental spectra can be observed separately. Nevertheless, the spectra obtained by the other experiments confirm the finding that the composition in the KASCADE-Grande energy range is dominated by the heavy component, and that the composition underlies a rather sharp change at around  $\log_{10}(E/\text{GeV}) \approx 8.0$ , what can be explained based on the results of this work by a knee-like feature in the flux of the mass group represented by iron nuclei. Hence, these two of the main results obtained in this thesis are already in agreement with the findings of other experiments.

Another main conclusion was that the spectra of protons, as well as of helium, carbon, and silicon nuclei are featureless in the framework of the given uncertainties. Only in case

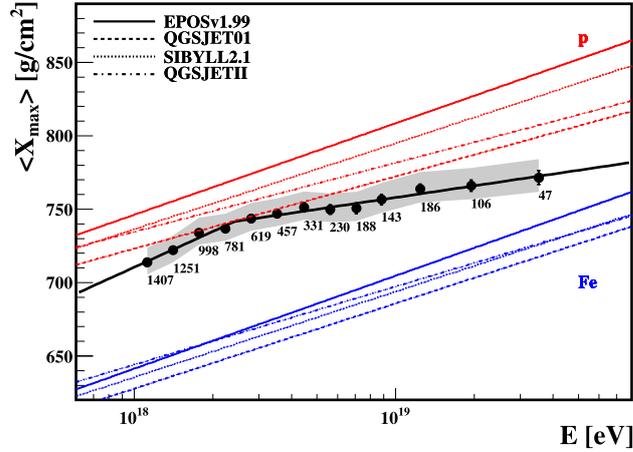


Figure 8.2: Mean  $X_{\max}$  in dependence on the primary energy, measured with the Pierre Auger Observatory (black markers, fitted by a two slope function). The error bars represent statistical uncertainties, the error bands systematic ones. Additionally shown are the expected mean  $X_{\max}$  values for either pure protons or pure iron nuclei, computed based on different interaction models (see legend). The figure was taken from [66].

of protons, there are slight indications that the proton spectrum possibly slightly recovers for energies  $\log_{10}(E/\text{GeV}) \gtrsim 7.4$ , what is, from a pure statistical point of view, without significance however. Nevertheless, adopting this impression hypothetically to be true, this would mean that the abundance of protons rises relatively in comparison to those of the other primaries. Since beyond the knee-like structure at about  $\log_{10}(E/\text{GeV}) \approx 8.0$  iron nuclei get less abundant progressively, an energy should exist where the fluxes of iron nuclei and protons cross. Extrapolating the unfolded fluxes of protons and iron nuclei shown in Fig. 8.4 by eye, predicts such a crossing at around  $\log_{10}(E/\text{GeV}) \approx 8.6$ . Beyond this energy, the lighter, or at least the intermediate mass groups, are expected to be dominating. This purely hypothetical finding can be compared to the result [66] of the Pierre Auger Observatory, depicted in Fig. 8.2. Shown there is the measured mean  $X_{\max}$  development in dependence on the primary energy compared to some predictions for either pure protons or pure iron nuclei, computed based on different interaction models. No matter what simulations are considered, the composition at above  $\log_{10}(E/\text{GeV}) \approx 9.0$  seems to be dominated by the light, or at least intermediate mass groups. However, following the trend of the shown fit down to lower energies, this result predicts a decreasing ratio of the light cosmic ray nuclei to the heavy ones. Hence, the composition gets less lighter, or, the other way around, gets heavier. This result would conform perfectly with that obtained in this work.

The finding that the knee positions of different primaries seem to scale with their charges cannot be verified based on other experiments. Such predictions are only possible if the knee positions of individual nuclei are determined with high precision. Up to now, only KASCADE and KASCADE-Grande together offer both the knee positions of heavy as well as of light primaries. However, due to the limited resolution of the respective unfolding procedures, only a limited number of nuclei can be considered, such that they represent mass groups rather than individual particles. This causes uncertainties in the predictions about the scaling of the knee positions. Hence, the charge dependent scaling, concluded in this work, was drawn under the assumption, that the mass groups

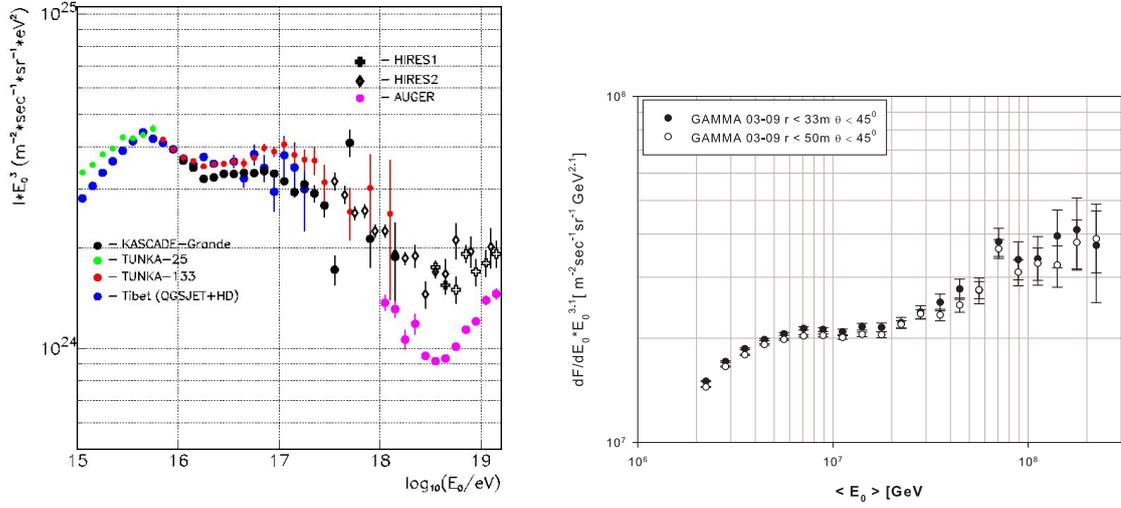


Figure 8.3: Left panel: Compilation of Kuzmichev [30] comprising all-particle spectra derived by the TUNKA experiment in comparison with those of diverse other experiments (see legend). Right panel: The recent all-particle spectra determined by the GAMMA experiment (taken from [133], see there for further details).

represented in this thesis by protons, as well as by helium or iron nuclei actually consist only, or at least primarily, of these respective primaries. Consequently, this result should be interpreted with caution, i.e. under keeping in mind that it is only true under certain circumstances.

The last main finding, i.e. that the all-particle spectrum determined in this thesis<sup>9</sup> exhibits a concave behaviour in the energy range from  $\log_{10}(E/\text{GeV}) = 6.9$  to  $\log_{10}(E/\text{GeV}) = 7.2$ , is compatible with results of other experiments, too. In Fig. 8.3, left panel, a compilation of Kuzmichev [30] on behalf of the TUNKA collaboration is depicted, and, on the right panel, the newest result of the GAMMA experiment [133]. Illustrated are all-particle spectra of diverse experiments. Both TUNKA results as well as the one of the Tibet experiment exhibit a concave structure at around  $\log_{10}(E/\text{GeV}) \approx 7.0$ . Furthermore, this structure can also be observed in the recent results of the GAMMA experiment. In addition, the new TUNKA-133 results also confirm the steepening in the all-particle flux at around  $\log_{10}(E/\text{GeV}) \approx 8.0$ , as can be seen in Fig. 8.3, left panel. In case of the GAMMA experiment (right panel), there could be a slight indication for a kink in the all-particle flux beyond  $\log_{10}(E/\text{GeV}) \approx 8.0$ , too.

### Brief interpretation of the results in context of some exemplarily selected models

After the previous discussions about the quality respectively reliability, the compatibility, and the importance of the obtained results, it will now be focused briefly on the interpretation of the findings. For this purpose, one can make use of theoretical models, which try to describe the shape of the spectra and the composition of cosmic rays (see Chapter 2). A comprehensive discussion of those models would go beyond the scope of this thesis. However, the results will be compared exemplarily to two well known models;

<sup>9</sup>But also in other analyses of our collaboration (cf. Section 7.3.3).

firstly, to the one developed by Hillas [91, 92]. It considers the results of direct cosmic ray measurements obtained with balloon or satellite based experiments. These results are extrapolated, and, based on KASCADE measurements, knee like structures are assigned to the elemental cosmic ray spectra. The assumption is that the galactic cosmic rays are accelerated in supernova remnants with a rigidity dependent maximal reachable energy, resulting in rather sharp knee-like structures. These cosmic rays are called “component A” by Hillas. However, the all-particle flux predicted by this component A differs from the one measured by diverse experiments. Hillas concluded that either the heavy component (especially iron nuclei) behaves differently as expected, or there is “component B” needed, which is an extension of the elemental cosmic ray fluxes beyond their knees. Furthermore, Hillas assumed that from energies of about  $E \approx 10^{17}$  eV extragalactic cosmic rays have to be considered, too, which consist mainly of protons. Taking into account components A and B, as well as the extragalactic one, allows a better description of the measured all-particle spectrum.

In Fig. 8.4, the all-particle spectrum derived in this work (“Grande...all-part”) is compared to the results of other experiments. Furthermore, the unfolded elemental spectra representing the light mass group (given by protons) of cosmic rays, the intermediate one (given by the sum of the individually unfolded spectra of helium, carbon, and silicon nuclei), as well as the heavy one (given by iron nuclei) are shown (“Grande...<mass group>”). For purpose of comparison, also the analogous elemental spectra obtained by KASCADE measurements [70] are depicted (“KASCADE...<mass group>”). In addition, some predictions based on Hillas’s model [91, 92] are shown (see legend). The assumed transition between the components A and B takes place at around  $E \approx 7 \times 10^{15}$  eV for protons, or at around  $E \approx 2 \times 10^{17}$  eV for iron nuclei.

If the absolute flux values of protons determined in this thesis are right, and not shifted systematically, the data points follow rather a pure extrapolation of component A, while a concave structure as seen in the combined component A+B at around  $E \approx 7 \times 10^{15}$  eV is not discernible. This is confirmed, as far as statistics allow such a conclusion, by the KASCADE results. However, beyond  $E \approx 2 \times 10^{16}$  eV, the KASCADE-Grande proton flux slightly recovers. This agrees very well with the extragalactic proton flux predicted by Hillas. However, in the “ankle model” (for details, see Section 2.3), favoured by Hillas, the transition from galactic to extragalactic cosmic rays should happen actually at higher energies, and thereby causes the ankle at around several  $10^{18}$  eV. The flux of iron nuclei, unfolded in this work, agrees very well with Hillas component A for iron nuclei, too. Due to a lack of statistics, the spectrum of iron nuclei could only be unfolded up to energies below the predicted transition of component A and B at around  $E \approx 2 \times 10^{17}$  eV, such that based on the spectrum of iron nuclei no statement about the necessity of a component B can be given. Analysing the all-particle spectrum obtained in this work gives indications that, also in this case, only the all-particle spectrum is needed that is predicted by Hillas based on the component A. Extrapolating the all-particle spectrum of pure component A to higher energies and considering the predicted extragalactic all-particle component, this conforms also well with the first shown markers representing the results of the Pierre Auger Observatory<sup>10</sup>, or with the spectrum of the HiRes-II experiment.

While, under the premise that the absolute flux values are not considerably shifted systematically, especially the all-particle and the proton spectrum determined in this thesis does not claim a necessity of a further component B, the observed concave behaviour

<sup>10</sup>Even when correcting the flux obtained by the Pierre Auger Observatory slightly to higher values, as sometimes stated to be possibly necessary (in the order of 10% to 15%).

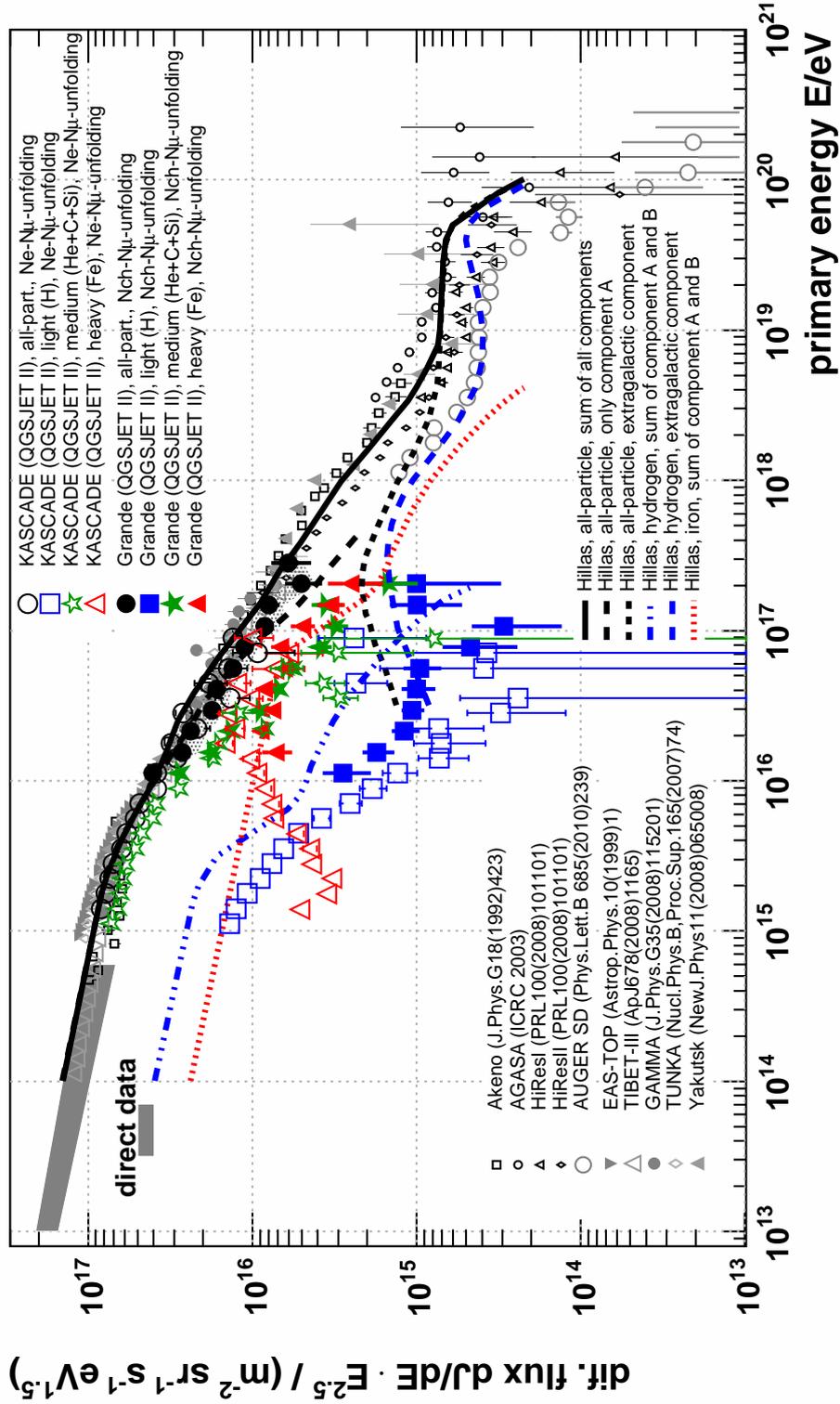


Figure 8.4: The all-particle spectrum, obtained in this work (“Grande ... Nch-Nu-unfolding”) based on the KASCADE-Grande experiment, and the one obtained in [70] (“KASCADE ... Ne-Nu-unfolding”) based on the KASCADE-Experiment (both based on the interaction models QGSJET-II-02 and FLUKA 2002.4), are compared to those obtained by other experiments (see legend for references). Additionally shown are some elemental spectra representing different mass groups (see legend), as well as a model of Hillas [91, 92]. The error bars represent statistical uncertainties, error bands the systematic ones (the latter ones are only shown for the results of this thesis).

in the all-particle spectrum at around  $E \approx 10^{16}$  eV and beyond would be compatible with the two-component model. This dip-like structure, which was confirmed by other analyses of our collaboration, as well as by results of the TUNKA and the GAMMA experiments, could indicate a transition between component A and B. However, there are also alternative attempts available that can explain such a concave behaviour. For example, in model calculations, De Donato and Medina-Tanco [57] assumed a pure rigidity dependence of galactic cosmic rays. As can be seen in Fig. 1 and 2 of [57], such a concave behaviour in the all-particle spectrum can simply be caused by a gap between the spectra of light primaries (“Z=1–5”, i.e. primarily protons and helium nuclei) and that of the next following mass group (“Z=6–12”, i.e. primarily carbon, nitrogen, and oxygen nuclei). Hence, the existence of the dip-like structure is not automatically an evidence for the two-component model.

A very promising alternative model is the “dip model” of Berezhinsky et al. [32] (for details, see Section 2.3). They assume that the dip in the energy range  $\sim 1 \times 10^{18}$  eV to  $\sim 4 \times 10^{19}$  eV is related to electron-positron pair production of extragalactic protons on cosmic microwave background photons. Thereby, the transition from galactic to extragalactic cosmic rays is assumed to happen already at energies close to that where the second knee is observed, in agreement with a pure rigidity based model for galactic cosmic rays. An additional component B like in case of Hillas’s model is not necessary. Furthermore, the pair production mechanism necessitates a dominance of protons in the composition of extragalactic cosmic rays. This would also agree with the finding of this work that the unfolded proton spectrum seems to recover at energies above  $E \gtrsim 2.5 \times 10^{16}$  eV, such that its relative abundance in the all-particle spectrum will again increase for higher energies. As mentioned above, by extrapolating the unfolded fluxes of protons and iron nuclei (shown in Fig. 8.4) by eye, already at around  $E \approx 4 \times 10^{17}$  eV a crossing of both spectra is expected. Beyond this energy, the lighter, or at least the intermediate mass groups are presumably dominating. To summarize, the model proposed by Berezhinsky et al. [32] seems to be compatible with the results obtained in this thesis within the uncertainties. Nevertheless, it has to be emphasized that a mixed model (cf. Section 2.3) could also be possible, such that the afore-performed comparison is only exemplary, and for sure not final.

The afore-shown comparisons are basing on only some exemplarily selected models. Due to the large amount of further alternative models, which are all trying to explain the spectrum of cosmic rays at the energies between the first knee and the ankle, a comprehensive and conclusive analysis will not be possible in this work. However, it could be shown that, under considering steadily the possibility of systematic uncertainties<sup>11</sup>, and hence exercising respective cautions, the results obtained in this thesis are also a powerful means to judge about the plausibility of diverse theoretical models.

### Closing statement

Beside the circumstance that the findings in this thesis allow to answer the long existing question about the existence of a second knee caused by a change in the fluxes of heavy cosmic ray primaries, they open new possibilities in evaluating the plausibility of diverse theoretical models used to describe the spectrum of cosmic rays at energies between the first knee and the ankle. Thereby, the obtained results enable, among other

<sup>11</sup>On the one hand the computed and given uncertainties, but also that caused by possible deficits in the used interaction models QGSJET-II-02 and FLUKA 2002.4 that could not be considered quantitatively in this work.

things, to draw inferences about the transition from galactic to extragalactic cosmic rays. Furthermore, the insights gained in this work allow on the one hand to judge roughly about the reliability of hadronic interaction models, and on the other hand even to give indications how the respective models could possibly be improved, as done exemplarily for EPOS 1.99. The obtained results appear to be of high quality; however, every interpretation has to be done under consideration of the given uncertainties, as well as under taking into account that all findings are based on the assumption that the used interaction models QGSJET-II-02 and FLUKA 2002.4 describe the physics of hadronic interactions at these energies reliably<sup>12</sup>. In the near future, new data of the LHC will help to improve the understanding of the hadronic interaction processes, and will finally allow to judge about the quality of the obtained results *ex post facto*, and, if necessary, to apply corrections.

Let us conclude this thesis with a statement Victor Franz Hess made in his Nobel Lecture, on December 12th, 1936, but what is still valid today:

“From a consideration of the immense volume of newly discovered facts in the field of physics, especially atomic physics, in recent years it might well appear to the layman that the main problems were already solved and that only more detailed work was necessary.

This is far from the truth, as will be shown by one of the biggest and most important newly opened fields of research, with which I am closely associated, that of cosmic rays.”

---

<sup>12</sup>Nevertheless, a data Monte Carlo comparison did not give any indication for serious deficits in these models, unlike it was the case for the combination EPOS 1.99 and FLUKA 2002.4.



## Appendix A

# Results of the Parametrization of the $\log_{10}N_{\text{ch}}^{\text{true}}$ Distribution

### A.1 Parameters of the Parametrization of the $\log_{10}N_{\text{ch}}^{\text{true}}$ Distribution

The  $\log_{10}N_{\text{ch}}^{\text{true}}$  distribution is parametrized based on Eq.(5.3). The fit parameters vary from primary to primary. They all show an energy dependence, except for the parameters  $p_3$  and  $p_5$  which can be kept constant over all energies without worsening the fit result. The latter parameter also does not change significantly with the particle type and can be set to  $p_5 = 2.3$ . The parameter  $p_0$  is a simple normalization factor and will thus not be listed. All determined parameters are compiled in Table A.1 and A.2.

particle, energy/eV	$p_3$	$p_5$
H, all energies	0.4	2.3
He, all energies	-4.4	2.3
C, all energies	-4.6	2.3
Si, all energies	-4.8	2.3
Fe, all energies	-5.0	2.3

Table A.1: Values of the parameters  $p_3$  and  $p_5$  of the parametrization of the  $\log_{10}N_{\text{ch}}^{\text{true}}$  distribution based on Eq.(5.3).

particle, energy/eV	$p_1$	$p_2$	$p_4$
H, 2.00e+15	5.1134±0.0056	0.1695±0.0050	6.1299±0.0083
H, 5.00e+15	5.5988±0.0051	0.1514±0.0047	6.5271±0.0066
H, 1.00e+16	5.9708±0.0066	0.1414±0.0058	6.8221±0.0106
H, 3.16e+16	6.5739±0.0065	0.1314±0.0057	7.3174±0.0090
H, 1.00e+17	7.1522±0.0069	0.1059±0.0062	7.8345±0.0109
H, 3.16e+17	7.7444±0.0073	0.0992±0.0064	8.3327±0.0112
H, 1.00e+18	8.3043±0.0089	0.0883±0.0078	8.8610±0.0149
H, 3.16e+18	8.9009±0.0128	0.0916±0.0099	9.3320±0.0176
He, 2.00e+15	5.2902±0.0094	0.1961±0.0039	5.9983±0.0223
He, 5.00e+15	5.7435±0.0075	0.1763±0.0036	6.3980±0.0147
He, 1.00e+16	6.0899±0.0101	0.1601±0.0046	6.6707±0.0208
He, 3.16e+16	6.6433±0.0092	0.1373±0.0046	7.2200±0.0217
He, 1.00e+17	7.2258±0.0086	0.1285±0.0046	7.7034±0.0159
He, 3.16e+17	7.7857±0.0106	0.1133±0.0055	8.2049±0.0192
He, 1.00e+18	8.3417±0.0112	0.1004±0.0060	8.7286±0.0208
He, 3.16e+18	8.8981±0.0172	0.0967±0.0092	9.2814±0.0316
C, 2.00e+15	5.1810±0.0047	0.1488±0.0025	5.7636±0.0078
C, 5.00e+15	5.6700±0.0060	0.1450±0.0030	6.1478±0.0089
C, 1.00e+16	6.0001±0.0058	0.1257±0.0032	6.5157±0.0112
C, 3.16e+16	6.5896±0.0076	0.1183±0.0038	7.0102±0.0131
C, 1.00e+17	7.1596±0.0090	0.1087±0.0046	7.5607±0.0159
C, 3.16e+17	7.7314±0.0090	0.0980±0.0047	8.0791±0.0144
C, 1.00e+18	8.3008±0.0092	0.0956±0.0051	8.6389±0.0133
C, 3.16e+18	8.8531±0.0119	0.0821±0.0065	9.1371±0.0172
Si, 2.00e+15	5.1038±0.0031	0.1159±0.0019	5.5741±0.0027
Si, 5.00e+15	5.5774±0.0034	0.1066±0.0021	6.0360±0.0040
Si, 1.00e+16	5.9261±0.0034	0.0973±0.0024	6.4357±0.0037
Si, 3.16e+16	6.5474±0.0061	0.1038±0.0031	6.8699±0.0079
Si, 1.00e+17	7.1182±0.0059	0.0942±0.0034	7.4485±0.0075
Si, 3.16e+17	7.6750±0.0049	0.0780±0.0032	8.0193±0.0058
Si, 1.00e+18	8.2523±0.0076	0.0710±0.0043	8.5233±0.0124
Si, 3.16e+18	8.8160±0.0115	0.0736±0.0063	9.0699±0.0160
Fe, 2.00e+15	5.0420±0.0029	0.0981±0.0017	5.4161±0.0035
Fe, 5.00e+15	5.5311±0.0033	0.0964±0.0019	5.8814±0.0045
Fe, 1.00e+16	5.8726±0.0030	0.0836±0.0021	6.2941±0.0027
Fe, 3.16e+16	6.4803±0.0044	0.0841±0.0025	6.8143±0.0075
Fe, 1.00e+17	7.0637±0.0051	0.0779±0.0030	7.3711±0.0085
Fe, 3.16e+17	7.6407±0.0059	0.0706±0.0034	7.9176±0.0095
Fe, 1.00e+18	8.2140±0.0058	0.0699±0.0036	8.4833±0.0076
Fe, 3.16e+18	8.7684±0.0089	0.0618±0.0054	9.0446±0.0182

Table A.2: Values of the parameters  $p_1$ ,  $p_2$  and  $p_4$  of the parametrization of the  $\log_{10} N_{\text{ch}}^{\text{true}}$  distribution based on Eq.(5.3).

## A.2 Energy Dependence of the Parameters of the Parametrization of the $\log_{10}N_{\text{ch}}^{\text{true}}$ Distribution – Parameters of the Interpolation

The energy dependence of the parameters  $p_1$ ,  $p_2$  and  $p_4$  of Eq.(5.3) is interpolated using polynomial functions of appropriate degree according to Eq.(5.4). The results are compiled in Table A.3. In Table A.4 the chi-squares per degree of freedom are summarized.

particle, par. $p_i$	parameters of the polynomial fits			
	$a_0$	$a_1$	$a_2$	$a_3$
H, $p_1$	$-5.178 \pm 1.490$	$2.176 \pm 0.584$	$-0.1120 \pm 0.0757$	$0.00408 \pm 0.00324$
H, $p_2$	$0.677 \pm 0.149$	$-0.1168 \pm 0.0391$	$0.00575 \pm 0.00253$	0
H, $p_4$	$2.266 \pm 2.192$	$0.068 \pm 0.858$	$0.117 \pm 0.111$	$-0.00481 \pm 0.00474$
He, $p_1$	$-0.877 \pm 2.222$	$0.722 \pm 0.866$	$0.058 \pm 0.112$	$-0.00265 \pm 0.00475$
He, $p_2$	$0.828 \pm 0.121$	$-0.1465 \pm 0.0321$	$0.00732 \pm 0.00209$	0
He, $p_4$	$0.326 \pm 0.512$	$0.821 \pm 0.133$	$0.01251 \pm 0.00856$	0
C, $p_1$	$-2.755 \pm 0.172$	$1.3354 \pm 0.0454$	$-0.01192 \pm 0.00295$	0
C, $p_2$	$0.4586 \pm 0.0914$	$-0.0679 \pm 0.0241$	$0.00301 \pm 0.00157$	0
C, $p_4$	$3.258 \pm 2.393$	$-0.564 \pm 0.939$	$0.207 \pm 0.121$	$-0.00873 \pm 0.00522$
Si, $p_1$	$-1.980 \pm 0.997$	$0.959 \pm 0.395$	$0.0411 \pm 0.0518$	$-0.00237 \pm 0.00224$
Si, $p_2$	$0.2392 \pm 0.0743$	$-0.0235 \pm 0.0197$	$0.00059 \pm 0.00129$	0
Si, $p_4$	$-12.795 \pm 1.156$	$5.456 \pm 0.463$	$-0.5441 \pm 0.0614$	$0.02235 \pm 0.00269$
Fe, $p_1$	$-1.934 \pm 0.915$	$0.905 \pm 0.362$	$0.0493 \pm 0.0474$	$-0.00274 \pm 0.00205$
Fe, $p_2$	$0.2522 \pm 0.0646$	$-0.0334 \pm 0.0171$	$0.00144 \pm 0.00112$	0
Fe, $p_4$	$-15.844 \pm 1.325$	$6.517 \pm 0.533$	$-0.6742 \pm 0.0710$	$0.02781 \pm 0.00312$

Table A.3: The parameters  $a_i$  of the polynomial fits according to Eq.(5.4) used to interpolate the energy dependence of the parameters  $p_1$ ,  $p_2$  and  $p_4$  of Eq.(5.3).

particle, energy/eV	$\chi^2/ndf$
H, 2.00e+15	81.41/65=1.3
H, 5.00e+15	73.85/61=1.2
H, 1.00e+16	52.98/52=1.0
H, 3.16e+16	53.57/48=1.1
H, 1.00e+17	53.89/42=1.3
H, 3.16e+17	74.06/36=2.1
H, 1.00e+18	27.84/31=0.9
H, 3.16e+18	22.41/25=0.9
He, 2.00e+15	72.92/51=1.4
He, 5.00e+15	58.93/49=1.2
He, 1.00e+16	84.58/43=2.0
He, 3.16e+16	40.48/38=1.1
He, 1.00e+17	25.51/36=0.7
He, 3.16e+17	34.51/29=1.2
He, 1.00e+18	31.07/27=1.2
He, 3.16e+18	26.39/22=1.2
C, 2.00e+15	47.73/48=1.0
C, 5.00e+15	63.34/41=1.5
C, 1.00e+16	36.22/41=0.9
C, 3.16e+16	49.88/33=1.5
C, 1.00e+17	27.88/29=1.0
C, 3.16e+17	30.05/25=1.2
C, 1.00e+18	34.15/24=1.4
C, 3.16e+18	19.41/19=1.0
Si, 2.00e+15	138.16/36=3.8
Si, 5.00e+15	88.20/35=2.5
Si, 1.00e+16	58.03/33=1.8
Si, 3.16e+16	78.16/29=2.7
Si, 1.00e+17	88.23/26=3.4
Si, 3.16e+17	22.29/21=1.1
Si, 1.00e+18	29.33/19=1.5
Si, 3.16e+18	21.56/17=1.3
Fe, 2.00e+15	47.38/30=1.6
Fe, 5.00e+15	125.16/30=4.2
Fe, 1.00e+16	92.46/28=3.3
Fe, 3.16e+16	71.71/26=2.8
Fe, 1.00e+17	26.79/25=1.1
Fe, 3.16e+17	27.33/19=1.4
Fe, 1.00e+18	21.82/20=1.1
Fe, 3.16e+18	7.38/16=0.5

Table A.4: Chi-squares per degree of freedom of the fits to the  $\log_{10}N_{\text{ch}}^{\text{true}}$  distribution based on Eq.(5.3). The parameters have been fixed to the results of the energy dependence interpolation according to Eq.(5.4).

## Appendix B

# Results of the Parametrization of the $\log_{10}N_{\mu}^{\text{true}}$ Distribution

### B.1 Parameters of the Parametrization of the $\log_{10}N_{\mu}^{\text{true}}$ Distribution

The  $\log_{10}N_{\mu}^{\text{true}}$  distribution is fitted with a function according to Eq.(5.11), which itself includes Eqs.(5.7) and (5.8). The fit parameters vary from primary to primary. In addition, the parameters  $p_6$  and  $b_0$  depend on the energy, while the others can be kept constant to an average value for all energies without worsening the fit result. All determined parameters are compiled in the Tables B.1 and B.2.

particle, energy/eV	$p_7$	$b_1$	$b_2$
H, all energies	0.20	-0.69	-0.54
He, all energies	0.06	-0.71	-0.53
C, all energies	0	-0.73	-0.45
Si, all energies	0	-0.75	-0.58
Fe, all energies	0	-0.75	-0.48

Table B.1: Values of the parameters  $p_7$ ,  $b_1$  and  $b_2$  of the parametrization of the  $\log_{10}N_{\mu}^{\text{true}}$  distribution based on Eq.(5.11) going along with Eqs.(5.7) and (5.8).

particle, energy/eV	$p_6$	$b_0$
H, 2.00e+15	0.4243±0.0063	-1.7181±0.0048
H, 5.00e+15	0.4716±0.0074	-2.7479±0.0057
H, 1.00e+16	0.4737±0.0088	-3.6521±0.0069
H, 3.16e+16	0.5381±0.0105	-5.3058±0.0087
H, 1.00e+17	0.5209±0.0122	-7.2407±0.0102
H, 3.16e+17	0.5435±0.0140	-9.3914±0.0120
H, 1.00e+18	0.5553±0.0171	-11.7673±0.0149
H, 3.16e+18	0.5797±0.0245	-14.4459±0.0217
He, 2.00e+15	0.2491±0.0042	-1.6931±0.0030
He, 5.00e+15	0.2597±0.0045	-2.7168±0.0035
He, 1.00e+16	0.2843±0.0054	-3.5903±0.0044
He, 3.16e+16	0.2983±0.0060	-5.2170±0.0050
He, 1.00e+17	0.3140±0.0071	-7.0640±0.0063
He, 3.16e+17	0.3257±0.0081	-9.1663±0.0073
He, 1.00e+18	0.3184±0.0096	-11.4916±0.0087
He, 3.16e+18	0.3182±0.0135	-14.0562±0.0123
C, 2.00e+15	0.0980±0.0039	-0.2909±0.0019
C, 5.00e+15	0.1279±0.0036	-1.0535±0.0022
C, 1.00e+16	0.1396±0.0040	-1.7159±0.0027
C, 3.16e+16	0.1498±0.0043	-2.9539±0.0031
C, 1.00e+17	0.1652±0.0050	-4.3930±0.0038
C, 3.16e+17	0.1778±0.0056	-6.0306±0.0045
C, 1.00e+18	0.1981±0.0068	-7.8793±0.0058
C, 3.16e+18	0.1837±0.0090	-9.9227±0.0076
Si, 2.00e+15	0.1639±0.0027	-2.9192±0.0020
Si, 5.00e+15	0.1765±0.0031	-4.1171±0.0024
Si, 1.00e+16	0.1782±0.0037	-5.1302±0.0029
Si, 3.16e+16	0.1860±0.0041	-7.0015±0.0033
Si, 1.00e+17	0.1972±0.0049	-9.1248±0.0041
Si, 3.16e+17	0.2015±0.0055	-11.5041±0.0048
Si, 1.00e+18	0.1950±0.0065	-14.1361±0.0056
Si, 3.16e+18	0.1945±0.0091	-17.0150±0.0079
Fe, 2.00e+15	0.0933±0.0023	-1.1109±0.0014
Fe, 5.00e+15	0.0992±0.0026	-1.9748±0.0016
Fe, 1.00e+16	0.1018±0.0031	-2.7194±0.0020
Fe, 3.16e+16	0.1106±0.0034	-4.1198±0.0023
Fe, 1.00e+17	0.1230±0.0039	-5.7248±0.0029
Fe, 3.16e+17	0.1310±0.0043	-7.5373±0.0034
Fe, 1.00e+18	0.1271±0.0051	-9.5665±0.0040
Fe, 3.16e+18	0.1300±0.0075	-11.7959±0.0059

Table B.2: Values of the parameters  $p_6$  and  $b_0$  of the parametrization of the  $\log_{10}N_{\mu}^{\text{true}}$  distribution based on Eq.(5.11) going along with Eqs.(5.7) and (5.8).

## B.2 Energy Dependence of the Parameters of the Parametrization of the $\log_{10}N_{\mu}^{\text{true}}$ Distribution – Parameters of the Interpolation

The energy dependence of the parameters  $p_6$  and  $b_0$  of Eq.(5.11) going along with Eqs.(5.7) and (5.8) is interpolated using polynomial functions of appropriate degree according to Eq.(5.4). The results are compiled in Table B.3. In Table B.4 the chi-squares per degree of freedom are summarized.

particle, par. $p_i$	parameters of the polynomial fits			
	$a_0$	$a_1$	$a_2$	$a_3$
H, $p_6$	$-4.357 \pm 2.296$	$1.711 \pm 0.911$	$-0.202 \pm 0.120$	$0.00802 \pm 0.00518$
H, $b_0$	$-5.802 \pm 0.217$	$3.7150 \pm 0.0583$	$-0.48672 \pm 0.00387$	0
He, $p_6$	$0.852 \pm 1.370$	$-0.358 \pm 0.541$	$0.0612 \pm 0.0708$	$-0.00310 \pm 0.00305$
He, $b_0$	$-5.062 \pm 0.130$	$3.4500 \pm 0.0349$	$-0.46271 \pm 0.00230$	0
C, $p_6$	$-0.547 \pm 0.120$	$0.1530 \pm 0.0310$	$-0.00800 \pm 0.00206$	0
C, $b_0$	$-1.971 \pm 0.703$	$2.046 \pm 0.279$	$-0.2408 \pm 0.0367$	$-0.00660 \pm 0.00159$
Si, $p_6$	$-0.224 \pm 0.105$	$0.0960 \pm 0.0281$	$-0.00546 \pm 0.00185$	0
Si, $b_0$	$-5.0862 \pm 0.0851$	$3.4934 \pm 0.0228$	$-0.49986 \pm 0.00150$	0
Fe, $p_6$	$-0.1130 \pm 0.0870$	$0.0454 \pm 0.0232$	$-0.00204 \pm 0.00153$	0
Fe, $b_0$	$-4.9570 \pm 0.0604$	$3.2309 \pm 0.0162$	$-0.41588 \pm 0.00107$	0

Table B.3: The parameters  $a_i$  of the polynomial fits according to Eq.(5.4) used to interpolate the energy dependence of the parameters  $p_6$  and  $b_0$  of Eq.(5.11) going along with Eqs.(5.7) and (5.8).

particle, energy/eV	$\chi^2/ndf$
H, 2.00e+15	43.28/42=1.0
H, 5.00e+15	35.89/41=0.9
H, 1.00e+16	44.40/32=1.4
H, 3.16e+16	30.19/33=0.9
H, 1.00e+17	21.80/27=0.8
H, 3.16e+17	28.73/26=1.1
H, 1.00e+18	28.59/25=1.1
H, 3.16e+18	13.25/20=0.7
He, 2.00e+15	23.85/26=0.9
He, 5.00e+15	21.60/21=1.0
He, 1.00e+16	12.50/20=0.6
He, 3.16e+16	12.36/18=0.7
He, 1.00e+17	15.54/17=0.9
He, 3.16e+17	7.90/13=0.6
He, 1.00e+18	6.11/14=0.4
He, 3.16e+18	2.85/11=0.3
C, 2.00e+15	18.25/19=1.0
C, 5.00e+15	29.18/16=1.8
C, 1.00e+16	20.36/16=1.3
C, 3.16e+16	14.18/13=1.1
C, 1.00e+17	12.47/12=1.0
C, 3.16e+17	7.86/10=0.8
C, 1.00e+18	5.86/10=0.6
C, 3.16e+18	7.08/8=0.9
Si, 2.00e+15	15.59/15=1.0
Si, 5.00e+15	9.44/14=0.7
Si, 1.00e+16	17.13/13=1.3
Si, 3.16e+16	5.88/11=0.5
Si, 1.00e+17	5.66/9=0.6
Si, 3.16e+17	3.53/9=0.4
Si, 1.00e+18	8.78/7=1.3
Si, 3.16e+18	3.52/7=0.5
Fe, 2.00e+15	14.28/13=1.1
Fe, 5.00e+15	19.53/13=1.5
Fe, 1.00e+16	8.64/12=0.7
Fe, 3.16e+16	10.93/10=1.1
Fe, 1.00e+17	6.06/8=0.8
Fe, 3.16e+17	7.30/9=0.8
Fe, 1.00e+18	10.34/8=1.3
Fe, 3.16e+18	2.47/6=0.4

Table B.4: Chi-squares per degree of freedom of the fits to the  $\log_{10}N_{\mu}^{\text{true}}$  distribution based on Eq.(5.11) going along with Eqs.(5.7) and (5.8). The parameters have been fixed to the results of the energy dependence interpolation according to Eq.(5.4).

## Appendix C

# Results of the Parametrization of the Efficiency

The parameters of the parametrization (Eq.(5.14)) of the combined trigger and reconstruction efficiency are listed in Table C.1.

part.	$p_0$	$p_1$	$p_2$	$p_3$	$\chi^2/ndf$
H	$2.400\pm 0.043$	$3.9841\pm 0.0067$	$3.302\pm 0.084$	$5.0269\pm 0.0051$	$872.2/465=1.9$
He	$3.210\pm 0.090$	$4.1038\pm 0.0087$	$3.026\pm 0.077$	$4.9907\pm 0.0074$	$802.5/323=2.5$
C	$3.046\pm 0.176$	$4.1634\pm 0.0126$	$3.056\pm 0.137$	$4.9848\pm 0.0092$	$578.7/246=2.4$
Si	$3.181\pm 0.088$	$4.3036\pm 0.0098$	$3.478\pm 0.326$	$4.8999\pm 0.0122$	$497.9/207=2.4$
Fe	$3.213\pm 0.111$	$4.3524\pm 0.0129$	$3.160\pm 0.114$	$4.8976\pm 0.0158$	$603.7/185=3.3$

Table C.1: Values of the parameters of the parametrization of the combined trigger and reconstruction efficiency based on Eq.(5.14).

## Appendix D

# Results of the Parametrization of the Systematic Bias in the Reconstruction Procedure

The parameters of the parametrization of the systematic bias in the charged particle or muon number reconstruction based on Eq.(5.15) and Eq.(5.16) are listed in Table D.1 and D.2.

parameter	value
$p_0$	$5.802 \pm 0.026$
$p_1$	$0.1294 \pm 0.0076$
$p_2$	$-0.0233 \pm 0.0010$
$p_3$	$-0.01273 \pm 0.00080$
$\chi^2/ndf$	$383.2/58=6.6$

Table D.1: Values of the parameters of the parametrization of the systematic bias in the charged particle number reconstruction based on Eq.(5.15).

parameter	value
$p_0$	$6.057 \pm 0.020$
$p_1$	$-0.0040 \pm 0.0014$
$p_2$	$0.2657 \pm 0.0088$
$p_3$	$0.295 \pm 0.013$
$p_4$	$0.0829 \pm 0.0049$
$p_5$	$0.0269 \pm 0.0033$
$\chi^2/ndf$	$109.1/57=1.9$

Table D.2: Values of the parameters of the parametrization of the systematic bias in the muon number reconstruction based on Eq.(5.16).

## Appendix E

# Results of the Parametrization of the Statistical Reconstruction Uncertainties for Charged Particles

### E.1 Parameters of the Parametrization of the Statistical Reconstruction Uncertainties for Charged Particles

The derived fit parameters, used in the parametrization (given by Eqs.(5.17) and (5.18)) of the statistical reconstruction uncertainties for charged particles, are listed in Table E.1.

$\log_{10}(N_{\text{ch}}^{\text{tru}})$ -interval	$p_1$	$p_2$	$p_3$
[4.5, 4.6]	0.0472±0.0001	0.0973±0.0359	-0.0950±0.1915
[4.6, 4.7]	0.0445±0.0001	0.1726±0.0000	-0.0950±0.1743
[4.7, 4.8]	0.0005±0.0049	0.1088±0.0044	-0.0891±0.0076
[4.8, 4.9]	-0.0124±0.0031	0.1087±0.0025	-0.0926±0.0049
[4.9, 5.0]	-0.0217±0.0023	0.1066±0.0019	-0.0942±0.0036
[5.0, 5.1]	-0.0236±0.0017	0.1031±0.0015	-0.0911±0.0027
[5.1, 5.2]	-0.0271±0.0016	0.1000±0.0014	-0.0972±0.0023
[5.2, 5.3]	-0.0269±0.0015	0.0938±0.0012	-0.1004±0.0020
[5.3, 5.4]	-0.0237±0.0014	0.0904±0.0012	-0.0921±0.0019
[5.4, 5.5]	-0.0188±0.0014	0.0893±0.0013	-0.0917±0.0020
[5.5, 5.6]	-0.0151±0.0015	0.0851±0.0014	-0.0888±0.0020
[5.6, 5.7]	-0.0128±0.0015	0.0857±0.0013	-0.0803±0.0021
[5.7, 5.8]	-0.0114±0.0015	0.0821±0.0014	-0.0769±0.0020
[5.8, 5.9]	-0.0117±0.0018	0.0816±0.0016	-0.0724±0.0025
[5.9, 6.0]	-0.0138±0.0017	0.0769±0.0015	-0.0675±0.0025
[6.0, 6.1]	-0.0106±0.0018	0.0768±0.0018	-0.0627±0.0029
[6.1, 6.2]	-0.0125±0.0021	0.0727±0.0018	-0.0657±0.0027
[6.2, 6.3]	-0.0052±0.0020	0.0737±0.0016	-0.0551±0.0029
[6.3, 6.4]	-0.0104±0.0021	0.0695±0.0017	-0.0489±0.0030
[6.4, 6.5]	-0.0072±0.0024	0.0704±0.0019	-0.0467±0.0044
[6.5, 6.6]	-0.0079±0.0024	0.0658±0.0020	-0.0463±0.0040
[6.6, 6.7]	-0.0082±0.0028	0.0667±0.0022	-0.0462±0.0044
[6.7, 6.8]	-0.0064±0.0024	0.0688±0.0018	-0.0234±0.0052
[6.8, 6.9]	-0.0076±0.0034	0.0666±0.0029	-0.0404±0.0071
[6.9, 7.0]	-0.0025±0.0031	0.0646±0.0027	-0.0386±0.0068
[7.0, 7.1]	-0.0058±0.0038	0.0614±0.0034	-0.0440±0.0067
[7.1, 7.2]	-0.0055±0.0035	0.0623±0.0027	-0.0304±0.0099
[7.2, 7.3]	0.0013±0.0037	0.0591±0.0025	-0.0017±0.0142
[7.3, 7.4]	0.0003±0.0050	0.0627±0.0037	0
[7.4, 7.5]	-0.0065±0.0056	0.0699±0.0050	0
[7.5, 7.6]	-0.0159±0.0051	0.0637±0.0040	0
[7.6, 7.7]	-0.0100±0.0058	0.0542±0.0048	0
[7.7, 7.8]	-0.0075±0.0068	0.0618±0.0060	0
[7.8, 7.9]	0.0140±0.0071	0.0624±0.0045	0
[7.9, 8.0]	-0.0153±0.0070	0.0647±0.0059	0
[8.0, 8.1]	-0.0026±0.0104	0.0707±0.0099	0
[8.1, 8.2]	0.0039±0.0072	0.0584±0.0048	0
[8.2, 8.3]	-0.0234±0.0168	0.0793±0.0161	0
[8.3, 8.4]	0.0149±0.0106	0.0615±0.0111	0
[8.4, 8.5]	-0.0399±0.0026	0.0700±0.0188	0
[8.5, 8.6]	-0.0591±0.0050	0.0700±0.0195	0
[8.6, 8.7]	-0.0457±0.0050	0.0700±0.0100	0
[8.7, 8.8]	0.0400±0.0039	0.0700±0.0142	0
[8.8, 8.9]	-0.0648±0.0035	0.0700±0.0109	0

Table E.1: Values of the parameters  $p_1$ ,  $p_2$  and  $p_3$  of the parametrization (according to Eqs.(5.17) and (5.18)) of the charged particle reconstruction uncertainty.

## E.2 Charged Particle Number Dependence of the Parameters of the Parametrization of the Statistical Reconstruction Uncertainties for Charged Particles – Parameters of the Interpolation

The dependence of the parameters  $p_1$ ,  $p_2$  and  $p_3$  – used in the parametrization (cf. Eqs.(5.17) and (5.18)) of the statistical uncertainties of the charged particle reconstruction – on the true charged particle number is interpolated by means of the functions given by Eq.(5.19). The results are summarized in Tables E.2, E.3 and E.4. In Table E.5 the chi-squares per degree of freedom are summarized in case of the fits (according to Eqs.(5.17) and (5.18)) with interpolated parameters (according to Eq.(5.19)) to the distributions of the charged particle reconstruction uncertainties.

parameter	value
$c_0$	$-0.0266000 \pm 0.0000023$
$c_1$	$0.0261 \pm 0.0017$
$c_2$	$5.160000 \pm 0.000020$
$c_3$	$-0.0081 \pm 0.0017$
$c_4$	$0.00009 \pm 0.00038$
$c_5$	$0.21200 \pm 0.00024$
$\chi^2/ndf$	$56.2/38=1.5$

Table E.2: Values of the parameters used in the parametrization (according to Eq.(5.19)) of the dependence of the parameter  $p_1$  (from Eqs.(5.17) and (5.18)) on the true charged particle number.

parameter	value
$d_0$	$0.0617 \pm 0.0013$
$d_1$	$-4.427 \pm 0.089$
$d_2$	$52.40 \pm 7.54$
$d_3$	$7.17 \pm 0.64$
$d_4$	$1.09 \pm 0.10$
$\chi^2/ndf$	$31.8/39=0.8$

Table E.3: Values of the parameters used in the parametrization (according to Eq.(5.19)) of the dependence of the parameter  $p_2$  (from Eqs.(5.17) and (5.18)) on the true charged particle number.

parameter	value
$e_0$	$-0.0917 \pm 0.0011$
$e_1$	$-9.83 \pm 4.69$
$e_2$	$5.40 \pm 0.15$
$e_3$	$-0.0043 \pm 0.0017$
$e_4$	$0.0403 \pm 0.0017$
$\chi^2/ndf$	$44.0/23=1.9$

Table E.4: Values of the parameters used in the parametrization (according to Eq.(5.19)) of the dependence of the parameter  $p_3$  (from Eqs.(5.17) and (5.18)) on the true charged particle number.

$\log_{10}(N_{\text{ch}}^{\text{tru}})$ -interval	$\chi^2/ndf$
[4.5, 4.6]	2.2/4=0.6
[4.6, 4.7]	16.8/12=1.4
[4.7, 4.8]	22.0/19=1.2
[4.8, 4.9]	25.9/29=0.9
[4.9, 5.0]	57.9/35=1.7
[5.0, 5.1]	67.7/30=2.3
[5.1, 5.2]	59.3/36=1.6
[5.2, 5.3]	67.9/33=2.1
[5.3, 5.4]	68.4/34=2.0
[5.4, 5.5]	88.3/34=2.6
[5.5, 5.6]	72.7/32=2.3
[5.6, 5.7]	49.1/30=1.6
[5.7, 5.8]	44.5/28=1.6
[5.8, 5.9]	53.8/28=1.9
[5.9, 6.0]	43.6/23=1.9
[6.0, 6.1]	35.2/20=1.8
[6.1, 6.2]	16.6/20=0.8
[6.2, 6.3]	9.8/18=0.5
[6.3, 6.4]	31.6/16=2.0
[6.4, 6.5]	9.1/15=0.6
[6.5, 6.6]	12.1/16=0.8
[6.6, 6.7]	6.5/13=0.5
[6.7, 6.8]	25.7/12=2.1
[6.8, 6.9]	12.5/12=1.0
[6.9, 7.0]	7.6/11=0.7
[7.0, 7.1]	11.1/11=1.0
[7.1, 7.2]	7.4/10=0.7
[7.2, 7.3]	15.0/10=1.5
[7.3, 7.4]	7.2/9=0.8
[7.4, 7.5]	7.7/9=0.9
[7.5, 7.6]	8.1/8=1.0
[7.6, 7.7]	8.2/8=1.0
[7.7, 7.8]	4.0/7=0.6
[7.8, 7.9]	28.8/8=3.6
[7.9, 8.0]	2.6/9=0.3
[8.0, 8.1]	9.5/8=1.2
[8.1, 8.2]	20.6/8=2.6
[8.2, 8.3]	6.1/7=0.9
[8.3, 8.4]	14.7/9=1.6
[8.4, 8.5]	1.1/4=0.3
[8.5, 8.6]	0.9/3=0.3
[8.6, 8.7]	0.6/2=0.3
[8.7, 8.8]	0/0
[8.8, 8.9]	0/0

Table E.5: Chi-squares per degree of freedom of the fits (according to Eqs.(5.17) and (5.18)) with interpolated parameters (according to Eq.(5.19)) to the distributions of the charged particle reconstruction uncertainties.

## Appendix F

# Results of the Parametrization of the Statistical Reconstruction Uncertainties for Muons

### F.1 Parameters of the Parametrization of the Statistical Reconstruction Uncertainties for Muons

The derived fit parameters, used in the parametrization (given by Eqs.(5.20) and (5.21)) of the statistical reconstruction uncertainties for muons, are listed in Table F.1.

$\log_{10}(N_{\mu}^{\text{tru}})$ -interval	$p_1$	$p_2$	$p_3$
[3.6, 3.7]	0.0000±0.1983	0.2850±0.0326	0.0500±0.0181
[3.7, 3.8]	-0.0932±0.1506	0.2850±0.0607	0.0500±0.0181
[3.8, 3.9]	-0.0025±0.1758	0.2850±0.0619	0.0500±0.0181
[3.9, 4.0]	0.0156±0.0206	0.2524±0.0200	0.1836±0.0584
[4.0, 4.1]	0.0045±0.0218	0.2427±0.0158	0.1529±0.0518
[4.1, 4.2]	0.0182±0.0050	0.2282±0.0036	0.1696±0.0067
[4.2, 4.3]	0.0374±0.0036	0.1997±0.0028	0.1779±0.0037
[4.3, 4.4]	0.0412±0.0026	0.1867±0.0021	0.1716±0.0029
[4.4, 4.5]	0.0394±0.0026	0.1691±0.0021	0.1772±0.0032
[4.5, 4.6]	0.0392±0.0024	0.1578±0.0020	0.1594±0.0029
[4.6, 4.7]	0.0277±0.0022	0.1456±0.0020	0.1391±0.0028
[4.7, 4.8]	0.0232±0.0021	0.1356±0.0019	0.1168±0.0027
[4.8, 4.9]	0.0150±0.0020	0.1258±0.0018	0.0981±0.0030
[4.9, 5.0]	0.0092±0.0020	0.1170±0.0018	0.0821±0.0032
[5.0, 5.1]	0.0021±0.0020	0.1020±0.0017	0.0748±0.0029
[5.1, 5.2]	0.0003±0.0019	0.0938±0.0016	0.0592±0.0033
[5.2, 5.3]	0.0020±0.0020	0.0832±0.0018	0.0527±0.0036
[5.3, 5.4]	0.0023±0.0019	0.0775±0.0017	0.0398±0.0048
[5.4, 5.5]	0.0027±0.0020	0.0726±0.0014	0.0310±0.0047
[5.5, 5.6]	0.0000±0.0022	0.0687±0.0017	0.0298±0.0058
[5.6, 5.7]	0.0007±0.0219	0.0636±0.0012	0.0089±0.0365
[5.7, 5.8]	-0.0015±0.0021	0.0545±0.0015	0
[5.8, 5.9]	-0.0043±0.0025	0.0517±0.0019	0
[5.9, 6.0]	0.0020±0.0025	0.0484±0.0021	0
[6.0, 6.1]	0.0010±0.0028	0.0455±0.0021	0
[6.1, 6.2]	0.0027±0.0033	0.0427±0.0029	0
[6.2, 6.3]	-0.0055±0.0044	0.0495±0.0033	0
[6.3, 6.4]	-0.0096±0.0044	0.0368±0.0029	0
[6.4, 6.5]	0.0055±0.0037	0.0334±0.0028	0
[6.5, 6.6]	-0.0025±0.0039	0.0391±0.0026	0
[6.6, 6.7]	0.0079±0.0034	0.0278±0.0023	0
[6.7, 6.8]	0.0017±0.0040	0.0305±0.0033	0
[6.8, 6.9]	-0.0004±0.0043	0.0269±0.0041	0
[6.9, 7.0]	0.0042±0.0059	0.0257±0.0049	0
[7.0, 7.1]	-0.0240±0.0504	0.0356±0.0298	0
[7.1, 7.2]	-0.0030±0.9932	0.0249±0.7268	0
[7.2, 7.3]	-0.0200±0.2768	0.0883±9.0902	0

Table F.1: Values of the parameters  $p_1$ ,  $p_2$  and  $p_3$  of the parametrization (according to Eqs.(5.20) and (5.21)) of the muon reconstruction uncertainty.

## F.2 Muon Number Dependence of the Parameters of the Parametrization of the Statistical Reconstruction Uncertainties for Muons – Parameters of the Interpolation

The dependence of the parameters  $p_1$ ,  $p_2$  and  $p_3$  – used in the parametrization (cf. Eqs.(5.20) and (5.21)) of the statistical uncertainties of the muon reconstruction – on the true muon number is interpolated by means of the functions given by Eq.(5.22). The results are summarized in Tables F.2, F.3 and F.4. In Table F.5 the chi-squares per degree of freedom are summarized in case of the fits (according to Eqs.(5.20) and (5.21)) with interpolated parameters (according to Eq.(5.22)) to the distributions of the muons reconstruction uncertainties.

parameter	value
$f_0$	0
$f_1$	$-0.023 \pm 0.068$
$f_2$	$5.12 \pm 0.14$
$f_3$	$0.17 \pm 0.12$
$f_4$	$0.168 \pm 0.038$
$\chi^2/ndf$	$33.3/33=1.0$

Table F.2: Values of the parameters used in the parametrization (according to Eq.(5.22)) of the dependence of the parameter  $p_1$  (from Eqs.(5.20) and (5.21)) on the true muon number.

parameter	value
$g_0$	$0.0249 \pm 0.0026$
$g_1$	$-2.92 \pm 0.14$
$g_2$	$14.29 \pm 2.25$
$g_3$	$3.50 \pm 0.45$
$g_4$	$0.650 \pm 0.079$
$\chi^2/ndf$	$51.4/32=1.6$

Table F.3: Values of the parameters used in the parametrization (according to Eq.(5.22)) of the dependence of the parameter  $p_2$  (from Eqs.(5.20) and (5.21)) on the true muon number.

parameter	value
$h_0$	$0.089 \pm 0.021$
$h_1$	$-1.26 \pm 0.61$
$h_2$	$4.91 \pm 0.16$
$h_3$	$0.084 \pm 0.036$
$h_4$	$-0.11 \pm 0.17$
$h_5$	$0.60 \pm 0.50$
$h_6$	$1.19 \pm 0.34$
$h_7$	$0.49 \pm 0.15$
$\chi^2/ndf$	$17.7/13=1.4$

Table F.4: Values of the parameters used in the parametrization (according to Eq.(5.22)) of the dependence of the parameter  $p_3$  (from Eqs.(5.20) and (5.21)) on the true muon number.

$\log_{10}(N_{\mu}^{\text{tru}})$ -interval	$\chi^2/ndf$
[3.6, 3.7]	0.7/2=0.4
[3.7, 3.8]	1.1/7=0.2
[3.8, 3.9]	15.6/16=1.0
[3.9, 4.0]	16.7/20=0.8
[4.0, 4.1]	55.4/25=2.2
[4.1, 4.2]	152.3/29=5.3
[4.2, 4.3]	241.8/29=8.3
[4.3, 4.4]	167.3/29=5.8
[4.4, 4.5]	81.9/26=3.2
[4.5, 4.6]	96.8/30=3.2
[4.6, 4.7]	56.2/28=2.0
[4.7, 4.8]	65.0/27=2.4
[4.8, 4.9]	28.6/22=1.3
[4.9, 5.0]	34.9/20=1.7
[5.0, 5.1]	24.8/21=1.2
[5.1, 5.2]	12.6/15=0.8
[5.2, 5.3]	32.6/17=1.9
[5.3, 5.4]	25.4/16=1.6
[5.4, 5.5]	18.8/16=1.2
[5.5, 5.6]	16.0/13=1.2
[5.6, 5.7]	16.8/10=1.7
[5.7, 5.8]	7.1/10=0.7
[5.8, 5.9]	8.8/7=1.3
[5.9, 6.0]	3.6/7=0.5
[6.0, 6.1]	0.4/6=0.1
[6.1, 6.2]	8.6/6=1.4
[6.2, 6.3]	15.4/5=3.1
[6.3, 6.4]	12.3/5=2.5
[6.4, 6.5]	10.6/5=2.1
[6.5, 6.6]	11.2/5=2.2
[6.6, 6.7]	8.6/4=2.2
[6.7, 6.8]	1.9/4=0.5
[6.8, 6.9]	0.9/3=0.3
[6.9, 7.0]	0.9/2=0.4
[7.0, 7.1]	0.1/1=0.1
[7.1, 7.2]	0.0/1=0.0
[7.2, 7.3]	0.4/1=0.4

Table F.5: Chi-squares per degree of freedom of the fits (according to Eqs.(5.20) and (5.21)) with interpolated parameters (according to Eq.(5.22)) to the distributions of the muons reconstruction uncertainties.

## Appendix G

# Entries of the Diagonal Matrix of the Response Matrix Factorization by SVD

The response matrix  $\mathbf{R}$ , defined in Eq.(5.26), was factorized by means of the SVD (cf. Eqs.(5.28) and (5.29)). The diagonal matrix  $\mathbf{D}$ , if the response matrix  $\mathbf{R}$  bases on all five primaries, is given by:

$$\begin{aligned}
D = \text{diag}( & 0.5422723667031193, 0.520103458246924, 0.49032600811023486, \\
& 0.45956993058320456, 0.42747353015486944, 0.3910380306670524, \\
& 0.35237908132838097, 0.3147429614398176, 0.3091031618504913, \\
& 0.2891468441149333, 0.2804111344910435, 0.2744934240985449, \\
& 0.26229102046318087, 0.2436729141806119, 0.24176842535923565, \\
& 0.2243505768178768, 0.2069164177080118, 0.20087508727396647, \\
& 0.17880489586724357, 0.1672052136581474, 0.1589656580233755, \\
& 0.15048884446946556, 0.139447274652344, 0.13747527300893722, \\
& 0.13619306978994655, 0.12780852385329297, 0.12259374664420425, \\
& 0.11750440480898337, 0.11241760351156814, 0.10394196794087762, \\
& 0.09493175375752856, 0.09345536180436431, 0.0841595281398968, \\
& 0.08207149713157465, 0.07576044369605106, 0.07155392351389586, \\
& 0.06602555575168623, 0.05949587457022103, 0.05441879253913666, \\
& 0.052989155404462904, 0.05005039614819152, 0.04601666928794832, \\
& 0.04437243533655123, 0.04370276335764391, 0.037810904095364174, \\
& 0.03622393260000482, 0.03417558784282055, 0.03161797648150007, \\
& 0.031353032122250525, 0.028066225920178177, 0.027339048243654993, \\
& 0.02612139255441941, 0.02253393846073224, 0.019620602776118518, \\
& 0.01949516874748848, 0.01728737165829651, 0.015202879561481837, \\
& 0.014570497074600662, 0.013249820107033931, 0.013166680832744125, \\
& 0.01252322491027643, 0.011650205170088019, 0.011434419496012893, \\
& 0.010961158540018362, 0.010127654467138442, 0.009713115723809628, \\
& 0.008885560042760306, 0.008054197526291003, 0.007367094517174988, \\
& 0.007274261244150086, 0.006748783495513325, 0.006045254909847775, \\
& 0.00527442156938896, 0.004593922691969088, 0.00426448121591894, \\
& 0.0032927445743306076, 0.002357553580842307, 0.0019607552812931205, \\
& 0.0012673756702215977, 0.0008082797395408309, 0.0005562283620021495, \\
& 0.0004253750365712129, 0.00021616084318648774, 0.00016454890749190165, \\
& 0.0000743082764299595, 0.0000446673297901509, 0.000024007604906245624, \\
& 0.000019911472889809562, 0.000014979659723874907, 0.000012913741341030732, \\
& 5.7639003397369064 \times 10^{-6}, 2.66475928364136 \times 10^{-6}, 1.9498722982342773 \times 10^{-6}, \\
& 7.448094851607356 \times 10^{-7}, 3.654329601037045 \times 10^{-7}, 1.7718407912301687 \times 10^{-7}, \\
& 1.3265426658523495 \times 10^{-7}, 2.257089572371777 \times 10^{-8}, 6.072746485998527 \times 10^{-9}, \\
& 2.0243870308037545 \times 10^{-9}, 1.4461609513714547 \times 10^{-9}, 2.5886593111196026 \times 10^{-10}, \\
& 9.169213831211122 \times 10^{-11}, 1.9396462672767594 \times 10^{-11}, 1.70657841869029 \times 10^{-11}, \\
& 2.632445240129791 \times 10^{-12}, 1.555982061690543 \times 10^{-12}, 4.619093332701857 \times 10^{-13}, \\
& 2.809692190026773 \times 10^{-13}, 1.1697048262018725 \times 10^{-13} ),
\end{aligned}
\tag{G.1}$$

before, and by:

$$\begin{aligned}
 \mathbf{D} = \text{diag}( & \\
 & 0.5410315651402106, 0.5192058500570461, 0.4897296699375651, \\
 & 0.45937495463932126, 0.42764652071028547, 0.3915330857066353, \\
 & 0.353083050839372, 0.31538748806031425, 0.309419023131552, \\
 & 0.2894793531411193, 0.2806426875054452, 0.274846580266015, \\
 & 0.26271865575312314, 0.24409979470818544, 0.24217891282061949, \\
 & 0.2247925592137016, 0.2072417193447789, 0.20121026301119, \\
 & 0.17915457868719556, 0.16732946331909762, 0.1589624937312841, \\
 & 0.15037288845655214, 0.13936513981540394, 0.13740055827964975, \\
 & 0.13618994546562013, 0.12781850725202834, 0.12260390488479823, \\
 & 0.11727375655469845, 0.11241177738741628, 0.10396309904439155, \\
 & 0.09448495036416728, 0.09340527243212254, 0.08406813428782303, \\
 & 0.08179295775131144, 0.07576321213684786, 0.071382421079445, \\
 & 0.06574215999765871, 0.05920718202019056, 0.05448516134447801, \\
 & 0.052702052119786756, 0.049865048303919565, 0.0461168885829859, \\
 & 0.0443735073337368, 0.04346470941693313, 0.03782172675941715, \\
 & 0.036172107253546124, 0.034165358085424626, 0.03170012090752251, \\
 & 0.03115689721226761, 0.02807398345805006, 0.027203011975072303, \\
 & 0.02607236257798598, 0.02252985488427554, 0.019494123420967818, \\
 & 0.019345744064621872, 0.017328540568230206, 0.015128223524993569, \\
 & 0.01459256957068736, 0.013299449553346557, 0.013146636421176696, \\
 & 0.012489479033455599, 0.011714548726624316, 0.011376029708766206, \\
 & 0.010748219025655583, 0.010090933556765356, 0.00974073036416497, \\
 & 0.008865754162545253, 0.008137496783872815, 0.007432787472103774, \\
 & 0.007299608808111442, 0.00675706911594217, 0.006072762666831561, \\
 & 0.005268643561534597, 0.004654824911593205, 0.004309572878309808, \\
 & 0.003356654488788354, 0.0026061748010492, 0.0020912319718943864, \\
 & 0.0014102208758410093, 0.0011156093206293699, 0.0008284191465039228, \\
 & 0.0007219049954235829, 0.000601919913399369, 0.0005849968976307148, \\
 & 0.00044795157265295276, 0.00037069364963798057, 0.0002864119499741611, \\
 & 0.0002606859245936317, 0.00006007260841748, 0.000040742627295298026, \\
 & 2.269220698818733 \times 10^{-7} ),
 \end{aligned} \tag{G.2}$$

after applying cuts to improve the condition number of the response matrix.

## Appendix H

# Values of the Elemental Energy Spectra

The differential flux values of the unfolded energy spectra for elemental groups of cosmic rays and their uncertainties, depicted in Fig. 7.3, are listed in Table H.1 to H.6.

energy $E/\text{GeV}$	$(dJ/dE \pm \sigma_{\text{stat.}} \pm \Delta_{\text{syst.}}) / (\text{m}^{-2}\text{sr}^{-1}\text{s}^{-1}\text{GeV}^{-1})$
$1.12 \times 10^7$	$(2.11 \pm 0.68 \pm 0.93) \times 10^{-16}$
$1.55 \times 10^7$	$(5.87 \pm 1.31 \pm 2.54) \times 10^{-17}$
$2.14 \times 10^7$	$(1.78 \pm 0.35 \pm 0.63) \times 10^{-17}$
$2.95 \times 10^7$	$(7.13 \pm 1.12 \pm 1.69) \times 10^{-18}$
$4.07 \times 10^7$	$(2.99 \pm 0.69 \pm 0.80) \times 10^{-18}$
$5.62 \times 10^7$	$(1.27 \pm 0.33 \pm 0.29) \times 10^{-18}$
$7.76 \times 10^7$	$(2.79 \pm 1.34 \pm 0.89) \times 10^{-19}$
$1.07 \times 10^8$	$(7.87 \pm 4.41 \pm 0.16) \times 10^{-20}$
$1.48 \times 10^8$	$(1.18 \pm 0.55 \pm 0.54) \times 10^{-19}$
$2.04 \times 10^8$	$(5.30 \pm 3.68 \pm 7.49) \times 10^{-20}$
$2.82 \times 10^8$	$(1.64 \pm 2.40 \pm 5.45) \times 10^{-20}$
$3.89 \times 10^8$	$(2.25 \pm 1.81 \pm 4.27) \times 10^{-20}$

Table H.1: Differential flux values  $dJ/dE$  of the unfolded energy spectrum of **protons** depicted in Fig. 7.3.  $\sigma_{\text{stat.}}$  respectively  $\Delta_{\text{syst.}}$  are the absolute statistical respectively systematic uncertainties. The result bases on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68].

energy $E/\text{GeV}$	$(dJ/dE \pm \sigma_{\text{stat.}} \pm \Delta_{\text{syst.}}) / (\text{m}^{-2}\text{sr}^{-1}\text{s}^{-1}\text{GeV}^{-1})$
$1.12 \times 10^7$	$(5.75 \pm 0.72 \pm 1.98) \times 10^{-16}$
$1.55 \times 10^7$	$(1.43 \pm 0.19 \pm 0.63) \times 10^{-16}$
$2.14 \times 10^7$	$(3.72 \pm 0.65 \pm 2.02) \times 10^{-17}$
$2.95 \times 10^7$	$(1.05 \pm 0.20 \pm 0.62) \times 10^{-17}$
$4.07 \times 10^7$	$(3.28 \pm 0.58 \pm 1.98) \times 10^{-18}$
$5.62 \times 10^7$	$(1.62 \pm 0.36 \pm 0.76) \times 10^{-18}$
$7.76 \times 10^7$	$(3.55 \pm 1.12 \pm 1.76) \times 10^{-19}$
$1.07 \times 10^8$	$(1.03 \pm 0.42 \pm 1.85) \times 10^{-19}$
$1.48 \times 10^8$	$(1.11 \pm 0.35 \pm 0.42) \times 10^{-19}$
$2.04 \times 10^8$	$(2.22 \pm 1.19 \pm 8.04) \times 10^{-20}$
$2.82 \times 10^8$	$(4.24 \pm 8.70 \pm 55.3) \times 10^{-21}$
$3.89 \times 10^8$	$(2.31 \pm 2.94 \pm 39.9) \times 10^{-21}$

Table H.2: Differential flux values  $dJ/dE$  of the unfolded energy spectrum of **helium nuclei** depicted in Fig. 7.3.  $\sigma_{\text{stat.}}$  respectively  $\Delta_{\text{syst.}}$  are the absolute statistical respectively systematic uncertainties. The result bases on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68].

energy $E/\text{GeV}$	$(dJ/dE \pm \sigma_{\text{stat.}} \pm \Delta_{\text{syst.}}) / (\text{m}^{-2}\text{sr}^{-1}\text{s}^{-1}\text{GeV}^{-1})$
$1.12 \times 10^7$	$(7.57 \pm 1.03 \pm 1.92) \times 10^{-16}$
$1.55 \times 10^7$	$(2.20 \pm 0.24 \pm 0.88) \times 10^{-16}$
$2.14 \times 10^7$	$(6.91 \pm 0.67 \pm 1.49) \times 10^{-17}$
$2.95 \times 10^7$	$(1.85 \pm 0.23 \pm 0.57) \times 10^{-17}$
$4.07 \times 10^7$	$(5.58 \pm 0.86 \pm 2.12) \times 10^{-18}$
$5.62 \times 10^7$	$(2.19 \pm 0.45 \pm 0.59) \times 10^{-18}$
$7.76 \times 10^7$	$(5.91 \pm 1.82 \pm 1.80) \times 10^{-19}$
$1.07 \times 10^8$	$(1.94 \pm 0.60 \pm 0.89) \times 10^{-19}$
$1.48 \times 10^8$	$(1.11 \pm 0.50 \pm 0.42) \times 10^{-19}$
$2.04 \times 10^8$	$(2.34 \pm 1.95 \pm 7.31) \times 10^{-20}$
$2.82 \times 10^8$	$(4.19 \pm 5.47 \pm 55.4) \times 10^{-21}$
$3.89 \times 10^8$	$(0.00 \pm 2.16 \pm 39.1) \times 10^{-21}$

Table H.3: Differential flux values  $dJ/dE$  of the unfolded energy spectrum of **carbon nuclei** depicted in Fig. 7.3.  $\sigma_{\text{stat.}}$  respectively  $\Delta_{\text{syst.}}$  are the absolute statistical respectively systematic uncertainties. The result bases on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68].

energy $E/\text{GeV}$	$(dJ/dE \pm \sigma_{\text{stat.}} \pm \Delta_{\text{syst.}}) / (\text{m}^{-2}\text{sr}^{-1}\text{s}^{-1}\text{GeV}^{-1})$
$1.55 \times 10^7$	$(2.32 \pm 0.27 \pm 1.12) \times 10^{-16}$
$2.14 \times 10^7$	$(9.79 \pm 0.80 \pm 7.30) \times 10^{-17}$
$2.95 \times 10^7$	$(3.10 \pm 0.28 \pm 2.41) \times 10^{-17}$
$4.07 \times 10^7$	$(1.19 \pm 0.11 \pm 0.97) \times 10^{-17}$
$5.62 \times 10^7$	$(3.83 \pm 0.52 \pm 2.54) \times 10^{-18}$
$7.76 \times 10^7$	$(1.35 \pm 0.26 \pm 0.79) \times 10^{-18}$
$1.07 \times 10^8$	$(5.35 \pm 1.14 \pm 3.33) \times 10^{-19}$
$1.48 \times 10^8$	$(2.04 \pm 0.70 \pm 1.07) \times 10^{-19}$
$2.04 \times 10^8$	$(3.49 \pm 1.67 \pm 7.33) \times 10^{-20}$
$2.82 \times 10^8$	$(1.47 \pm 1.07 \pm 5.62) \times 10^{-20}$
$3.89 \times 10^8$	$(2.35 \pm 5.47 \pm 39.7) \times 10^{-21}$

Table H.4: Differential flux values  $dJ/dE$  of the unfolded energy spectrum of **silicon nuclei** depicted in Fig. 7.3.  $\sigma_{\text{stat.}}$  respectively  $\Delta_{\text{syst.}}$  are the absolute statistical respectively systematic uncertainties. The result bases on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68].

energy $E/\text{GeV}$	$(dJ/dE \pm \sigma_{\text{stat.}} \pm \Delta_{\text{syst.}}) / (\text{m}^{-2}\text{sr}^{-1}\text{s}^{-1}\text{GeV}^{-1})$
$1.55 \times 10^7$	$(2.43 \pm 0.35 \pm 1.33) \times 10^{-16}$
$2.14 \times 10^7$	$(1.34 \pm 0.11 \pm 0.80) \times 10^{-16}$
$2.95 \times 10^7$	$(5.09 \pm 0.42 \pm 2.58) \times 10^{-17}$
$4.07 \times 10^7$	$(2.56 \pm 0.21 \pm 1.03) \times 10^{-17}$
$5.62 \times 10^7$	$(8.58 \pm 1.00 \pm 3.07) \times 10^{-18}$
$7.76 \times 10^7$	$(4.05 \pm 0.46 \pm 1.08) \times 10^{-18}$
$1.07 \times 10^8$	$(1.33 \pm 0.21 \pm 0.41) \times 10^{-18}$
$1.48 \times 10^8$	$(4.01 \pm 0.77 \pm 0.99) \times 10^{-19}$
$2.04 \times 10^8$	$(1.35 \pm 0.50 \pm 0.28) \times 10^{-19}$
$2.82 \times 10^8$	$(1.04 \pm 0.27 \pm 0.26) \times 10^{-19}$
$3.89 \times 10^8$	$(6.45 \pm 8.80 \pm 39.3) \times 10^{-21}$

Table H.5: Differential flux values  $dJ/dE$  of the unfolded energy spectrum of **iron nuclei** depicted in Fig. 7.3.  $\sigma_{\text{stat.}}$  respectively  $\Delta_{\text{syst.}}$  are the absolute statistical respectively systematic uncertainties. The result bases on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68].

energy $E/\text{GeV}$	$(dJ/dE \pm \sigma_{\text{stat.}} \pm \Delta_{\text{syst.}}) / (\text{m}^{-2}\text{sr}^{-1}\text{s}^{-1}\text{GeV}^{-1})$
$8.13 \times 10^6$	$(1.02 \pm 0.12 \pm 0.23) \times 10^{-14}$ *unreliable*
$1.12 \times 10^7$	$(3.03 \pm 0.20 \pm 0.66) \times 10^{-15}$
$1.55 \times 10^7$	$(8.97 \pm 0.55 \pm 1.72) \times 10^{-16}$
$2.14 \times 10^7$	$(3.57 \pm 0.19 \pm 0.74) \times 10^{-16}$
$2.95 \times 10^7$	$(1.18 \pm 0.07 \pm 0.25) \times 10^{-16}$
$4.07 \times 10^7$	$(4.94 \pm 0.34 \pm 0.89) \times 10^{-17}$
$5.62 \times 10^7$	$(1.75 \pm 0.16 \pm 0.33) \times 10^{-17}$
$7.76 \times 10^7$	$(6.62 \pm 0.69 \pm 1.18) \times 10^{-18}$
$1.07 \times 10^8$	$(2.24 \pm 0.28 \pm 0.44) \times 10^{-18}$
$1.48 \times 10^8$	$(9.45 \pm 1.37 \pm 1.86) \times 10^{-19}$
$2.04 \times 10^8$	$(2.68 \pm 0.69 \pm 0.48) \times 10^{-19}$
$2.82 \times 10^8$	$(1.43 \pm 0.39 \pm 0.26) \times 10^{-19}$
$3.89 \times 10^8$	$(3.36 \pm 2.10 \pm 3.86) \times 10^{-20}$

Table H.6: Differential flux values  $dJ/dE$  of the unfolded **all-particle** energy spectrum depicted in Fig. 7.3.  $\sigma_{\text{stat.}}$  respectively  $\Delta_{\text{syst.}}$  are the absolute statistical respectively systematic uncertainties. The result bases on the interaction models QGSJET-II-02 [144, 145] and FLUKA 2002.4 [24, 67, 68]. As discussed in Section 7.4.1, the flux value of the all-particle spectrum at  $8.13 \times 10^6$  GeV has to be interpreted with caution, since its reliability within the given uncertainties is not guaranteed.



# Bibliography

- [1] A.A. Abdo et al. Observations of the young supernova remnant RX J1713.7–3946 with the Fermi Large Area Telescope. *Astrophys. J.*, 734(1):28, 2011. doi: 10.1088/0004-637X/734/1/28.
- [2] F.A. Aharonian et al. High-energy particle acceleration in the shell of a supernova remnant. *Nature*, 432:75–77, 2004. doi: 10.1038/nature02960.
- [3] F. Aharonian et al. Primary particle acceleration above 100 TeV in the shell-type supernova remnant RX J1713.7–3946 with deep HESS observations. *Astron. Astrophys.*, 464:235–243, 2007. doi: 10.1051/0004-6361:20066381.
- [4] H.S. Ahn et al. Discrepant hardening observed in cosmic-ray elemental spectra. *Astrophys. J. Lett.*, 714(1):L89–L93, 2010. doi: 10.1088/2041-8205/714/1/L89.
- [5] D. Allard, A.V. Olinto, and E. Parizot. Signatures of the extragalactic cosmic-ray source composition from spectrum and shower depth measurements. *Astron. Astrophys.*, 473(1):59–66, 2007. doi: 10.1051/0004-6361:20077478.
- [6] D. Allard, E. Parizot, and A.V. Olinto. On the transition from galactic to extragalactic cosmic-rays: Spectral and composition features from two opposite scenarios. *Astropart. Phys.*, 27(1):61–75, 2007. doi: 10.1016/j.astropartphys.2006.09.006.
- [7] D. Allard et al. UHE nuclei propagation and the interpretation of the ankle in the cosmic-ray spectrum. *Astron. Astrophys.*, 443(3):L29–L32, 2005. doi: 10.1051/0004-6361:200500199.
- [8] T. Antoni et al. Measurements of attenuation and absorption lengths with the KASCADE experiment. *Astropart. Phys.*, 19(6):703–714, 2003. doi: 10.1016/S0927-6505(03)00116-6.
- [9] T. Antoni et al. KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems. *Astropart. Phys.*, 24(1–2):1–25, 2005. doi: 10.1016/j.astropartphys.2005.04.001.
- [10] T. Antoni et al. The cosmic-ray experiment KASCADE. *Nucl. Instrum. Methods, A* 513(3):490–510, 2003. doi: 10.1016/S0168-9002(03)02076-X.
- [11] V.B. Anykeyev, A.A. Spiridonov, and V.P. Zhigunov. Comparative investigation of unfolding methods. *Nucl. Instrum. Methods, A* 303(2):350–369, 1991. doi: 10.1016/0168-9002(91)90802-W.
- [12] W.D. Apel et al. Energy spectra of elemental groups of cosmic rays: Update on the KASCADE unfolding analysis. *Astropart. Phys.*, 31(2):86–91, 2009. doi: 10.1016/j.astropartphys.2008.11.008.

- [13] W.D. Apel et al. Kneelike structure in the spectrum of the heavy component of cosmic rays observed with KASCADE-Grande. *Phys. Rev. Lett.*, 107(17):171104, 2011. doi: 10.1103/PhysRevLett.107.171104.
- [14] W.D. Apel et al. The KASCADE-Grande experiment. *Nucl. Instrum. Methods, A* 620(2–3):202–216, 2010. doi: 10.1016/j.nima.2010.03.147.
- [15] J.C. Arteaga-Velázquez et al. The KASCADE-Grande experiment: measurements of the all-particle energy spectrum of cosmic rays. In *Proc. XVI ISVHECRI*, Batavia, USA, 2010. arXiv:1009.4716.
- [16] R.C. Aster, B. Borchers, and C.H. Thurber. *Parameter Estimation and Inverse Problems*. Elsevier Academic Press, 2005.
- [17] K.E. Atkinson. *The Numerical Solution of Integral Equations of the Second Kind*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, United Kingdom, 1997.
- [18] P. Auger, R. Maze, and T. Grivet-Meyer. Grandes gerbes cosmiques atmosphériques contenant des corpuscles ultrapénétrants. *C. R. Acad. Sci.*, 206:1721–1723, 1938.
- [19] W.I. Axford, E. Leer, and G. Skadron. The acceleration of cosmic rays by shock waves. In *Proc. 15th ICRC*, volume 11, pages 132–137. Bulgarian Acad. Sci., 1978.
- [20] W. Baade and F. Zwicky. On super-novae. In *Proc. Natl. Acad. Sci. USA*, volume 20, pages 254–259, 1934. doi: 10.1073/pnas.20.5.254.
- [21] W. Baade and F. Zwicky. Cosmic rays from super-novae. In *Proc. Natl. Acad. Sci. USA*, volume 20, pages 259–263, 1934. doi: 10.1073/pnas.20.5.259.
- [22] A. Bamba et al. Small-scale structure of non-thermal X-rays in historical SNRs. *Adv. Space Res.*, 37(8):1439–1442, 2006. doi: 10.1016/j.asr.2005.08.011.
- [23] R. Barceló, M. Masip, and I. Mastromatteo. Cosmic ray knee and new physics at the TeV scale. *J. Cosmo. Astropart. Phys.*, 2009(06):027, 2009. doi: 10.1088/1475-7516/2009/06/027.
- [24] G. Battistoni et al. The FLUKA code: Description and benchmarking. In M. Albrow and R. Raja, editors, *Proc. Hadronic Shower Simulation Workshop 2006*, volume 896, pages 31–49. AIP, 2007. doi: 10.1063/1.2720455.
- [25] A.R. Bell and S.G. Lucek. Cosmic ray acceleration to very high energy through the non-linear amplification by cosmic rays of the seed magnetic field. *Mon. Not. R. Astron. Soc.*, 321(3):433–438, 2001. doi: 10.1046/j.1365-8711.2001.04063.x.
- [26] A. R. Bell. The acceleration of cosmic rays in shock fronts – I. *Mon. Not. R. Astron. Soc.*, 182:147–156, 1978.
- [27] A. R. Bell. The acceleration of cosmic rays in shock fronts – II. *Mon. Not. R. Astron. Soc.*, 182:443–455, 1978.
- [28] E.G. Berezhko and L.T. Ksenofontov. Composition of cosmic rays accelerated in supernova remnants. *Zh. Éksp. Teor. Fiz.*, 116:737–759, 1999 [in *J. Exp. Theor. Phys.* (Engl. transl.), 89(3):391–403, 1999]. doi: 10.1134/1.558996.

- [29] S.F. Berezhnev et al. Tunka-133: Primary cosmic ray mass composition in the energy range  $6 \cdot 10^{15}$ – $10^{18}$  eV. In *Proc. 32nd ICRC (Talk V. Prosin, 15th August 2011)*, Beijing, China, 2011.
- [30] S.F. Berezhnev et al. Tunka-133: Primary cosmic ray energy spectrum in the energy range  $6 \cdot 10^{15}$ – $10^{18}$  eV. In *Proc. 32nd ICRC (Talk L.A. Kuzmichev, 15th August 2011)*, Beijing, China, 2011.
- [31] V. Berezhinsky. Transition from galactic to extragalactic cosmic rays. In *Proc. 30th ICRC*, Merida, Mexico, 2007. arXiv:0710.2750v2.
- [32] V. Berezhinsky, A. Gazizov, and S. Grigorieva. On astrophysical solution to ultrahigh energy cosmic rays. *Phys. Rev. D*, 74(4):043005, 2006. doi: 10.1103/PhysRevD.74.043005.
- [33] M. Bertina et al. The cosmic ray energy spectrum in the range  $10^{16}$ – $10^{18}$  eV measured by KASCADE-Grande. *Astrophys. Space Sci. Trans.*, 7(2):229–234, 2011. doi: 10.5194/astra-7-229-2011.
- [34] M. Bertina et al. A study of the mass composition of cosmic rays based on an event-by-event assignment with KASCADE-Grande data. In *Proc. 32nd ICRC*, Beijing, China, 2011. arXiv:1111.5436.
- [35] P.L. Biermann and V. de Souza. On a common origin of galactic and extragalactic cosmic rays. arXiv:1106.0625, 2011.
- [36] D. Bindig. Calculation of atmospheric neutrino- and muonfluxes with respect to the cosmic ray composition. Diploma thesis WU D 10-09, University of Wuppertal, Germany, 2010.
- [37] R. Blandford and D. Eichler. Particle acceleration at astrophysical shocks: A theory of cosmic ray origin. *Phys. Rep.*, 154(1):1–75, 1987. doi: 10.1016/0370-1573(87)90134-7.
- [38] W.E. Blass and G. Halsey. *Deconvolution of Absorption Spectra*. Academic Press, New York, 1981.
- [39] V. Blobel. Unfolding methods in high-energy physics experiments. In *Proc. 8th CERN School Comput.*, pages 88–127, Aiguablava, Spain, 1985. CERN-85-09 and DESY-84-118.
- [40] V. Blobel. The unfolding problem in data analysis. In *ZEUS seminar*, DESY, Hamburg, 2000.
- [41] V. Blobel and E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. B. G. Teubner, Stuttgart, 1998.
- [42] J. Blümer, R. Engel, and J.R. Hörandel. Cosmic rays from the knee to the highest energies. *Astropart. Phys.*, 63(2):293–338, 2009. doi: 10.1016/j.ppnp.2009.05.002.
- [43] W. Bothe and W. Kolhörster. Das Wesen der Höhenstrahlung. *Z. Phys. A*, 56 (11–12):751–777, 1929. doi: 10.1007/BF01340137.

- [44] I.M. Brancus et al. Event-by-event studies of the discriminative features of the LDF of charged EAS particles: observable correlations and non-parametric analyses of multivariate distributions. In *Proc. 29th ICRC*, volume 00, pages 101–106, Pune, India, 2005.
- [45] R. Brun and F. Rademakers. ROOT – an object oriented data analysis framework. *Nucl. Instrum. Methods, A* 389(1–2):81–86, 1997. doi: 10.1016/S0168-9002(97)00048-X.
- [46] R. Brun et al. GEANT3. Report CERN DD/EE/84-1, CERN, Geneva, Switzerland, Sep 1987.
- [47] P. Campisi and K. Egiazarian, editors. *Blind Image Deconvolution: Theory and Applications*. CRC Press, Boca Raton, USA, 2007.
- [48] J. Candia, L.N. Epele, and E. Roulet. Cosmic ray photodisintegration and the knee of the spectrum. *Astropart. Phys.*, 17(1):23–33, 2002. doi: 10.1016/S0927-6505(01)00131-1.
- [49] E. Cantoni. *Studies of the Cosmic Rays Chemical Composition with the KASCADE-Grande Experiment*. PhD thesis, University of Turin, Italy, 2010.
- [50] E. Cantoni et al. The cosmic ray elemental composition based on measurement of the  $N_p/N_{ch}$  ratio with KASCADE-Grande. In *Proc. 32nd ICRC*, Beijing, China, 2011. arXiv:1111.5436.
- [51] A. Castellina and F. Donato. Astrophysics of galactic charged cosmic rays. In T.D. Oswalt et al., editors, *to appear in Planets, Stars and Stellar Systems*, volume 5. Springer, 2012. arXiv:1110.2981v1.
- [52] J. Clay. Ultra Radiation (penetrating radiation). III. Annual variation and variation with the geographical latitude. In *Proc. K. Ned. Akad. Wet.*, volume 33, pages 711–718, 1930.
- [53] A.K. Cline. An estimate for the condition number of a matrix. *SIAM J. Numer. Anal.*, 16(2):368–375, 1979. doi: 10.1137/0716029.
- [54] G. Cowan. *Statistical Data Analysis*. Oxford University Press, New York, 1998.
- [55] G. Cowan. A survey of unfolding methods for particle physics. In *IPPP Work. Adv. Stat. Tech. Part. Phys.*, pages 248–257, Durham, UK, 2002.
- [56] G. D’Agostini. A multidimensional unfolding method based on Bayes’ theorem. *Nucl. Instrum. Methods, A* 362(2-3):487 – 498, 1995. doi: 10.1016/0168-9002(95)00274-X.
- [57] C. De Donato and G.A. Medina-Tanco. Experimental constraints on the astrophysical interpretation of the cosmic ray Galactic–extragalactic transition region. *Astropart. Phys.*, 32(5):253–268, 2009. doi: 10.1016/j.astropartphys.2009.09.004.
- [58] H.P. Dembinski et al. Latest results from the Pierre Auger Observatory. In *J. Phys. Conf. Ser.*, volume 337, page 012068, 2012. doi: 10.1088/1742-6596/337/1.

- [59] D. d’Enterria et al. Constraints from the first LHC data on hadronic event generators for ultra-high energy cosmic-ray physics. *Astropart. Phys.*, 35(2):98–113, 2011. doi: 10.1016/j.astropartphys.2011.05.002.
- [60] C.D. Dermer. Gamma Ray Bursts, Supernovae, and Cosmic Ray Origin. In M.M. Shapiro, T. Stanev, and J.P. Wefel, editors, *Proc. 13th Course Int. School Cosmic Ray Astrophys.*, pages 189–198. World Scientific Publishing Co. Pte. Ltd., 2004. doi: 10.1142/9789812702968\_0012.
- [61] E.A. Dorfi. Gamma rays and cosmic rays in supernova remnants with radiative cooling. *Astron. Astrophys.*, 251(2):597–610, 1991.
- [62] M.T. Dova, L.N. Epele, and J.D. Swain. Massive relic neutrinos in the galactic halo and the knee in the cosmic ray spectrum. arXiv:astro-ph/0112191v1, 2001.
- [63] L. O’C. Drury. Particle acceleration mechanisms for cosmic rays. In *Proc. Astron. Sci. Group Ireland, Meeting, Galway, Ireland, 1986*, volume 18, pages 28–33. Irish Astron. J., 1987.
- [64] A.D. Erlykin and A.W. Wolfendale. A single source of cosmic rays in the range  $10^{15} - 10^{16}$  eV. *J. Phys. G: Nucl. Part. Phys.*, 23(8):979–989, 1997. doi: 10.1088/0954-3899/23/8/012.
- [65] A.D. Erlykin and A.W. Wolfendale. Models for the origin of the knee in the cosmic-ray spectrum. *Adv. Space Res.*, 27(4):803–812, 2001. doi: 10.1016/S0273-1177(01)00125-9.
- [66] P. Facal San Luis. The distribution of shower maxima of UHECR air showers. In *Proc. 32nd ICRC*, Beijing, China, 2011.
- [67] A. Fassò et al. FLUKA: Status and prospective for hadronic applications. In A. Kling et al., editors, *Proc. MonteCarlo 2000 Conference, Lisbon*, pages 955–960, Berlin, Germany, 2001. Springer-Verlag.
- [68] A. Fassò et al. FLUKA: A multi-particle transport code. Technical report CERN-2005-10, INFN/TC\_05/11, SLAC-R-773, Stanford Linear Accelerator Center (SLAC), 2005. doi: 10.2172/877507.
- [69] E. Fermi. On the origin of the cosmic radiation. *Phys. Rev.*, 75(8):1169–1174, 1949. doi: 10.1103/PhysRev.75.1169.
- [70] M. Finger. *Reconstruction of energy spectra for different mass groups of high-energy cosmic rays*. PhD thesis, Karlsruher Institut für Technologie, Germany, 2011.
- [71] I. Fredholm. Sur une classe d’équations fonctionnelles. *Acta Math.*, 27(1):365–390, 1903. doi: 10.1007/BF02421317.
- [72] D. Fuhrmann et al. Muonic Component of Air Showers Measured by the KASCADE-Grande Experiment. In *Proc. 31st ICRC*, Łódź, Poland, 2009. arXiv:0906.4007.
- [73] D. Fuhrmann et al. Muonic component of air showers measured by the KASCADE-Grande experiment. In *Proc. 31st ICRC*, Łódź, Poland, 2009. arXiv:0906.4007.

- [74] T.K. Gaisser. *Cosmic Rays and Particle Physics*. Cambridge University Press, New York, USA, 1990.
- [75] T.K. Gaisser and A.M. Hillas. Reliability of the method of constant intensity cuts for reconstructing the average development of vertical showers. In *Proc. 15th ICRC*, volume 8, pages 353–357, 1977.
- [76] S. Giani et al. GEANT detector description and simulation tool. CERN program library long writeup W5013, CERN, Geneva, Switzerland, March 1994.
- [77] R. Gold. An iterative unfolding method for response matrices. *Report ANL-6984*, Argonne, USA, 1964.
- [78] R. Gold and E.F. Bennett. Error estimates for iterative unfolding. *J. Comput. Phys.*, 3(2):167–175, 1968. doi: 10.1016/0021-9991(68)90014-4.
- [79] G. Golub and W. Kahan. Calculating the singular values and pseudo-inverse of a matrix. *SIAM J. Numer. Anal.*, B2(2):205–224, 1965. doi: 10.1137/0702016.
- [80] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, USA, 3rd edition, 1996.
- [81] K. Greisen. The extensive air showers. In J.G. Wilson, editor, *Prog. Cosmic Ray Phys.*, volume 3, chapter 1, pages 1–141. North-Holland Pub. Co., Amsterdam, Netherlands, 1956.
- [82] S. Hannestad. Neutrinos in cosmology. *New J. Phys.*, 6:108, 2004. doi: 10.1088/1367-2630/6/1/108.
- [83] P.C. Hansen. *Discrete Inverse Problems – Insight and Algorithms*. Fundamentals of Algorithms. Society for Industrial and Applied Mathematics, 2010.
- [84] P.M. Hansen, J. Alvarez-Muñiz, and R.A. Vázquez. A comprehensive study of shower to shower fluctuations. *Astropart. Phys.*, 34(6):503–512, 2011. doi: 10.1016/j.astropartphys.2010.11.001.
- [85] W.M. Hasan and E. Viola. Use of the singular value decomposition method to detect ill-conditioning of structural identification problems. *Computers & Structures*, 63(2):267–275, 1997. doi: 10.1016/S0045-7949(96)00361-6.
- [86] A. Haungs, H. Rebel, and M. Roth. Energy spectrum and mass composition of high-energy cosmic rays. *Rep. Prog. Phys.*, 66(7):1145–1206, 2003. doi: 10.1088/0034-4885/66/7/202.
- [87] D. Heck and J. Knapp. Upgrade of the Monte Carlo code CORSIKA to simulate extensive air showers with energies  $> 10^{20}$  eV (corrected version from Sep. 5, 2003). Report FZKA 6097, Forschungszentrum Karlsruhe, Germany, 1998.
- [88] D. Heck et al. CORSIKA: A Monte Carlo code to simulate extensive air showers. Report FZKA 6019, Forschungszentrum Karlsruhe, Germany, 1998.
- [89] S. Heinz and R. Sunyaev. Cosmic rays from microquasars: A narrow component to the CR spectrum? *Astron. Astrophys.*, 390(2):751–766, 2002. doi: 10.1051/0004-6361:20020615.

- [90] V.F. Hess. Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten. *Phys. Z.*, 13:1084–1091, 1912.
- [91] A.M. Hillas. Where do  $10^{19}$  eV cosmic rays come from? *Nucl. Phys. (Proc. Suppl.)*, B 136:139–146, 2004. doi: 10.1016/j.nuclphysbps.2004.10.004.
- [92] A.M. Hillas. Can diffusive shock acceleration in supernova remnants account for high-energy galactic cosmic rays? *J. Phys. G*, 31(5):R95–R131, 2005. doi: 10.1088/0954-3899/31/5/R02.
- [93] A.M. Hillas et al. In *Proc. 12th ICRC*, volume 3, page 1001, Hobart, 1971.
- [94] J.R. Hörandel. Models of the knee in the energy spectrum of cosmic rays. *Astropart. Phys.*, 21(3):241–265, 2004. doi: 10.1016/j.astropartphys.2004.01.004.
- [95] J.R. Hörandel. On the knee in the energy spectrum of cosmic rays. *Astropart. Phys.*, 19:193–220, 2003. doi: 10.1016/S0927-6505(02)00198-6.
- [96] B.R. Hunt. Bayesian methods in nonlinear digital image restoration. *IEEE Trans. Comput.*, C-26(3):219–229, 1977. doi: 10.1109/TC.1977.1674810.
- [97] Google Inc. *Google Maps*, 12th May 2011. URL <http://maps.google.de/>.
- [98] F. James. MINUIT minimization package. CERN program library long writeup D506, CERN, Geneva, Switzerland, 1994.
- [99] F. James and M. Roos. MINUIT – a system for function minimization and analysis of the parameter errors and correlations. *Comput. Phys. Commun.*, 10(6):343–367, 1975. doi: 10.1016/0010-4655(75)90039-9.
- [100] P.A. Jansson, editor. *Deconvolution – With Application in Spectroscopy*. Academic Press, London, 1984.
- [101] P.A. Jansson. Method for determining the response function of a high-resolution infrared spectrometer. *J. Opt. Soc. Am.*, 60(2):184–190, 1970. doi: 10.1364/JOSA.60.000184.
- [102] E.T. Jaynes. Prior probabilities. *IEEE Trans. Syst. Sci. Cyb.*, SSC-4(3):227–241, 1968. doi: 10.1109/TSSC.1968.300117.
- [103] K. Kamata and J. Nishimura. The lateral and the angular structure functions of electron showers. *Prog. Theor. Phys. Suppl.*, 6:93–155, 1958. doi: 10.1143/PTPS.6.93.
- [104] K.-H. Kampert. Personal communication, April 2009. University of Wuppertal, Germany.
- [105] K.-H. Kampert and M. Unger. Measurements of the cosmic ray composition with air shower experiments. arXiv:1201.0018, 2011. Preprint.
- [106] S. Karakula and W. Tkaczyk. The formation of the cosmic ray energy spectrum by a photon field. *Astropart. Phys.*, 1(2):229–237, 1993.
- [107] D. Kazanas and A. Nicolaidis. Cosmic ray “knee”: A herald of new physics? arXiv:astro-ph/0103147v2, 2001.

- [108] D. Kazanas and A. Nicolaidis. Letter: Cosmic rays and large extra dimensions. *Gen. Relativ. Gravit.*, 35(6):1117–1123, 2003. doi: 10.1023/A:1024077103557.
- [109] T.J. Kennett, W.V. Prestwich, and A. Robertson. Bayesian deconvolution I: Convergent properties. *Nucl. Instrum. Methods*, 151(1–2):285–292, 1978. doi: 10.1016/0029-554X(78)90502-5.
- [110] T.J. Kennett, W.V. Prestwich, and A. Robertson. Bayesian deconvolution II: Noise properties. *Nucl. Instrum. Methods*, 151(1–2):293–301, 1978. doi: 10.1016/0029-554X(78)90503-7.
- [111] T.J. Kennett et al. Bayesian deconvolution III: Applications and algorithm implementation. *Nucl. Instrum. Methods*, 153(1):125–135, 1978. doi: 10.1016/0029-554X(78)90628-6.
- [112] T.H. Kerr. The proper computation of the matrix pseudoinverse and its impact in MVRO filtering. *IEEE Aero. El. Sys. Mag.*, AES-21(5):711–724, 1985. doi: 10.1109/TAES.1985.310599.
- [113] V. Klema and A. Laub. The singular value decomposition: Its computation and some applications. *IEEE Trans. Automat. Control*, 25(2):164–176, 1980. doi: 10.1109/TAC.1980.1102314.
- [114] K. Kobayakawa, Y.S. Honda, and T. Samura. Acceleration by oblique shocks at supernova remnants and cosmic ray spectra around the knee region. *Phys. Rev. D*, 66(8):083004, 2002. doi: 10.1103/PhysRevD.66.083004.
- [115] W. Kolhörster. Messungen der durchdringenden Strahlung im Freiballon in grösseren Höhen. *Phys. Z.*, 14:1153–1156, 1913.
- [116] W. Kolhörster, I. Matthes, and E. Weber. Gekoppelte Höhenstrahlen. *Naturwissenschaften*, 26(35):576, 1938. doi: 10.1007/BF01773491.
- [117] A.N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *G. Ist. Ital. Attuari*, 4:83, 1933.
- [118] G.F. Krymskii. A regular mechanism for the acceleration of charged particles on the front of a shock wave. *Dok. Akad. Nauk SSSR*, 234:1306–1308, 1977 [in *Sov. Phys. Dokl.* (Engl. transl.), 22:327, 1977].
- [119] G.V. Kulikov and G.B. Khristiansen. On the size spectrum of extensive air showers. *Zh. Éksp. Teor. Fiz.*, 35:635–640, 1958 [in *Sov. Phys. J. Exp. Theor. Phys.* (Engl. transl.), 8(3):441–444, 1959].
- [120] P.O. Lagage and C.J. Cesarsky. The maximum energy of cosmic rays accelerated by supernova shocks. *Astron. Astrophys.*, 125(2):249–257, 1983.
- [121] A.A. Lagutin and R.I. Raikin. Lateral distribution of electrons in EAS at superhigh energies: predictions and experimental data. *Nucl. Phys. (Proc. Suppl.)*, B 97(1–3):274–277, 2001. doi: 10.1016/S0920-5632(01)01282-8.
- [122] A.A. Lagutin, Yu.A. Nikulin, and V.V. Uchaikin. The “knee” in the primary cosmic ray spectrum as consequence of the anomalous diffusion of the particles in the fractal interstellar medium. *Nucl. Phys. (Proc. Suppl.)*, B 97(1–3):267–270, 2001. doi: 10.1016/S0920-5632(01)01280-4.

- [123] M. Lemoine and R. Benoît. Relativistic fermi acceleration with shock compressed turbulence. *Mon. Not. R. Astron. Soc.*, 366(2):635–644, 2006. doi: 10.1111/j.1365-2966.2005.09912.x.
- [124] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quart. Appl. Math.*, 2:164–168, 1944.
- [125] R.E. Lingenfelter, J.C. Higdon, and R. Ramaty. Cosmic ray acceleration in superbubbles and the composition of cosmic rays. In R.A. Mewaldt et al., editors, *Acceleration and Transport of Energetic Particles Observed in the Heliosphere: ACE 2000 Symposium*, volume 528, pages 375–382. AIP Conf. Proc., 2000. doi: 10.1063/1.1324342.
- [126] J. Linsley and L. Scarsi. Arrival times of air shower particles at large distances from the axis. *Phys. Rev.*, 128(5):2384–2392, 1962. doi: 10.1103/PhysRev.128.2384.
- [127] J. Linsley, L. Scarsi, and B. Rossi. Suppl. *J. Phys. Soc. Jpn.*, 17(A-III):91, 1962.
- [128] M.S. Longair. *High Energy Astrophysics – Volume 1: Particles, photons and their detection*. Cambridge University Press, New York, USA, 2nd edition, 1992.
- [129] M.S. Longair. *High Energy Astrophysics – Volume 2: Stars, the Galaxy and the interstellar medium*. Cambridge University Press, New York, USA, 2nd edition, 1994.
- [130] L.B. Lucy. An iterative technique for the rectification of observed distributions. *Astron. J.*, 79(6):745–754, 1974. doi: 10.1086/111605.
- [131] G. Maier. *Untersuchung von atmosphärischen Einflüssen auf ausgedehnte Luftschauer und Bestimmung von Abschwächlängen*. Diploma thesis, University of Karlsruhe, Germany, 2000.
- [132] D.W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 11(2):431–441, 1963. doi: 10.1137/0111030.
- [133] R.M. Martirosov et al. Energy spectrum and mass composition of primary cosmic radiation in the region above the knee from the GAMMA experiment. In *Proc. 32nd ICRC*, Beijing, China, 2011.
- [134] A. Mastichiadis. On the high energy non-thermal emission from shell-type supernova remnants. *Astron. Astrophys.*, 305:L53–L56, 1996. arXiv:astro-ph/9601132.
- [135] M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.*, 8(1):3–30, 1998. doi: 10.1145/272991.272995.
- [136] M. Mendès and C. De Polignac. Unfolding of physical spectra by sequential bayes methods part II: Estimation of parameters in non-linear models. *Nucl. Instrum. Methods*, 127(3):413–419, 1975. doi: 10.1016/S0029-554X(75)80014-0.
- [137] M. Mendès, C. De Polignac, and C. Delestre. Unfolding of physical spectra by sequential bayes methods part I: Linear spline models. *Nucl. Instrum. Methods*, 127(3):405–411, 1975. doi: 10.1016/S0029-554X(75)80013-9.

- [138] R. Minkowski. Optical investigations of radio sources. Introductory lecture. In H.C. Van de Hulst, editor, *Proc. 4th IAU Symp.*, pages 107–122. Cambridge University Press, 1957.
- [139] J.E. Mottershead and C.D. Foster. On the treatment of ill-conditioning in spatial parameter estimation from measured vibration data. *Mech. Syst. Signal Process.*, 5(2):139–154, 1991. doi: 10.1016/0888-3270(91)90020-6.
- [140] R. Narayan and R. Nityananda. Maximum entropy image restoration in astronomy. *Annu. Rev. Astron. Astrophys.*, 24:127–170, 1986. doi: 10.1146/annurev.aa.24.090186.001015.
- [141] S.I. Nikolsky. Break in the cosmic ray spectrum or confinement violation near  $10^6$  TeV? *Nucl. Phys. (Proc. Suppl.)*, B 39(1):228–234, 1995. doi: 10.1016/0920-5632(95)00025-5.
- [142] S. Ogio and F. Kakimoto. Advective diffusion propagation model for galactic cosmic rays above  $10^{12}$  eV. In *Proc. 28th ICRC*, volume 1, page 315, Trukuba, Japan, 2003.
- [143] E.E. Osborne. On least squares solutions of linear equations. *J. ACM*, 8(4):628–636, 1961. doi: 10.1145/321088.321100.
- [144] S. Ostapchenko. QGSJET-II: Towards reliable description of very high energy hadronic interactions. *Nucl. Phys. (Proc. Suppl.)*, B 151(1):143–146, 2006. doi: 10.1016/j.nuclphysbps.2005.07.026.
- [145] S. Ostapchenko. Nonlinear screening effects in high energy hadronic interactions. *Phys. Rev. D*, 74(1):014026, 2006. doi: 10.1103/PhysRevD.74.014026.
- [146] E. Parizot et al. Observational constraints on energetic particle diffusion in young supernovae remnants: amplified magnetic field and maximum energy. *Astron. Astrophys.*, 453(2):387–395, 2006. doi: 10.1051/0004-6361:20064985.
- [147] S.N. Patnaik, J.D. Guptill, and L. Berke. Singularity in structural optimization. *Int. J. Num. Meth. Eng.*, 36(6):931–944, 1993. doi: 10.1002/nme.1620360604.
- [148] B. Peters. Primary cosmic radiation and extensive air showers. *Il Nuovo Cimento*, 22(4):800–819, 1961. doi: 10.1007/BF02783106.
- [149] Y.P. Petrov and V.S. Sizikov. *Well-Posed, Ill-Posed, and Intermediate Problems with Applications*. Inverse and Ill-Posed Problems. VSP, Netherlands, 2005.
- [150] R. Plaga. A possible universal origin of hadronic cosmic rays from ultrarelativistic ejecta of bipolar supernovae. *New Astron.*, 7(6):317–336, 2002. doi: 10.1016/S1384-1076(02)00135-5.
- [151] V. Ptuskin, V. Zirakashvili, and E.-S. Seo. Spectrum of galactic cosmic rays accelerated in supernova remnants. *Astrophys. J.*, 718(1):31–36, 2010. doi: 10.1088/0004-637X/718/1/31.
- [152] V.S. Ptuskin et al. Diffusion and drift of very high energy cosmic rays in galactic magnetic fields. *Astron. Astrophys.*, 268:726–735, 1993.

- [153] H. Rebel et al. Studies of different LDFs for primary energy estimation and mass discrimination of cosmic rays by the EAS lateral charged particle distribution as observed by KASCADE-Grande. In *Proc. 29th ICRC*, volume 6, pages 297–300, Pune, India, 2005.
- [154] S.P. Reynolds and J.W. Keohane. Maximum energies of shock-accelerated electrons in young shell supernova remnants. *Astrophys. J. Lett.*, 525(1):368–374, 1999. doi: 10.1086/307880.
- [155] W.H. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.*, 62(1):55–59, 1972. doi: 10.1364/JOSA.62.000055.
- [156] E. Rothwell and B. Drachman. A unified approach to solving ill-conditioned matrix problems. *Int. J. Num. Meth. Eng.*, 28(3):609–620, 1989. doi: 10.1002/nme.1620280309.
- [157] E. Roulet. Astroparticle theory: Some new insights into high energy cosmic rays. *Int. J. Mod. Phys. A*, 19(7):1133–1141, 2004. doi: 10.1142/S0217751X04019044.
- [158] R. Schlickeiser. *Cosmic Ray Astrophysics*. Springer Verlag, Berlin, Germany, 2003.
- [159] M. Schmelling. The method of reduced cross-entropy: A general approach to unfold probability distributions. *Nucl. Instrum. Methods, A* 340(2):400–412, 1994. doi: 10.1016/0168-9002(94)90119-8.
- [160] C.E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27: 379–423 and 623–656, 1948.
- [161] M.M. Shapiro, T. Stanev, and J.P. Wefel, editors. *Neutrinos and Explosive Events in the Universe*. NATO Science Series: II. Mathematics, Physics and Chemistry – Vol. 209. Springer, Dordrecht, Netherlands, 2005.
- [162] I.S. Shklovskii. On the nature of the optical emission from the Crab Nebula. *Dok. Akad. Nauk SSSR*, 90:983–986, 1953 [Engl. transl. in K.R. Lang and O. Gingerich, editors, *Source Book Astron. Astrophys., 1900–1975*, Harvard Univ. Press, 1979].
- [163] N.V. Smirnov. Tables for estimating the goodness of fit of empirical distributions. *Ann. Math. Statist.*, 19:279–281, 1948.
- [164] T. Stanev. *High Energy Cosmic Rays*. Springer-Verlag, Berlin, Germany, 2nd edition, 2010.
- [165] T. Stanev, P.L. Biermann, and T.K. Gaisser. Cosmic rays – IV. The spectrum and chemical composition above  $10^4$  GeV. *Astron. Astrophys.*, 274:902–908, 1993. arXiv:astro-ph/9303006.
- [166] J.-L. Starck and F. Murtagh. *Astronomical Image and Data Analysis*. Springer-Verlag, Berlin, Germany, 2nd edition, 2006.
- [167] G.W. Stewart. *Introduction to Matrix Computations (Computer Science and Applied Mathematics)*. Academic Press, New York, 1973.
- [168] L.G. Sveshnikova. The knee in the galactic cosmic ray spectrum and variety in supernovae. *Astron. Astrophys.*, 409(3):799–807, 2003. doi: 10.1051/0004-6361:20030909.

- [169] S.P. Swordy. In *Proc. 24th ICRC*, volume 2, page 697, Rome, Italy, 1995.
- [170] T. Tanaka et al. Gamma-ray observations of the supernova remnant RX J0852.0–4622 with the Fermi Large Area Telescope. *Astrophys. J.*, 740(2):L51, 2011. doi: 10.1088/2041-8205/740/2/L51.
- [171] The ROOT team. *ROOT – An Object Oriented Data Analysis Framework: Users Guide 5.26*. CERN, Geneva, Switzerland, December 2009.
- [172] A.N. Tikhonov. On the stability of inverse problems. *Proc. USSR Acad. Sci.*, 39(5):195–198, 1943.
- [173] A.N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Sov. Math. Dokl.*, 4(4):1035–1038, 1963.
- [174] A.N. Tikhonov. Regularization of incorrectly posed problems. *Sov. Math. Dokl.*, 4(6):1624–1627, 1963.
- [175] A.N. Tikhonov and V.Y. Arsenin. Solutions of ill-posed problems. *SIAM Rev.*, 21(2):266–267, 1977. doi: 10.1137/1021044.
- [176] W.M. To. Application of singular value decomposition to direct matrix update method. *AIAA J.*, 32(10):2124–2126, 1994. doi: 10.2514/3.12265.
- [177] G. Toma. *The KASCADE-Grande Experiment: Study of the Lateral Distributions and of the Arrival Times*. PhD thesis, University of Bucharest, Romania, 2010.
- [178] G. Toma et al. Primary energy reconstruction from the S(500) observable recorded with the KASCADE-Grande detector array. In *Proc. 31st ICRC*, Łódź, Poland, 2009. arXiv:0906.4007.
- [179] H. Ulrich. *Untersuchung atmosphärischer Einflüsse auf die Entwicklung ausgedehnter Luftschauer anhand von Simulationsrechnungen*. Diploma thesis, University of Karlsruhe, Germany, 1997.
- [180] H. Ulrich. *Untersuchungen zum primären Energiespektrum der kosmischen Strahlung im PeV-Bereich mit dem KASCADE-Experiment*. PhD thesis, University of Karlsruhe, Germany, 2003.
- [181] H. Ulrich, K.-H. Kampert, and H.O. Klages. Effect of different atmospheric profiles to EAS observables. AUGER internal note GAP-1998-043, 1998.
- [182] M. Unger. Cosmic Rays above the Knee. In *Proc. 21st ECRS*, Košice, Slovakia, 2008. arXiv:0812.2763.
- [183] United States National Oceanic and Atmospheric Administration. U.S. standard atmosphere: 1976. National Oceanic and Atmospheric Administration, National Aeronautics and Space Administration, United States Air Force, Washington D.C., 1976.
- [184] J. van Buren. *Untersuchung der myonischen Komponente ausgedehnter Luftschauer mit dem KASCADE-Grande Experiment*. PhD thesis, University of Karlsruhe, Germany, 2005.

- 
- [185] P.H. van Cittert. Zum Einfluß der Spaltbreite auf die Intensitätsverteilung in Spektrallinien. II. *Z. Phys.*, A 69(5):298–308, 1931. doi: 10.1007/BF01391351.
- [186] H.J. Völk and P.L. Biermann. Maximum energy of cosmic-ray particles accelerated by supernova remnant shocks in stellar wind cavities. *Astrophys. J. Lett.*, 333: L65–L68., 1988. doi: 10.1086/185289.
- [187] H.J. Völk and V.N. Zirakashvili. Cosmic ray acceleration by spiral shocks in the galactic wind. *Astron. Astrophys.*, 417(3):807–817, 2004. doi: 10.1051/0004-6361:20040018.
- [188] K. Werner. The hadronic interaction model EPOS. *Nucl. Phys. (Proc. Suppl.)*, B 175-176:81–87, 2008. doi: 10.1016/j.nuclphysbps.2007.10.012.
- [189] S.D. Wick, C.D. Dermer, and A. Atoyan. High-energy cosmic rays from  $\gamma$ -ray bursts. *Astropart. Phys.*, 21(2):125–148, 2004. doi: 10.1016/j.astropartphys.2003.12.008.
- [190] R. Wigmans. PeV cosmic rays: a window on the leptonic era? *Astropart. Phys.*, 19 (3):379–392, 2003. doi: 10.1016/S0927-6505(02)00242-6.
- [191] C. Xu, I. Aissaoui, and S. Jacquy. Algebraic analysis of the Van Cittert iterative method of deconvolution with a general relaxation factor. *J. Opt. Soc. Am.*, A 11 (11):2804–2808, 1994. doi: 10.1364/JOSAA.11.002804.
- [192] A. Yushkov et al. Measurements of the muon shower content at the Pierre Auger Observatory. In *Proc. Int. Symp. Future Directions UHECR Phys.*, CERN, Geneva, Switzerland, 2012.