

Lateral Distribution Function and Energy Spectrum for the 750 m Array of the Pierre Auger Observatory

Master's Thesis of

Philipp David Meder

At the KIT Department of Physics
Institute for Astroparticle Physics (IAP)

First examiner: Prof. Dr. Ralph Engel
Second examiner: Priv.-Doz. Dr. Roger Wolf
First advisor: Dr. Darko Veberič
Second advisor: Dr. David Schmidt

6th September 2023 – 5th September 2024

Karlsruher Institut für Technologie
Fakultät für Physik
76128 Karlsruhe

Lateral Distribution Function and Energy Spectrum for the 750 m Array of the Pierre Auger Observatory *(Master's Thesis)*

I declare that I have developed and written the enclosed thesis completely by myself. I have not used any other than the aids that I have mentioned. I have marked all parts of the thesis that I have included from referenced literature, either in their original wording or paraphrasing their contents. I have followed the by-laws to implement scientific integrity at KIT.

Karlsruhe,

.....
(Philipp David Meder)

.....
(Prof. Dr. Ralph Engel)

Abstract

There are many known sources in the universe, e.g. supernovae explosions, which produce and accelerate highly energetic charged particles. These so-called *cosmic rays* propagate through space where they are affected by magnetic fields and other propagation effects. By measuring the energy spectrum of cosmic rays upon their arrival at Earth, one can get not only a better understanding of astrophysical objects and galactic magnetic fields but also an insight into particle physics at energies far beyond the energies achieved by modern particle colliders.

Large ground-based experiments like the Pierre Auger Observatory measure the energy spectrum in the range of the highest energies where the flux of cosmic rays is low by detecting the secondary particles originating from primary cosmic rays inducing a particle shower cascade in the atmosphere. While originally designed to measure the changes in the spectrum – so-called features – for energies above 10^{18} eV with a detector array of 1500 m spacing, a denser extension with 750 m spacing was designed to measure the spectrum at lower energies.

In this thesis, I will present an update of the event reconstruction using new triggers to lower the overall energy threshold of the 750 m array. I suggest a new approach to the likelihood used during the shower reconstruction procedure. New selection conditions for events where all necessary parameters can be fitted freely are introduced and used to find a stable parameterisation of the lateral distribution function of the shower. Therefore, the reconstruction of events detected between 2014 and 2021 by the 750 m array is performed, including low-energetic events that could not be reconstructed satisfyingly without this parameterisation. An updated correction for atmospheric effects using the constant intensity cut is provided as well as a new energy calibration using SD-FD hybrids. We will briefly explore the detector efficiency in a data-driven way to decide down to which energy threshold low-energetic events can be used for a spectrum analysis.

Including these low-energy events leads to an updated energy spectrum that allows us to investigate the spectrum properties in the energy range of a feature known as the *second knee* as reported by different cosmic ray experiments. I will argue that the second knee is measured by the 750 m array for the first time by showing that the fitted features of the spectrum are stable for different fit configurations.

Zusammenfassung

Es existiert eine Vielzahl bekannter Quellen im Universum, wie z.B. Supernova-Explosionen, die hochenergetische geladene Teilchen erzeugen und beschleunigen. Diese Teilchen werden als Kosmische Strahlung bezeichnet und propagieren unter dem Einfluss von Magnetfeldern und anderen Propagationseffekte durch das Weltall. Die Messung des Energiespektrums der Kosmischen Strahlung bei ihrer Ankunft auf der Erde bzw. der Erdatmosphäre erlaubt Rückschlüsse auf die astrophysikalischen Quellen und die galaktischen Magnetfelder. Darüber hinaus erlaubt dies einen Einblick in teilchenphysikalische Prozesse auf Energieskalen weit jenseits der mit modernen Teilchenbeschleunigern erreichbaren Maximalenergien.

Große Experimente auf der Erdoberfläche, wie das Pierre Auger Observatorium, messen das Energiespektrum der Kosmischen Strahlung im Bereich der höchsten beobachteten Energien durch die Detektion von Sekundärteilchen. Diese entstehen, wenn Primärteilchen der Kosmischen Strahlung mit der Atmosphäre wechselwirken und dabei durch teilchenphysikalische Prozesse sogenannte *ausgedehnte Luftschauer* auslösen. Obwohl das Pierre Auger Observatorium ursprünglich für die Messung des Spektrums oberhalb von 10^{18} eV konzipiert wurde, einem Energiebereich, in dem der Fluss der Kosmischen Strahlung sehr gering ist, existiert eine Erweiterung des Detektors, die das Messen des Spektrums bei niedrigeren Energien erlaubt. Diese Erweiterung besteht aus Detektorstationen, die im Gegensatz zu den ursprünglichen 1500 m einen Abstand von 750 m zueinander haben.

Diese Masterarbeit beschäftigt sich mit der Rekonstruktion von Ereignissen, die mit dieser Erweiterung, genannt SD-750, gemessen wurden. Dabei wird die Rekonstruktion so angepasst, dass Daten, die mit neuen, niederschweligen Triggern gemessen wurden, korrekt in der Auswertung berücksichtigt werden. Dadurch wird die Effizienz des Detektors bei niedrigeren Energien erhöht. Es werden neue Auswahlregeln für Ereignisse, welche eine vollständige Rekonstruktion aller nötigen Parameter ermöglichen, eingeführt. Diese Auswahlregeln werden anschließend benutzt, um eine Parameterisierung der Lateralverteilung zu finden, so dass auch Ereignisse, die diese Auswahlregeln nicht erfüllen, rekonstruiert werden können. Die Korrektur des Einflusses der Atmosphäre für nicht-vertikale Schauer wird diskutiert. Durch das Verwenden von Ereignissen, die sowohl mit dem Oberflächendetektor als auch mit Fluoreszenzteleskopen beobachtet wurden, wird eine Energiekalibrierung für den SD-750-Detektor durchgeführt. Zusätzlich wird die Effizienz des Detektors für niedere Energien auf Echtdatenbasis untersucht.

Die erwähnten Schritte sind notwendig, um die untere Energieschwelle des Detektors herabzusetzen. Dadurch kann das mit dem SD-750-Detektor zwischen 2014 und 2021 gemessene Energiespektrum im Bereich kleinerer Energien untersucht werden. Dies erlaubt die erste vollständige Messung des *zweiten Knies der Kosmischen Strahlung* mit dem SD-750-Detektor.

Contents

Abstract	i
Zusammenfassung	iii
1 Cosmic Rays and Extensive Air Showers	1
1.1 Composition and Propagation	1
1.2 Energy Spectrum	1
1.3 Extensive Air Showers	2
2 The Pierre Auger Observatory and Thesis Objectives	5
2.1 The SD-750 Detector	5
2.2 Lowering the Energy Threshold with New Triggers	6
2.3 Reconstructing SD Events	7
3 Investigating the Signal Variance for Small Signals	9
3.1 Simulation Setup	9
3.2 Vertical Showers	9
3.3 Inclined Showers	10
3.4 Implications for Stations with Small Signals	13
4 Updating the LDF Likelihood	17
4.1 Current Likelihood Setup	17
4.2 Criticism of the Current Setup	18
4.3 Investigating the Distribution of Small Signals	19
4.4 Updating the Likelihood Using the Trigger Probability	22
4.5 Likelihood Contributions for the LDF Slope Parameters	23
5 New Lever-Arm Criteria to Fit the LDF Slope Parameters	27
5.1 Current Lever-Arm Criteria	28
5.2 New Criteria	28
5.2.1 Linear Fit in Log-Log-Space	29
5.2.2 Selection Conditions	30
5.2.3 Additional Conditions	31
5.3 Comparison	32
5.3.1 Events Satisfying the Old but not the New Criteria	33
5.3.2 Events Satisfying the New but not the Old Criteria	33
5.4 Summarising the Lever-Arm Criteria Studies	33
6 Parameterising the Lateral Distribution Function	39
6.1 Proposed Changes	39
6.2 LDF Parameterisation	40
6.2.1 Selecting Events for the Parameterisation	40
6.2.2 Rewriting the Slope Parameter	41
6.2.3 Parameterising the Slope Parameter	42

6.3	Validating the Parameterisation	43
6.4	Comparing the New LDF Parameterisation with the Old Parameterisation . .	47
6.5	Reconstructing Events with the New Parameterisation	47
6.5.1	Comparison against the Standard Reconstruction on the Event Level .	47
6.5.2	Comparison against the Standard Reconstruction on the Station Level	49
6.5.3	Reconstruction Quality	50
6.6	Relevance of the New Parameterisation	51
7	Correcting for Atmospheric Attenuation Effects Using the CIC Procedure	59
7.1	CIC Procedure	60
7.2	Investigating the Attenuation-Correction	61
8	Energy Calibration using SD-FD Hybrids and Array Efficiency	65
8.1	Energy Calibration	65
8.2	Energy Bias and Resolution	66
8.3	Investigating the Array Efficiency	67
8.3.1	Isotropy Condition	68
8.3.2	Uniformity of Impact Point Distributions	69
8.3.3	Full-Efficiency Thresholds	70
9	Energy Spectrum of Cosmic Rays and Fit of its Features	73
9.1	Exposure	73
9.2	Forward-Folding Procedure and Fit Model	75
9.3	Investigating the Influence of the Full-Efficiency Threshold	76
9.4	Fitting the Second Knee and the Ankle	77
9.5	Comparing the Measurements with other Publications	81
10	Conclusion and Outlook	83
	Acknowledgments	85
	Bibliography	87
A	Constant Intensity Cut (CIC) for 40°	91
B	Energy Spectrum	93
B.1	Calculating the Hexagon Area for the Exposure	93
B.2	Investigating the Influence of the Full-Efficiency Threshold	93
B.3	Data Used for the Energy Spectrum	93

CHAPTER 1

Cosmic Rays and Extensive Air Showers

Since their discovery in 1912 [1], cosmic rays have been the subject of extensive studies. They originate from different galactic and extragalactic sources and propagate through the universe while being influenced by magnetic fields and other effects. Galactic cosmic rays are thought to be primarily produced by supernova remnants where they are accelerated to very high energies by shock waves [2]. Other sources and acceleration mechanisms are explained in Ref. [3]. On the other hand, it is assumed that extragalactic cosmic rays originate from other high-energy astrophysical phenomena like active galactic nuclei [4]. The sources and acceleration mechanisms as well as the transition from galactic to extragalactic components in the energy spectrum of cosmic rays continue to be the subject of ongoing research.

1.1 Composition and Propagation

The main components of cosmic rays are protons and heavier nuclei, with further contributions of anti-protons as well as electrons and positrons [5, p. 535]. Gamma rays and neutrinos were measured by various experiments at low and moderately large energies [5, pp. 545–552]. However, this thesis does not cover them since there are no reported measurements in the energy range of interest in the discussed experiment.

Depending on their energy, charged cosmic rays can either be trapped in the galaxy by magnetic fields or leave the galaxy for larger energies resulting in a loss of high-energy particles [6, p. 803]. Further propagation effects, e.g. radioactive decay and nuclear fragmentation, are explained in Ref. [7, 8].

1.2 Energy Spectrum

In principle, there are three observables for the measurement of cosmic rays: the particle type, arrival direction, and energy. While the arrival direction and particle type can be used to estimate the location of the source by using appropriate propagation models and the knowledge of galactic magnetic fields, the energy together with the particle type gives an insight into the mechanisms of cosmic accelerators.

Measuring a large number of cosmic rays over time is equivalent to measuring the flux $J(E)$. Understanding this flux is crucial since it is the main source for any deduction of models for cosmic accelerators. In other words, the measured flux is a window for testing and understanding astrophysical models.

The overall flux is a steeply falling power law $J(E) \propto E^{-\gamma}$ with the energy-dependent spectral index γ . To this day, many features in the spectrum have been identified and measured by different experiments. The most important features are [5, pp. 535, 542–545]:

- The *knee* at $E \approx 5 \text{ PeV} \approx 10^{15.7} \text{ eV}$ where γ changes from ~ 2.7 to ~ 3 . One possible explanation is that galactic sources reach their mass-dependent maximum energy to

which they can accelerate particles. In this scenario, the cutoff occurs at lower energies for lighter particles, followed by the cutoff of heavier elements.

- The *second knee* at $E \approx 100 \text{ PeV} = 10^{17.0} \text{ eV}$ where γ changes to ~ 3.3 . In the scenario described for the knee, this corresponds to heavier components, esp. iron, cutting off.
- The *ankle* at $E \approx 5 \text{ EeV} \approx 10^{18.7} \text{ eV}$ where γ changes to ~ 2.5 . A possible interpretation is the dominance of an extragalactic component.
- The *instep* at $E \approx 10 \text{ PeV} = 10^{19.0} \text{ eV}$ where γ changes to ~ 3.1 . It could be explained by the interplay of the helium and carbon-nitrogen-oxygen components together with a photodisintegration effect [9].
- A *suppression* at $E \approx 50 \text{ EeV} \approx 10^{19.7} \text{ eV}$ where the spectrum is highly suppressed by γ changing to ~ 5.2 . It is believed that cosmic accelerators reach their maximum energy and that propagation effects, e.g. the GZK effect, further suppress the spectrum [9].

Even though many explanations for the origin of these features exist, this still is a subject of current investigations and final conclusions have not been made. The spectrum of cosmic rays with its features is shown in Fig. 1.1. Since investigations of the mass composition are still ongoing, the measured energy spectrum is a combination of an energy-dependent mixture of different primary particles that is not fully understood.

This thesis focuses on the energy spectrum around and above the energy region of the second knee. An update of the reconstruction for events measured with the surface detector of the Pierre Auger Observatory with 750 m spacing is presented. It is argued that this allows an investigation of the second knee. Together with ongoing mass composition studies, this is an important step on the way to understanding the spectrum and physics in the region below the energy of the ankle.

1.3 Extensive Air Showers

While cosmic rays can be measured directly with satellite or balloon experiments for smaller energies [10], investigations of the high energy range of the spectrum are more difficult. Due to the steeply falling spectrum, the flux at these high energies is very low [6]. Therefore, huge detectors are needed to measure *ultra-high-energy cosmic rays* (UHECRs) with sufficient statistics. This can only be done with large ground-based experiments that measure not the primary particle, but the footprint of the secondary particle shower induced by the primary.

UHECRs penetrating the atmosphere interact with molecules of the air. This creates charged and neutral pions π^\pm and π^0 . The neutral pions decay into two photons. This starts a cycle of pair-production \rightarrow bremsstrahlung \rightarrow pair-production \rightarrow ... leading to an electromagnetic shower until the energy becomes too low to produce more particles [11, pp. 232–235]. The charged pions trigger a similar hadronic shower, producing more pions and nuclear fragments. For smaller energies where the distance to the next interaction point is larger than the decay length, the charged pions decay into muons and neutrinos [12, pp. 8–10]. This leads to a large shower footprint on the ground of sometimes several kilometers and, therefore, the name *extensive air showers* (EAS). Ref. [6, pp. 804–813] presents a more detailed explanation of EAS phenomenology. A schematic view of air showers is shown in Fig. 1.2.

While the longitudinal development of the shower can be observed with fluorescence telescopes measuring the excitations of air molecules by the shower particles, the lateral shower footprint of secondary particles can be measured by a grid of surface detector stations [6, pp. 814–824]. One experiment using this technique is the Pierre Auger Observatory which is discussed in the next chapter.

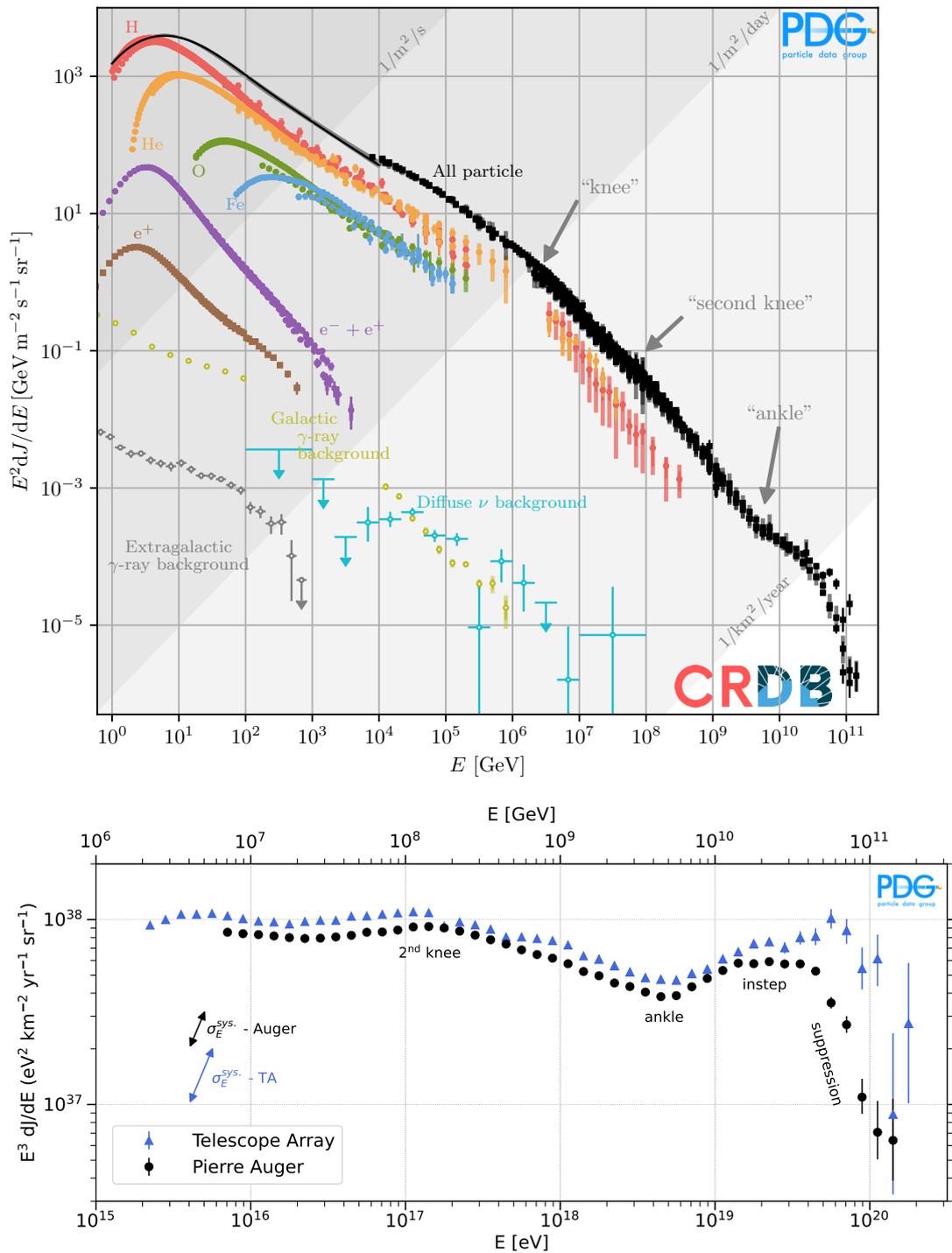


Figure 1.1: The spectrum of cosmic rays. *Top:* The spectrum for charged and neutral particles for low to high energies. The differential flux was multiplied by $(E/\text{eV})^2$ to make the features visible. Taken from Ref. [5, p. 536]. *Bottom:* The spectrum of ultra-high-energy cosmic rays as reported by the Pierre Auger Observatory and Telescope Array. The differential flux was multiplied by $(E/\text{eV})^3$. Taken from Ref. [5, p. 543].

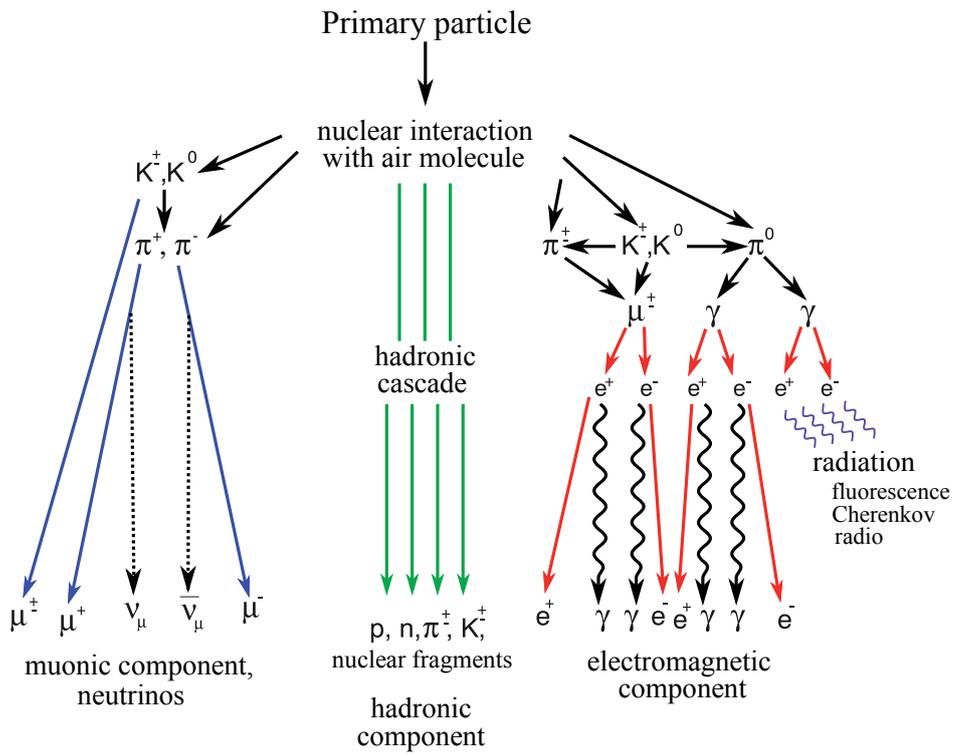


Figure 1.2: Schematic view of an extensive air shower (EAS). The primary particle interacts with the molecules of the atmosphere, producing an EAS due to cascade processes. A larger footprint of secondary particles can then be measured on the ground level. Taken from Ref. [13, p. 3], originally from Ref. [14].

CHAPTER 2

The Pierre Auger Observatory and Thesis Objectives

Set in Malargüe, Argentina, the Pierre Auger Observatory uses a hybrid approach to detect the secondary particles of cosmic rays. The longitudinal shower profile is observed with 4 fluorescence detectors (FD) consisting of a total of 24 fluorescence telescopes which allows direct observation of the shower development and gives an insight into its energy. Three additional telescopes overlook an additional detector array described in Section 2.1. Since the FD can only operate during moonless nights, the duty cycle is only $\sim 15\%$ [15].

In addition, a grid of 1660 surface detector (SD) stations with a spacing of 1.5 km is used to observe the lateral shower profile with a duty cycle of almost 100%. It spans over more than 3000 km^2 on an almost flat area. Each station consists of a *water-Cherenkov detector* (WCD) tank to measure the Cherenkov light produced by the shower particles going through the station and the necessary read-out electronics [15]. During the AugerPrime upgrade, each station was additionally equipped with a *surface scintillator detector* (SSD) and further detectors, i.e. radio antennas and underground muon chambers, were deployed [16]. However, this thesis focuses on the period before the upgrade was performed.

Important results observed in the analysis of the experiment's data are a dipole structure [17], the change of the mass composition of cosmic rays at higher energies [18], and the measurement of the features of the UHECR energy spectrum [9].

2.1 The SD-750 Detector

The footprint of EAS typically spans over a large area. Its size depends on the shower energy, i.e. the shower footprint is smaller for smaller energies. This means that the surface detector has a lower energy limit – a so-called *full-efficiency threshold* – where it cannot measure the lateral profile sufficiently enough to reconstruct the shower since the footprint does not trigger enough stations. This threshold obviously depends on the array spacing.

Therefore, the surface detector of the Pierre Auger Observatory does not only consist of the array mentioned above with 1500 m spacing (SD-1500) but also of a smaller array with 750 m spacing (SD-750) [15]. The SD-1500 together with the SD-750 and the FD are shown in Fig. 2.1. An even smaller array with 433 m (SD-433) was deployed in recent years but is not of further interest to this thesis [19].

These arrays lower the mentioned energy threshold allowing further investigation of the energy spectrum of UHECRs. One of the main goals is to measure the *second knee* of the spectrum which lies in the range of $\sim 10^{17}$ eV. While a measurement with the SD-433 has been presented recently [20], the possibility of a measurement with the SD-750 has only been suggested [21]. The main reason for this is that the original data-taking limited the energy threshold to a value above the position of the second knee.

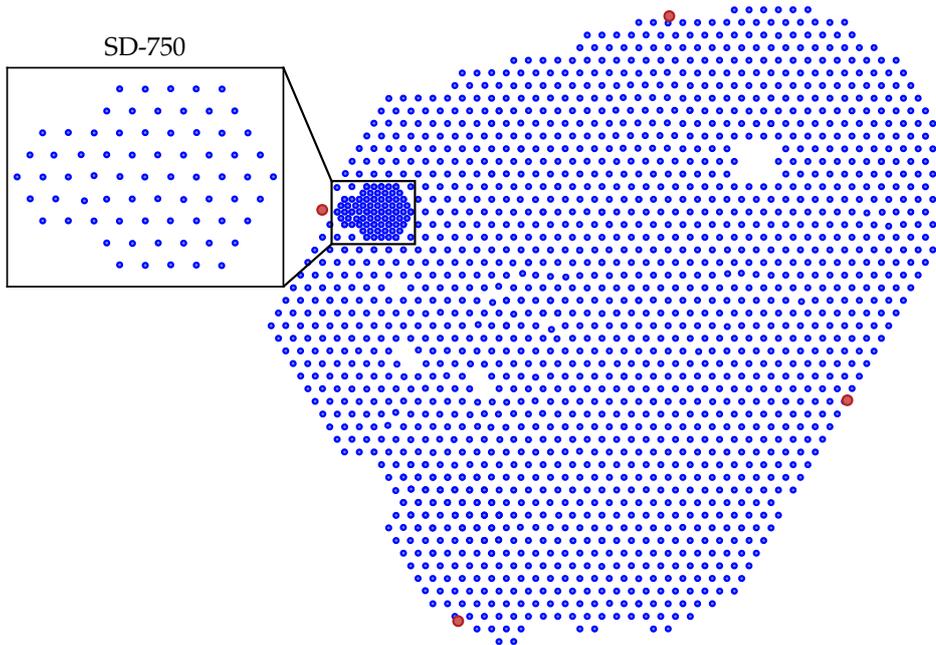


Figure 2.1: The SD-1500 and SD-750 arrays (blue) of the surface detector of the Pierre Auger Observatory. The red points represent the fluorescence detectors.

2.2 Lowering the Energy Threshold with New Triggers

As laid out in Ref. [22, pp. 50–57], two new station trigger algorithms called after their respective working principles *multiplicity of positive steps* (MoPS) and *time-over-threshold deconvoluted* (TOTd) were installed in 2014. These algorithms are sensitive to the electromagnetic component of the shower that dominates at large distances from the shower core. Therefore, they are sensitive to smaller signals leading to a reduction of the station trigger threshold.

In addition to the station triggers, a physics trigger called T4 selects events where the combined data of all triggered stations hints at a real physics event [15]. By reducing the station trigger threshold with the MoPS and TOTd triggers, the overall array trigger efficiency p_{T4} improves for lower energies. Since we are interested in good-quality events that are sufficiently sampled by the detector, we usually will use only events with the so-called 6T5 trigger for most studies. It requires that the station with the largest signal is surrounded by six operating neighbour stations to ensure that the impact point of the shower core can be reconstructed accurately [15].

The data-taking is usually divided into two phases, namely phase I from the beginning of data-taking until the end of 2021, when new read-out electronics were deployed during the AugerPrime upgrade, and phase II from 2022 onwards. However, the implementation of the new triggers divides phase I into two subphases. Phase Ia spans from the beginning of phase I to the end of 2013. Phase Ib spans from 2014 to the end of 2021 and contains more low-energy data due to the new triggers.

Including stations only triggered by MoPS and TOTd correctly in the reconstruction is essential when trying to lower the overall array trigger threshold. Even though the influence of the new triggers has been investigated in Ref. [22], the new-trigger-only stations are still not included in the standard reconstruction software `Offline` [23] due to the lack of knowledge about the signal distributions for low signals. The study presented in Ref. [21] used the information of these stations but with a simplified fitting model.

This thesis will present a full study on how the stations only triggered by MoPS and/or TOTd should be used in the reconstruction. I will show that this allows reducing the energy threshold so that *the second knee can be measured with the SD-750 for the first time*.

2.3 Reconstructing SD Events

This section briefly summarises the more detailed description of the standard reconstruction procedure using the `Offline` software package described in Ref. [23].

The important data provided by a station is its position, the measured signal S in units of a *vertical equivalent muon* (VEM), i.e. the amount of signal one vertical muon would deposit in the tank, and the time t when it triggered. Stations that did not trigger, so-called *silent stations*, can be included with the knowledge of the trigger probability as further information into the reconstruction. The reconstruction procedure is then performed as follows:

1. Initial estimation of the impact point (x_c, y_c, z_c) of the shower core by calculating the barycenter of the signals.
2. Estimation of the axis \hat{a} of the shower core using the position and timing information of the stations assuming a plane shower front. The zenith angle θ necessary for the following steps is the angle between this axis and a line perpendicular to the ground.
3. A likelihood fit of the *lateral distribution function* (LDF), which is described in detail in Chapter 4 and Chapter 6, by fitting a function $S(r)$ to the measured signals S_i at distances r from the shower axis. In an iterative process, the so-called shower size S_{ref} and the impact point $(x_c, y_c, z_c \approx 0)$ are fitted. This step requires a good knowledge of the underlying signal distribution and the trigger probability of each station, an issue which is addressed and solved in Chapters 3 and 4. The LDF also depends on a set of slope parameters that cannot be fitted for most low energetic events due to the lack of enough triggered stations to satisfy the necessary number of degrees of freedom. Therefore, a parameterisation of these slope parameters as functions of the other shower quantities is needed. A new parameterisation is shown in Chapters 5 and 6.
4. The axis \hat{a} with the new impact point is refitted with a curved shower front.
5. The shower size S_{ref} is used as an energy estimator and converted to an energy E of the primary particle. This energy conversion is done by cross-calibrating with the FD since this detector provides an energy measurement. The procedure to find an updated energy calibration is presented in Chapters 7 and 8.

Finally, the reconstruction of all observed events together with the knowledge of the detector properties then allows the measurement of the energy spectrum of UHECRs. This final step together with the fit of the spectrum features is discussed in Chapters 8 and 9.

Some chapters of this thesis have been previously made available as internal notes – so-called GAP notes – inside the Pierre Auger Collaboration.

An example of a measured shower footprint together with the fitted LDF using the current standard reconstruction is shown in Fig. 2.2. The LDF overshoots the data points for distances far from the shower core. The reason for this is discussed in the following chapters and a solution is presented by including the correct trigger probability and underlying signal distribution, updating the functional form of the LDF and using the additional data from stations only triggered by MoPS and/or TOTd. Note that the example shown in Fig. 2.2 is one of the rare high-energy events and that most events have much less triggered stations.

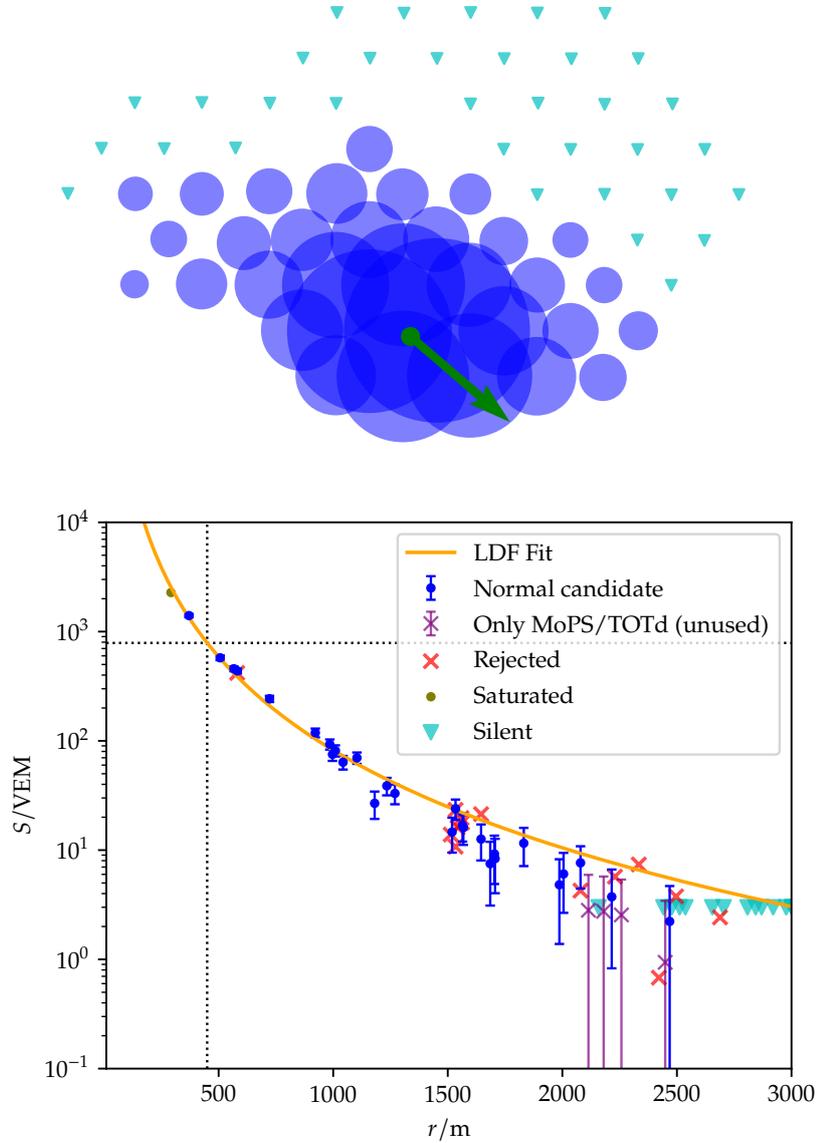


Figure 2.2: Example of a measured and reconstructed shower with the SD-750. *Top:* Shower footprint where the size of the blue dots corresponds to the measured signal. Stations that have not triggered are shown as triangles. The green dot is the reconstructed impact point of the shower core. The green arrow shows the projection of the reconstructed shower axis \hat{a} into the ground plane. *Bottom:* Measured signal over the distance to the reconstructed shower core together with the fit of the lateral distribution function. Red crosses represent stations that were not used in the fitting procedure since they were rejected. This can happen for multiple reasons, e.g. the station was malfunctioning or is part of a different detector array. The dashed lines represent the position of the shower size S_{ref} and the reference distance $r_{\text{ref}} = 450$ m.

CHAPTER 3

Investigating the Signal Variance for Small Signals

This chapter was first made available as GAP–2023–55 [24].

In order to include stations that were only triggered by MoPS or TOTd triggers in the SD reconstruction of the LDF, the signal variance model has to be revisited and validated for small signals. This is necessary since these two triggers were designed to be sensitive to small signals, a fact which results in the lowering of the over-all trigger threshold. In other words, we are interested in signals on the level of around and below 1 VEM, but still larger than, for example, 0.1 VEM.

For this task, simulations with an artificially lowered trigger threshold and placement of dense rings at several large distances from the shower core – where small signals are expected – were performed. Showers with different zenith angles were used to give a first impression on whether an update to the signal variance models described in GAP–2012–12 [25] and GAP–2014–35 [26] is needed or not.

3.1 Simulation Setup

To examine fluctuations of small signals, we need simulations with artificially lowered trigger thresholds using the detector simulation module of `Offline`. To collect data of stations where only a few particles arrive, the trigger threshold for the stations was reduced to ~ 0.1 VEM.¹ Using the current LDF parameterisation, one can find the distances to the shower core where small signals are expected. For completeness, also distances with larger signals were included. Dense rings, all with 24 stations with equally-spaced azimuth angles ψ_i , were set up at the following radial distances to the core, chosen quite *ad hoc*,

$$r = (350, 450, 550, 700, 880, 970, 1130, 1620, 1920, \text{ and } 2800) \text{ m.} \quad (3.1)$$

Simulated showers for zenith angles $\theta \in \{0^\circ, 38^\circ, 56^\circ\}$ from the CORSIKA fixed library for protons were used for this analysis. The interaction model used was QGSJetII-04.

3.2 Vertical Showers

The signals obtained from the dense rings in simulations of vertical showers are shown in Fig. 3.1. For each dense ring in a simulation the mean and the standard deviation of the signals,

$$\bar{S} = \frac{1}{N_{\text{ring}}} \sum_i S_i \quad \text{and} \quad \sigma^2(S) = (\text{std}(S))^2 = \frac{1}{N_{\text{ring}} - 1} \sum_i (S_i - \bar{S})^2, \quad (3.2)$$

were obtained, where N_{ring} is the number of stations in the given ring and $\text{std}(S)$ is the sample standard deviation of signals. The standard deviation $\sigma(S) = \text{std}(S)$ of the signals on each

¹The `SdTraceCalibrator` threshold was set to 0.5 VEM and the T1 threshold in the `TankTriggerSimulator` module of `Offline` was set to 0.1 VEM.

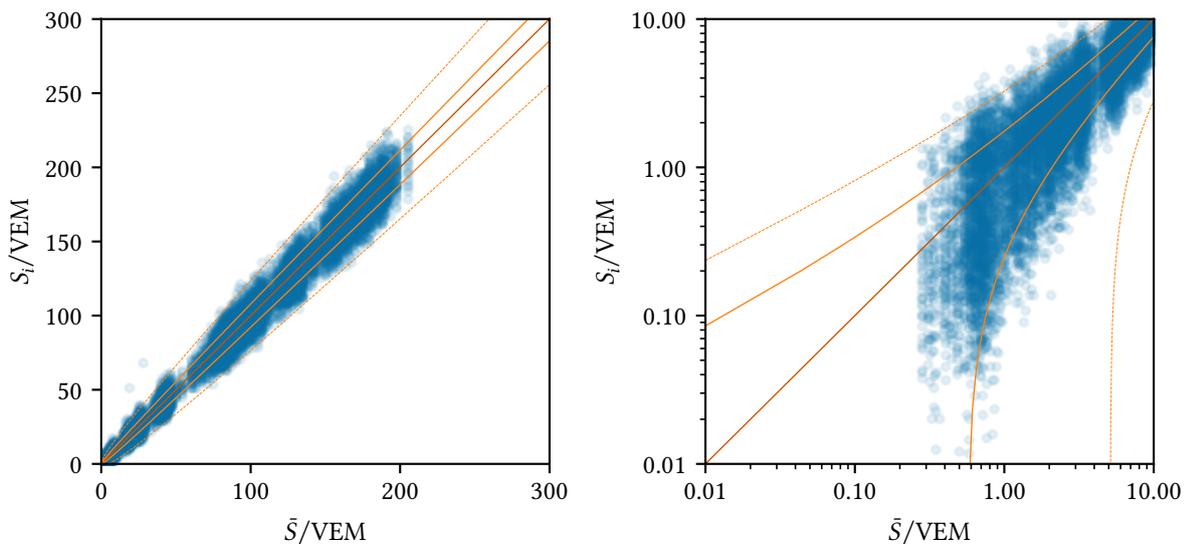


Figure 3.1: Simulated data for showers with $\theta = 0^\circ$. For each dense ring, station signals S_i are plotted as a function of the mean signal \bar{S} . The dark orange line represents the mean of the signal distributions and the straight light orange outlines the 1σ -errorband as given by the current signal variance model described in GAP-2014-35 [26]. The dotted light-orange lines represent the 3σ -errorband. The right plot shows a zoomed-in version of the left plot at small signals and in log scale.

ring will be used as a measure of the signal fluctuations and will be compared to the existing signal variance models. The results are shown in Fig. 3.2, where we can see that for the vertical showers the current signal variance models included in Offline agree well with the simulated data for large and small signals. Note that we are interested in the validity of the signal variance models only for signals with $\lg(S/\text{VEM}) \gtrsim -0.5$, i.e. $S \gtrsim 0.3 \text{ VEM}$.

3.3 Inclined Showers

Repeating the same procedure as for vertical also for inclined showers, the plots in Fig. 3.3 are obtained. The deviations are expected since for inclined showers asymmetry effects in the dense rings have to be considered. For inclined showers the signals S_i of the stations in the dense ring can be in the first order described as

$$S_i = \bar{S}(1 + a \cos \psi_i), \quad (3.3)$$

where \bar{S} is now the asymmetry-corrected mean signal of the dense ring, a is the asymmetry amplitude, and ψ_i is the shower-plane azimuth of the station i in the ring. The mean \bar{S} and the amplitude a are fitted to the signals S_i obtained from simulations. Additionally, it is convenient to introduce asymmetry-corrected signal residuals

$$S'_i = S_i - \bar{S}(1 + a \cos \psi_i). \quad (3.4)$$

The signal fluctuations are then estimated as the standard deviation of these residuals,

$$\sigma(S') = \text{std}(S'_i). \quad (3.5)$$

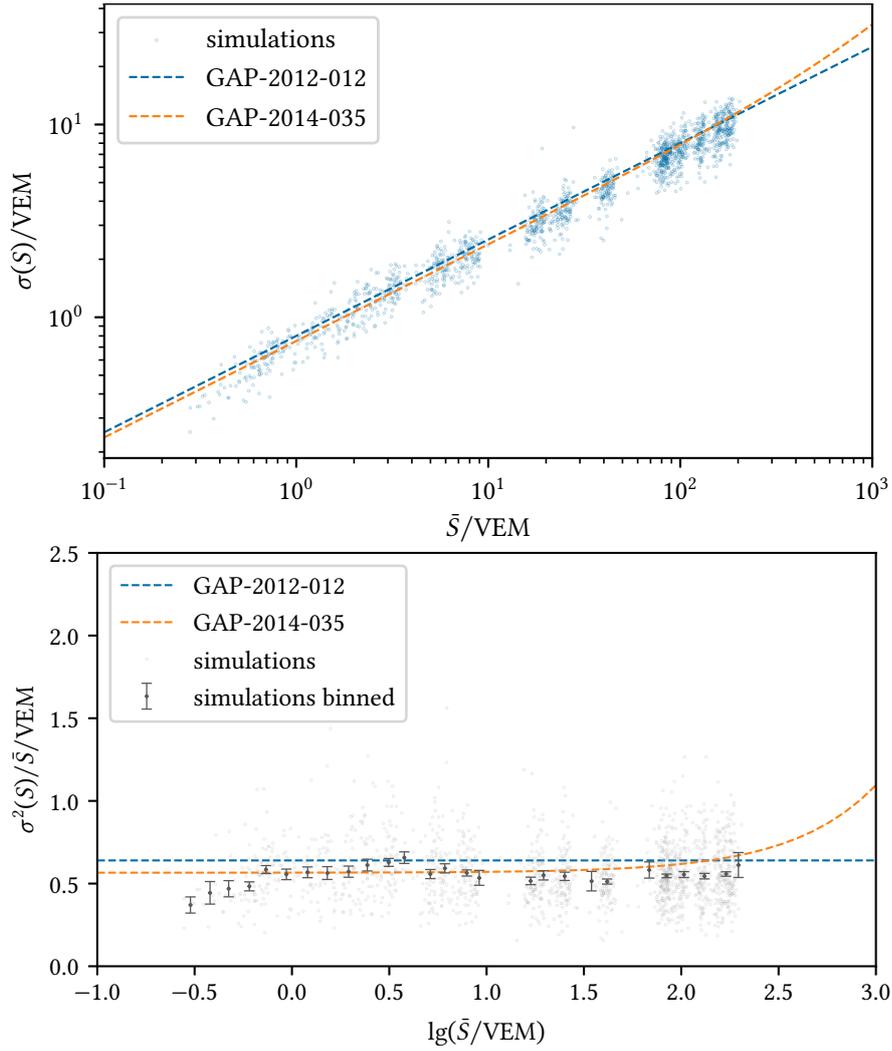


Figure 3.2: Signal fluctuations for simulated events with zenith angle $\theta = 0^\circ$. *Top:* Scatter plot of signal fluctuations $\sigma(S)$ as a function of the mean signal \bar{S} on the dense ring. The two existing models implemented in `Offline` are also shown. *Bottom:* The same data as above but plotted instead as the relative variance $\sigma^2(S)/\bar{S}$ to highlight departures from the pure Poissonian behaviour.

Using these data and a fit quality cut on the asymmetry requiring $a \in [0, 1]$ and that its error $\sigma_a < 0.5$ leads to data shown in Fig. 3.4. Obviously, the data differs from the models for small signals. However, this behaviour vanishes if a stricter asymmetry cut with $a \in [0, 1]$ and $\sigma_a < 0.2$ is used, as shown in Fig. 3.5. For signals $\lg(S/\text{VEM}) \gtrsim 1.0$, the models seem to overestimate the variance. These deviations are likely an effect of the naive asymmetry fitting since the current variance models were developed and validated using data of doublet stations with signals in this range. A further look into this is presented in the appendix of [27]. Two conclusions can be drawn:

- The asymmetry-corrected simulation data confirms the current signal variance models for showers with $\theta = 38^\circ$.

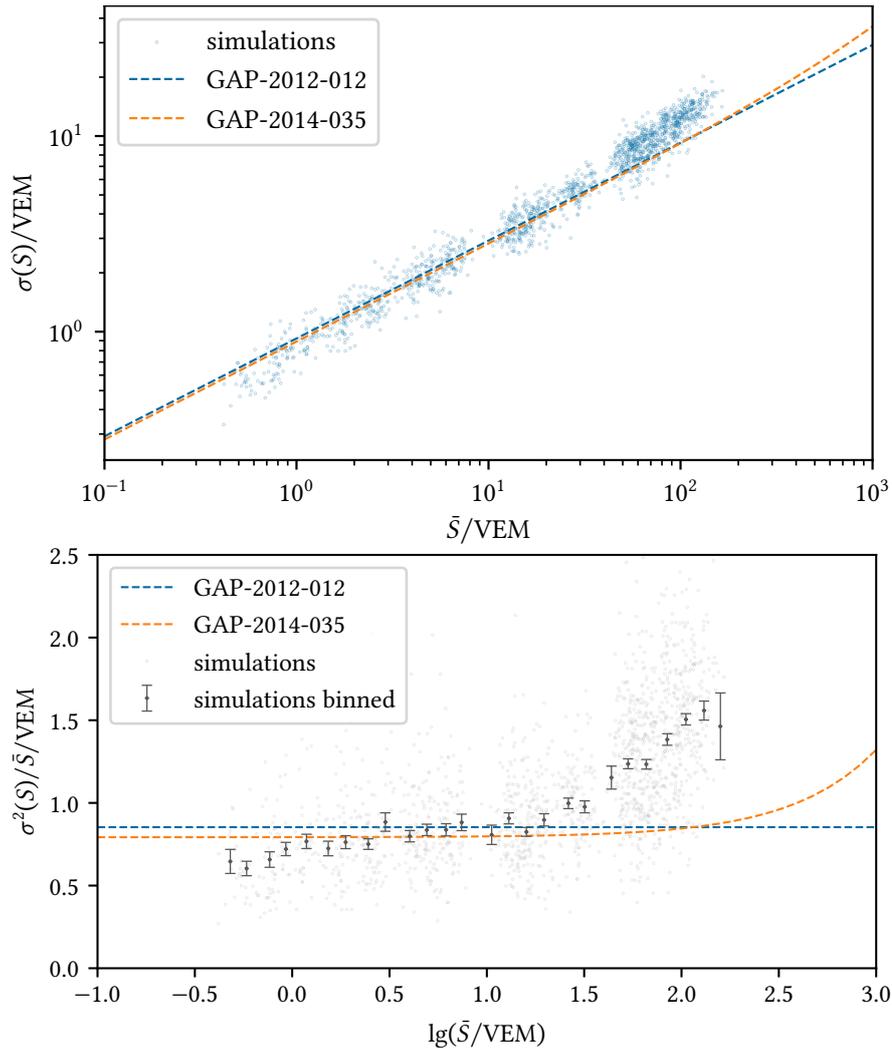


Figure 3.3: Signal fluctuations for simulated events with zenith angle $\theta = 38^\circ$. *Top:* Scatter plot of the signal fluctuation $\sigma(S)$ as a function of the mean signal \bar{S} on the dense ring. The two existing models implemented in Offline are also shown. *Bottom:* The same data but plotted as the relative variance $\sigma^2(S)/\bar{S}$ which emphasises deviations from the pure Poissonian behaviour.

- The naive asymmetry fit for each dense ring fails for small signals since the fluctuations are larger than the expected asymmetry effects.² Therefore, the asymmetry correction is not applied for small signals.

Finally, examining very inclined showers with $\theta = 56^\circ$ we still see acceptable compatibility with the current signal variance models (see Fig. 3.6). Large deviations mainly arise in bins with only a few data points, while the discrepancy with respect to the models in bins with enough data points are likely an effect of an inadequate asymmetry fit. Nevertheless, a better estimation of the asymmetry is beyond the scope of this thesis.

²Asymmetry becomes statistically observable only when a lot of particles are depositing the signal.

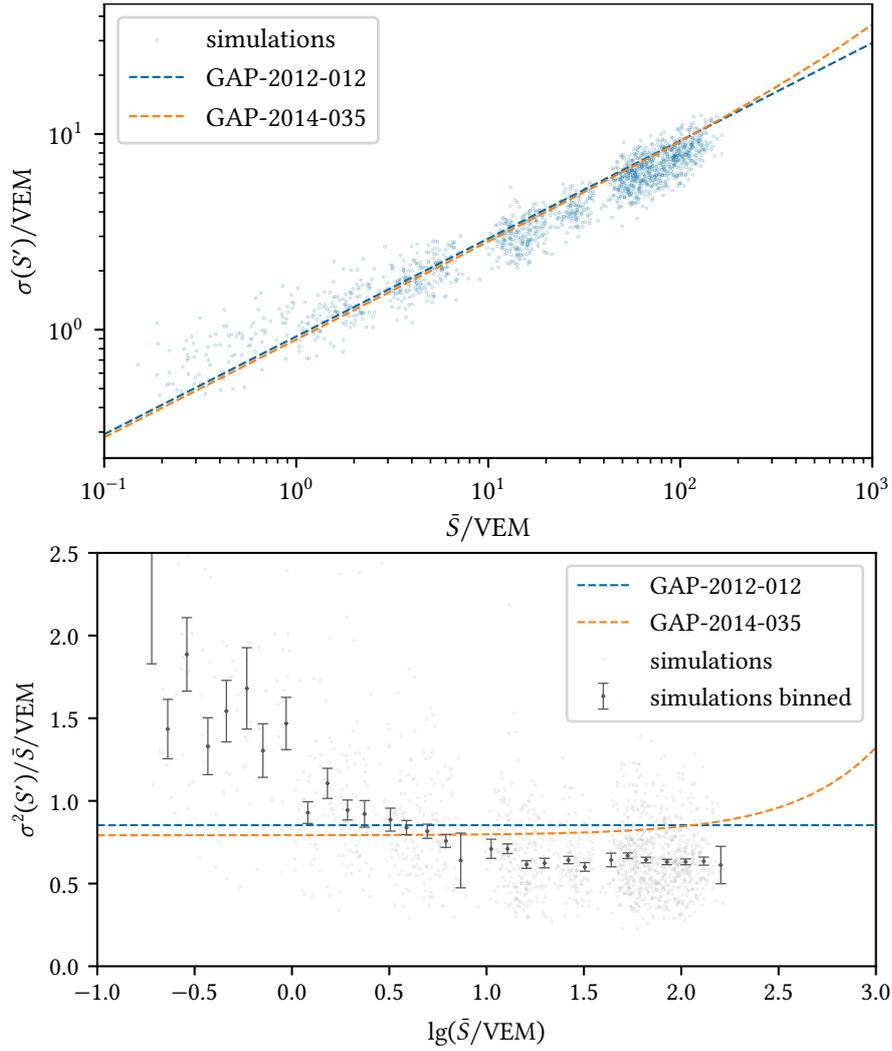


Figure 3.4: Signal fluctuations for simulated events with zenith angle $\theta = 38^\circ$ (the same data as in Fig. 3.3) with asymmetry-corrected signals S'_i instead of S_i . The applied asymmetry-fit cuts are $a \in [0, 1]$ and $\sigma_a < 0.5$. *Top:* Scatter plot of the fluctuations $\sigma(S')$ as a function of the mean signal \bar{S} on a dense ring. The two existing models implemented in Offline are also shown. *Bottom:* The same data but in the Poissonian representation with the relative variance $\sigma^2(S')/\bar{S}$.

3.4 Implications for Stations with Small Signals

The studies presented in this chapter did not find any evidence that the current signal variance models would not be appropriate for signals below their original range of validity, i.e. at the small(er) signals, which are, nevertheless, still large enough, i.e. for $S > 0.5$ VEM or even ~ 0.3 VEM. Therefore, the current signal variance models can be used in the SD LDF reconstruction for small signals, where stations are predominantly triggered only by the MoPS and/or the TOTd triggers.

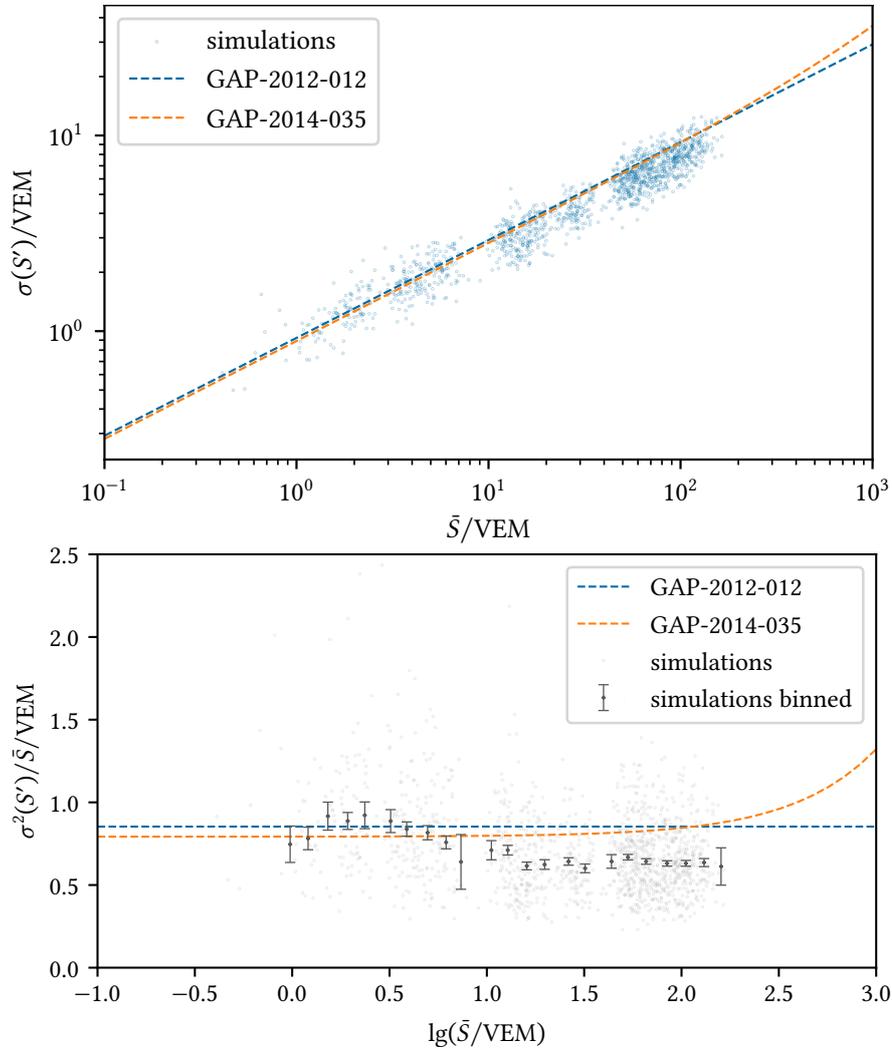


Figure 3.5: Signal fluctuations for simulated events with zenith angle $\theta = 38^\circ$ (the same data as in Fig. 3.3 and Fig. 3.4) with asymmetry-corrected signals S'_i and stricter cuts on the asymmetry-fit. The applied asymmetry-fit cuts are $a \in [0, 1]$ and $\sigma_a < 0.2$. *Top:* Scatter plot of the fluctuations $\sigma(S')$ as a function of the mean signal \bar{S} on a dense ring. The two existing models implemented in Offline are also shown. *Bottom:* The same data but in the Poissonian representation with the relative variance $\sigma^2(S')/\bar{S}$.

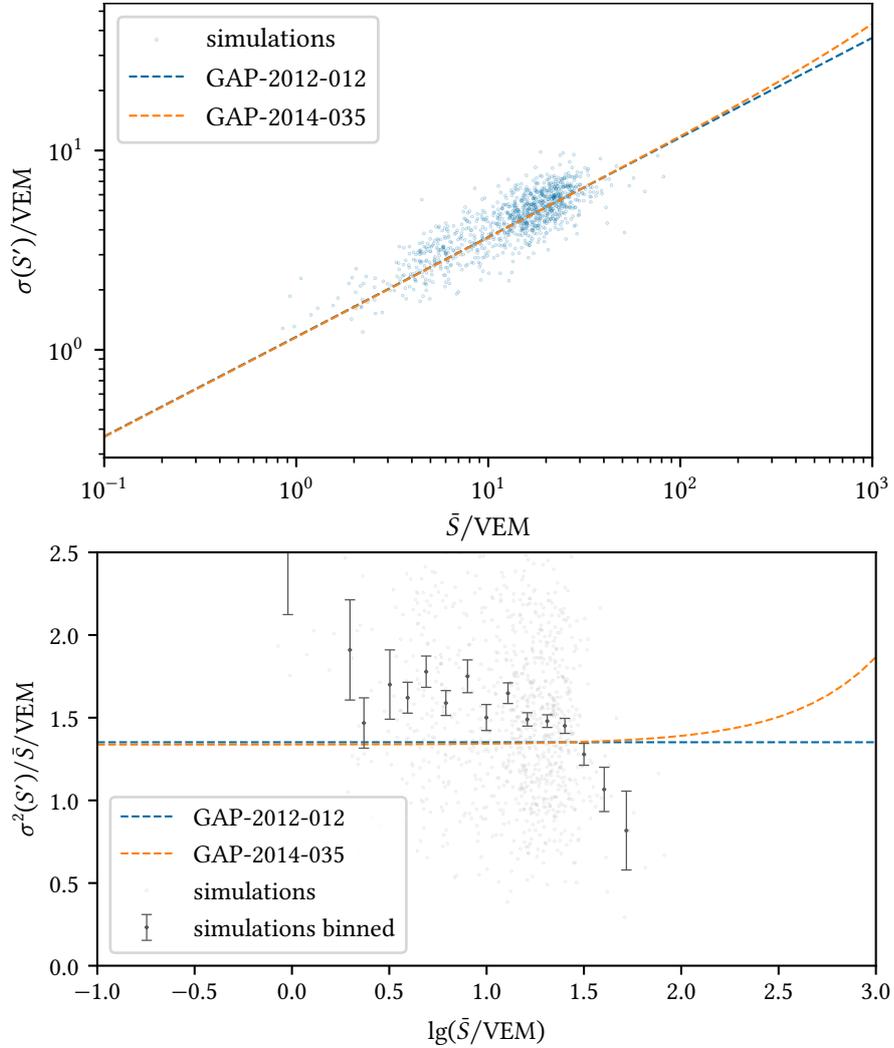


Figure 3.6: Signal fluctuations for simulated events with zenith angle $\theta = 56^\circ$ with asymmetry-corrected signals S'_i . The applied asymmetry-fit cuts are $a \in [0, 1]$ and $\sigma_a < 0.2$. *Top:* Scatter plot of the fluctuations $\sigma(S')$ as a function of the mean signal \bar{S} on a dense ring. The two existing models implemented in Offline are also shown. *Bottom:* The same data but in the Poissonian representation with the relative variance $\sigma^2(S')/\bar{S}$.

CHAPTER 4

Updating the LDF Likelihood

This chapter was first made available as GAP-2024-40 [28].

When reconstructing SD events, we have to fit the lateral distribution function (LDF) using a maximum likelihood method. The current likelihood setup uses a signal-to-particle conversion to decide whether a Gaussian or a Poissonian likelihood contribution should be used for a triggered station. However, the underlying distribution needs to be understood when using low-signal stations triggered by only the muon-insensitive MoPS and TOTd triggers, because the validity of the Poissonian particle-counting assumption is no longer clear. This chapter will present a new likelihood that tries to solve all these issues, while including the correct station trigger probabilities.

4.1 Current Likelihood Setup

For triggered stations with measured signal S and given signal uncertainty σ_S , the normal signal PDF, $\mathcal{N}(S; S_{\text{pred}}, \sigma_S)$, is first converted into a Poissonian-like PDF, which is scaled with a factor k so that the expected value and the variance of the resulting PDF match the signal and the signal variance model, respectively. The equivalent number of observed particles is thus $n = k(S/\text{VEM})$ and its expected value $\nu = k(S_{\text{pred}}/\text{VEM})$, where the scaling factor k is obtained from the variance requirement $\sigma_n^2 = n$, i.e. $k = (S/\text{VEM})/(\sigma_S/\text{VEM})^2$. Since the signal uncertainty for the surface detector is modelled with a Poissonian-like uncertainty factor f_{σ_S} ,

$$\sigma_S/\text{VEM} = f_{\sigma_S}(\theta)\sqrt{S/\text{VEM}} \quad (4.1)$$

we finally get $k = (f_{\sigma_S}(\theta))^{-2}$. In addition, in the current version of Offline the scaling factor must be strictly greater than 1, i.e. $k = \max(1, (f_{\sigma_S}(\theta))^{-2})$. For the uncertainty factor, we can use either

$$f_{\sigma_S}(\theta) = 0.34 + 0.46 \sec \theta \quad (4.2)$$

from Ref. [25] or a more recent and more sophisticated uncertainty

$$\begin{aligned} \left[\frac{\sigma_S}{\text{VEM}}\right]^2 &= [f_{\sigma_S}(\theta)]^2 \frac{S}{\text{VEM}} + \left[\frac{0.023S}{\text{VEM}}\right]^2 \quad \text{with} \\ f_{\sigma_S}(\theta) &= 0.865[1 + 0.593(\sec \theta - \sec 35^\circ)] \end{aligned} \quad (4.3)$$

from Ref. [26], which is valid for signals down to 0.5 VEM. The signal uncertainty in Offline is calculated using the *predicted signal* S_{pred} instead of the measured signal S .

In the current likelihood, for $n > 30$ a Gaussian PDF is used for each station,

$$p_{\text{high}}(n; \nu, \sigma_n) = \mathcal{N}(n; \nu, \sigma_n) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{1}{2}\left(\frac{n-\nu}{\sigma_n}\right)^2\right], \quad (4.4)$$

while for signals with $n \leq 30$ a scaled Poissonian is used instead,

$$p_{\text{low}}(n; \nu) = \text{Poi}(n; \nu) = e^{-\nu} \frac{\nu^n}{n!} = \frac{e^{n \ln \nu - \nu}}{\Gamma(n+1)}. \quad (4.5)$$

For the zero-signal stations, the probability of not observing any signal is the complement of the trigger probability, $p_{\text{zero}}(\nu) = 1 - p_{\text{trig}}(\nu)$. The trigger probability is modelled using a Poissonian PDF to observe more than 3 particles, i.e.

$$p_{\text{zero}}(\nu) = 1 - \text{Poi}(n > 3; \nu) = \text{Poi}(n \leq 3; \nu) = \sum_{j=0}^3 \text{Poi}(j; \nu) = e^{-\nu} \left[1 + \nu + \frac{\nu^2}{2} + \frac{\nu^3}{6} \right]. \quad (4.6)$$

An additional treatment is necessary for stations where the true signal could not be measured due to saturation effects of the detector. In this case, we have two possibilities. A recovery algorithm can estimate a possible measured signal S_{rec} and its uncertainty σ_{rec} in many cases. If we decide to use this recovery, we can simply use a Gaussian PDF term, i.e.

$$p_{\text{rec}}(n_{\text{rec}}; \nu, \sigma) = \mathcal{N}(n_{\text{rec}}; \nu, \sigma) \quad \text{where} \quad \sigma^2 = \sigma_S^2 + \sigma_{\text{rec}}^2. \quad (4.7)$$

If we do not want to use the recovery or if it is not available, the measured saturated signal will be used only as a lower limit for the predicted signal. The likelihood contribution is therefore

$$p_{\text{sat}}(n; \nu, \sigma_n) = \text{Sat}(n; \nu, \sigma_n) := \text{erfc} \left(\frac{n - \nu}{\sqrt{2} \sigma_n} \right). \quad (4.8)$$

If a recovered signal is available, n in Eq. (4.8) is replaced by $k(S_{\text{rec}}/\text{VEM} - \sigma_{\text{rec}}/\text{VEM})$.

The final log-likelihood $\mathcal{L} = \ln L$ is thus assembled as

$$\mathcal{L} = \ln \prod_i^{\text{stations}} p_i = \sum_i \ln \left(\begin{array}{ll} \text{Sat}(n_i; \nu_i, \sigma_i) & ; \text{non-recovered saturated stations} \\ \mathcal{N}(n_i; \nu_i, \sigma_i) & ; n_i > 30 \\ \text{Poi}(n_i; \nu_i) & ; n_i \leq 30 \\ p_{\text{zero}}(\nu_i) & ; n_i \equiv 0 \end{array} \right), \quad (4.9)$$

where in station i the measured signal is S_i and the LDF prediction is $S_{\text{pred},i}$, which converts to the observed number of particles $n_i = k(S_i/\text{VEM})$ and the predicted number of particles $\nu_i = k(S_{\text{pred},i}/\text{VEM})$. σ_i is the to $S_i/S_{\text{pred},i}$ corresponding signal uncertainty. In the case of recovered or saturated stations the replacements as described above have to be made.

4.2 Criticism of the Current Setup

As mentioned in Section 4.1, the original dependency of σ on the measured signal S should rather be replaced by the predicted signal S_{pred} since S might be just an upward fluctuation. This change was already applied in Offline.

A self-consistent likelihood has to also satisfy the requirement that the probability of observing any value of S given the prediction S_{pred} should be 1, or formally,

$$\int_0^\infty p(S; S_{\text{pred}}, \sigma_{S_{\text{pred}}}) dS \equiv 1 \quad \text{for any } S_{\text{pred}} > 0. \quad (4.10)$$

This is not the case for the current setup since the terms for stations with non-zero signals do not include a trigger probability $p_{\text{trig}}(S_{\text{pred}})$ suppression factor that is required to fulfil the aforementioned normalisation.

It is also worth noting that while the estimated trigger probability as given in Eq. (4.6) was a reasonable choice at the time of its implementation, we should nowadays use a data-driven model as discussed in Section 4.4.

In addition, the Poisson contribution in Eq. (4.5) is a mass distribution function instead of a PDF and is only well-defined for positive integer arguments. Since neither $n!$ nor its replacement $\Gamma(n + 1)$ can be simply evaluated with real arguments (the signal and therefore the effective particle number is indeed not an integer), the solution in the past was to use the rounded effective particle number $\lfloor n \rfloor$ instead. Clearly, this is not a viable solution. Furthermore, such a function has a complicated and non-analytical normalisation and is not very practical to use. One could argue that the number of injected particles in a tank is a discrete quantity, but – even when neglecting the fact that different types of particles will deposit strongly varying signals – the produced signal is still subject to a detector response. The Poisson process would thus have to be convolved with a normal distribution, effectively converting the Poissonian mass distribution into a probability distribution function for the signal.

In the end, we need to better understand the underlying distribution of a signal in a station. Using this knowledge, it becomes clear how to set up this likelihood correctly.

4.3 Investigating the Distribution of Small Signals

To unify the distinct treatment of high and low signals, it is necessary to find the underlying distribution for small signals.

For this task, the simulations presented in Chapter 3 are used. Two possible continuous distributions were fitted to the obtained simulated data. The first choice was a truncated normal distribution. However, this has a non-trivial normalisation. Therefore, the second choice, a log-normal distribution, is preferred, meaning a normal distribution for the logarithm of the signal (while ignoring silent stations since they are treated separately in the likelihood). It is practical to use the log quantities

$$s := \ln(S/\text{VEM}) \quad \text{and} \quad \mu := \ln(S_{\text{pred}}/\text{VEM}) \quad (4.11)$$

of the signal S and the predicted signal S_{pred} . The signal uncertainty $\tilde{\sigma}(S_{\text{pred}}, \theta)$, as modelled in Eq. (4.3), can be propagated into the log-space to the first order of accuracy¹ and is then given by

$$\sigma := \sigma(\mu, \theta) = \frac{d\mu}{dS_{\text{pred}}} \tilde{\sigma}(S_{\text{pred}}, \theta) = \frac{\tilde{\sigma}(S_{\text{pred}}, \theta)}{S_{\text{pred}}} \quad (4.12)$$

since all values are positive anyway. This propagation is used for signal uncertainties of saturated and recovered stations, too.

Looking at Fig. 4.1 justifies our assumption of a log-normal distribution. The log-normal distribution on the left matches the distribution of the (log) signals very well all the way down to the very small signals. Around $s = 0$ ($S \approx 1$ VEM), a second peak becomes visible in some distributions. Since this is – by definition – the expected signal of a vertical muon, this behaviour was expected for vertical showers. Therefore, the peak is an artifact of counting single muons. As shown in Chapter 3, asymmetry corrections should be applied for inclined showers. Since the asymmetry fit, described in Chapter 3, fails for signals $S \lesssim 3$ VEM, we will use the original signals below this threshold. The signal distributions for $\theta = 38^\circ$ are shown in Fig. 4.2 where the asymmetry-corrected signals were used for the first three rows. The proposed log-normal distribution fits the data well, while the discussed muon peak is smeared out.

¹We used $\ln(1 + x) \approx x$ for $\sigma^2 = \ln(1 + \tilde{\sigma}^2/S_{\text{pred}}^2)$ (see Ref. [29, pp. 34–35] with $\tilde{\sigma}^2 = V[x]$ and $S_{\text{pred}} = E[x]$).

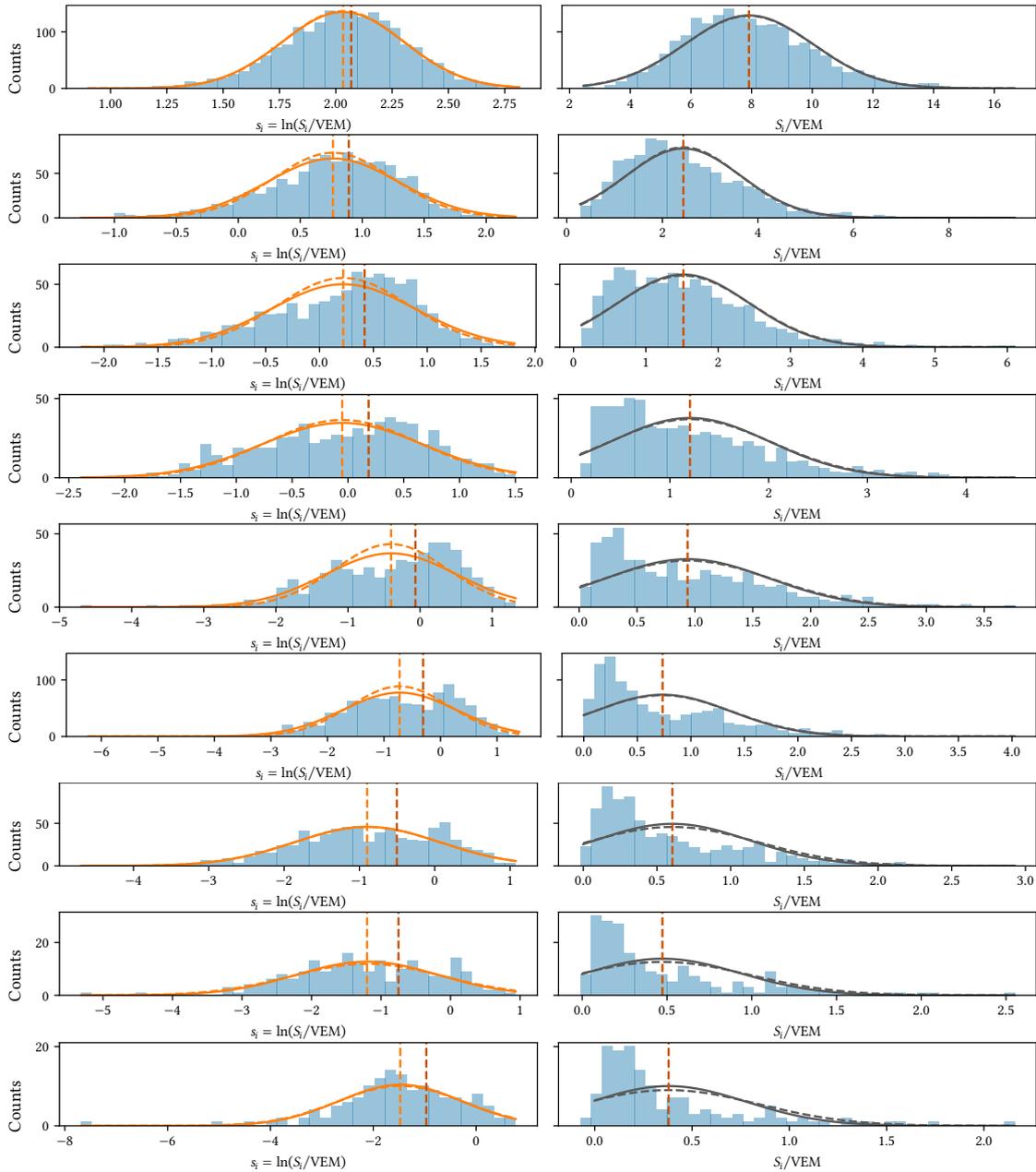


Figure 4.1: *Left:* Distributions of the log signals s_i for $\theta = 0^\circ$. The full-line envelope represents a Gaussian with \bar{s}_i and $\sigma = \text{std}(s_i)$. The dashed envelope shows the same, but using the (propagated) σ given by the variance model. The orange vertical line represents the mean \bar{s}_i , while the red line represents $\ln \bar{S}_i$. *Right:* Distributions of signals S_i for $\theta = 0^\circ$. The full-line envelope represents a truncated Gaussian with \bar{S}_i and $\sigma = \text{std}(S_i)$. The dashed envelope shows the same, but using the σ given by the variance model. The vertical red line represents the mean \bar{S}_i .

The signal distributions with fixed x -axes are shown in Fig. 4.4 for $\theta = 0^\circ$ and in Fig. 4.5 for $\theta = 38^\circ$ at the end of the chapter.

4.3 Investigating the Distribution of Small Signals

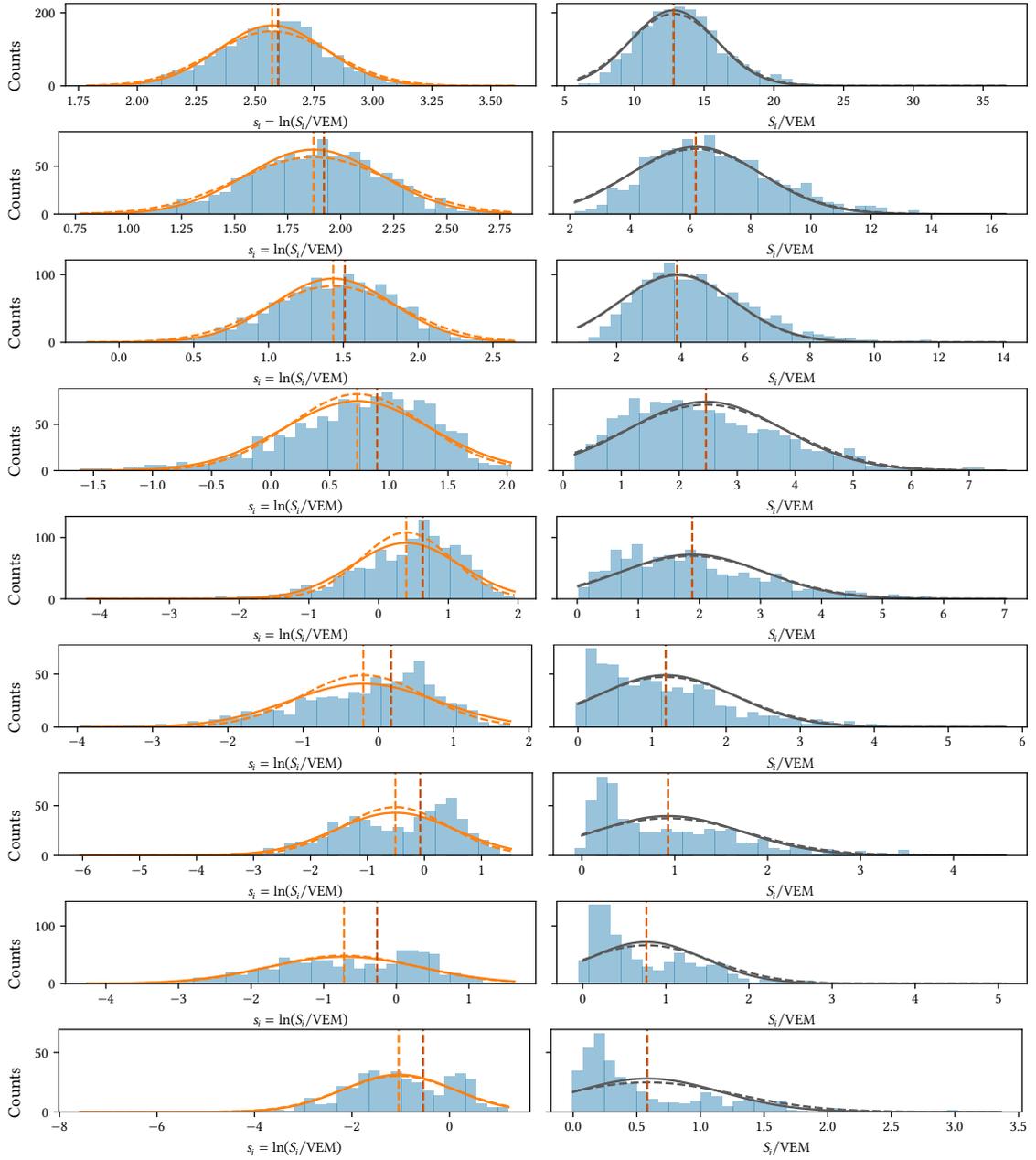


Figure 4.2: *Left:* Distributions of the log signals s_i for $\theta = 38^\circ$. The full-line envelope represents a Gaussian with \bar{s}_i and $\sigma = \text{std}(s_i)$. The dashed envelope shows the same but using the (propagated) σ given by the variance model. The vertical orange line represents the mean \bar{s}_i while the red one represents $\ln \bar{S}_i$. *Right:* Distributions of the signals S_i for $\theta = 38^\circ$. The full-line envelope represents a truncated Gaussian with \bar{S}_i and $\sigma = \text{std}(S_i)$. The dashed envelope shows the same but using the σ given by the variance model. The vertical red line represents the mean \bar{S}_i .

4.4 Updating the Likelihood Using the Trigger Probability

For the final likelihood, the trigger probability should be correctly included. The trigger probability, as modeled in Eq. (4.6), can nowadays be replaced with the trigger probability obtained directly from Auger data; see e.g. Ref. [30]. The trigger probability given in Ref. [22] using the log quantities then reads as

$$p_{\text{trig,new}}(\mu, \theta) = T_N \exp\left(-\frac{(\mu - T_\mu)^2}{T_\sigma^2}\right) + \text{Fer}\left(\frac{\mu - T_X}{T_\Delta}\right) \quad (4.13)$$

where $\text{Fer}(x) = (1 + e^{-4x})^{-1}$ is the standardised Fermi function, and the parameters $T_i = T_i(\theta)$ are given by

$$T_N(\theta) = (0.091_{\pm 0.084} - 0.252_{\pm 0.049} z), \quad (4.14)$$

$$T_\mu(\theta) = (0.261_{\pm 0.002} + 0.174_{\pm 0.039} z) \ln 10, \quad (4.15)$$

$$T_\sigma(\theta) = (0.097_{\pm 0.007}) \ln 10, \quad (4.16)$$

$$T_X(\theta) = (0.173_{\pm 0.006} + 0.170_{\pm 0.004} z - 0.246_{\pm 0.048} z^2) \ln 10, \quad (4.17)$$

$$T_\Delta(\theta) = -4(0.225_{\pm 0.006}) \ln 10, \quad (4.18)$$

where $z = \sin^2 \theta$. The values in brackets are taken from Ref. [22] with the parameter definition being updated to match our definition of the Fermi function and the log signal.

Alternatively, we can also use a simple function like

$$p_{\text{trig,simp}}(\mu) = \text{Fer}\left(\frac{\mu - \mu_{1/2}}{w}\right) \quad (4.19)$$

where the standardised Fermi function was used to fit² the data from Ref. [30]. The obtained parameters are $\mu_{1/2} = \ln 1.54844$ and $w = 0.921006 \ln 10$ for the data with the new triggers and $\mu_{1/2} = \ln 2.7692$ and $w = 0.8645 \ln 10$ for the data where only old triggers are available. The two fits are shown as red and green lines in Fig. 4.3, respectively.

Since $p_{\text{trig,new}}(\mu_i, \theta)$ as given in Eq. (4.13) is only valid for events where MoPS and TOTd triggers were already available, the trigger probability is

$$p_{\text{trig}}(\mu_i, \theta) = \begin{cases} p_{\text{trig,simp}}(\mu_i) & \text{for periods with old triggers only} \\ p_{\text{trig,new}}(\mu_i, \theta) & \text{for periods with new triggers} \end{cases}. \quad (4.20)$$

Using the correct values for S_0 and w as shown above then gives the correct trigger probability without the new triggers.

We can now write the new log-likelihood \mathcal{L} simply as

$$\mathcal{L} = \sum_i \ln \left(\begin{cases} p_{\text{trig}}(\mu_i, \theta) \text{Sat}(s_i; \mu_i, \sigma_i) & \text{for non-recovered saturated stations} \\ p_{\text{trig}}(\mu_i, \theta) \mathcal{N}(s_i; \mu_i, \sigma_i) & \text{for non-saturated or recovered stations} \\ 1 - p_{\text{trig}}(\mu_i, \theta) & \text{for silent stations} \end{cases} \right). \quad (4.21)$$

²The Fermi function fitted best compared to other sigmoidal functions like $[1 + \text{erf}(\sqrt{\pi}x)]/2$ and $[\pi/2 + \arctan(\pi x)]/\pi$.

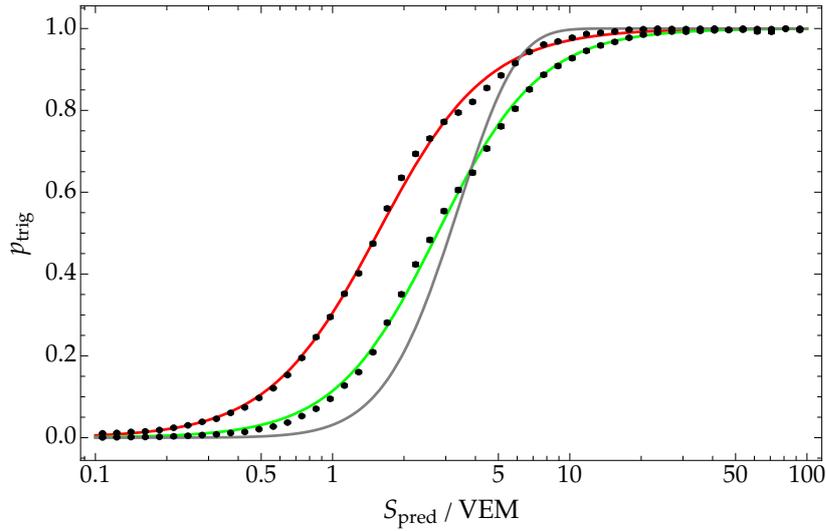


Figure 4.3: Trigger probability $p_{\text{trig,simp}}(S_{\text{pred}})$. Black points are data from Ref. [30] while the fit to the data is shown in red for the new+old triggers and in green for the old triggers only. The current Offline trigger probability $1 - p_{\text{zero}}(S_{\text{pred}})$ from Eq. (4.6) is shown in gray. Provided by Darko Veberič [31].

4.5 Likelihood Contributions for the LDF Slope Parameters

Since most events do not have enough degrees of freedom to fit all the parameters of the LDF, a parameterisation of the slope parameters is usually employed when reconstructing events. These parameterisations are obtained by fitting the slope parameters freely for many events with enough stations and using these values for a global fit of the slope parameters as a function of S_{ref} and θ . While Offline currently uses an LDF with two slope parameters, β and γ , the more general case is having n slope parameters

$$\alpha = \{\alpha_i : i = 1 \dots n\}. \quad (4.22)$$

Instead of strictly setting α_i to the parameterised “mean” value $\bar{\alpha}_i$ like it is done in the current reconstruction, one can allow the slope parameters to vary assuming a normal distribution $\mathcal{N}(\alpha_i, \bar{\alpha}_i, \sigma_{\alpha_i})$. For this case, both the $\bar{\alpha}_i$ and σ_{α_i} are parameterised as functions of S_{ref} and θ . Adding corresponding constraints to the log-likelihood then results in

$$\mathcal{L}' = \mathcal{L} + \sum_i^n \mathcal{L}_{\alpha_i} = \mathcal{L} + \sum_i^n \ln \mathcal{N}(\alpha_i, \bar{\alpha}_i, \sigma_{\alpha_i}), \quad (4.23)$$

where \mathcal{L} is the initial log-likelihood described by Eq. (4.21). Minimising Eq. (4.23) should fix the fit parameters α_i to the mean values $\bar{\alpha}_i$ for events with not enough stations while allowing α_i to vary around $\bar{\alpha}_i$ using the parameterised σ_{α_i} for events where more stations are available.

This chapter introduced a new likelihood setup for the LDF fit that unifies the treatment of high and low signals by switching to a log-normal distribution as the base for the likelihood contributions. This change makes the conversion of the measured signal to particles obsolete. By correctly including the trigger probability, a self-consistent likelihood capable of correctly reconstructing events with MoPS-only and/or TOTd-only triggered stations was obtained. Furthermore, the inclusion of slope-parameter constraints into the log-likelihood allows for variations of the parameters around their parameterised mean values during the reconstruction.

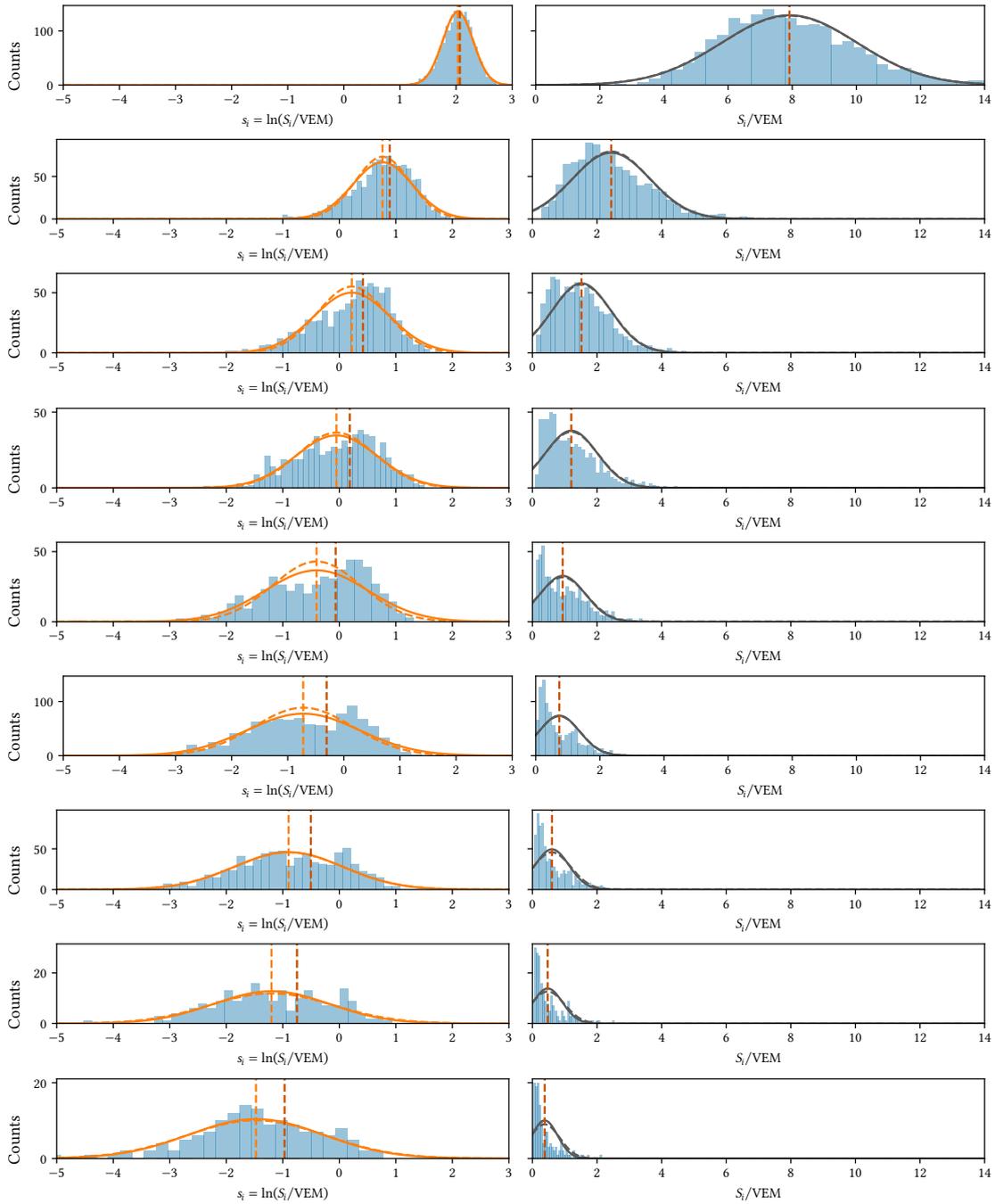


Figure 4.4: *Left:* Distributions of the log signals s_i for $\theta = 0^\circ$. The full-line envelope represents a Gaussian with \bar{s}_i and $\sigma = \text{std}(s_i)$. The dashed envelope shows the same, but using the (propagated) σ given by the variance model. The orange vertical line represents the mean \bar{s}_i , while the red line represents $\ln \bar{S}_i$. *Right:* Distributions of signals S_i for $\theta = 0^\circ$. The full-line envelope represents a truncated Gaussian with \bar{S}_i and $\sigma = \text{std}(S_i)$. The dashed envelope shows the same but using the σ given by the variance model. The vertical red line represents the mean \bar{S}_i .

4.5 Likelihood Contributions for the LDF Slope Parameters

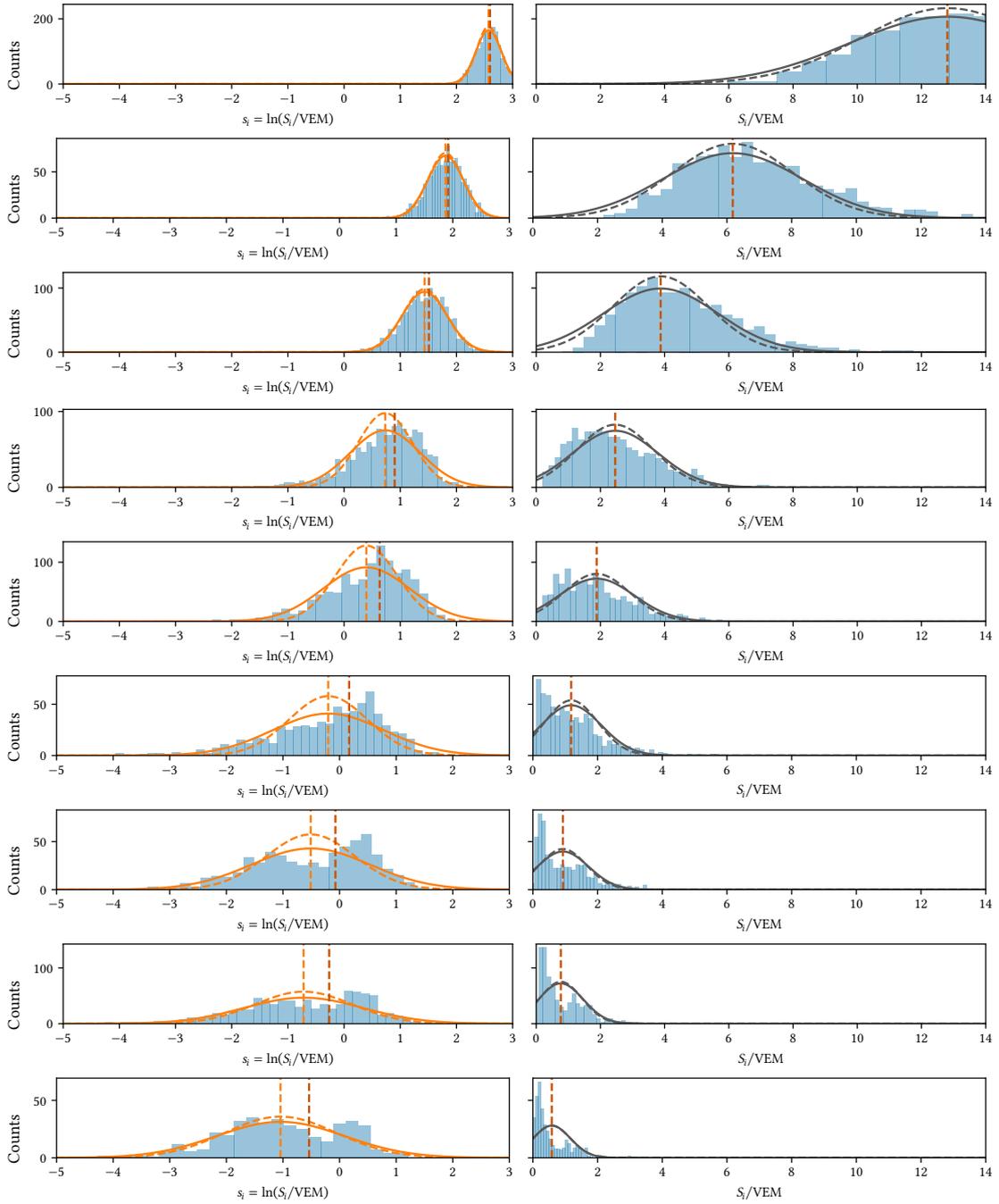


Figure 4.5: *Left:* Distributions of the log signals s_i for $\theta = 38^\circ$. The full-line envelope represents a Gaussian with \bar{s}_i and $\sigma = \text{std}(s_i)$. The dashed envelope shows the same but using the (propagated) σ given by the variance model. The vertical orange line represents the mean \bar{s}_i while the red one represents $\ln \bar{S}_i$. *Right:* Distributions of the signals S_i for $\theta = 38^\circ$. The full-line envelope represents a truncated Gaussian with \bar{S}_i and $\sigma = \text{std}(S_i)$. The dashed envelope shows the same but using the σ given by the variance model. The vertical red line represents the mean \bar{S}_i .

CHAPTER 5

New Lever-Arm Criteria to Fit the LDF Slope Parameters

This chapter was first made available as GAP–2024–33 [32].

During the reconstruction of SD events (see Ref. [23]), the *lateral distribution function* (LDF) is fitted to the signals measured in the stations. In Offline, the LDF $S(r)$ is currently described by a modified NKG [33, 34, 35] function $f_{\text{NKG}}(r)$ as

$$S(r) = S_{\text{ref}} f_{\text{NKG}}(r), \quad \text{where} \quad f_{\text{NKG}}(r) = \left(\frac{r}{r_{\text{ref}}} \right)^{\beta} \left(\frac{r + r_{\text{scale}}}{r_{\text{ref}} + r_{\text{scale}}} \right)^{\beta + \gamma}. \quad (5.1)$$

The LDF depends on the *shower size* parameter $S_{\text{ref}} = S(r_{\text{ref}})$, i.e. LDF signal at a certain reference distance r_{ref} , and the two shape or slope parameters β and γ of the NKG power laws. While the optimal choice [36] for the reference distance r_{ref} depends on the array type and spacing (see Table 5.1), the scaling parameter is usually fixed to $r_{\text{scale}} = 700$ m. Since the variable r describes the shower-plane distance of the observed point from the shower core, the LDF thus depends on the reconstructed shower axis and the impact point $(x_c, y_c, z_c \approx 0)$ of the shower core at the ground.

Therefore, once the axis is determined from the timing-geometry fit, it is clear that at least three stations with measured signal are required to fit S_{ref} , x_c , and y_c . However, the LDF slope parameters β and γ are needed as well. This requires at least five stations for one-slope and six for two-slope parameter fitting. Since most of the collected events do not have such large station multiplicities, β and γ are usually parameterised as functions of the other reconstructed shower quantities, primarily as functions of the zenith angle θ (from the shower axis) and S_{ref} , so that the reconstruction of low-multiplicity events is still possible. To find such a parameterisation it is important to use events of good quality that have (a) enough stations and, therefore, degrees of freedom to allow for the fitting of the slope parameters β and γ , and (b) that the radial distribution of stations allows for an accurate estimation of the LDF slope parameter β and, potentially, also γ , if possible. From now on, we will refer to these quality conditions as *lever-arm criteria*.

The current set of lever-arm criteria used in Offline to select appropriate events was developed by Pierre Billoir. Unfortunately, these criteria were described only in a not-so-well-documented form of Ref. [37]. For these criteria, an event has to have a certain number of stations at the right distances to the core, as described in Section 5.1.

Therefore, a different and more intuitive approach to select events suitable for fitting the LDF slope parameters is presented in Section 5.2. Without loss of generality, here we limit ourselves to the selection of events suitable for fitting β only, since the influence of γ on the shape of the LDF is smaller.

The study presented in this chapter focuses on the SD-750 array. However, the new criteria can be generalised for the other SD arrays of the Pierre Auger Observatory.

Table 5.1: Current values for the β lever-arm criteria in `Offline`. Note that the lever-arm criteria are fulfilled if the distance-range condition is satisfied by at least n_{\min} stations and at least one pair of these stations satisfies the corresponding d_{\min} criterion.

array	r_{ref}/m	$[r_{\min}, r_{\max}]/\text{m}$	n_{\min}	d_{\min}/m
SD-1500	1000	[400, 1600]	2	900
			3	800
			4	700
SD-750	450	[180, 720]	2	405
			3	360
			4	315
SD-433	300	[100, 400]	2	225
			3	200
			4	175

5.1 Current Lever-Arm Criteria

The current lever-arm criteria to select events where β can be fitted aim at ensuring the existence of enough stations around r_{ref} to have an accurate estimate of the slope of the LDF (see Ref. [23]). The criteria require at least n_{\min} stations in a shower-plane radius range $[r_{\min}, r_{\max}]$ around the reference distance r_{ref} (i.e. $r_{\min} \leq r_{\text{ref}} \leq r_{\max}$), while additionally at least two stations i and j must fulfill the requirement $|r_i - r_j| \geq d_{\min}$. The current values for these condition parameters are shown in Table 5.1.

At first glance, the choice of the values for these parameters seems unclear but were for sure optimally chosen by Pierre Billoir and/or Xavier Bertou. The original `Offline` commit c55c195e (15 April 2019) mentions the CDAS software as the source of these numbers. The values for the SD-1500 array were introduced into CDAS by Xavier Bertou (with a comment mentioning Carla Bonifazi) in the commit a37f75e1 (09 April 2010). However, the CDAS values for the SD-750 array introduced by Bonifazi, Aublin, and Münchmeyer (commit dc3bb570, 08 December 2010) are very different than the values in `Offline`, where the values for the SD-750 and SD-433 arrays are just scaled-down versions of the SD-1500 numbers. Since the origin of all these numbers is not well documented, one can speculate that the values were approximately chosen for an $E = 10^{19}$ eV shower that has $S(r_{\min}) \approx 1000$ VEM at $r_{\min} = 440$ m and $S(r_{\max}) \approx 10$ VEM at $r_{\max} = 1620$ m, so that stations, which are probably saturated or below a certain trigger threshold, are excluded.

This chapter explores the possibility of a more mathematically-oriented approach and introduces a new set of criteria for the SD-750 array that can be adjusted for the other arrays in the future.

5.2 New Criteria

The idea behind the new approach is the following: In the distance interval, where the LDF is dominated by β , the LDF is in the log-log-space approximately a linear function with a slope $m \approx \beta$. A linear fit of the log signals $\ln S$ (with their corresponding uncertainties) in log distance $\ln r$ should give us the slope m and its uncertainty σ_m . The magnitude of the

uncertainty σ_m is used as a suitable indicator of whether or not the slope β can be freely fitted in the reconstruction of the LDF of an event. A large σ_m namely indicates that the slope is not constrained well by the available station data and thus β cannot be freely fitted. This new criterion incorporates the main qualitative goals of the current lever-arm criteria, but uses a more quantitative approach. The condition is mainly expressed in only one quantity, σ_m , which serves as a quality indicator for the potential β fit. The details of the new procedure are given below.

5.2.1 Linear Fit in Log-Log-Space

The log quantities s_i , $\sigma(s_i)$, and ρ_i for a station i are defined as

$$s_i = \ln(S_i/\text{VEM}) \quad \text{with} \quad \sigma(s_i) = \frac{d \ln S_i}{d S_i} \sigma(S_i) = \frac{\sigma(S_i)}{S_i} \quad \text{and} \quad \rho_i = \ln(r_i/m). \quad (5.2)$$

To decorrelate the slope and offset of the linear fit later, we introduce centered quantities \tilde{s}_i , $\sigma(\tilde{s}_i)$, and $\tilde{\rho}_i$ such that

$$\tilde{s}_i = s_i - \langle s \rangle, \quad \sigma(\tilde{s}_i) \approx \sigma(s_i) = \frac{\sigma(S_i)}{S_i}, \quad \text{and} \quad \tilde{\rho}_i = \rho_i - \langle \rho \rangle, \quad (5.3)$$

where the averages are running over all selected stations described below. The final linear fit function is

$$\tilde{s}(\tilde{\rho}) = m \tilde{\rho} + c, \quad (5.4)$$

where m is the slope and c the offset of the linear function. Alternatively, when a previous LDF reconstruction already exists, the shower-size parameter S_{ref} (see Table 5.1) can be used as a suitable decorrelation method instead, i.e.

$$\tilde{s}_i = \ln(S_i/S_{\text{ref}}) \quad \text{and} \quad \tilde{\rho}_i = \ln(r_i/r_{\text{ref}}). \quad (5.5)$$

Note that this choice does not change the values obtained for m and c , while reducing their mutual correlation.

To reduce the potential biases encountered in this fit, we have to carefully select the stations that are used in the m fit in Eq. (5.4) for each event:¹

- Saturated stations, including recovered stations, are ignored. Saturated stations normally lie close to the core, the rest of the first-crown stations will thus cluster around the same distance.
- Stations that are not *very* likely to trigger are ignored, i.e. the corresponding trigger probability as given in Eq. (4.19) is required to be

$$p_{\text{trig}}(S_{\text{pred}}(r_i)) \geq 0.9 \quad (5.6)$$

The predicted signal $S_{\text{pred}}(r_i)$ at a distance r_i is calculated using the (parameterised) LDF values from an earlier reconstruction stage. This ensures that the m fit and thus the new selection criteria are not dominated by low-signal fluctuations.

¹This station selection is only used for the lever-arm criteria. A subsequent event reconstruction should use all stations.

To obtain a meaningful fit of m , at least *three* stations have to meet the selection conditions described above.

The precision σ_m of the slope m can be extracted from the fit covariance matrix. Another way (see Ref. [38, §15.2]) is to calculate it as

$$\sigma_m^2 = \frac{W(1)}{W(1)W(\tilde{\rho}^2) - (W(\tilde{\rho}))^2} \quad (5.7)$$

where $W(x) = \sum_i x_i / \sigma_i^2$ is a short-hand notation for a weighted sum. The uncertainty σ_m determines how accurately m can be estimated from a given configuration of $(\tilde{s}_i, \tilde{\rho}_i)$ data for a given event. Since $\beta \approx m$ and thus $\sigma_\beta \approx \sigma_m$, this also tells us how accurate a potential LDF fit with free β will be in such an event.

5.2.2 Selection Conditions

As mentioned above, a large σ_m implies that the event data do not allow a reliable LDF fit with free β . For this reason, we thus impose an upper limit on σ_m when selecting events for the relaxed- β fit, i.e.

$$\sigma_m \leq \sigma_m^{\max}. \quad (5.8)$$

Another useful quantity is the standard deviation of the station distances $\text{std}(\tilde{\rho})$. If $\text{std}(\tilde{\rho}) \rightarrow 0$, the stations cluster around the same distance, meaning that the fit is thus dominated by only a narrow range of available radii. Such degenerate events can be excluded by imposing a lower limit on $\text{std}(\tilde{\rho})$, i.e.

$$\text{std}(\tilde{\rho}) \geq \text{std}(\tilde{\rho})_{\min}. \quad (5.9)$$

In contrast to the old lever-arm criteria, the new setup imposes some conditions on the progression of signals in an event. From a shower point of view, the slope m cannot be positive since this would imply a nonphysical LDF where the signals increase with distance. Therefore, all events with $m > 0$ are discarded. Lowering the m limit even further eliminates also events with very *flat* or too *steep* LDFs, i.e.

$$m_{\min} \leq m \leq m_{\max}. \quad (5.10)$$

As mentioned before, the LDF in log-log-space is to a good approximation just a linear function of the distance. How well the signals are (anti)correlated with the distance and how well the points in an event are aligned to this line can be easily measured with the Pearson correlation coefficient

$$\mathcal{R}(\tilde{s}, \tilde{\rho}) = \frac{\sum_i (\tilde{s}_i - \langle \tilde{s} \rangle) (\tilde{\rho}_i - \langle \tilde{\rho} \rangle)}{\sqrt{\sum_i (\tilde{s}_i - \langle \tilde{s} \rangle)^2 \sum_i (\tilde{\rho}_i - \langle \tilde{\rho} \rangle)^2}}. \quad (5.11)$$

An upper limit (since signal and distance are anti-correlated) on the Pearson correlation coefficient in the form of

$$\mathcal{R}(\tilde{s}, \tilde{\rho}) \leq \mathcal{R}(\tilde{s}, \tilde{\rho})_{\max} \quad (5.12)$$

excludes events where the data does not follow the required LDF trend narrow enough. This is mainly the case for events that are not close to the full trigger efficiency of the array.

The chosen values for all of these conditions above are shown in Table 5.2.

Table 5.2: Values for the selection parameters of the conditions in the new lever-arm criteria. Note that the SD-1500 and SD-433 values were not verified with data and are just informed suggestions.

array	σ_m^{\max}	$[m_{\min}, m_{\max}]$	$\text{std}(\tilde{\rho})_{\min}$	$\mathcal{R}(\tilde{\xi}, \tilde{\rho})_{\max}$
SD-1500	0.5	$[-4, -1]$	0.2	-0.7
SD-750	0.5	$[-4, -1]$	0.2	-0.7
SD-433	0.5	$[-4, -1]$	0.2	-0.7

Table 5.3: Values for the T4 trigger efficiency parameters of Eq. (5.14) as given for proton primaries using EPOS in Ref. [22, p. 119].

a_0	a_1	a_2	a_3	b_0	b_1
16.482 ± 0.011	0.25 ± 0.14	0.24 ± 0.48	1.01 ± 0.47	0.218 ± 0.006	0.036 ± 0.015

5.2.3 Additional Conditions

Events that satisfy the lever-arm criteria should additionally fulfil some event-level quality cuts when used for an LDF parameterisation.

Minimum station multiplicity 5. The conditions of the new lever-arm criteria are enough to reliably fit β when the core position is reliably known. However, this is not always the case. As mentioned before, the core position is fitted during the reconstruction as well, consuming two degrees of freedom from the station data. Therefore, a *minimum number* of at least $n_{\min} = 5$ stations is needed.

Event quality condition: One should require the 6T5 event trigger and the zenith angle in the full-efficiency region to ensure only good-quality events are used for the β fitting in search for a global β parameterisation, thus

$$\theta \leq \theta_{\max}. \quad (5.13)$$

Using the studies presented in Ref. [39], $\theta_{\max} = 55^\circ$ seems appropriate for the SD-750 array.

Base condition: This condition requires the 6T5 event trigger, a fitted curvature, a minimal value $p_{T4}^{\min} = 0.99$ of the T4 trigger efficiency at the given energy and no saturated stations without recovery. It is described in Ref. [40] and was used to select good-quality events with the old lever-arm criteria for the current β parameterisation for the SD-750 in [Offline](#). The T4 trigger efficiency for the SD-750 array including the data from the new triggers, as given in Ref. [22, pp. 118–119], reads as

$$p_{T4}(E, \theta) = \frac{1}{2} \left[1 + \text{erf} \left(\frac{\lg(E/\text{eV}) - a(\theta)}{b_0 + b_1 \sin^2 \theta} \right) \right], \quad \text{where} \quad a(\theta) = \sum_{i=0}^3 a_i \sin^{2i} \theta. \quad (5.14)$$

The values of the parameters a_i and b_i are shown in Table 5.3.

In addition, all events where one or more stations were rejected by lightning are ignored.

Table 5.4: Performance comparison of the old vs. new lever-arm criteria for events with at least $n_{\min} = 5$ stations. The \neg symbol denotes the complement operation “not”.

condition	number of events
old, but failed linear fit	21 274
old	45 139
\neg old	71 021
new	67 449
\neg new	48 711
new & \neg old	35 425
old & \neg new	13 115
new & old	32 024

Table 5.5: Performance comparison of the old vs. new lever-arm criteria for events with at least $n_{\min} = 5$ stations that additionally satisfy the quality and base conditions.

condition	number of events
old, but failed linear fit	10706
old	29037
\neg old	42118
new	43224
\neg new	27931
new & \neg old	21498
old & \neg new	7311
new & old	21726

5.3 Comparison

Table 5.4 shows a comparison of selected events of both sets of lever-arm criteria on a test sample (2017) of the SD-750 array, which all have at least five stations. Fig. 5.1 shows the all possible 2D projections of the multidimensional data in the space of parameters m , $\lg \sigma_m$, $\lg \text{std}(\tilde{\rho})$, and $\mathcal{R}(\tilde{s}, \tilde{\rho})$ of the new lever-arm criteria. Table 5.5 and Fig. 5.2 show the same but using only those events where all additional conditions described in Section 5.2.3 are met.

Examining the Fig. 5.1, it is obvious that the old criteria selected many events with a questionable signal progression when no other quality conditions were imposed (e.g. conditions excluding increasing LDF with distance). This is the reason why from now on only the data described in Table 5.5 will be used, i.e. only events from the test sample which fulfil *all additional conditions* as laid out in Section 5.2.3.

According to the Table 5.5, using the new criteria to select events leads to an increase of

$$\Delta n_{\beta} = n_{\text{new}} - n_{\text{old}} = 43\,224 - (29\,037 + 10\,706) = 3481 \quad (5.15)$$

events where a free β fit could be performed. Here, n_{old} (n_{new}) is the number of events satisfying the old (new) criteria. The number of events for a free β fit thus increases by

$$\Delta \epsilon = \frac{\Delta n_{\beta}}{n_{\text{old}}} = 8.76\%. \quad (5.16)$$

5.3.1 Events Satisfying the Old but not the New Criteria

When using the new criteria, we want to keep as many events that satisfy the old criteria as possible. However, excluding events that may bias the results is desirable.

A selection of events that satisfy the old criteria but not the new criteria (old & \neg new) is shown in Fig. 5.3. We see that there are three reasons for the events not passing the new criteria.

The error σ_m is too large. When removing stations with an expected signal below the 90% trigger efficiency, σ_m can drastically increase leading to the event not being selected. Otherwise, a subsequent LDF fit with free β would be dominated by low-signal stations, imposing a bias on β . Not including these events is thus desired and is an advantage of the new criteria.

The fit looks reasonable but the limits on m are exceeded. These events have a very steep or very flat LDF, which is for hadronic showers highly unlikely. This case is very rare.

The linear fit fails. This occurs when not enough stations satisfy the station selection conditions. As pointed out before, we do not want to select such events since an LDF fit with free β would be biased by low-signal fluctuations.

5.3.2 Events Satisfying the New but not the Old Criteria

One goal of the new criteria is to select additional events that were discarded by the old methodology but still look like β can be fitted freely. In Fig. 5.4 some of these (new & \neg old) events are shown. These events do not satisfy the strict distance conditions of the old criteria. However, they seem reasonable to fit β .

5.4 Summarising the Lever-Arm Criteria Studies

This chapter presented an alternative to the existing lever-arm criteria, which was used until now to select events where a β fit could be attempted. Comparing the plots and the number of events, we can conclude that the new criteria increase the number of eligible events while also successfully removing events where the free β fit probably would not be accurate enough. Therefore, the inclusion of these new criteria in the Offline framework is suggested so that it can be used as the new default for future parameterisations of the slope parameters of the LDF.

Even though the comparison of the old and the new criteria in the former section was done using data from the SD-750 array, the new lever-arm criteria should work similarly for the other arrays as well. Since the station spacing for SD-433 and SD-1500 is different, the value of $\text{std}(\tilde{\rho})_{\min}$ might change with the array. Nevertheless, since scaling with the array spacing is expected and since these are logarithmic quantities where scaling is just an additive shift, the values for these selection parameters should not change much. However, this needs to be studied in detail and is beyond the scope of this thesis.

In addition, further studies should clarify the influence of the different criteria on the LDF parameterisation. Criteria to select events where also a free γ fit can be performed should be developed in the future.

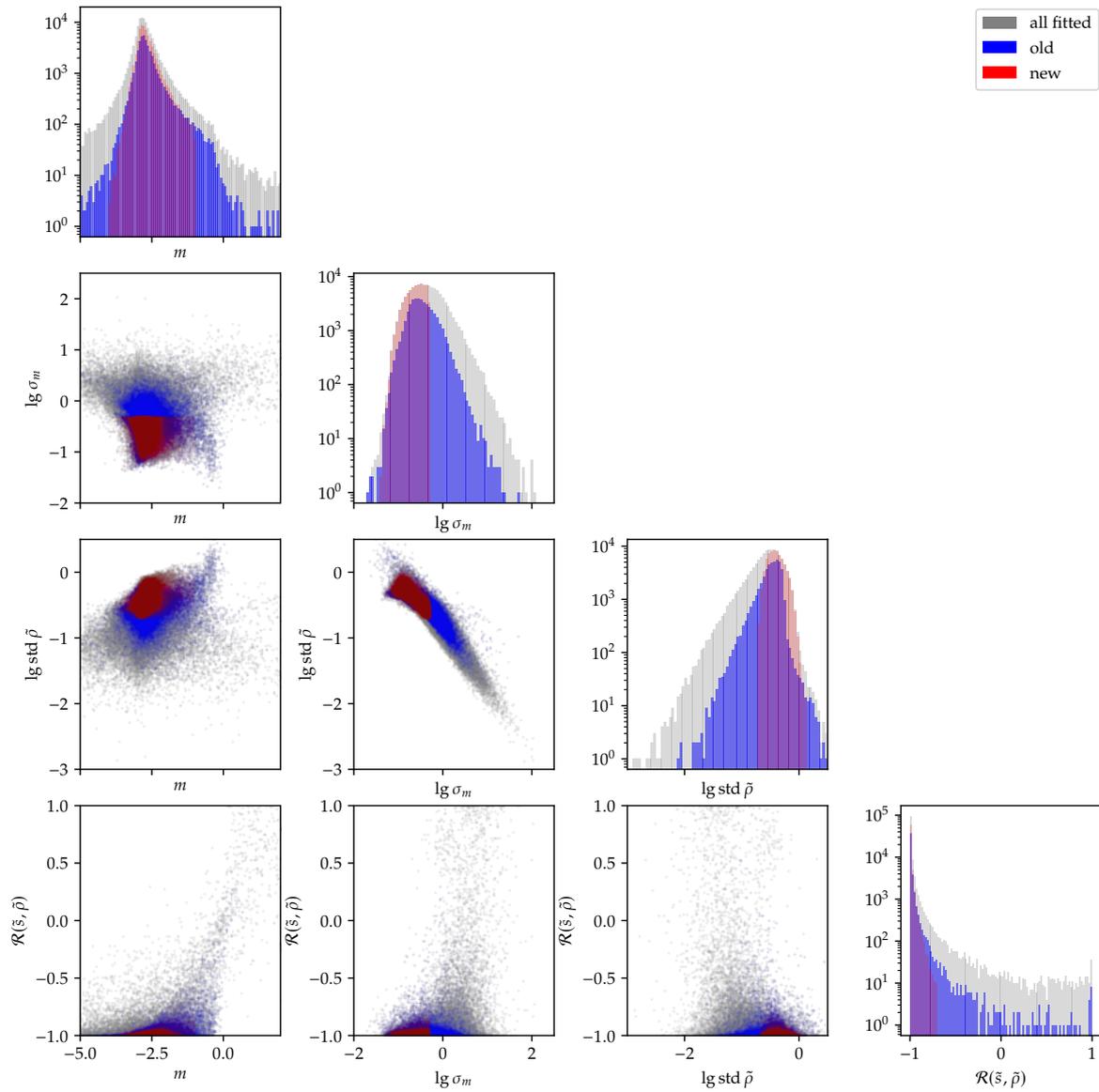


Figure 5.1: Comparison of events with at least five stations and a successful linear m fit. All events are shown in grey while blue (red) corresponds to the events selected by the old (new) lever-arm criteria.

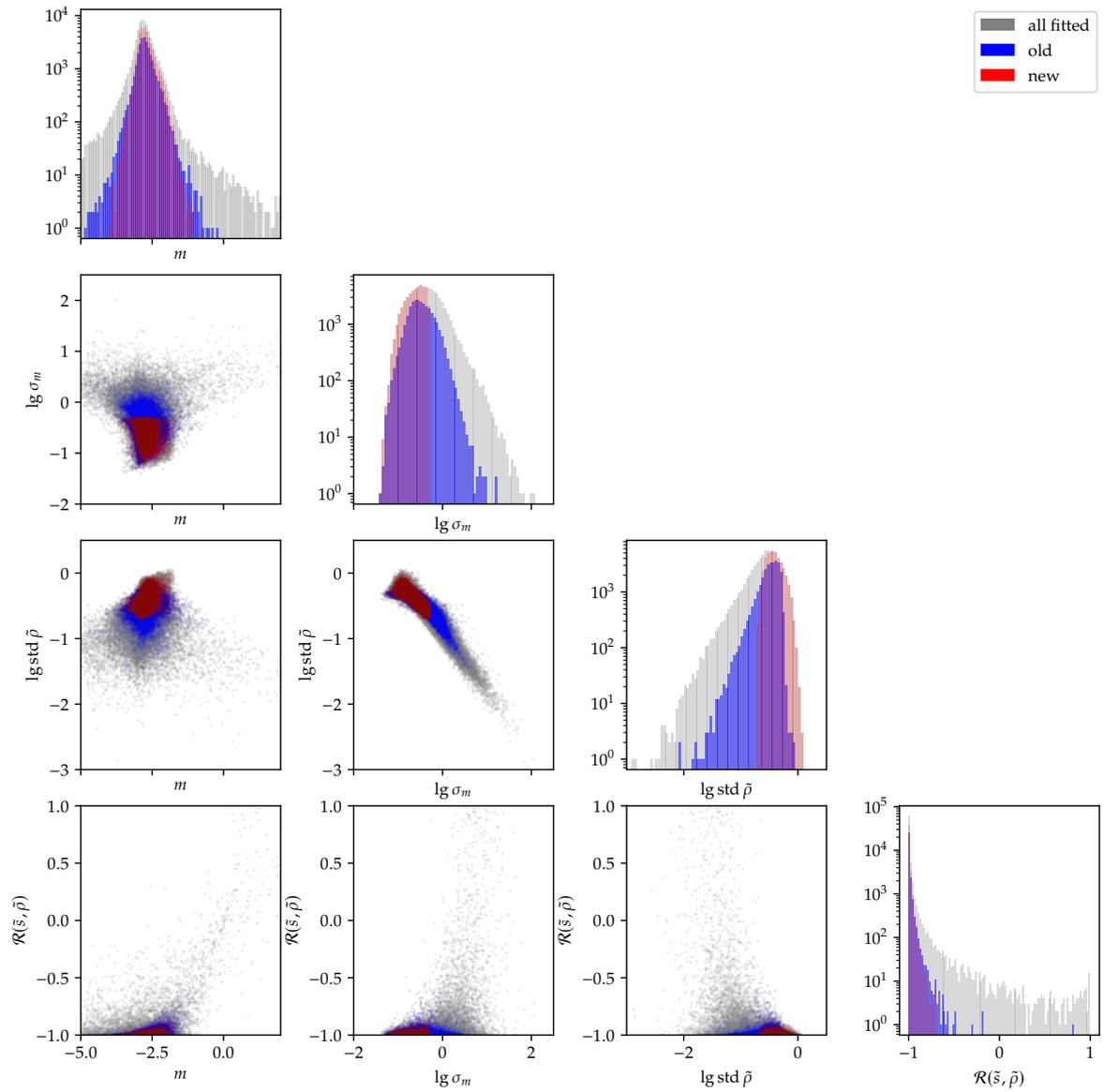


Figure 5.2: Comparison of events with a successful linear m fit where all additional conditions are satisfied. All events are shown in grey while blue (red) corresponds to the events selected by the old (new) lever-arm criteria.

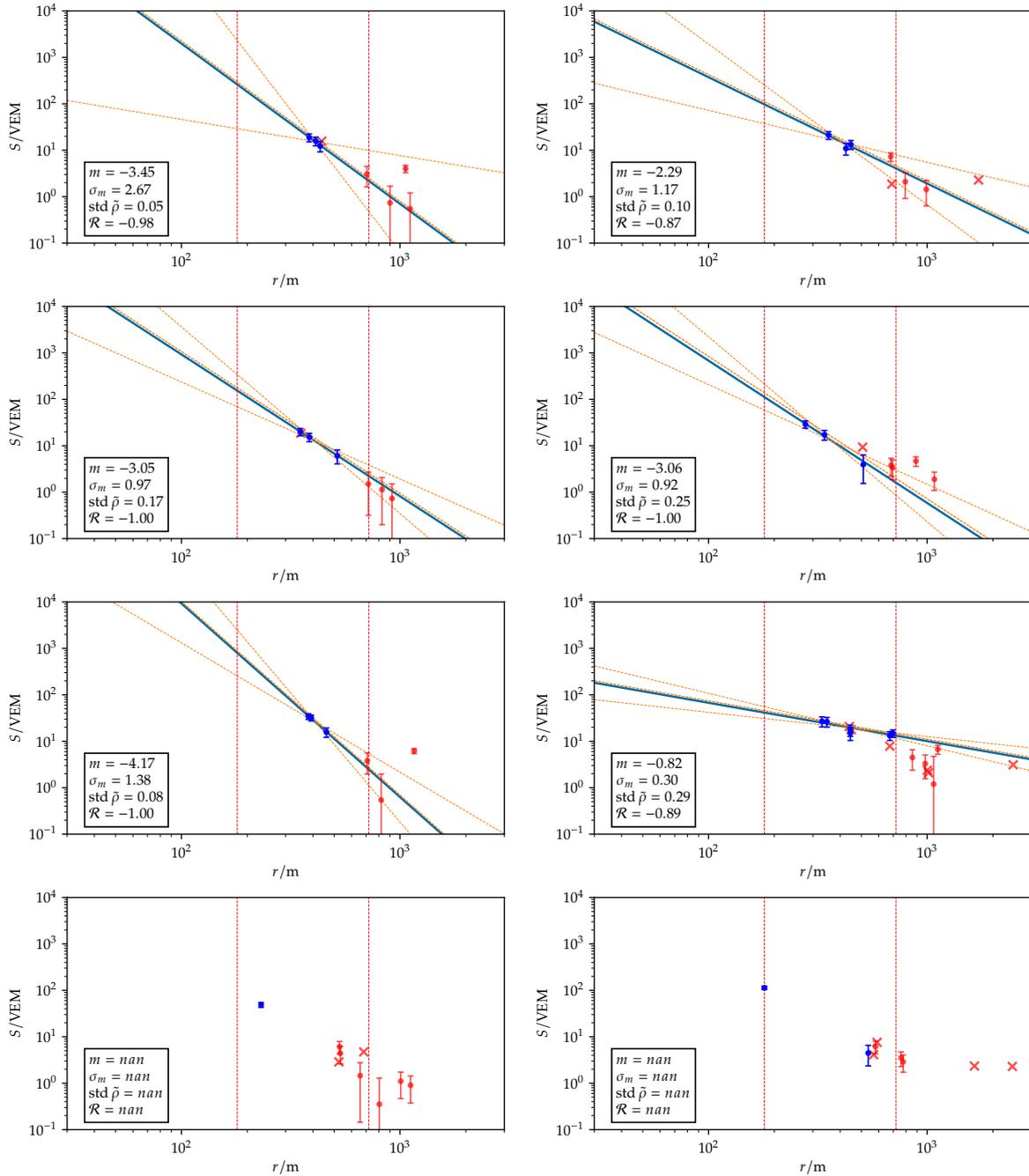


Figure 5.3: Events satisfying the old but not the new criteria (old & \neg new). Stations used for the linear fit are shown as blue dots. The red dots represent stations with $p_{\text{trig}}(S_{\text{pred}}) < 0.9$ that were not used for the linear fit. Rejected stations are represented by red crosses. The linear fit (when possible) is shown as a blue line with its variations ($m \pm \sigma_m, c \pm \sigma_c$) as dashed orange lines. The red vertical lines represent $[r_{\text{min}}, r_{\text{max}}]$ of the old lever-arm criteria for the SD-750 array.

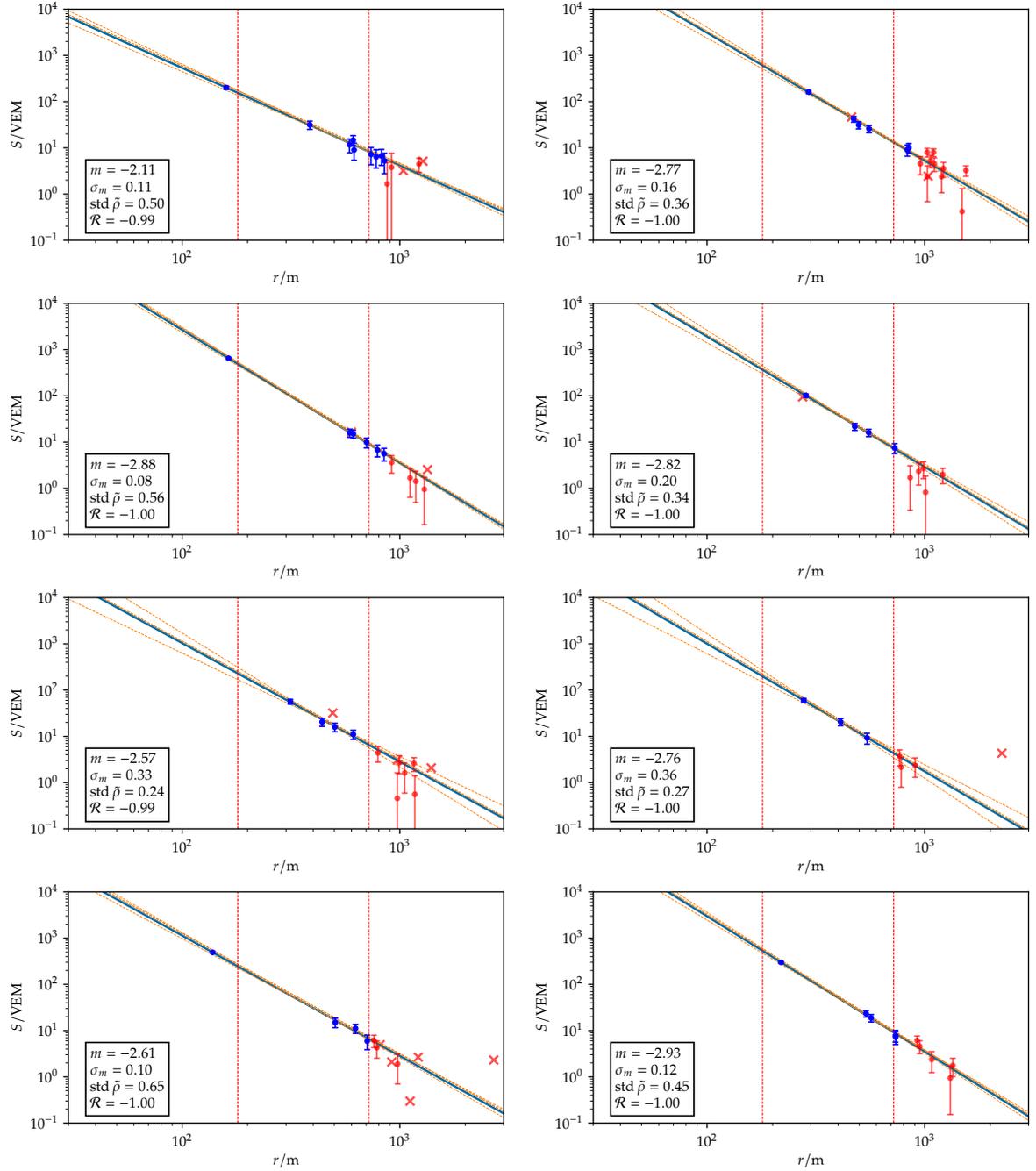


Figure 5.4: Events satisfying the new but not the old criteria (new & \neg old). Stations used for the linear fit are shown as blue dots. The red dots represent stations with $p_{\text{trig}}(S_{\text{pred}}) < 0.9$ that were not used for the linear fit. Rejected stations are represented by red crosses. The linear fit is shown as a blue line with its variations ($m \pm \sigma_m, c \pm \sigma_c$) as dashed orange lines. The red vertical lines represent $[r_{\text{min}}, r_{\text{max}}]$ of the old lever-arm criteria for the SD-750 array.

CHAPTER 6

Parameterising the Lateral Distribution Function

As mentioned in Chapter 5, the *lateral distribution function* (LDF) is fitted to the measured signals in the stations during the reconstruction of SD events with Offline. The LDF $S(r)$ is currently described by a modified NKG function $f_{\text{NKG}}(r)$ [33, 34, 35],

$$S(r) = S_{\text{ref}} f_{\text{NKG}}(r), \quad \text{where} \quad f_{\text{NKG}}(r) = \left(\frac{r}{r_{\text{ref}}} \right)^{\beta} \left(\frac{r + r_{\text{scale}}}{r_{\text{ref}} + r_{\text{scale}}} \right)^{\beta + \gamma}, \quad (6.1)$$

which depends on the so-called shower size $S_{\text{ref}} = S(r_{\text{ref}})$ and the two slope parameters β and γ . While the optimal choice [36] for the reference distance r_{ref} depends on the array spacing,¹ the scaling parameter is fixed to $r_{\text{scale}} = 700$ m. Since the variable r describes the perpendicular distance of the observed point to the shower core, the LDF thus depends on the reconstructed axis and impact point of the shower. The shower core is described by the impact point (x_c, y_c, z_c) of the shower on the ground, the zenith angle θ , and the azimuth angle ϕ of the shower axis. All these parameters have to be fitted during the reconstruction process. With a suitable setup of the ground plane, the z_c parameter can be eliminated, since in this plane it becomes just a constant. The axis \hat{a} , pointing towards the arrival direction of the shower, is estimated in the geometry fit, which relies on the station-timing information, with a pair of directional cosines u and v so that $\hat{a} = (u, v, \sqrt{1 - u^2 - v^2})$. It is therefore clear that at least three stations are required for the fit to cover the three degrees of freedom S_{ref} , x_c , and y_c . The z -component of the impact point of the core is kept the same as in a signal-weighted barycenter of the participating stations [23]. However, for the LDF fit, the LDF slope parameters β and γ are also needed, requiring at least 5 stations for one-slope and 6 for two-slope parameter fitting. Using the criteria described in Chapter 5, β and γ are usually parameterised as functions of θ and S_{ref} to be able to reconstruct also events with a smaller station multiplicity by eliminating the slopes as free fit parameters.

Including the MoPS and TOTd triggers into the reconstruction, and the studies described in Chapter 4, requires a reparameterisation of the mentioned slope parameters. In this chapter, the LDF given in Eq. (6.1) will be revisited and its functional form updated. Finally, an up-to-date LDF parameterisation for the SD-750 array will be presented. This concludes the event reconstruction study started in Chapter 3 and Chapter 4. The combined results should be included in the standard reconstruction of Offline in the future.

6.1 Proposed Changes

As pointed out in Ref. [41], the NKG-like LDF overestimates the signals at large distances from the core. Therefore, an *exponentially suppressed power law* (ESPL) will be used, i.e.

$$f_{\text{LDF}}(r) = \left(\frac{r}{r_{\text{ref}}} \right)^{\beta_{\text{ESPL}}} \frac{\exp\left(-\left(\frac{r}{r_{\text{scale}}}\right)^{\gamma_{\text{ESPL}}}\right)}{\exp\left(-\left(\frac{r_{\text{ref}}}{r_{\text{scale}}}\right)^{\gamma_{\text{ESPL}}}\right)} = \left(\frac{r}{r_{\text{ref}}} \right)^{\beta_{\text{ESPL}}} \exp\left(\frac{r_{\text{ref}}^{\gamma_{\text{ESPL}}} - r^{\gamma_{\text{ESPL}}}}{r_{\text{scale}}^{\gamma_{\text{ESPL}}}}\right), \quad (6.2)$$

¹1000 m, 450 m, and 300 m for the SD-1500, SD-750, and SD-433 array respectively.

as it was proposed in Ref. [41]. The same $r_{\text{scale}} = 700$ m as for the NKG-like LDF is used.

A quick study on real data showed that $\gamma_{\text{ESPL}} \approx 1$ for the SD-750 array.² Therefore, it is sufficient setting $\gamma_{\text{ESPL}} \equiv 1$, leading to the simpler expression

$$f_{\text{LDF}}(r) = \left(\frac{r}{r_{\text{ref}}}\right)^{\beta_{\text{ESPL}}} \exp\left(\frac{r_{\text{ref}} - r}{r_{\text{scale}}}\right). \quad (6.3)$$

For the saturation-recovery check in Offline,³ the second derivative is needed. Using the product and the chain rule, we obtain

$$\begin{aligned} f'_{\text{LDF}}(r) &= \frac{\beta_{\text{ESPL}}}{r_{\text{ref}}} \left(\frac{r}{r_{\text{ref}}}\right)^{-1} f_{\text{LDF}} + \left(-\frac{1}{r_{\text{scale}}}\right) f_{\text{LDF}} = \left(\frac{\beta_{\text{ESPL}}}{r} - \frac{1}{r_{\text{scale}}}\right) f_{\text{LDF}} \\ &= \underbrace{\left(\frac{\beta_{\text{ESPL}} r_{\text{scale}} - r}{r r_{\text{scale}}}\right)}_{=:g(r)} f_{\text{LDF}}. \end{aligned} \quad (6.4)$$

Finding the derivative of this expression leads to

$$f''_{\text{LDF}}(r) = g'(r) f_{\text{LDF}} + g(r) f'_{\text{LDF}} \stackrel{(6.4)}{=} [g'(r) + g^2(r)] f_{\text{LDF}}, \quad (6.5)$$

where the derivative of $g(r)$ is given by

$$g'(r) = -\frac{1}{r r_{\text{scale}}} - r_{\text{scale}} \frac{\beta_{\text{ESPL}} r_{\text{scale}} - r}{r^2 r_{\text{scale}}^2} = -\frac{\beta_{\text{ESPL}}}{r^2}. \quad (6.6)$$

As mentioned before, we need a parameterisation of β_{ESPL} in terms of S_{ref} and θ to be able to reconstruct events with low-station-multiplicity. Since the MoPS-only and TOTd-only triggered stations should be used for the reconstruction in the future, we will use the updated log-likelihood including the correct trigger probability, which is appropriate for small signals given in Chapter 4.

6.2 LDF Parameterisation

6.2.1 Selecting Events for the Parameterisation

To find good quality events for the parameterisation, we follow the procedure described in Chapter 5. As mentioned there, we need a previous reconstruction for the event selection. Since the likelihood and the LDF were updated, the old Offline reconstructions cannot be used. Therefore, an initial reconstruction of the events with the new setup is needed before performing the selection criteria for the parameterisation. This requires an initial parameterisation of β_{ESPL} . Since there is already a parameterisation for β_{NKG} and γ_{NKG} , we will try to find an expression which allows us an estimation of β_{ESPL} as a function of the NKG parameters. The idea is to match the derivative of f_{NKG} with the derivative of f_{ESPL} at low distances where the LDF is roughly given by a power law. The derivative of f_{ESPL} was already calculated in Eq. (6.4). The derivative of f_{NKG} is

$$f'_{\text{NKG}}(r) = \frac{\beta_{\text{NKG}}(r + r_{\text{scale}}) + r(\beta_{\text{NKG}} + \gamma_{\text{NKG}})}{r(r + r_{\text{scale}})} f_{\text{NKG}}. \quad (6.7)$$

²This should be investigated in more detail in the future.

³This is used to decide if the recovered signal of a saturated station is reasonable.

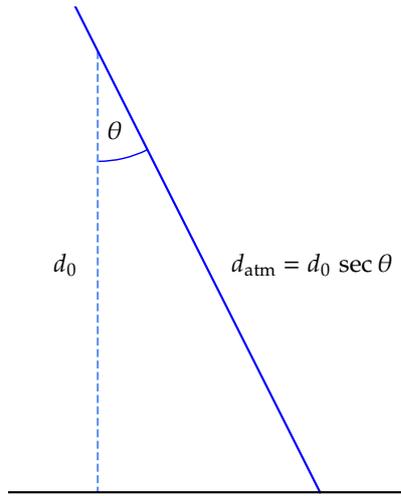


Figure 6.1: The distance d_{atm} the shower has to move through the atmosphere depends on $\sec \theta$.

Matching Eqs. (6.4) and (6.7) leads to

$$\beta_{\text{ESPL}} = \frac{r \ln(r/r_{\text{ref}})/r_{\text{scale,ESPL}} + W(z(r))}{\ln(r/r_{\text{ref}})} \quad (6.8)$$

with the Lambert W function [42] and

$$z(r) := \exp\left(\frac{r - r_{\text{ref}}}{r_{\text{scale,ESPL}}}\right) \left(\frac{r}{r_{\text{ref}}}\right)^{-\frac{r}{r_{\text{scale,ESPL}}}} \left(\beta_{\text{NKG}} + \frac{r(\beta_{\text{NKG}} + \gamma_{\text{NKG}})}{r + r_{\text{scale}}}\right) \ln\left(\frac{r}{r_{\text{ref}}}\right) f_{\text{NKG}}. \quad (6.9)$$

The r_{scale} parameter for $f_{\text{ESPL}}(r)$ can in principle be different from the one in $f_{\text{NKG}}(r)$. For now, we will set $r_{\text{scale,ESPL}} = r_{\text{scale}} = 700$ m and evaluate Eq. (6.8) always for $r = 100$ m.

Using this initial parameterisation, all SD-750 events from January 2014 to the end of 2021 were reconstructed.⁴ Then the selection procedure described Chapter 5 was applied to select all events that can be used for the parameterisation.

6.2.2 Rewriting the Slope Parameter

For physical solutions, β_{ESPL} has to be negative (upward-going LDFs are not physical!). Since minimisers work best for unbound variables, rewriting

$$\beta_{\text{ESPL}} = -\exp \beta \quad \text{with} \quad \beta \in \mathbb{R} \quad (6.10)$$

allows the use of the unbound parameter β while β_{ESPL} will be strictly negative since the exponential function is strictly positive. Rewriting Eq. (6.3) yields

$$f_{\text{LDF}}(r) = \left(\frac{r}{r_{\text{ref}}}\right)^{-\exp \beta} \exp\left(\frac{r_{\text{ref}} - r}{r_{\text{scale}}}\right). \quad (6.11)$$

⁴First day with the new triggers working properly to the end of phase I.

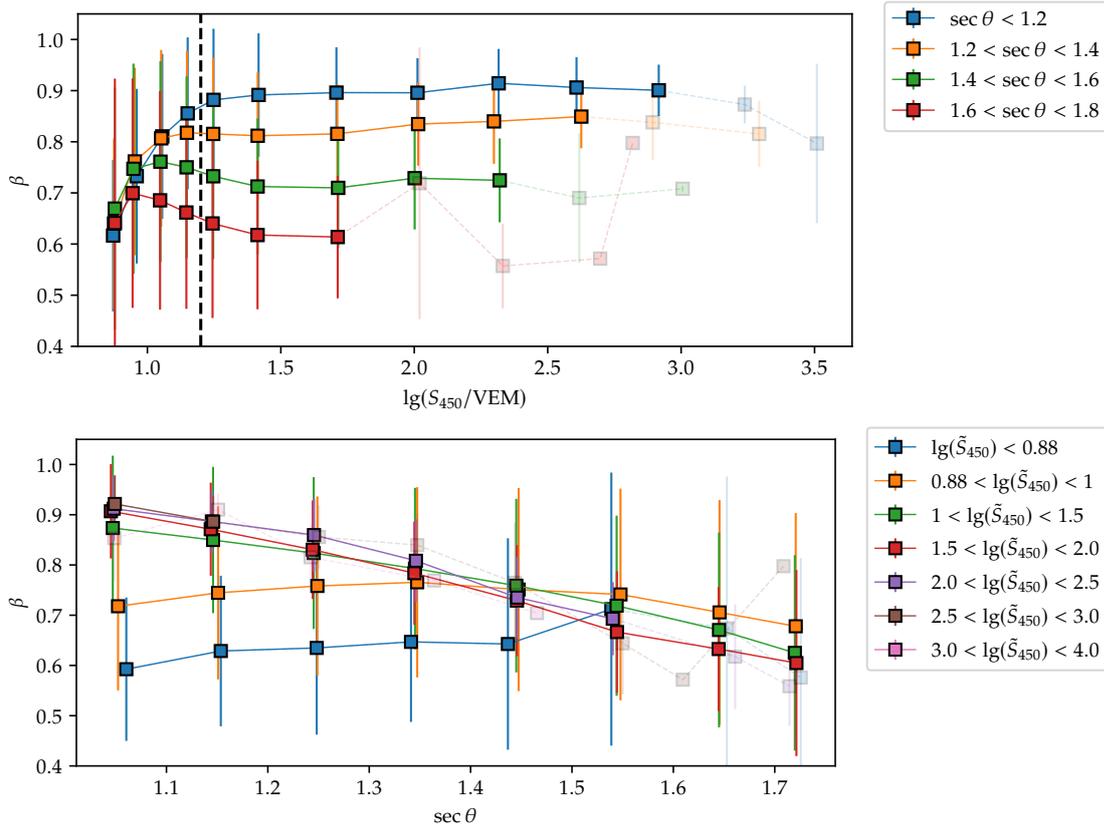


Figure 6.2: *Top:* Dependency of β on $\lg(S_{450}/\text{VEM})$ for different $\sec \theta$ bins. The dotted line represents the $\lg(S_{450}/\text{VEM}) = 1.2$ threshold. *Bottom:* Dependency of β on $\sec \theta$ for different $\lg \tilde{S}_{450} := \lg(S_{450}/\text{VEM})$ bins. Bins with less than 30 entries are plotted with transparency.

6.2.3 Parameterising the Slope Parameter

All selected events were reconstructed while also fitting β freely. Since the LDF shape depends on the energy and the traversed atmosphere, β is expected to be dependent on the energy estimator S_{450} and $\sec \theta$ (see Fig. 6.1). The dependencies of the fitted β on $\lg \tilde{S}_{450} := \lg(S_{450}/\text{VEM})$ and $\sec \theta$ is shown in Fig. 6.2. For $\lg \tilde{S}_{450} \geq 1.2$, a linear dependency of β on both variables is visible. For $\lg \tilde{S}_{450} < 1.2$ this behaviour changes.

Since the error bars in this region are quite large, the simple form

$$\beta_{\text{simple}}(\tilde{S}_{450}, \theta) = a + b \lg \tilde{S}_{450} + c \sec \theta. \quad (6.12)$$

can be used. The parameters a , b , and c of Eq. (6.12) were fitted once to all data and once to the data with $\lg \tilde{S}_{450} \geq 1.2$ using an unbinned least-squares fit. As pointed out in Chapter 4, a parameterisation of the variance of β is also desirable. The variance $(\beta - \beta_{\text{param}})^2$ was parameterised according to⁵

$$\sigma_{\beta}^2(\tilde{S}_{450}, \theta) = \exp(s_1 + s_2 \ln \tilde{S}_{450} + s_3 \sec \theta) \quad (6.13)$$

⁵We use the same trick as in Section 6.2.2 to ensure a positive variance.

Table 6.1: Values for the simple parameterisation in Eqs. (6.12) and (6.13). For “all” all data was used while for “cut” only the data with $\lg \tilde{S}_{450} \geq 1.2$ was used.

	a	b	c
Simple, all	1.2255 ± 0.0019	0.0267 ± 0.0006	-0.3484 ± 0.0014
Simple, cut	1.3246 ± 0.0022	0.0087 ± 0.0007	-0.4073 ± 0.0017
	s_1	s_2	s_3
Simple, all	-2.2312 ± 0.0240	-1.8648 ± 0.0102	0.5784 ± 0.0157
Simple, cut	-2.5725 ± 0.0364	-1.6828 ± 0.0163	0.6322 ± 0.0249

using an unbinned likelihood fit by minimising the log-likelihood

$$\ell = -2 \ln L = \sum_i \left(\frac{(\beta_i - \beta_{\text{param},i})^2}{\sigma_{\beta}^2(\tilde{S}_{450,i}, \theta_i)} + \ln \sigma_{\beta}^2(\tilde{S}_{450,i}, \theta_i) \right). \quad (6.14)$$

The values for the parameters of the two parameterisations of β_{simple} and their variances are listed in Table 6.1.

6.3 Validating the Parameterisation

Using Ockham’s razor, the simple parameterisation given in Eq. (6.12) is chosen. The version where only data with $\lg \tilde{S}_{450} \geq 1.2$ was used is preferred since the behaviour below this threshold is not understood and could be some artificial effect. Fig. 6.3 shows the chosen parameterisation compared to the data.

Using the parameterisation, a subset of SD-750 events was reconstructed. Since the fitted shower size S_{450} is used for the energy estimation of the primary particle, we need to ensure that reconstructions with the new parameterisation lead to a correct S_{450} . It is practical to look at all stations in a distance interval of (450 ± 20) m. In a purely Gaussian case, looking at the *residuals* $(S - S_{\text{pred}})/S_{\text{pred}}$ and the so-called *pulls* $(S - S_{\text{pred}})/\sigma$, we can identify possible biases in the signal predictions. Since according to Chapter 4 we have a more complicated likelihood than just simple Gaussians, we need to derive a correct goodness of fit from the likelihood itself. Due to the usage of the log-normal distribution for the signals, we can use the residuals and pulls of the log signals

$$s = \ln(S/\text{VEM}) \quad \text{and} \quad s_{\text{pred}} = \ln(S_{\text{pred}}/\text{VEM}). \quad (6.15)$$

However, the trigger-probability terms in the likelihood complicate this slightly. Only in the range where $p_{\text{trig}}(S_{\text{pred}}, \theta) \approx 1$ the Gaussian assumption is valid for the log signals and the residuals and pulls are approximately a valid measure for the goodness of fit. According to the trigger probability as given in Ref. [30], this is the case for $S_{\text{pred}} \gtrsim 8 \text{ VEM}$ where $p_{\text{trig}} \geq 0.96$. The corresponding residuals and pull plots for the chosen parameterisation are shown in Fig. 6.4. From the residuals plot, we can see a small bias $< 2\%$. As mentioned above, this estimation of the goodness of fit is only valid approximately.

Another possibility is the definition of a more general goodness of fit (GoF). The likelihood contribution \mathcal{L}_i of a station in the LDF fit can be compared to the best possible fit meaning the

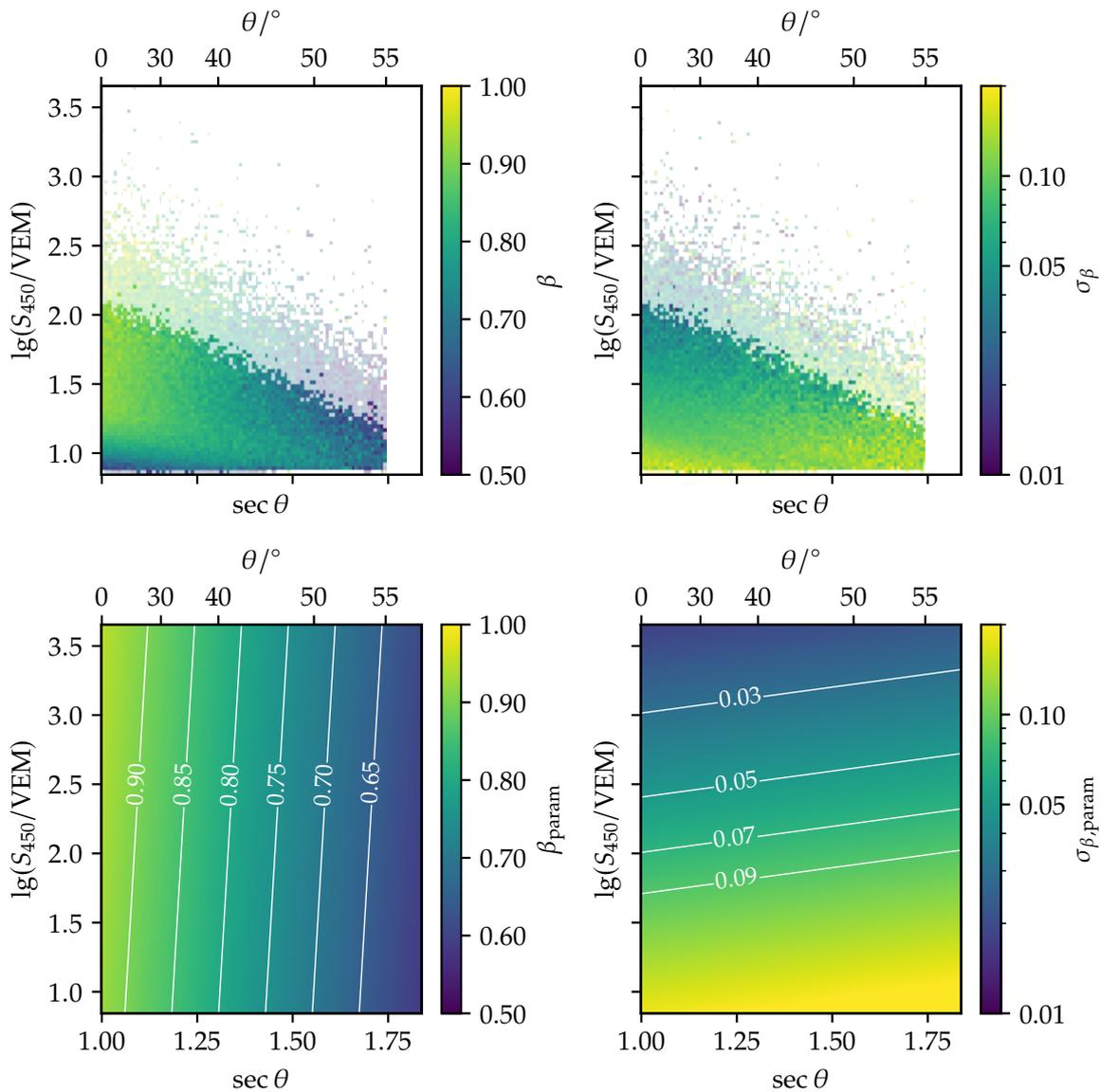


Figure 6.3: Comparison of data and the simple parameterisation where data with $\lg \tilde{S}_{450} \geq 1.2$ was used. Bins with less than 30 entries are plotted with transparency. *Top:* freely fitted β and σ_β (as in Fig. 6.2). *Bottom:* parameterised β and σ_β .

best likelihood contribution the station could give. We can calculate this so-called *saturated model* by minimising the log-likelihood contribution for the station $\mathcal{L}_{i, \text{sat}}$. The goodness of fit for the prediction μ_i is then given by the log-likelihood ratio

$$\text{GoF}(\mu_i) = -2 \ln \frac{L_i(\mu_i)}{L_{i, \text{sat}}} = -2(\mathcal{L}_i(\mu_i) - \mathcal{L}_{i, \text{sat}}), \quad (6.16)$$

where the likelihood $\mathcal{L}_{i, \text{sat}}$ of the saturated model is obtained as

$$\mathcal{L}_{i, \text{sat}} = \max_{\mu_i} \mathcal{L}_i(\mu_i). \quad (6.17)$$

A value close to 0 means that the model fits the data very well. In a purely Gaussian case, we have $\text{GoF} \rightarrow \chi^2$ [29, pp. 89–92]. This goodness of fit is more flexible and robust than the

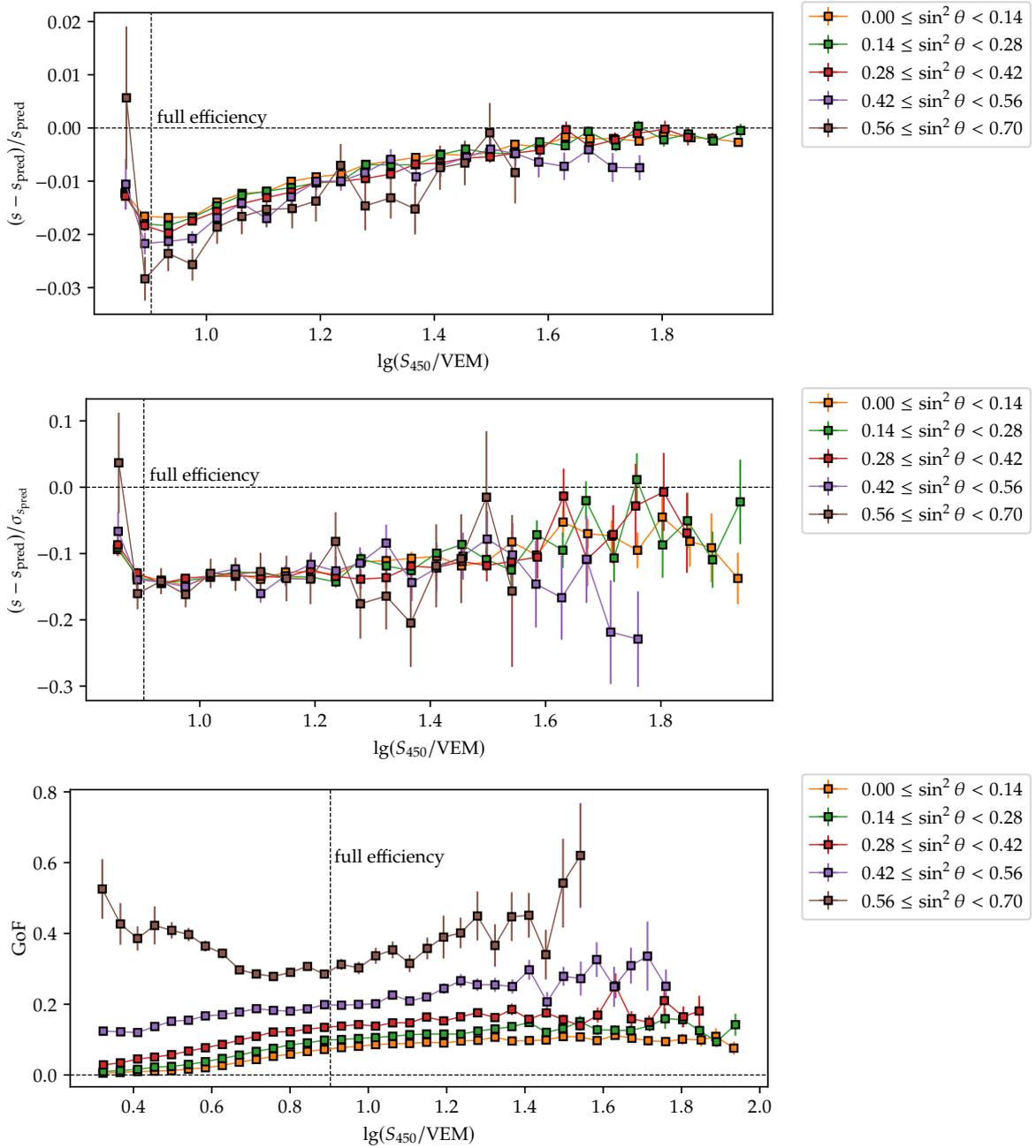


Figure 6.4: Different comparisons of the measured signal for stations at (450 ± 20) m with the predicted signal. The predicted signal was calculated using the simple parameterisation (see Eq. (6.12)) where only data with $\lg(S_{450}/\text{VEM}) \geq 1.2$ was used. *Top:* Residuals $(s - s_{\text{pred}})/s_{\text{pred}}$ for the log signals s for triggered stations with $S_{\text{pred}} \geq 8 \text{ VEM}$. *Middle:* Pull plot for the log signals, $(s - s_{\text{pred}})/\sigma_{\text{pred}}$, for triggered stations with $S_{\text{pred}} \geq 8 \text{ VEM}$. *Bottom:* Goodness of fit (GoF) for all triggered stations at (450 ± 20) m.

described pulls and residuals of the log signals. Fig. 6.4 shows that the parameterisation estimates S_{450} very well.

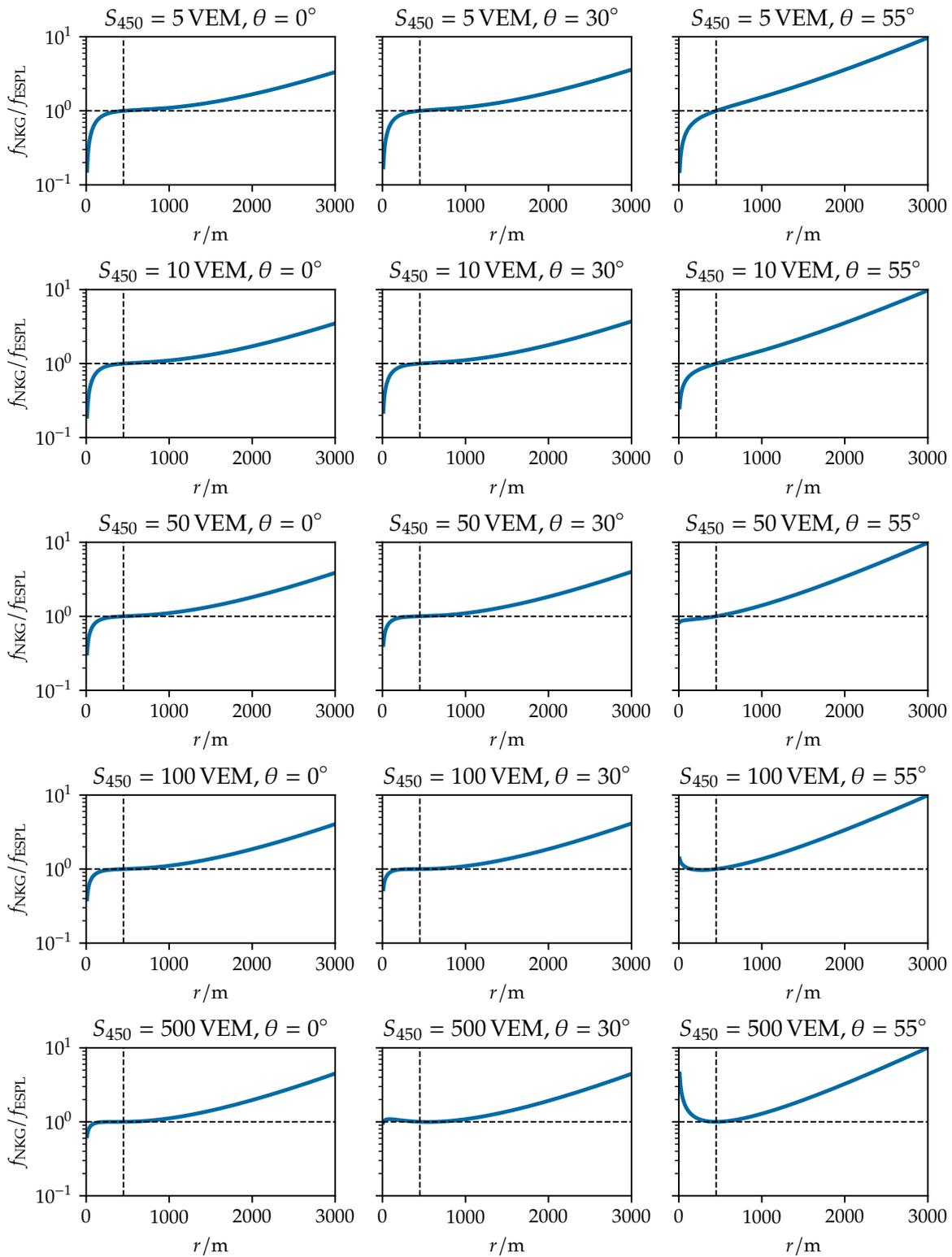


Figure 6.5: The ratio of the parameterisations of the NKG LDF and the ESPL LDF for different shower sizes and zenith angles.

6.4 Comparing the New LDF Parameterisation with the Old Parameterisation

Before reconstructing the measured data and comparing the old and the new reconstruction, we will compare the current parameterisation of the NKG LDF with our new parameterisation of the ESPL LDF. We will look at the ratio $f_{\text{NKG}}(r)/f_{\text{ESPL}}(r)$. As mentioned before, the NKG function is not flexible enough to tune it independently to threshold fit the data of stations close and far from the core. This results in an undershoot at small and an overshoot at large distances r from the core. Since the ESPL is more flexible and should catch the behaviour better, we expect the mentioned ratio of the LDFs to be smaller than one for small r and larger than one for large r .

The ratio plots in Fig. 6.5 for different S_{450} and θ show that these expectations are indeed fulfilled. Only for large S_{450} and large θ we see that the ratio for small distances is larger than one. However, the shape of the LDF in this range is not known since it is anyway very likely that all stations in this range are saturated for these high energies.

6.5 Reconstructing Events with the New Parameterisation

Using the chosen parameterisation and the new likelihood setup given in Chapter 4 all SD-750 events until the end of 2021 were reconstructed.

As mentioned in Chapter 4, the new setup allows the reconstructed β to vary around the β_{param} constraint assuming a normal distribution. This assumption is justified by evaluating the comparison of β and β_{param} in Fig. 6.6. Only for events with 3 triggered stations, the distribution looks different. Since the influence of silent stations in the likelihood is stronger for those events, we expected a different behaviour. Comparing the data from 2011 and 2016, we can see an increase in events with 4 or more stations due to the new triggers.

6.5.1 Comparison against the Standard Reconstruction on the Event Level

This chapter introduced many changes in the reconstruction chain compared to the current standard reconstruction. The new reconstruction assumes a log-normal distribution for the signals while Offline currently uses either a normal or a Poisson distribution. The trigger probability was included in the likelihood and the functional form of the LDF was changed. Therefore, a quick comparison between the new and the standard reconstruction is necessary.

Figs. 6.7 and 6.8 show that both reconstructions agree well in reconstructing the shower axis and the impact point of the core. Since the new reconstruction was developed to be more accurate for showers with smaller energies, changes in the shower sizes for small S_{450} are expected. On the other hand, both reconstructions should yield similar shower sizes for larger values of S_{450} . Fig. 6.9 shows that these expectations are fulfilled.

The differences in the shower sizes should also depend on the zenith angle. Assuming an isotropic flux of incoming cosmic rays, without atmospheric effects we expect a flat distribution of the measured shower sizes in $\sin^2 \theta$. However, the amount of atmosphere that the showers have to traverse while developing depends on the zenith angle, leading to a non-flat distribution of shower sizes in $\sin^2 \theta$ when detected showers are considered. Therefore, the differences in the shower sizes of the two reconstructions should also have a non-flat distribution in $\sin^2 \theta$. Fig. 6.10 validates this assumption.

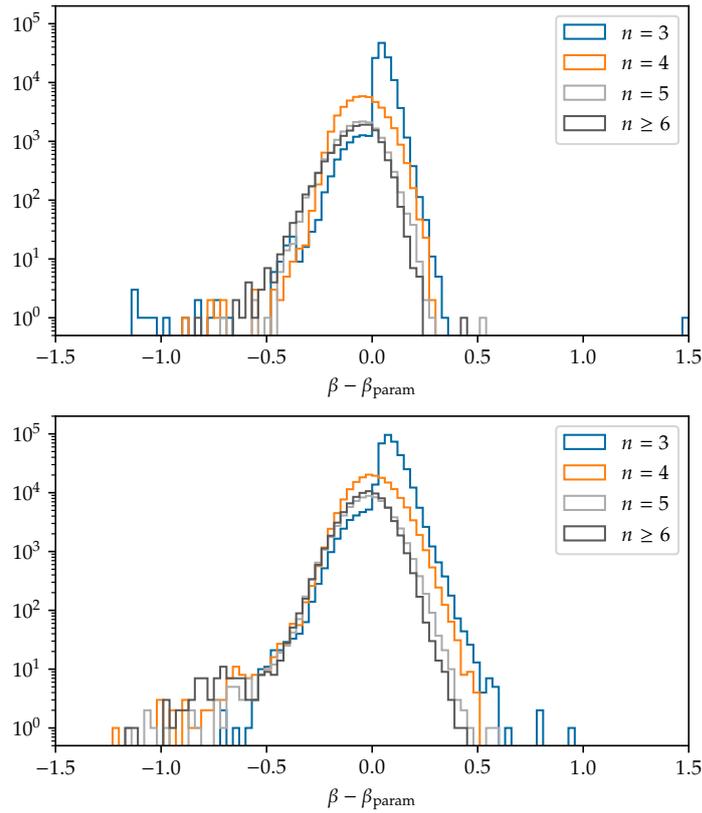


Figure 6.6: Comparison of the parameterised β_{param} and the reconstructed β (varied around β_{param} according to Chapter 4) for events with n triggered stations. Only 6T5 events with $\theta \leq 55^\circ$ were used. *Top:* Data from 2011 (before the new triggers were installed). *Bottom:* Data from 2016 (with new triggers).

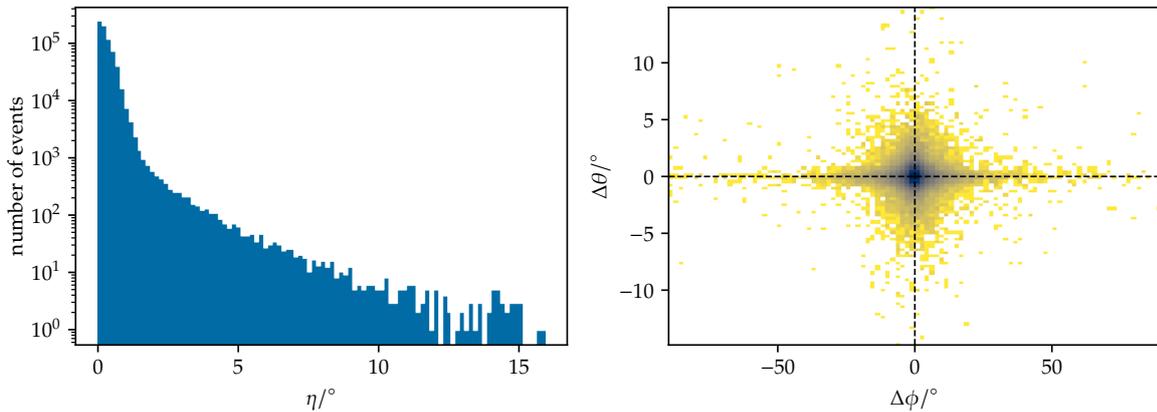


Figure 6.7: Comparison of the shower axes reconstructed by the standard reconstruction using Offline and the new reconstruction for the 2016 SD-750 dataset. η is the opening angle between both axes. $\Delta\theta$ and $\Delta\phi$ are the differences of the zenith and the azimuth angle between the Offline and the new reconstruction.

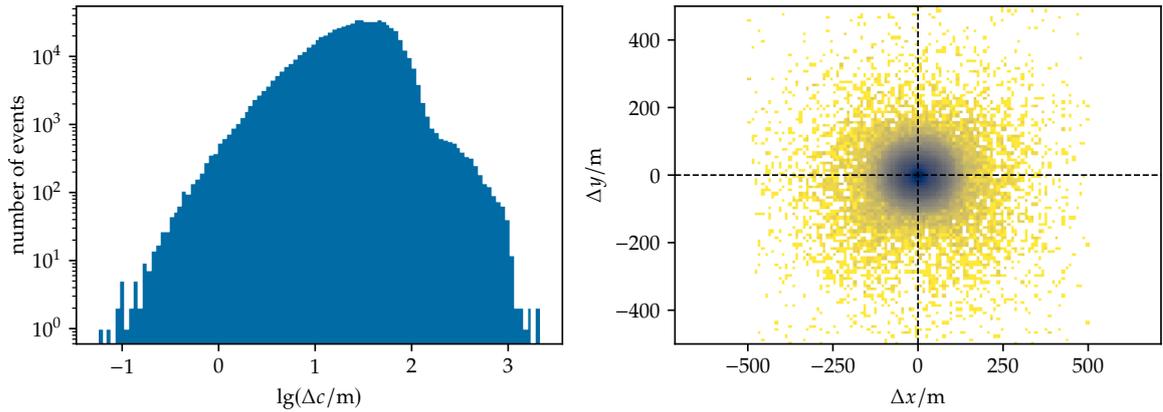


Figure 6.8: Comparison of the impact point of the shower core reconstructed by the standard reconstruction using Offline and the new reconstruction for the 2016 SD-750 dataset. The difference of the impact points is $\Delta c = |\vec{c}_{\text{Offline}} - \vec{c}_{\text{new}}| = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$.

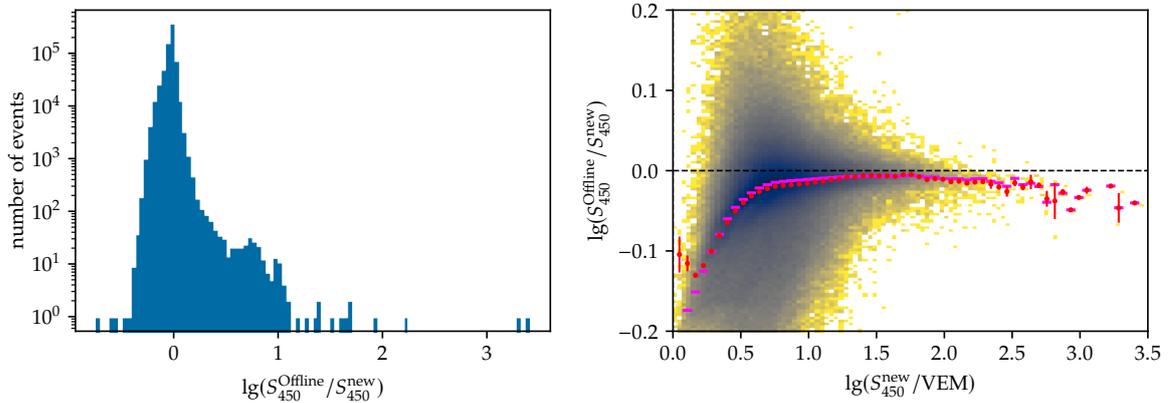


Figure 6.9: Comparison of the shower sizes reconstructed by the standard reconstruction using Offline and the new reconstruction for the 2016 SD-750 dataset as a function of the shower size of the new reconstruction. The pink dashes represent the median and the red dots represent the mean with its standard error as error bars.

6.5.2 Comparison against the Standard Reconstruction on the Station Level

In addition to comparing the event level, the two reconstructions can be compared on a station level for many events. In other words, we will look at the differences between the predicted and measured signals for many stations.

We will compare the *residuals* $(x - x_{\text{pred}})/x_{\text{pred}}$ and the so-called *pulls*, meaning the quantity $(x - x_{\text{pred}})/\sigma$ which measures the deviation of a measured x from its predicted value x_{pred} in units of σ . If x is normally distributed, these quantities should be zero for a perfect prediction model. Two important conclusions arise from this. First, we need to compare the residuals and pulls of the log signals since this reflects the underlying signal distribution. Second, we expect deviations from zero for small signals because the influence of the trigger probability has to pull the predicted signal away from the measured signal.

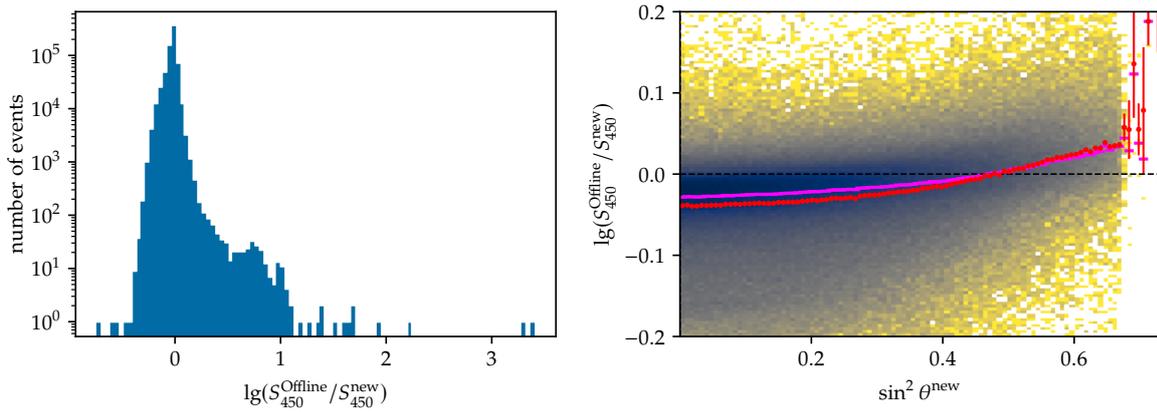


Figure 6.10: Comparison of the shower sizes reconstructed by the standard reconstruction using Offline and the new reconstruction for the 2016 SD-750 dataset as a function of the zenith of the new reconstruction. The pink dashes represent the median and the red dots represent the mean with its standard error as error bars.

Fig. 6.11 shows the pulls for both reconstructions for stations with $p_{\text{trig}} \geq 99\%$. The new reconstruction is closer to the zero line. The downward tendency of the new reconstruction for small signals is somewhat unexpected. Since the stations trigger only on upward fluctuations for small predicted signals, an increase in the pulls for small signals is expected. This can be observed for very small signals in Fig. 6.15. However, the origin of the negative pulls for moderately small signals remains unclear. However, the residuals in Fig. 6.12 shows that the offset from zero in percent is negligible.

The advantage of the new reconstruction becomes visible when looking at Fig. 6.13 and Fig. 6.14 where we show the pulls and residuals for stations $p_{\text{trig}} \geq 90\%$. For $\lg(S/\text{VEM}) \approx 1.25$ we see a jump of the pulls. This happens because Offline treats signals larger than this threshold as normally distributed but signals smaller than this threshold as Poisson distributed in the likelihood fit.⁶ The new reconstruction instead shows smooth behaviour since the assumption of a log-normal distribution is valid down to low signals.

6.5.3 Reconstruction Quality

One can check for the reconstruction quality of an event using the T5 posterior trigger described in Ref. [43]. As it turns out, the fit of the core is wrong in some cases. Since β is fixed to β_{param} in the first reconstruction steps, the minimiser will compensate for deviations from β_{param} by shifting the core instead. If the minimiser gets stuck in a local minimum, the event reconstruction will be wrong. Therefore, a workaround is suggested. If the reconstruction of an event fails or if the T5 posterior trigger indicates a poor reconstruction, a new reconstruction of the event should be attempted, this time by shifting $\beta_{\text{param}} \rightarrow \beta_{\text{param}} \pm \sigma_{\beta}$.

However, this does not solve the problem for all events. An example is shown in Fig. 6.16 where the likelihood was minimised by fitting S_{450} for different fixed core positions. The found minimum was then compared with the reconstructed core of the event. The reconstructed core is far away from the global minimum of the log-likelihood because the minimiser got stuck in a local minimum during the event reconstruction.

⁶ $\lg(S/\text{VEM}) \approx 1.25 \Rightarrow S = 17.78 \text{ VEM}$ corresponds to an estimation of $n = k(0^\circ) S \approx 31$ particles (see Section 4.1) which is approximately the threshold where Offline switches the treatment of the signals.

It is important to note that this problem is not an issue of the new reconstruction but can also appear for the standard reconstruction. This emphasises the importance of finding good initial values for the minimiser to prevent it from falling into local minima.

6.6 Relevance of the New Parameterisation

This chapter introduced a new LDF for the SD-750 array, i.e.

$$f_{\text{LDF}}(r) = \left(\frac{r}{r_{\text{ref}}} \right)^{-\exp \beta} \exp \left(\frac{r_{\text{ref}} - r}{r_{\text{scale}}} \right). \quad (6.18)$$

An updated likelihood as given in Chapter 4 and an updated event selection as described in Chapter 5 were used to reconstruct events which can be used to parameterise the slope parameter of the LDF. Using Ockham's razor, we found a simple parameterisation for

$$\beta = a + b \lg \tilde{S}_{450} + c \sec \theta \quad \text{and} \quad \sigma_{\beta}^2 = \exp (s_1 + s_2 \lg \tilde{S}_{450} + s_3 \sec \theta), \quad (6.19)$$

where the values of the parameters are shown in Table 6.2.

The correct estimation of the energy estimator S_{450} was validated using a test data set. This parameterisation combined with the presented update to the likelihood allows the reconstruction of events down to low energies leading the way to investigations of the low-energy part of the SD-750 energy spectrum.

A selection of LDFs obtained with the standard reconstruction using Offline and the new reconstruction for the same events is shown in Fig. 6.17.

Table 6.2: Values for the parameterisation in Eq. (6.19) using data with $\lg \tilde{S}_{450} \geq 1.2$.

a	b	c
1.3246 ± 0.0022	0.0087 ± 0.0007	-0.4073 ± 0.0017
s_1	s_2	s_3
-2.5725 ± 0.0364	-1.6828 ± 0.0163	0.6322 ± 0.0249

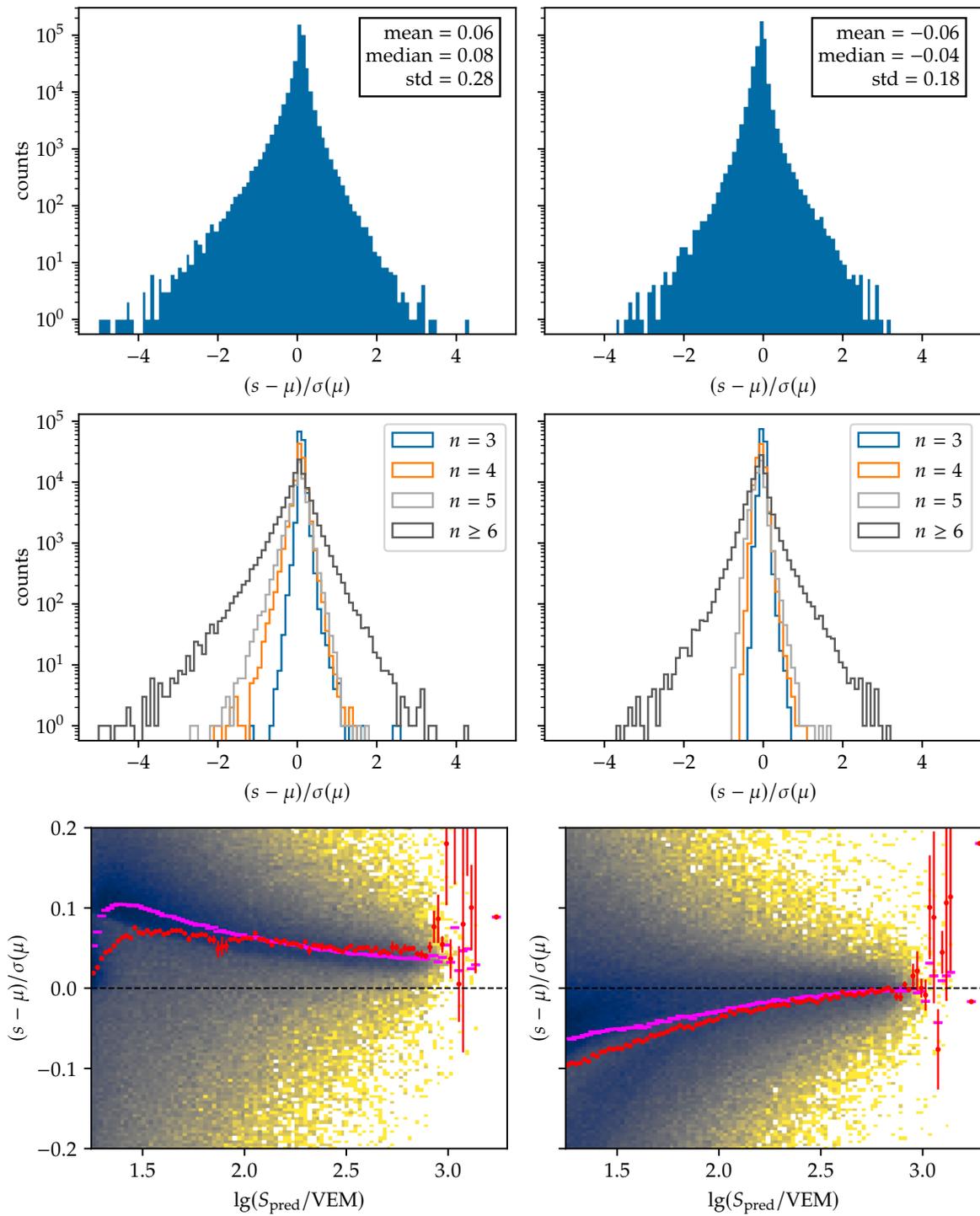


Figure 6.11: Comparison of the standard reconstruction using Offline (left) and the new reconstruction (right) for the 2016 SD-750 dataset. All triggered, unsaturated stations with $p_{\text{trig}} \geq 99\%$ from 6T5 events with $\theta \leq 55^\circ$ were used. *Top:* Histogram of the pulls. *Middle:* Histogram of pulls for different numbers of triggered stations n in the events. *Bottom:* Pulls over the predicted signal. The pink dashes represent the median and the red dots represent the mean with its standard error as error bars.

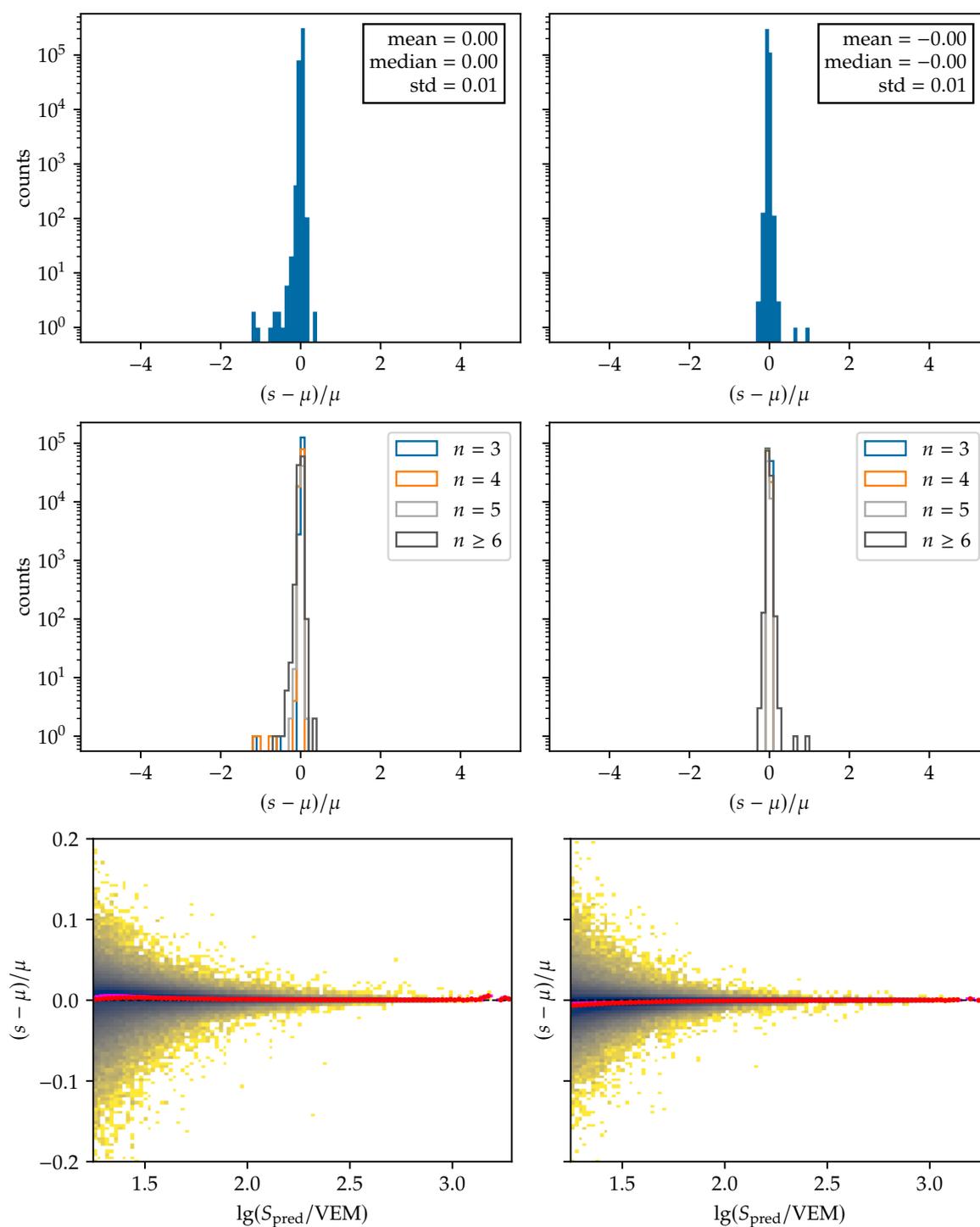


Figure 6.12: Comparison of the standard reconstruction using Offline (*left*) and the new reconstruction (*right*) for the 2016 SD-750 dataset. All triggered, unsaturated stations with $p_{\text{trig}} \geq 99\%$ from 6T5 events with $\theta \leq 55^\circ$ were used. *Top:* Histogram of the residuals. *Middle:* Histogram of residuals for different numbers of triggered stations n in the events. *Bottom:* Residuals over the predicted signal. The pink dashes represent the median and the red dots represent the mean with its standard error as error bars.

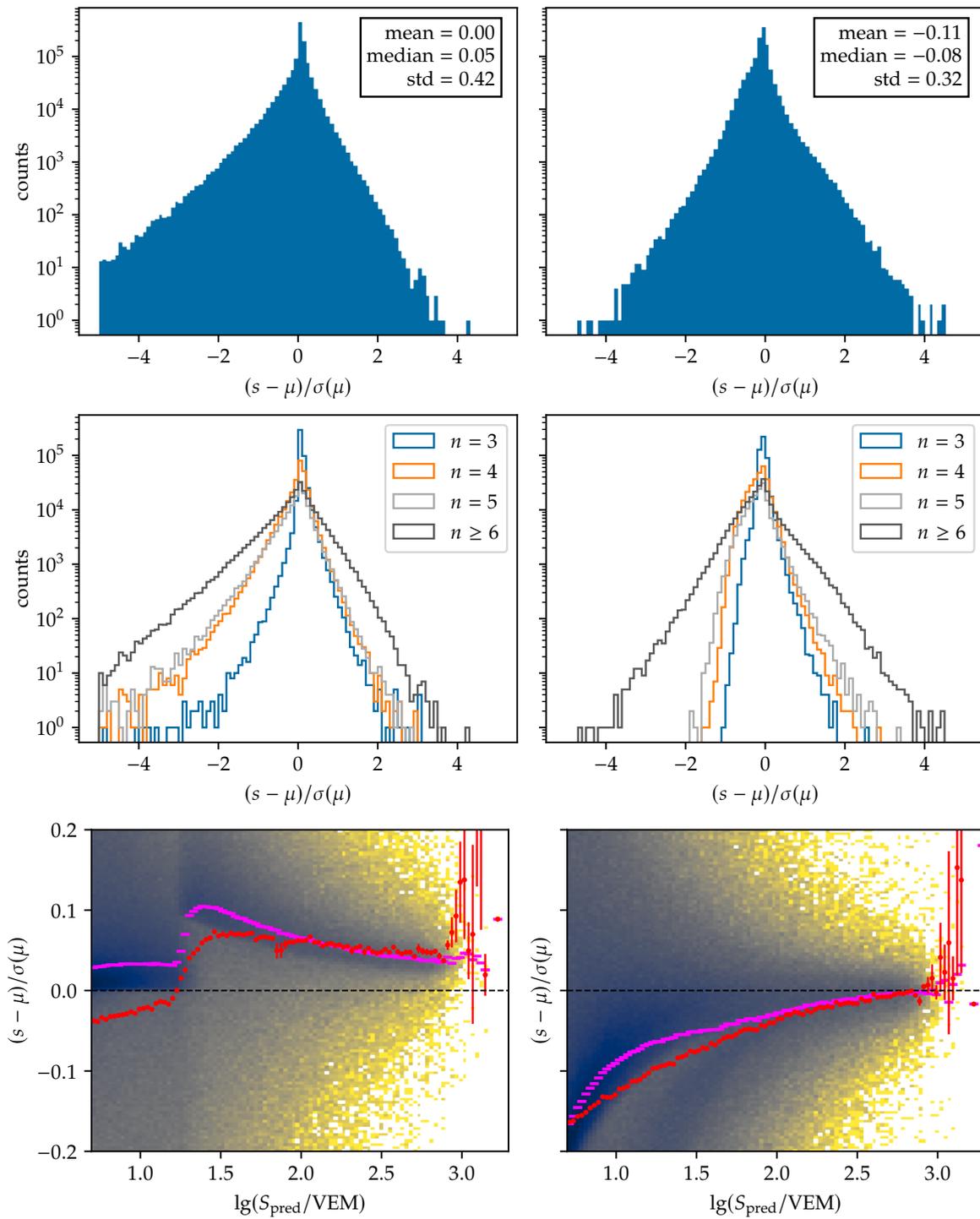


Figure 6.13: Comparison of the standard reconstruction using Offline (left) and the new reconstruction (right) for the 2016 SD-750 dataset. All triggered, unsaturated stations with $p_{\text{trig}} \geq 90\%$ from 6T5 events with $\theta \leq 55^\circ$ were used. *Top:* Histogram of the pulls. *Middle:* Histogram of pulls for different numbers of triggered stations n in the events. *Bottom:* Pulls over the predicted signal. The pink dashes represent the median and the red dots represent the mean with its standard error as error bars.

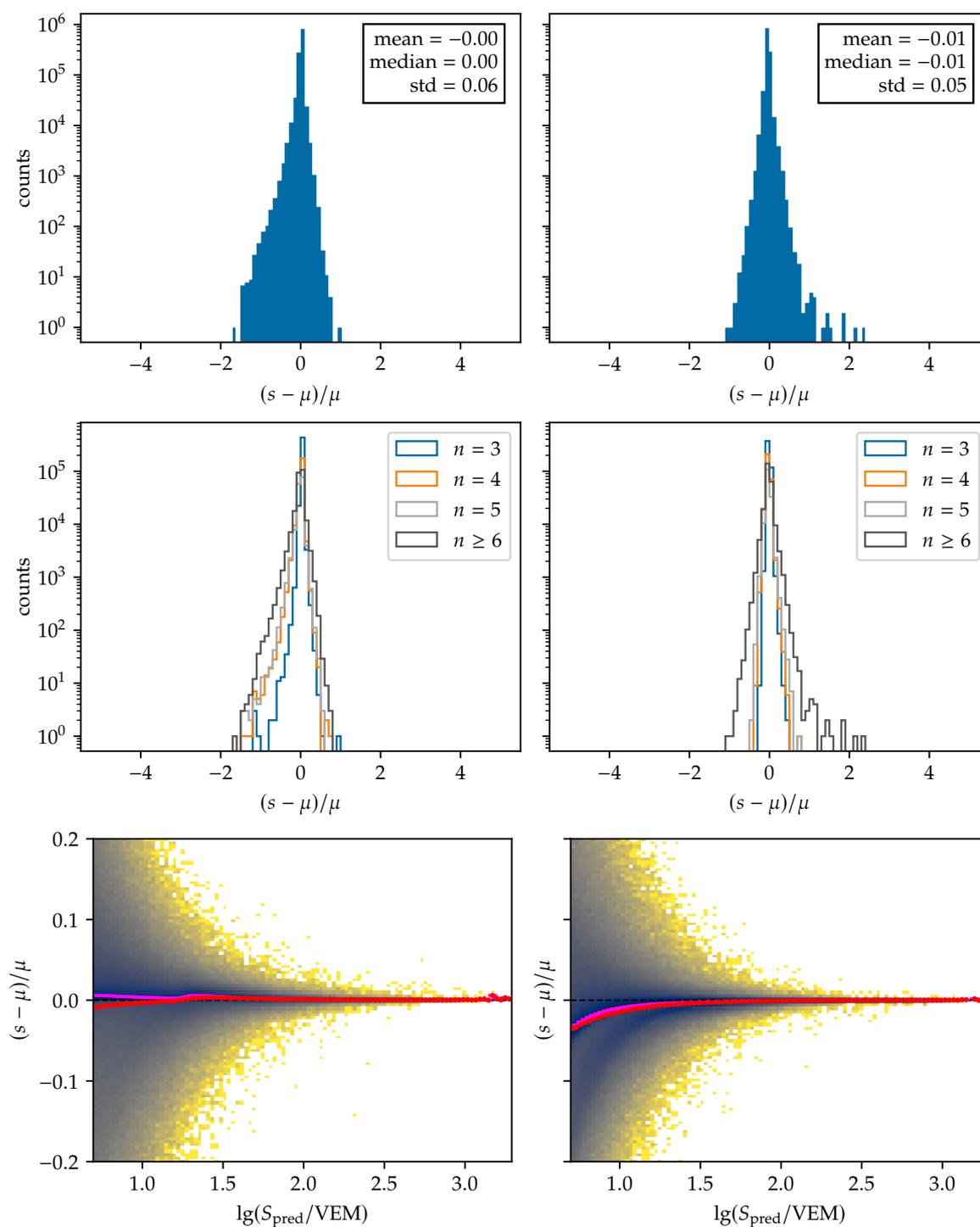


Figure 6.14: Comparison of the standard reconstruction using Offline (left) and the new reconstruction (right) for the 2016 SD-750 dataset. All triggered, unsaturated stations with $p_{\text{trig}} \geq 90\%$ from 6T5 events with $\theta \leq 55^\circ$ were used. *Top:* Histogram of the residuals. *Middle:* Histogram of residuals for different numbers of triggered stations n in the events. *Bottom:* Residuals over the predicted signal. The pink dashes represent the median and the red dots represent the mean with its standard error as error bars.

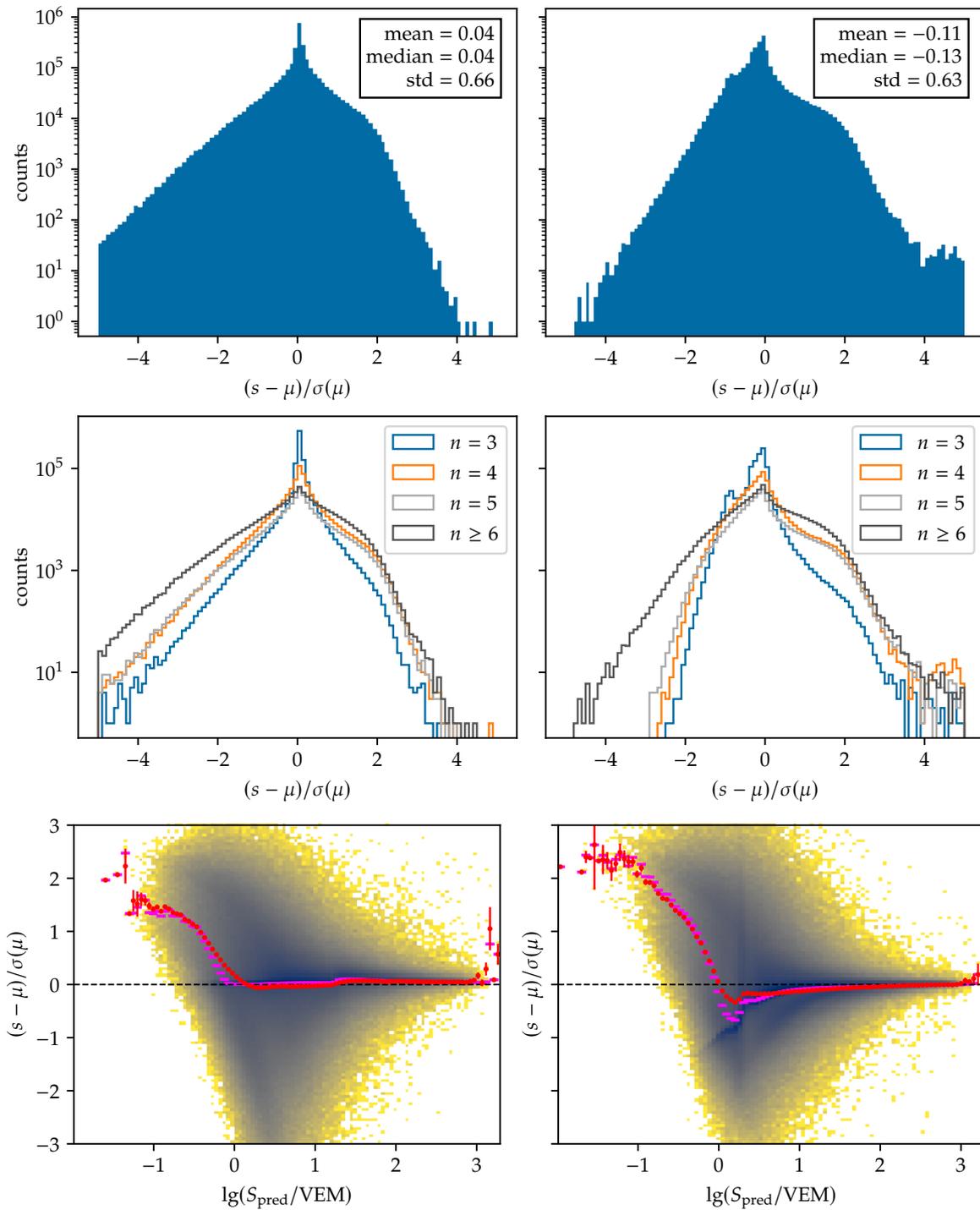


Figure 6.15: Comparison of the standard reconstruction using Offline (left) and the new reconstruction (right) for the 2016 SD-750 dataset. All triggered, unsaturated stations from 6T5 events with $\theta \leq 55^\circ$ were used. *Top:* Histogram of the pulls. *Middle:* Histogram of pulls for different numbers of triggered stations n in the events. *Bottom:* Pulls over the predicted signal. The pink dashes represent the median and the red dots represent the mean with its standard error as error bars.

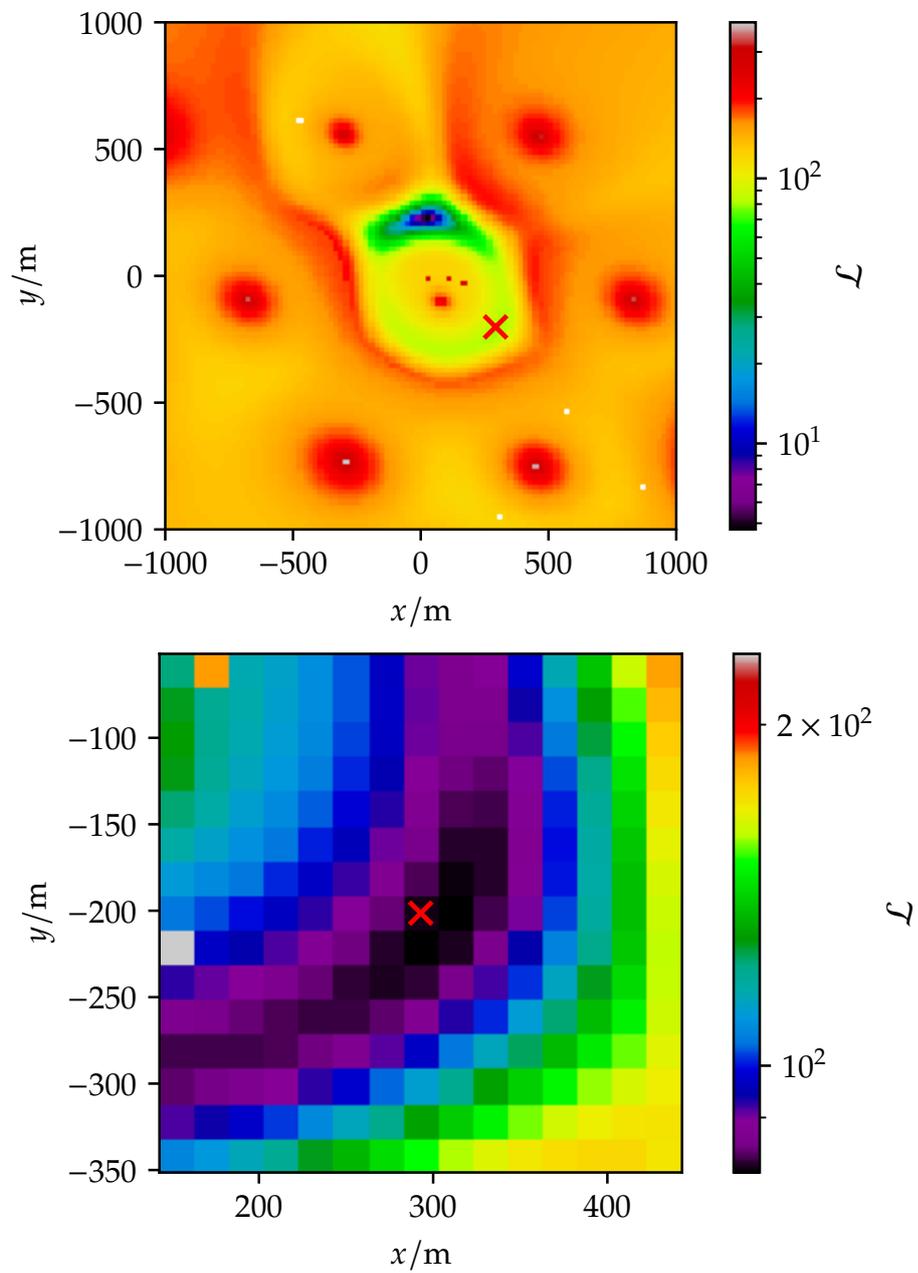


Figure 6.16: Likelihood scan for an event with a wrong reconstruction. The point $(0,0)$ marks the barycenter of the signals while a red cross represents the reconstructed core. *Top:* Reconstructed core position and the global minimum of the log-likelihood \mathcal{L} where the core should be. *Bottom:* Scan around the reconstructed core position. Obviously, the minimiser got stuck in a local minimum.

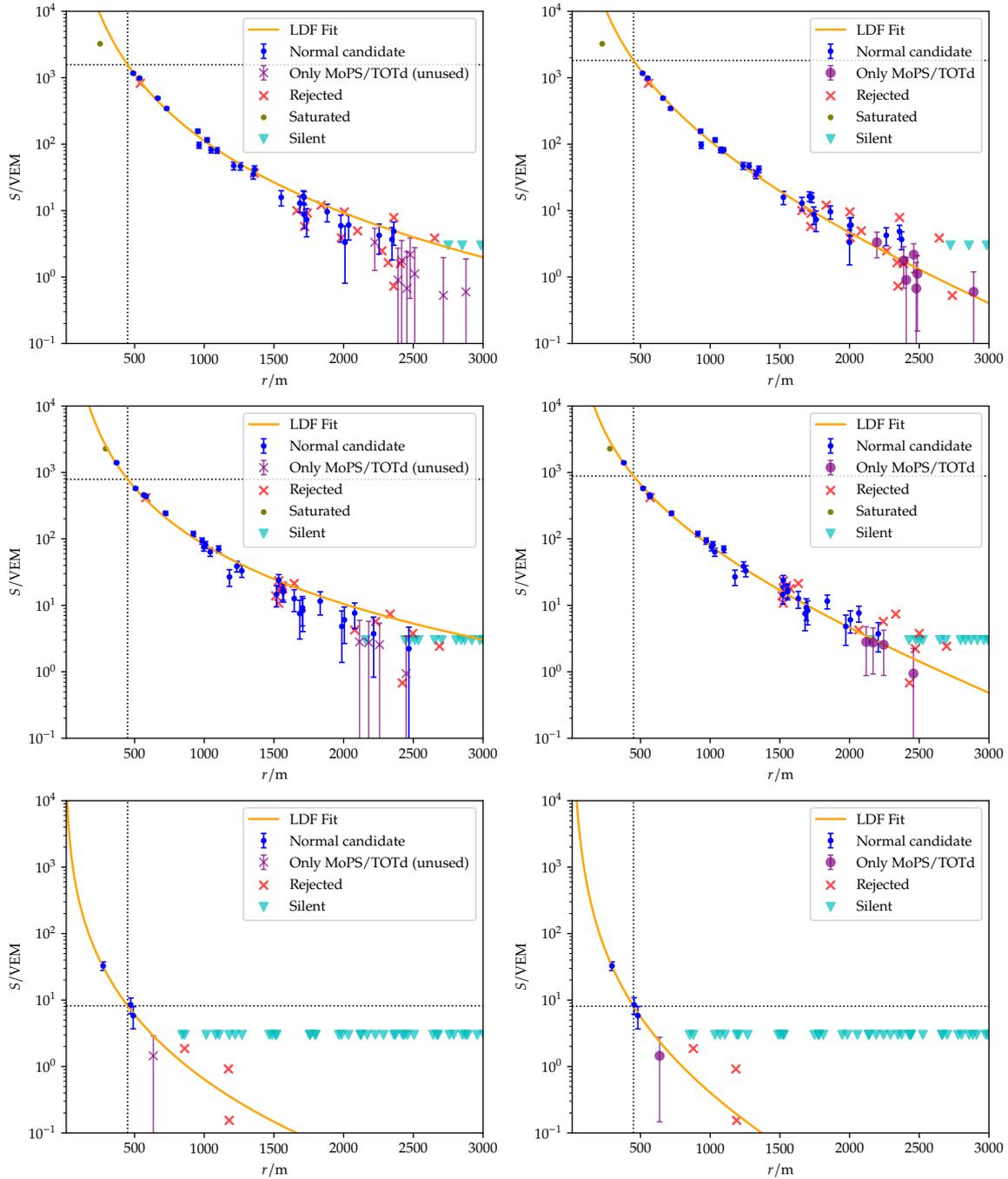


Figure 6.17: Collection of LDFs where the standard reconstruction is shown on the left and the new reconstruction for the same event is shown on the right.

CHAPTER 7

Correcting for Atmospheric Attenuation Effects Using the CIC Procedure

Assuming that the spectrum of cosmic rays at our observation point does not depend on the observation angle, the observed flux of primary particles should be isotropic. However, the atmosphere attenuates the shower of secondary particles leading to a zenith dependence of the observed shower size S_{ref} . The *constant intensity cut* (CIC) tries to correct this attenuation effect to find the »true« shower size. The attenuation-corrected shower size can then be converted to the energy of the primary particle by using an energy calibration obtained from *golden hybrid events*, meaning events which were observed by the SD and FD detector.

Fig. 7.1 shows the shower size which is not constant in $\sin^2 \theta$ and, therefore, violates the isotropy assumption and how it should look after the attenuation correction.

In this chapter, an update to the CIC for the SD-750 array with data obtained by the reconstruction chain described in the previous chapters is presented.

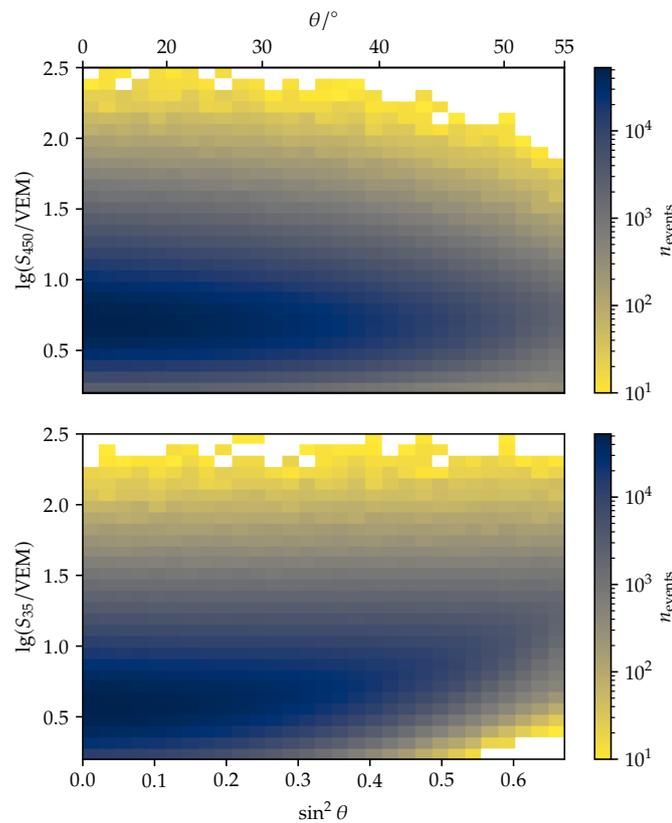


Figure 7.1: The shower size S_{450} (*top*) and the attenuation-corrected shower size S_{35} (*bottom*) over $\sin^2 \theta$. S_{450} is not constant in $\sin^2 \theta$ in the full efficiency range of the detector, i.e. the isotropy assumption is violated due to attenuation effects.

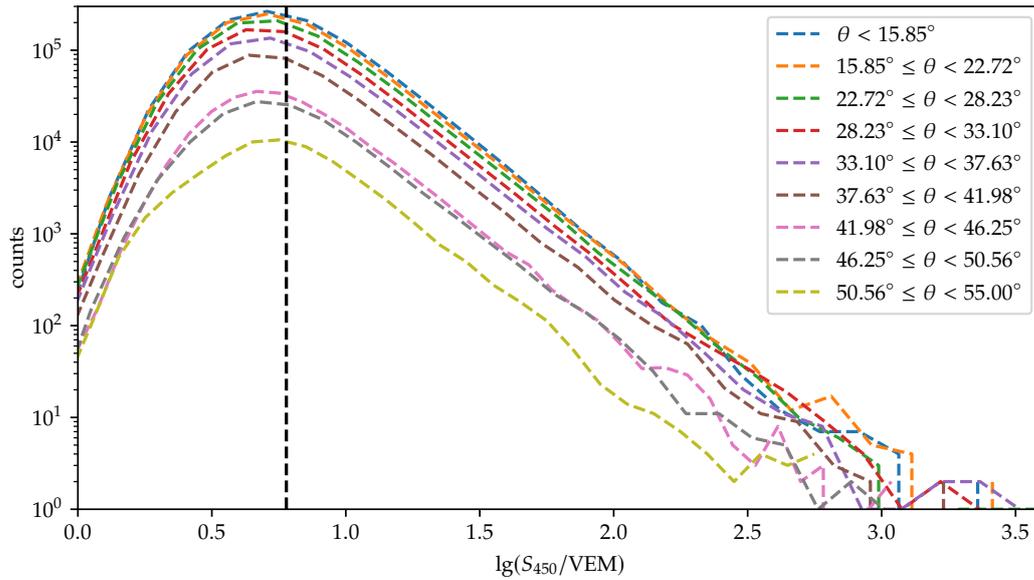


Figure 7.2: Raw S_{450} spectra for different zenith bins equally distributed in $\sin^2 \theta$. The spectra do not align because the attenuation of the shower in the atmosphere leads to a violation of isotropy. The dashed vertical black line marks the point where according to Eq. (5.14) the array is approximately fully efficient.

7.1 CIC Procedure

This chapter follows the procedure explained in Ref. [44] adapted for the SD-750 array.

Translating the isotropy assumption to equations, for each zenith bin $d\theta$ the number of events dN should follow

$$dN \propto \cos \theta \sin \theta d\theta = \frac{1}{2} d \sin^2 \theta, \quad (7.1)$$

or in other words

$$\frac{dN}{d \sin^2 \theta} = \text{const.} \quad (7.2)$$

according to Ref. [44]. Fig. 7.2 shows that the number of events is not isotropic in the measured shower size S_{450} and Eq. (7.2) is violated. This is due to the attenuation of the shower in the atmosphere. Since inclined showers have to travel longer through the atmosphere (see Fig. 6.1), their shower size is reduced compared to the shower size of the same shower entering the atmosphere more vertically.

This effect can be corrected in a data-driven way using the so-called *constant intensity cut* (CIC) analysis. Following Ref. [44], the attenuation-corrected shower size is

$$S_{\text{CIC}} = \frac{S_{\text{ref}}}{f_{\text{att}}(\theta)} \quad \text{with} \quad f_{\text{att}}(\theta) = 1 + ax + bx^2 + cx^3, \quad (7.3)$$

where $x(\theta) = \cos^2 \theta - \cos^2 \theta_{\text{ref}} = \sin^2 \theta_{\text{ref}} - \sin^2 \theta$ and constants a, b, c .

Assuming that the attenuation depends not only on the zenith angle but also on the energy [21], the attenuation is also a function of the attenuation-corrected shower size, i.e. $f_{\text{att}}(\theta) \rightarrow f_{\text{att}}(\theta, S_{\text{CIC}})$. In this case, we have

$$\alpha = k_{\alpha,0} + k_{\alpha,1} y + k_{\alpha,2} y^2, \quad \text{where} \quad y = \lg \left(\frac{S_{\text{CIC}}}{50 \text{ VEM}} \right) \quad (7.4)$$

Table 7.1: Parameters for the attenuation functions in Eqs. (7.3) and (7.4) for $\theta_{\max} = 55^\circ$.

	energy-independent	energy-dependent		
		$k_{\alpha,0}$	$k_{\alpha,1}$	$k_{\alpha,2}$
a	1.751	1.627	0.016	-0.803
b	-1.346	-1.546	-0.249	-2.265
c	-2.286	-1.685	-2.637	6.616

for $\alpha \in \{a, b, c\}$. Since the right-hand side of Eq. (7.3) now depends on the left side, the attenuation-corrected shower size has to be calculated iteratively until the value of S_{CIC} converges.

The reference angle is chosen to be $\theta_{\text{ref}} = 35^\circ$ since it is close to the median of the zenith angle distribution. Using this reference angle, we will write $S_{35} := S_{\text{CIC}}$ from now on.

The CIC procedure fits the attenuation function so that the intensity at a chosen value S_{35}^{cut} becomes constant in $\sin^2 \theta$. Since it was used multiple times in Auger analyses in the past, it is usually performed by collaboration members assigned to this task. Based on the updated reconstruction data set presented in this thesis, an energy-dependent CIC using the countdown method [44, p. 5] was obtained [45]. In addition, an energy-independent CIC was fitted [46] using binned and unbinned tests of uniformity [44, pp. 5–10] to fit the attenuation function. The parameters for both CICs fitted to data up to $\theta_{\max} = 55^\circ$ are listed in Table 7.1.¹ Appendix A present an energy-dependent and an energy-independent CIC fitted to data up to $\theta_{\max} = 40^\circ$, which are not used in this thesis. However, future studies might benefit from having direct access to this CIC when limiting the spectrum studies to zenith angles $\theta \leq 40^\circ$.

7.2 Investigating the Attenuation-Correction

The raw S_{35} spectra with equal, exclusive bins in $\sin^2 \theta$ for the energy-dependent and energy-independent CIC are shown in Fig. 7.4. the raw S_{35} spectra for different exclusive bins in $\sin^2 \theta$ should align if the isotropy condition is fulfilled. We still see some deviations from isotropy, mainly in the bin containing $\theta = 45^\circ$. The reason for this is not understood. Unknown inefficiencies in the detector or the reconstruction may explain this behaviour. However, the integral spectra – meaning that bins for smaller $\lg(S_{35}/\text{VEM})$ contain all the counts from bins with larger S_{35} – in Fig. 7.5 align well. This means that the constant intensity requirement is fulfilled and the CICs can be used.

For all further studies presented in this thesis, the energy-independent CIC will be used since the spread of the spectra is smaller. Furthermore, Fig. 7.3 shows that the intensity at S_{35}^{cut} of the energy-independent CIC is flat in $\sin^2 \theta$ while this is not the case for the energy-dependent.

¹Uncertainties were not provided by the sources.

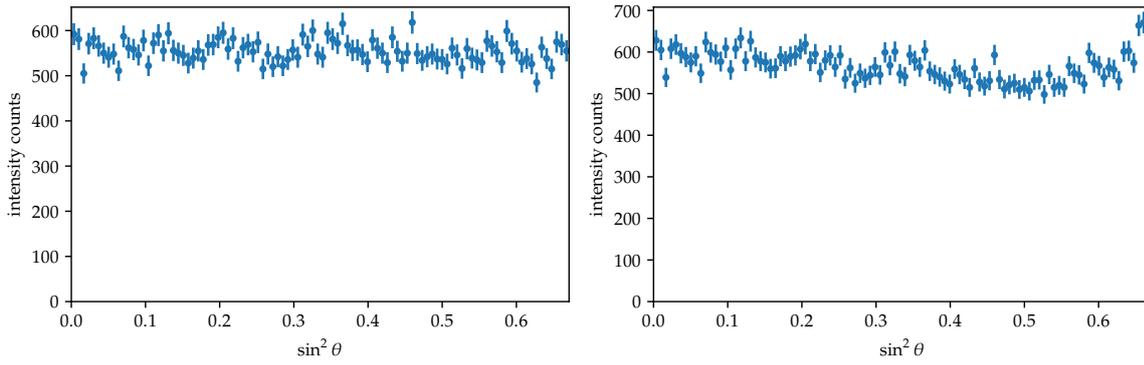


Figure 7.3: Intensity at S_{35}^{cut} for zenith bins equally distributed in $\sin^2 \theta$ using the CICs from Table 7.1. *Left:* energy-independent CIC. *Right:* energy-dependent CIC.

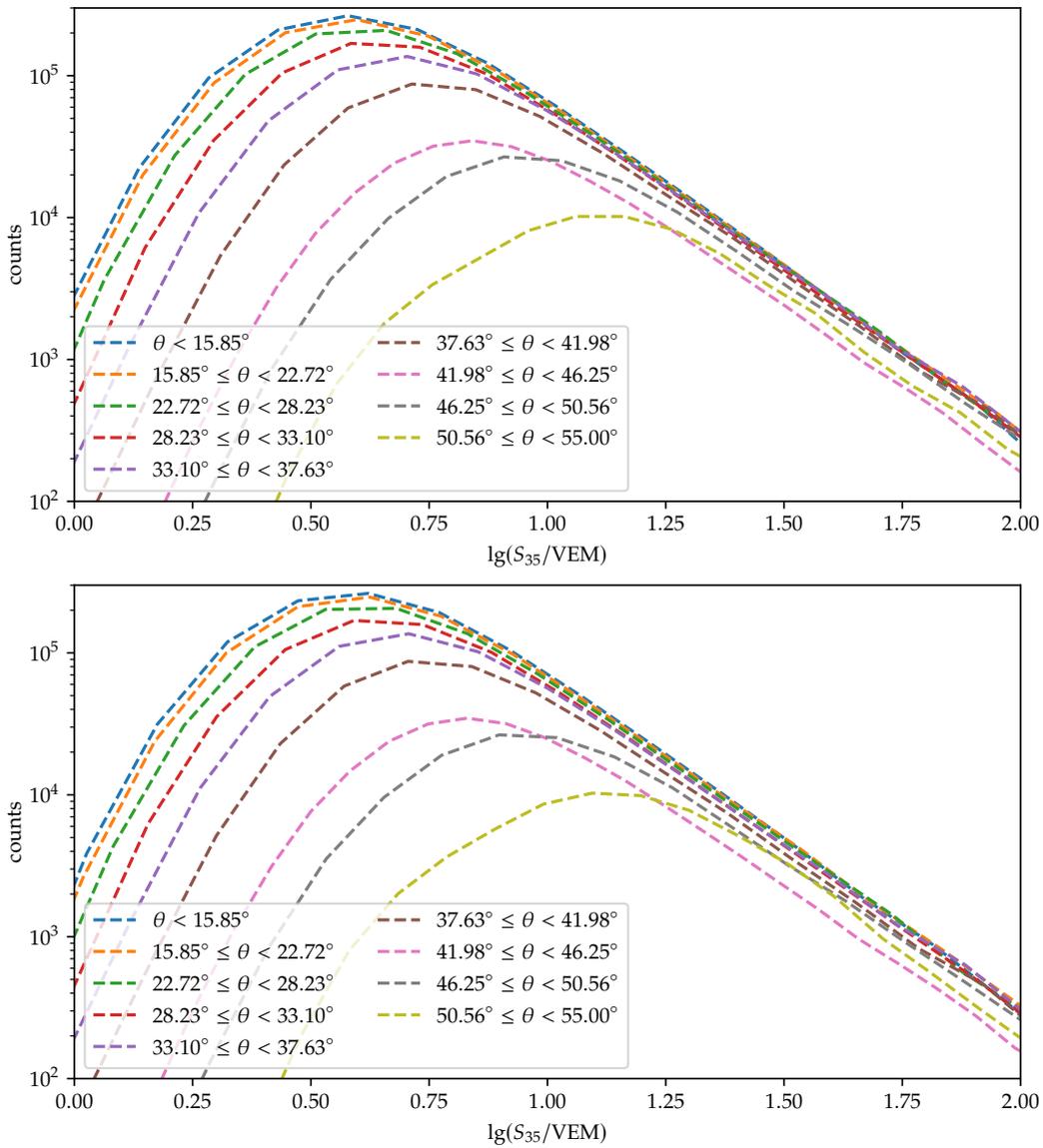


Figure 7.4: Raw S_{35} spectra for zenith bins equally distributed in $\sin^2 \theta$ using the CICs from Table 7.1. *Top:* energy-independent CIC. *Bottom:* energy-dependent CIC.

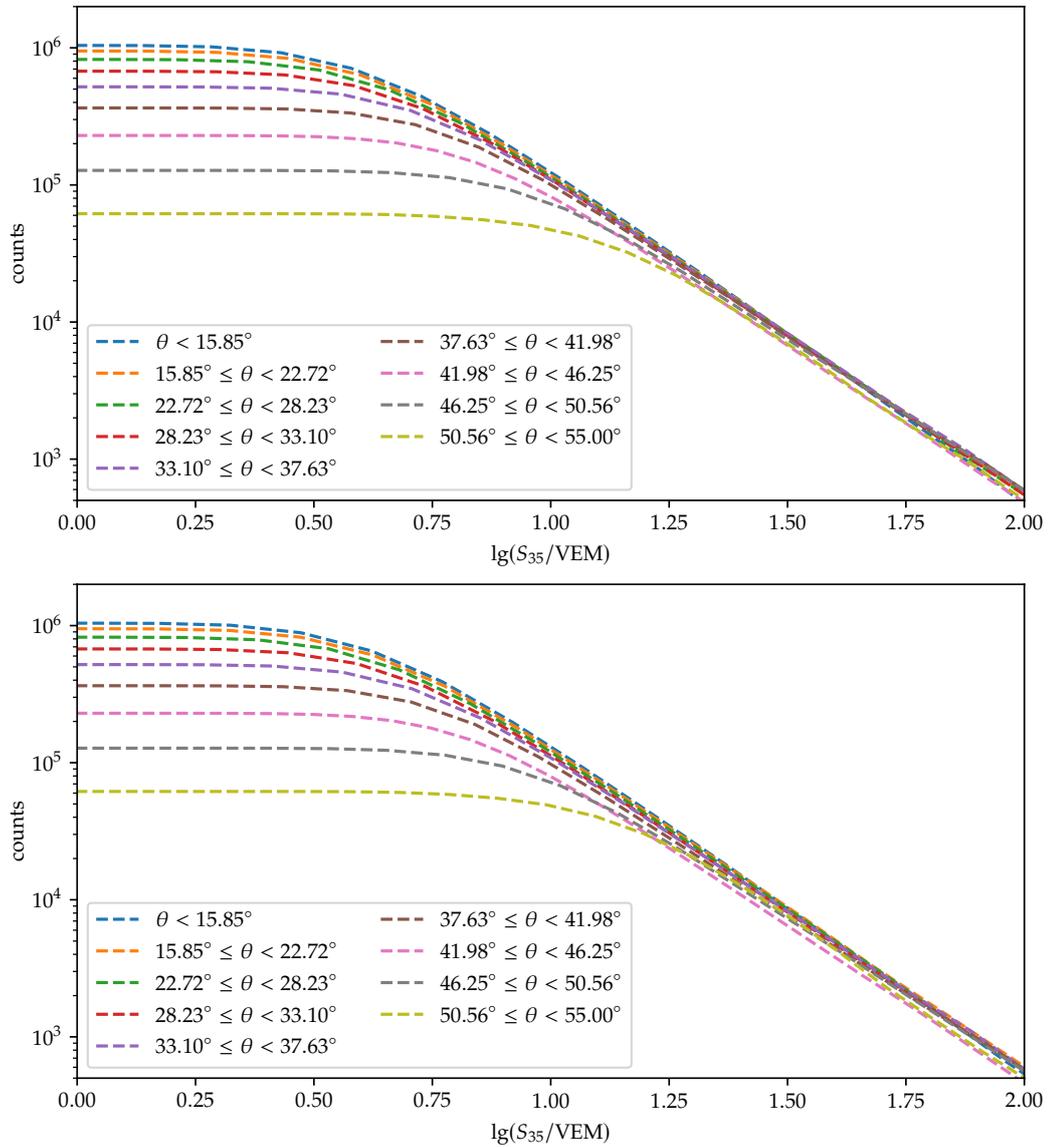


Figure 7.5: Integral raw S_{35} spectra for zenith bins equally distributed in $\sin^2 \theta$ using the CICs from Table 7.1. *Top:* energy-independent CIC. *Bottom:* energy-dependent CIC.

CHAPTER 8

Energy Calibration using SD-FD Hybrids and Array Efficiency

Using the SD-FD hybrid setup of the Pierre Auger Observatory, the attenuation-corrected shower size S_{35} of the surface detector needs to be cross-calibrated to the energy measured by the fluorescence telescopes. In this way, a reconstruction of the energy of the primary particle is still possible for events measured by only the SD and not the FD. This is important since the FD can only operate in dark moonless nights while the surface detector has an uptime of almost 100%. Therefore, it is necessary to use events measured only by the SD for the energy spectrum since the flux of cosmic rays at high energies is already very small.

This chapter presents an energy calibration for the updated reconstruction and chosen CIC to be used to measure an energy spectrum of cosmic rays in the following chapters. Since the measured spectrum also contains detector effects, we need to find parameterisation of e.g. the detector efficiency to correct the measured spectrum. A brief study on this is also presented in this chapter.

8.1 Energy Calibration

Using phase Ib SD-FD hybrid events – meaning the events observed simultaneously with both, the SD and the FD detector – an energy calibration for the SD-750 array can be found. Data from bad periods, non-6T5 events, and events with a failed T5 posterior trigger (reconstruction quality) are excluded. The calibration procedure is described in detail in Ref. [22, pp. 97–100]. While the surface detector only measures the attenuation-corrected shower size S_{35} , the FD measures the energy of the shower and therefore of the primary particle. Fig. 8.1 shows that the relation between the energy and S_{35} can be described by the simple form

$$E(S_{35}) = A \left(\frac{S_{35}}{\text{VEM}} \right)^B. \quad (8.1)$$

2513 hybrid events with $17.2 \leq \lg(E/\text{eV}) \leq 19$ were used in the calibration. The parameters of Eq. (8.1) were found to be [47]

$$A = (13.91 \pm 0.29) \text{ PeV} \quad \text{and} \quad B = 0.986 \pm 0.006. \quad (8.2)$$

The goodness of fit is $\chi^2/n_{\text{dof}} = 2612/2511$ ($p = 0.008$). The used events together with the calibration function are shown in Fig. 8.1. Since there are not enough hybrid events in the range $\lg(E/\text{eV}) < 17.2$,¹ the calibration is extrapolated into this region.

According to Ref. [21], the main uncertainty of 10% on the energy arises from the calibration of the FD, which, after including other effects, leads to an overall systematic uncertainty on the energy scale of 14%. This uncertainty is energy-independent. Since the uncertainty on

¹This is the threshold where the FD events are dominated by Cherenkov instead of fluorescence light. Potentially, it could be reduced by restricting the whole study to, e.g. $\theta \leq 40^\circ$.

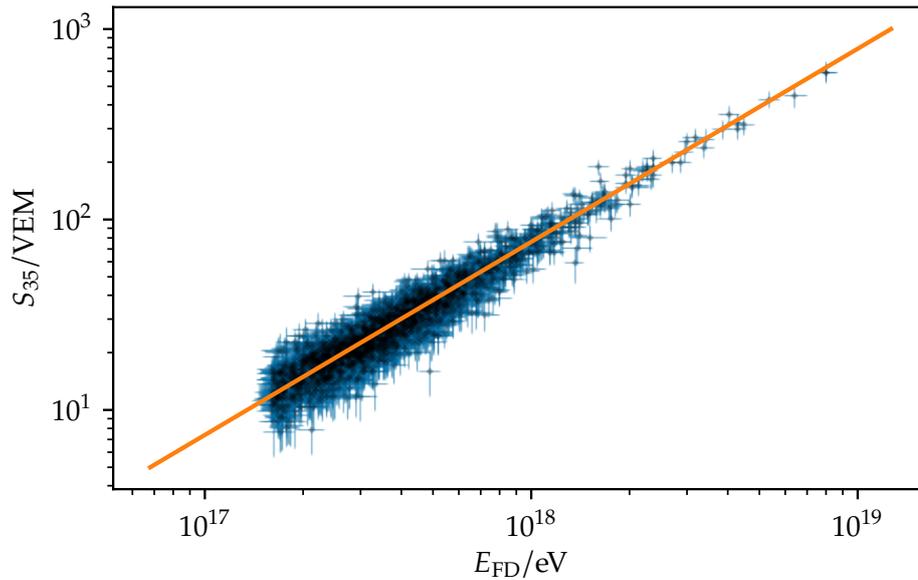


Figure 8.1: Energy calibration with 2513 SD-FD hybrid events. The events are shown as black dots with their blue error bars showing the uncertainties on S_{35} and E_{FD} . The orange line shows the energy calibration fitted to these data points.

the energy scale is large, other uncertainties arising from, e.g. the new reconstruction and the CIC, will be neglected.²

8.2 Energy Bias and Resolution

In a proper spectrum study, it is important to understand the possible differences between the estimated and the *true* energy.

As explained in Ref. [21], the energy resolution can be estimated in a data-driven way by analysing the distribution of the ratio between the energy obtained from the SD, using the calibration from Eqs. (8.1) and (8.2), and the measured FD energy. Since σ_{FD} is known, σ_{SD} can be fitted by minimising the log-likelihood assuming a *ratio distribution* for E_{SD}/E_{FD} .³

Using this method and the code provided by Ref. [21], the energy resolution is modeled as

$$\frac{\sigma_{SD}(E)}{E} = a_0 + a_1 \sqrt{\frac{E \text{eV}}{E}}, \quad (8.3)$$

where the values of the parameters a_i are given in Table 8.1.

In addition to resolution effects, we need to consider also the energy bias. In Ref. [21] the main reasons for their bias are explained by unknown changes in the mass composition and overestimation of the shower size for small energies. However, the reconstruction presented in this thesis uses the probability for stations to trigger. Therefore, it is reasonable to assume that the overestimation effect should be smaller leading to a smaller overall bias compared to Ref. [21].

Unfortunately, the time constraints of this thesis prevented a proper bias study. The bias can be studied with time-consuming simulations or by cross-checking with the SD-433. Anyway,

²This is the usual procedure, see e.g. Ref. [21] or Ref. [22, pp. 101–102].

³A more detailed discussion can be found in Ref. [21].

Table 8.1: Values for the parameters of the energy resolution and bias [21].

	energy resolution				energy bias		
	a_0	a_1	b_0	b_1	b_2	b_3	b_4
value	0.0850	0.0319	-3	26	0.35	12.7	-0.0039
uncertainty	0.0157	0.0089	1	4	0.02	0.1229	0.0008

since the energy scale of the SD-433 was calibrated with the SD-750 energy scale, such a bias study seems unreliable.

As a workaround, the energy bias

$$b_{\text{SD}}(E) = b_0(\lg(E/\text{eV}) - b_1) \exp(-b_2(\lg(E/\text{eV}) - b_3)^2) + b_4, \quad (8.4)$$

with the values of the parameters b_i as given in Table 8.1, from Ref. [21] will be used as an additional uncertainty in the spectrum study. In other words, we assume a bias of exactly 0 to find the parameters for the features of the spectrum. Repeating the analysis including the bias from Eq. (8.4) then allows estimating the influence σ_{bias} on these parameters. Therefore, the results will consist of the fitted value, its statistical uncertainty σ_{stat} , its systematic uncertainty σ_{sys} coming from the uncertainty on the energy scale, and its systematic uncertainty σ_{bias} coming from the lack of knowledge of the energy bias. To compare the found parameters to the results of other publications, it is sufficient to combine σ_{sys} and σ_{bias} in quadrature to find a total systematic uncertainty

$$\sigma_{\text{sys}}^{\text{tot}} = \sqrt{\sigma_{\text{sys}}^2 + \sigma_{\text{bias}}^2}. \quad (8.5)$$

8.3 Investigating the Array Efficiency

As mentioned before, we need to understand the efficiency of the detector array when performing studies on the measured energy spectrum.

The detector efficiency strongly depends on the energy E and the zenith angle θ . Since the surface detector does not measure the energy directly but rather the shower size S_{450} which becomes S_{35} after correction of the attenuation effects, the efficiency of the detector can be investigated in terms of S_{35} and does not depend on the energy calibration with the SD-FD hybrids.

The efficiency of the SD-750 was investigated and parameterised, e.g. Eq. (5.14), based on simulations in the past. The reason for this is that the efficiency can only be parameterised by comparing the number of measured showers with the "true" number of showers arriving when the latter is known. Unfortunately, the "true" number of showers is not known for a detector setup with uncontrolled sources. Therefore, we need to rely on simulation studies. Since the interpretation of the measured energy spectrum heavily relies on knowing the detector properties, this might seem unsatisfying since the validity of simulations is not clear.

However, even though obtaining an efficiency parameterisation in a data-driven way seems impossible, the data still allows us to test assumptions of full efficiency. The usual definition for full efficiency within the Pierre Auger Collaboration is $p_{\text{T4}}(E, \theta) \geq 0.97$. A brief study on this threshold is presented in this chapter.

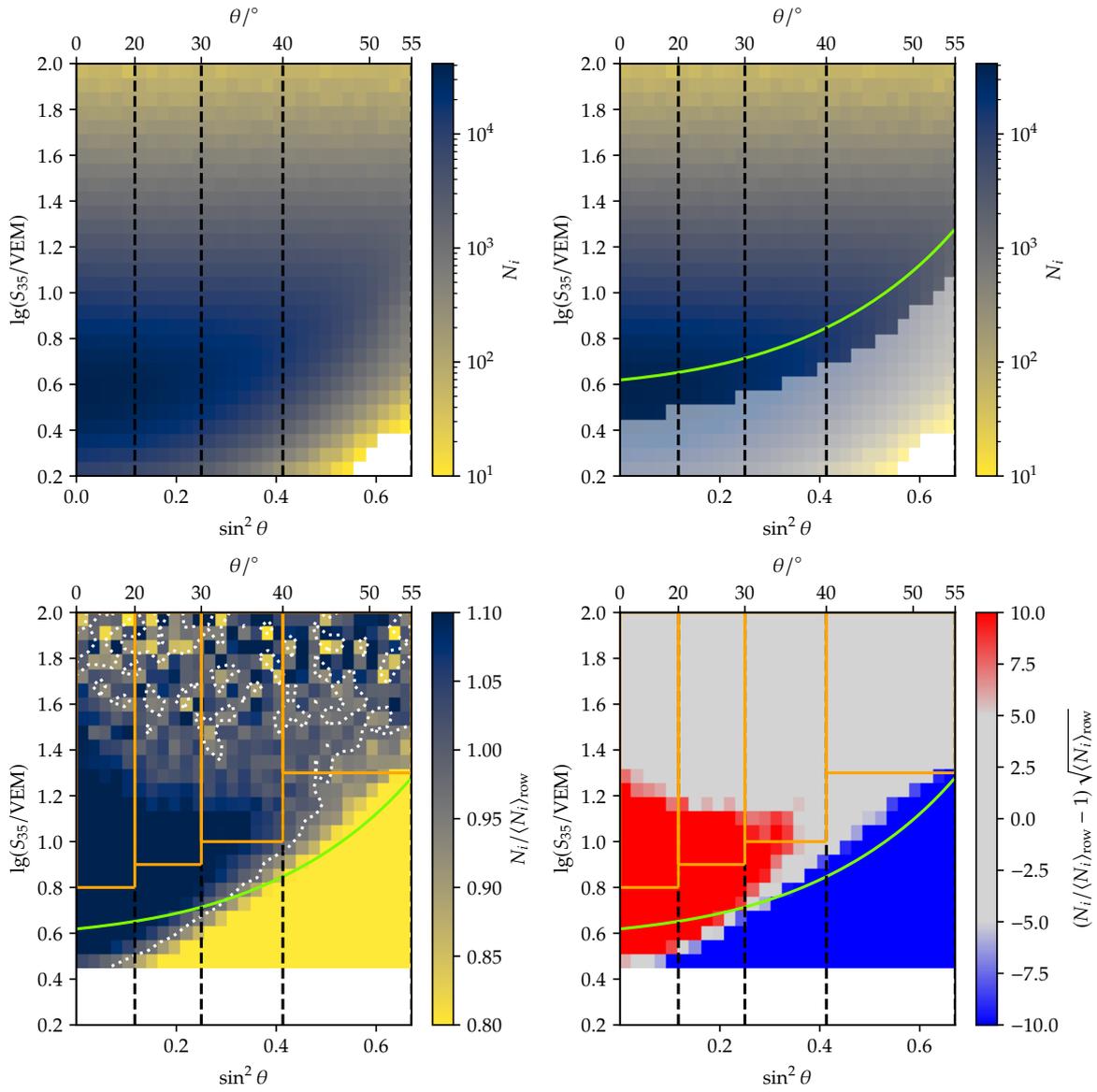


Figure 8.2: *Upper left:* two dimensional histogram of number of events N_i per $(S_{35}, \sin^2 \theta)$ bin i . *Upper right:* efficiency parameterisation from Eq. (5.14) as a green curve. The non-transparent bins were selected on a parallel to this curve and were used to calculate the row mean $\langle N_i \rangle_{\text{row}}$. *Lower left:* number of events divided by the mentioned row mean. The dotted line represents the 0.97 contour of this ratio. The orange boxes are suggested to be regions of full efficiency. *Lower right:* roughly the deviations in terms of Poisson fluctuations from the calculated row mean. This is only valid for the box including 55° since for a correct estimation the row mean should stop at the $\sin^2 \theta$ limit of each box.

8.3.1 Isotropy Condition

The distribution of events as functions of S_{35} and $\sin^2 \theta$ in Fig. 8.2 shows that fewer events are recorded for small values of S_{35} and high zenith angles. This hints at a drop in the detector efficiency. Even though we have a parameterisation for the efficiency from simulations as

given in Eq. (5.14) (originally given in Ref. [22]), i.e.

$$p_{T4}(E, \theta) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\lg(E/\text{eV}) - a(\theta)}{b_0 + b_1 \sin^2 \theta} \right) \right], \quad \text{where} \quad a(\theta) = \sum_{i=0}^3 a_i \sin^{2i} \theta \quad (8.6)$$

and the a_i and b_i are shown in Table 5.3, we want to investigate this efficiency with data in more detail.

Calculating the mean $\langle N_i \rangle_{\text{row}}$ of the number of events for each S_{35} row over the zenith bins i , we can then show the deviations from a flat distribution – which one would expect for regions of full-efficiency – by plotting the histogram with $N_i / \langle N_i \rangle_{\text{row}}$ on the colour axis. While implemented, a χ^2 test as presented in Section 8.3.2 was not finished due to time constraints. For large S_{35} (meaning high energies) the deviations arise from statistical effects since the number of events decreases strongly for higher energies. Fig. 8.2 shows that the parameterised efficiency is too optimistic. Therefore, we should use the shown boxes (where full efficiency is always fulfilled) when creating a spectrum.

8.3.2 Uniformity of Impact Point Distributions

For full efficiency, we expect the reconstructed impact points to be equally distributed around the hottest station since the flux of cosmic rays is isotropic. Therefore, plotting this distribution for different S_{35} bins allows a data-driven insight into the estimation of full efficiency. Fig. 8.3 shows an example for the case of full efficiency, where the impact points are distributed uniformly, and for the case where the impact point distributions are not uniform, suggesting that the detector is not fully efficient. This S_{35} range corresponds mainly to events with only 3 triggered stations leading to the reconstructed impact point being pulled into the middle of these stations. Therefore, the distribution is not uniform meaning that the array or the reconstruction is not fully efficient. A collection of impact point distributions for different intervals of S_{35} is shown in Fig. 8.4 for events with $\theta \leq 30^\circ$ including the new triggers. It illustrates the influence of efficiency on the uniformity of impact point distributions for smaller shower sizes.

We can test the uniformity of core distributions using a χ^2 -like approach on the bins of the inner hexagon. According to Ref. [44, p. 9], we can define a χ^2 -like quantity for Poisson distributions with expected value ν and k measured data points n_i as

$$\chi_{\text{poi}}^2 = 2 \sum_i^k (\nu - n_i(1 + \ln \nu) + n_i \ln n_i). \quad (8.7)$$

Note, that for $n_i \rightarrow 0$ (empty bins) $n_i \ln n_i \rightarrow 0$ according to the rule of l'Hôpital. Since we estimated ν directly from the data, the number of degrees of freedom is $n_{\text{dof}} = k - 1$.

As an example, the uniformity test for events with $\theta \leq 30^\circ$ using the reduced chi-square $\chi_{\text{poi}}^2 / n_{\text{dof}}$ for different S_{35} bins is shown in Fig. 8.5. The influence of the new triggers, lowering the full-efficiency threshold, is clearly visible. We will accept the hypothesis of uniformity of the impact point distributions for $\chi_{\text{poi}}^2 / n_{\text{dof}} \leq 2$.

However, the reconstructed shower size might still be valid even when the impact points are not uniformly distributed since the shower size and impact point are fitted independently in the reconstruction. Therefore, it is reasonable to slightly reduce the full-efficiency threshold obtained with the uniformity study.

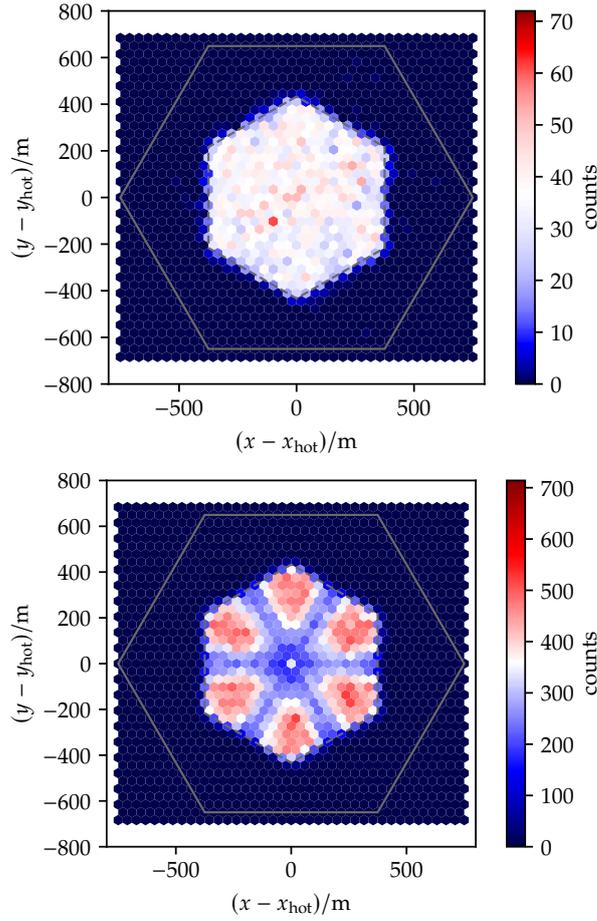


Figure 8.3: Example of impact point distributions around the hottest station. The neighbouring stations are located on the corners of the outer grey hexagon. White bins correspond to the number of events being roughly in the range $n_{\text{expected}} \pm 3\sigma_{\text{poi}}$. *Top:* example for the detector being fully efficient. The impact points for $\theta \leq 30^\circ$ and $S_{35} \in [13, 13.5]$ VEM are shown. *Bottom:* example of the detector not being fully efficient. The impact points for $\theta \leq 30^\circ$ and $S_{35} \in [6, 6.5]$ VEM are shown.

8.3.3 Full-Efficiency Thresholds

As shown above, the threshold of full efficiency can be estimated based on data for different maximum zenith angles and is given by a cut in S_{35} . Table 8.2 shows the results of this study together with the corresponding energy threshold using the calibration found in Section 8.1.

Table 8.2: Thresholds for full efficiency for different maximum zenith angles.

$\theta_{\text{max}}/^\circ$	$\lg(S_{35}/\text{VEM})$	S_{35}/VEM	$\lg(E/\text{eV})$
20	0.8	6.31	16.9
30	0.9	9	17.0
40	1.0	10	17.1
55	1.3	19.95	17.4

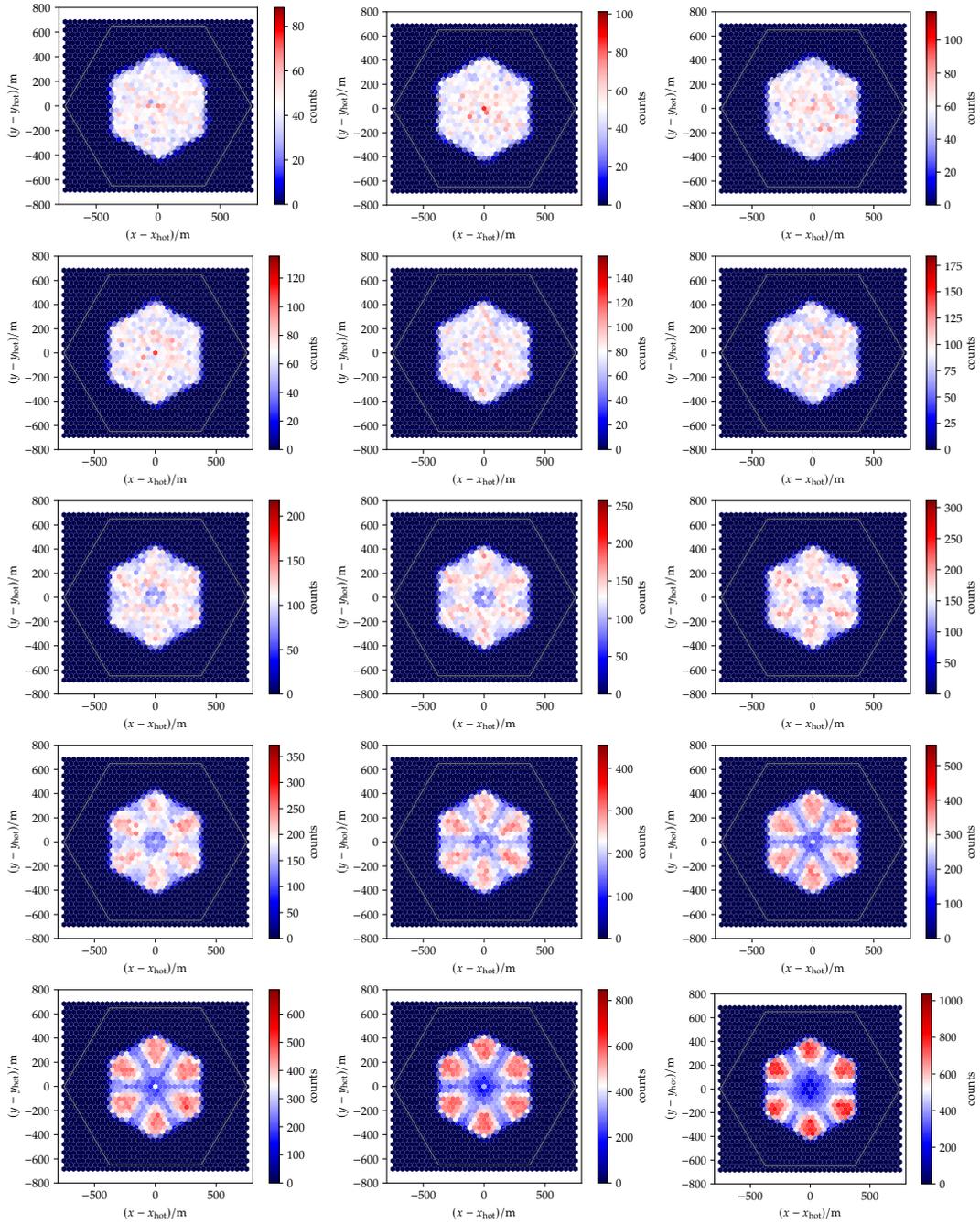


Figure 8.4: Impact point distributions around the hottest station for events with $\theta \leq 30^\circ$ for intervals of S_{35} from $[12, 12.5]$ VEM (*upper left*) to $[5, 5.5]$ VEM (*lower right*) in steps of 0.5 VEM. The neighbouring stations are located on the corners of the outer grey hexagon. White bins correspond to bins where the number of events is roughly in the range $n_{\text{expected}} \pm 3\sigma_{\text{poi}}$. The uniformity of core distributions is violated for small S_{35} .

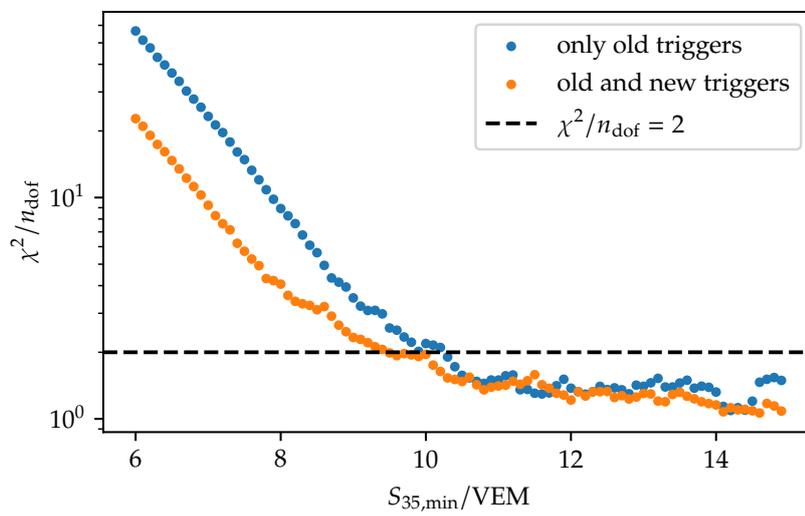


Figure 8.5: Uniformity test of the impact point distributions in the inner hexagon for $\theta \leq 30^\circ$, i.e. the reduced $\chi_{\text{poi}}^2/n_{\text{dof}}$ for different S_{35} bins ($S_{35,\text{min}}, S_{35,\text{min}} + 0.5 \text{ VEM}$).

CHAPTER 9

Energy Spectrum of Cosmic Rays and Fit of its Features

As shown in Section 8.3.3, restricting the spectrum study to lower zenith angles reduces the full-efficiency threshold down to lower energies. The disadvantage is that this decreases the total number of events and, therefore, the statistics. This is a more prominent problem for higher energies since the spectrum of cosmic rays is steeply falling. Therefore, the spectrum study will be performed for different zenith ranges. Fig. 9.1 shows the comparison of the raw spectra for $\theta_{\max} \in [20^\circ, 30^\circ, 40^\circ, 55^\circ]$. The influence of the zenith-dependent efficiency becomes visible for small energies while for higher energies the mentioned effect of low statistics is apparent.

Even though the raw spectra can already show the features, a proper investigation needs to take into account the detector effects. In addition, the number of events is not meaningful since it depends e.g. on the detector size. The spectrum should rather show the flux over energy including the correct exposure of the detector.

9.1 Exposure

The exposure of the array is [22, pp. 121–123]

$$\varepsilon = \int dt \mathcal{A}(t, E, \theta) \quad \text{with the aperture} \quad \mathcal{A}(E, \theta) = \Omega A_{\text{eff}}(t, \theta) p_{\text{T4}}(E, \theta), \quad (9.1)$$

where $A_{\text{eff}}(t, \theta)$ is the effective array area depending on the number of working stations at a certain time t and Ω the solid angle of the observation depending on θ_{\max} . It is not practical to have an efficiency-dependent exposure since the efficiency is not fully understood. Requiring full efficiency $p_{\text{T4}}(E, \theta) \approx 1$ for the dataset, the aperture simplifies to $\mathcal{A} = \Omega A_{\text{eff}}(t, \theta)$.

The effective area $A_{\text{eff}}(t, \theta)$ is the sum of the effective areas of all working hexagons at time t , i.e.

$$A_{\text{eff}}(t, \theta) = n_{\text{hex}}(t) A_{\text{hex}} \cos \theta \quad \text{with} \quad A_{\text{hex}} = \frac{\sqrt{3}}{2} d^2, \quad (9.2)$$

where $n_{\text{hex}}(t)$ is the number of working hexagons and d the distance between the stations. A hexagon is defined by the area around a station, shown as the blue area in Fig. 9.2. If the central station and all of its six neighbours are working, the hexagon is counted as working. For the SD-750 with $d = 750$ m, we have $A_{\text{hex}} \approx 0.487$ km².

Together with the solid angle

$$\Omega = \int d\varphi \int d\theta \sin \theta \quad (9.3)$$

the aperture is given by

$$\mathcal{A}(t) = n_{\text{hex}}(t) \int d\varphi \int d\theta \sin \theta \cos \theta A_{\text{hex}} = n_{\text{hex}}(t) \pi A_{\text{hex}} \sin^2 \theta_{\max}. \quad (9.4)$$

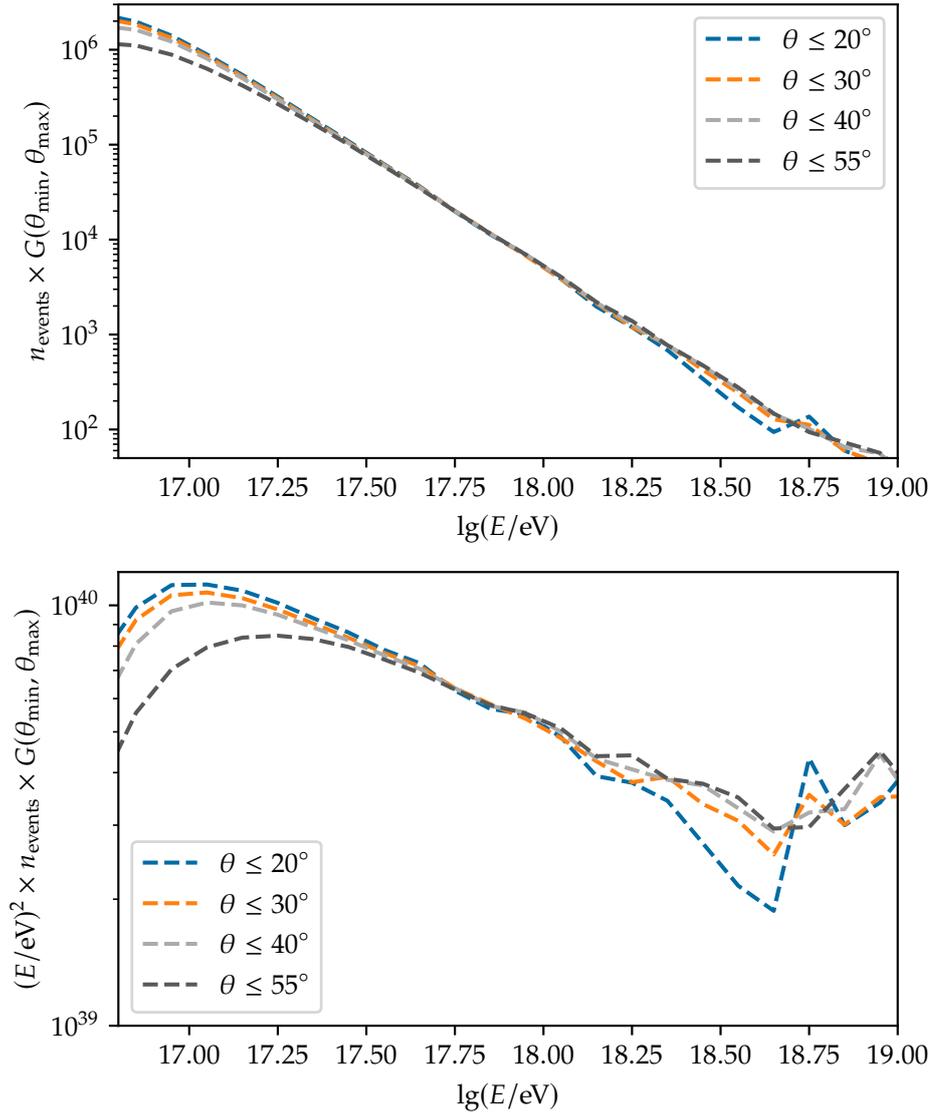


Figure 9.1: Number of events corrected by the geometry factor $G(\theta_{\min}, \theta_{\max}) = 1/(\sin^2 \theta_{\max} - \sin^2 \theta_{\min})$ over the reconstructed energy. The lines connect the midpoints of an underlying histogram with a bin width of $\Delta \lg(E/\text{eV}) = 0.1$. *Top:* Raw spectrum. *Bottom:* Raw spectrum scaled by the square of the energy.

Therefore, the exposure is

$$\varepsilon = A_{\text{hex}} \pi \sin^2 \theta_{\max} \int dt n_{\text{hex}}(t). \quad (9.5)$$

Calculating the exposure requires a good knowledge of the operating status of the detector. For the Pierre Auger Observatory, the number of working hexagons in each array is counted every minute. Excluding bad periods (BP) where the detector was not working properly, the exposure can be written as

$$\varepsilon = A_{\text{hex}} \pi \sin^2 \theta_{\max} \Delta t \sum_{i: t_i \notin \text{BP}} n_{\text{hex}, i}, \quad (9.6)$$

where $\Delta t = 1 \text{ min} = 525960^{-1} \text{ yr}$.

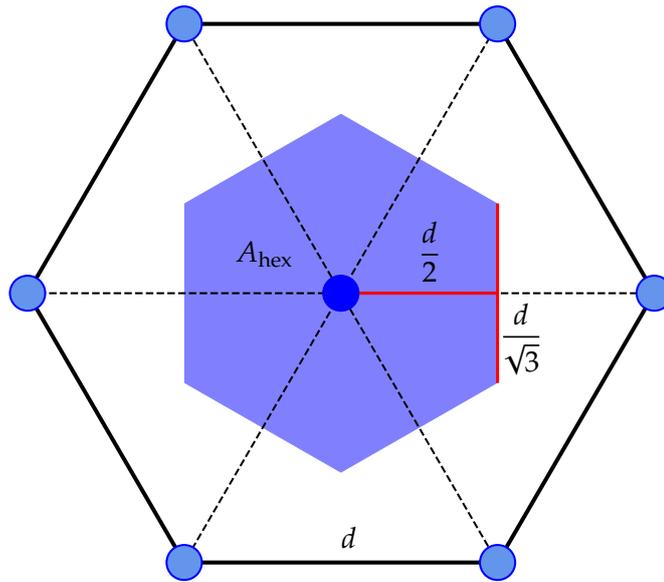


Figure 9.2: If the station in the middle and all of its six neighbours are working, the hexagon denoted by the blue area A_{hex} is considered working.

For the SD-750 array, phase I describes the time between the deployment of the array in 2008 to the end of 2021 when the read-out electronics were changed. As pointed out in Section 2.2, this phase is divided into phase Ia, spanning from 2008 to 2013 where only the old triggers were available, and phase Ib, spanning from 2014 to 2021 when the MoPS and TOTd triggers were implemented.

The exposure of the SD-750 array for phase I is

$$\varepsilon(\theta_{\max}) = k \sin^2 \theta_{\max}, \quad \text{with} \quad \frac{k}{\text{km}^2 \text{ sr yr}} = \begin{cases} 167.536 & ; \text{phase Ia} \\ 402.745 & ; \text{phase Ib} \\ 570.281 & ; \text{phase I} \end{cases} \quad (9.7)$$

The flux is then given by

$$J_{\text{meas}}(E) = \frac{n_{\text{events}}(E, \theta_{\max})}{\Delta E \varepsilon(\theta_{\max})} \quad (9.8)$$

with the observed number of events $n_{\text{events}}(E, \theta_{\max})$ with $\theta \leq \theta_{\max}$ in the energy bin with midpoint E and the bin width ΔE .¹

9.2 Forward-Folding Procedure and Fit Model

As explained before, the detector does not measure the true spectrum $J_{\text{true}}(E)$ but its convolution $J_{\text{meas}}(E)$ with the trigger efficiency $p_{T4}(E, \theta)$ and a response function $\mathcal{R}(E, \theta)$ which describes the smearing from the true energy E' to the measured energy E using the energy resolution described above.

¹Note that the bin width of $\Delta \lg(E/\text{eV}) = 0.1$ has to be converted.

Assuming $J_{\text{true}}(E)$ is described by a model $J_{\text{model}}(E, \vec{\alpha})$ with a set of parameters $\vec{\alpha}$, the expected true number of events in bin j is [21]

$$\mu_j(\vec{\alpha}) = \varepsilon \int_{\bar{E}_j}^{E_j + \Delta E_j} dE' J_{\text{model}}(E', \vec{\alpha}). \quad (9.9)$$

The bin-to-bin migration is represented by a matrix R_{ij} that is calculated using the response function $\mathcal{R}(E, \theta)$ and $p_{T4}(E, \theta)$ [21, 48]. Therefore, the expected measured number of events in bin i is

$$\nu_i(\vec{\alpha}) = \sum_j R_{ij} \mu_j(\vec{\alpha}). \quad (9.10)$$

By minimising a log-likelihood based on a Poissonian probability of observing $n_{\text{events}}(E_i)$ events when $\nu_i(\vec{\alpha})$ are expected, the optimal spectrum parameters $\hat{\alpha}$ can be found [21]. Therefore, the flux in bin i is

$$J(E_i) = c_i J_{\text{meas}}(E_i) \quad \text{with the correction factors} \quad c_i = \frac{\mu_i(\hat{\alpha})}{\nu_i(\hat{\alpha})}. \quad (9.11)$$

Since the suppression starts dominating at around $\lg(E/\text{eV}) = 19.5$ this value will be used as an upper limit for the data points used in the fitting procedure. A broken power law with soft transition is used as the fit model [21, p. 12], i.e.

$$J_{\text{model}}(E, \vec{\alpha}) = J_0 \left(\frac{E}{100 \text{ PeV}} \right)^{-\gamma_0} \prod_{i=0} \left(1 + \left(\frac{E}{E_{ij}} \right)^{\frac{1}{\omega_{ij}}} \right)^{(\gamma_i - \gamma_j) \omega_{ij}} \quad \text{with} \quad j = i + 1. \quad (9.12)$$

In this model, the fit parameters $\vec{\alpha}$ have to be interpreted as follows:

- J_0 is the over-all normalisation constant.
- E_{01} is the position of the second knee with the transition width ω_{01} .
- E_{12} is the position of ankle with the transition width ω_{12} .
- γ_0 and γ_1 are the spectral indices before and after the second knee. γ_2 is the spectral index after the ankle.

Note that the transition widths will be fixed to the values given in Ref. [49].

9.3 Investigating the Influence of the Full-Efficiency Threshold

Table 9.1 shows the fit parameters for the spectrum with $\theta \leq 20^\circ$ including the statistical and the bias uncertainties for different thresholds of full efficiency. It is important to note that the fit parameters for the second knee agree within their uncertainties for all thresholds of full efficiency.

This strongly indicates that the fitted feature is not an efficiency effect but rather a feature of the energy spectrum itself. Therefore, the *second knee can be measured* with the SD-750.

The low statistics dominate the uncertainties for the fit of the ankle. This is the result of restricting the spectrum to a very narrow zenith range to lower the full-efficiency threshold below the energy of the second knee. In contrast, the unknown energy bias increases the uncertainty of the fit parameters for the second knee. Future studies could solve this by investigating and incorporating the bias into the fitting procedure.

Similar studies on the influence of the full-efficiency threshold on the fit parameters of the spectrum for other zenith ranges were performed and are shown in Appendix B.2.

Table 9.1: Fit parameters for the spectrum with $\theta \leq 20^\circ$ for different full-efficiency thresholds E_{full} . As shown in Section 8.3.3, the array is fully efficient for all of these thresholds in this zenith range.

parameter $\lg(E_{\text{full}}/\text{eV})$	value $\pm \sigma_{\text{stat}} \pm \sigma_{\text{bias}}$		
	16.9	17.0	17.1
$\frac{J_0 \times 10^{13}}{\text{km}^{-2} \text{yr}^{-1} \text{sr}^{-1} \text{eV}^{-1}}$	$1.21 \pm 0.04 \pm 0.14$	$1.20 \pm 0.03 \pm 0.17$	$1.20 \pm 0.01 \pm 0.16$
$\lg(E_{01}/\text{eV})$	$17.07 \pm 0.03 \pm 0.02$	$17.08 \pm 0.08 \pm 0.10$	$17.08 \pm 0.04 \pm 0.33$
γ_0	$2.83 \pm 0.07 \pm 0.17$	$2.87 \pm 0.13 \pm 0.02$	$2.85 \pm 0.03 \pm 0.28$
γ_1	$3.41 \pm 0.01 \pm 0.03$	$3.41 \pm 0.02 \pm 0.01$	$3.41 \pm 0.02 \pm 0.03$
ω_{01} (fixed)	0.25	0.25	0.25
$\lg(E_{12}/\text{eV})$	$18.66 \pm 0.08 \pm 0.02$	$18.65 \pm 0.08 \pm 0.01$	$18.65 \pm 0.08 \pm 0.05$
γ_2	$2.65 \pm 0.27 \pm 0.01$	$2.65 \pm 0.27 \pm 0.00$	$2.65 \pm 0.29 \pm 0.08$
ω_{12} (fixed)	0.05	0.05	0.05

9.4 Fitting the Second Knee and the Ankle

The spectra for different zenith ranges were investigated using the full-efficiency thresholds obtained in Section 8.3.3. The unfolded spectra together with the fit and the σ_{bias} contours are shown in Figs. 9.3 and 9.4. Table 9.2 shows the fit parameters with the systematic uncertainties being calculated by shifting the energy by $\pm 14\%$ as mentioned in Section 8.1. For the spectrum with $\theta \leq 55^\circ$ the values for the second knee were fixed to the mean of the values obtained from the other spectra where the features could be fitted freely.

Except for the value of the spectral index γ_1 after the second knee which is a little high for the spectrum with $\theta \leq 20^\circ$, the obtained spectrum features are consistent within their uncertainties for all different zenith ranges. A visual comparison of the fits and their parameters for the different zenith ranges is shown in Fig. 9.5.

This supports the statement from Section 9.3: The SD-750 allows the measurement of the second knee. A comparison between the presented measurements of the spectrum features with the measurements of other publications will be presented in Section 9.5.

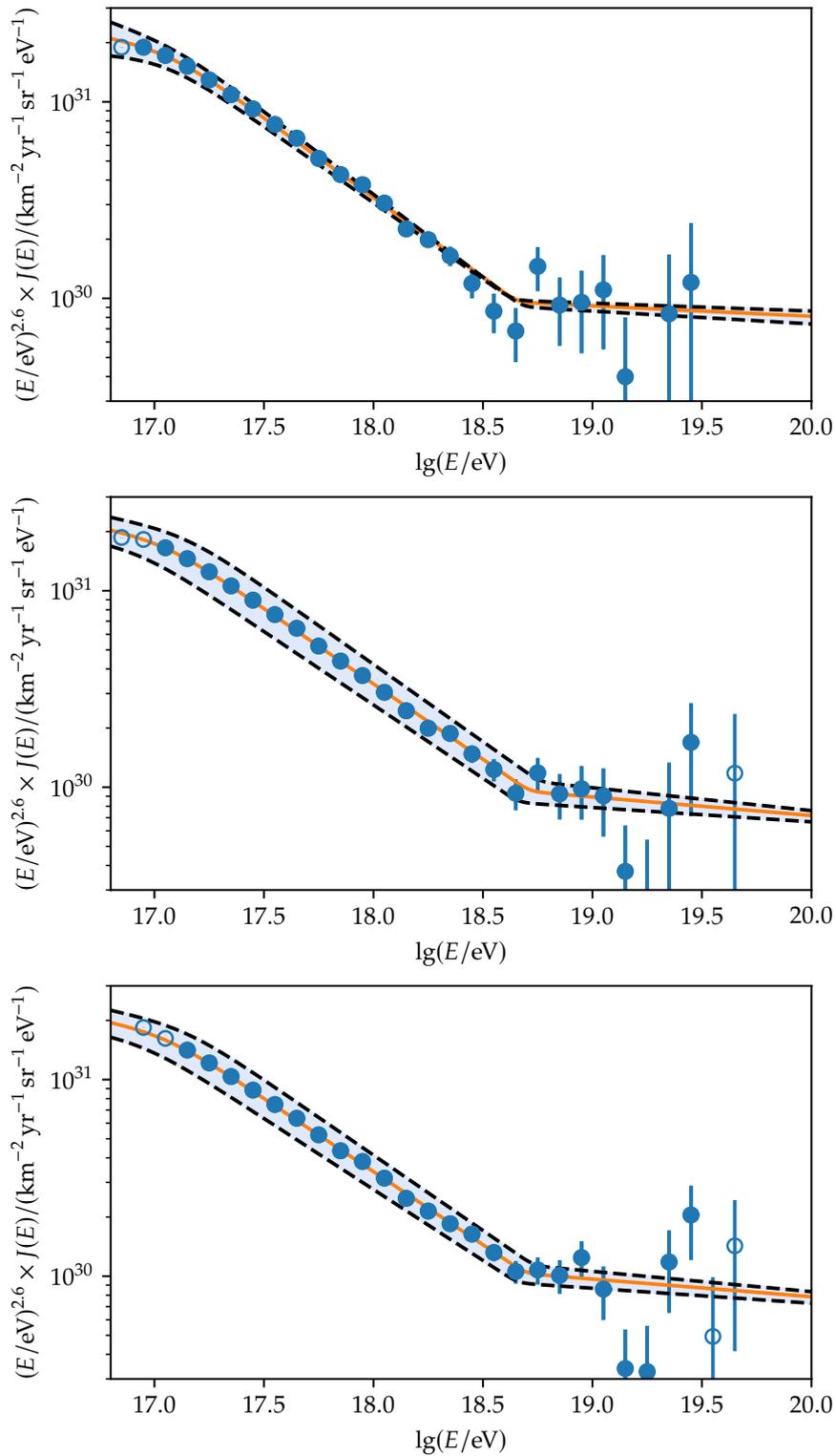


Figure 9.3: Unfolded spectra multiplied with $(E/\text{eV})^{2.6}$. Filled dots represent the data points with Poissonian errors used for the fit (orange). Empty dots were not used. The black dashed lines represent the contour of $\pm\sigma_{\text{bias}}$. *Top:* Spectrum for $0 \leq \theta \leq 20^\circ$. *Middle:* Spectrum for $0 \leq \theta \leq 30^\circ$. *Bottom:* Spectrum for $0 \leq \theta \leq 40^\circ$.

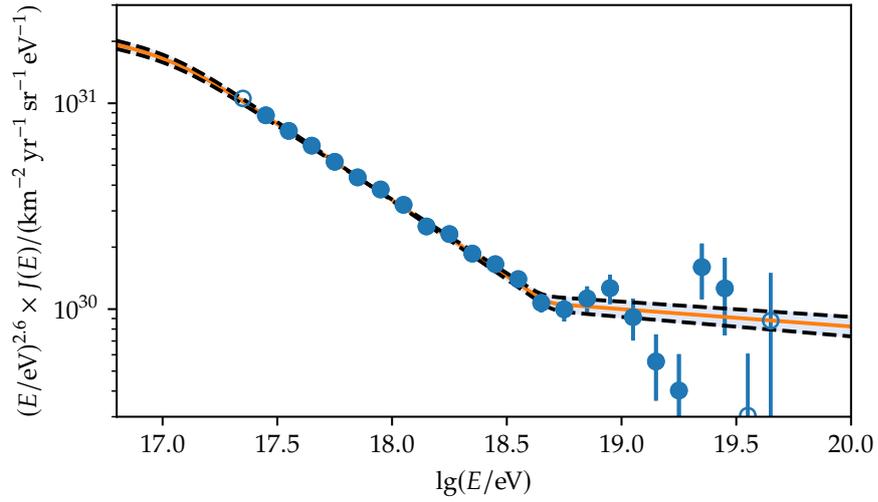


Figure 9.4: Unfolded spectrum for $\theta \leq 55^\circ$ multiplied with $(E/\text{eV})^{2.6}$. Filled dots represent the data points with Poissonian errors used for the fit (orange). Empty dots were not used. The black dashed lines represent the contour of $\pm\sigma_{\text{bias}}$. The second knee was fixed to the mean of the results from the fits shown in Fig. 9.3.

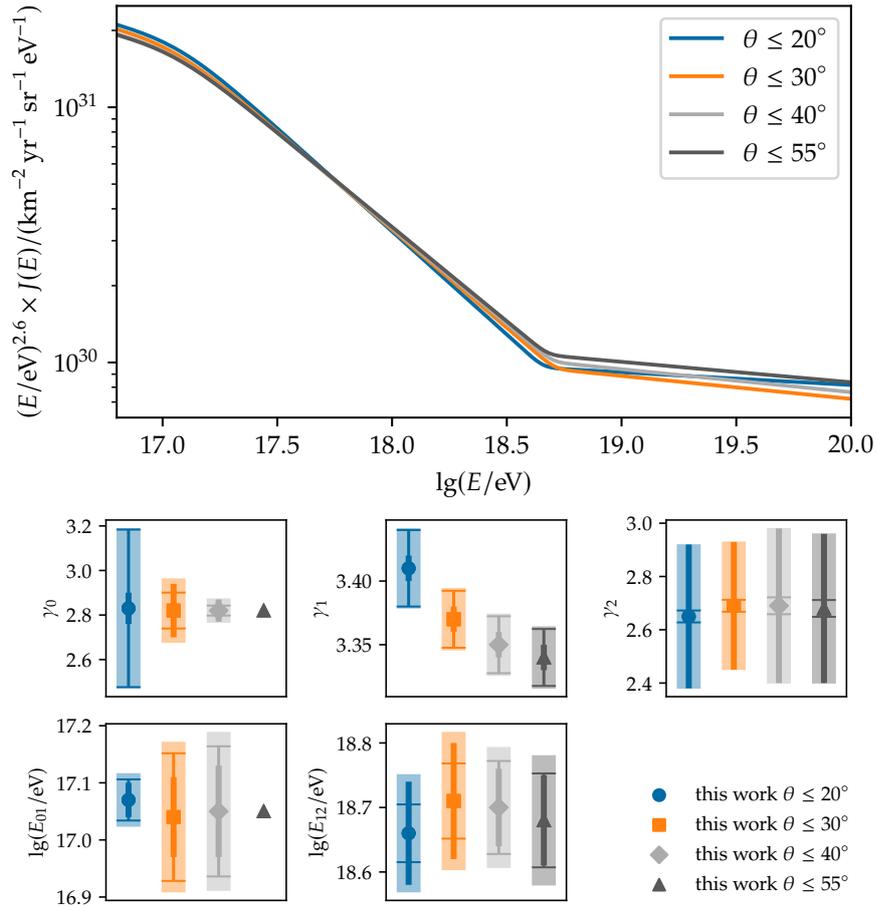


Figure 9.5: The fits of the unfolded spectrum for different zenith ranges. *Top:* Fitted models. *Bottom:* Parameters for the features with σ_{stat} (bold), $\sigma_{\text{sys}}^{\text{tot}}$ (thin with caps, includes σ_{bias}), and $\sigma_{\text{tot}}^2 = \sum_i \sigma_i^2$ (bold with transparency).

Table 9.2: Fit parameters for different spectra. Values without uncertainties were fixed during the fitting procedure.

parameter	20	30	40	55
$\theta_{\max}/^\circ$	20	30	40	55
$\lg(E_{\text{full}}/\text{eV})$	16.9	17.0	17.1	17.4
$\frac{J_0 \times 10^{13}}{\text{km}^{-2} \text{yr}^{-1} \text{sr}^{-1} \text{eV}^{-1}}$	$1.21 \pm 0.04 \pm 0.14 \pm 0.43$	$1.17 \pm 0.03 \pm 0.18 \pm 0.31$	$1.12 \pm 0.01 \pm 0.16 \pm 0.28$	$1.11 \pm 0.01 \pm 0.05 \pm 0.34$
$\lg(E_{01}/\text{eV})$	$17.07 \pm 0.03 \pm 0.02 \pm 0.03$	$17.04 \pm 0.07 \pm 0.10 \pm 0.05$	$17.05 \pm 0.08 \pm 0.09 \pm 0.07$	17.05
γ_0	$2.83 \pm 0.07 \pm 0.17 \pm 0.31$	$2.82 \pm 0.12 \pm 0.01 \pm 0.08$	$2.82 \pm 0.05 \pm 0.01 \pm 0.02$	2.82
γ_1	$3.41 \pm 0.01 \pm 0.03 \pm 0.00$	$3.37 \pm 0.01 \pm 0.02 \pm 0.01$	$3.35 \pm 0.01 \pm 0.02 \pm 0.01$	$3.34 \pm 0.01 \pm 0.02 \pm 0.01$
ω_{01}	0.25	0.25	0.25	0.25
$\lg(E_{12}/\text{eV})$	$18.66 \pm 0.08 \pm 0.02 \pm 0.04$	$18.71 \pm 0.09 \pm 0.05 \pm 0.03$	$18.70 \pm 0.06 \pm 0.04 \pm 0.06$	$18.68 \pm 0.07 \pm 0.02 \pm 0.07$
γ_2	$2.65 \pm 0.27 \pm 0.01 \pm 0.02$	$2.69 \pm 0.24 \pm 0.02 \pm 0.01$	$2.69 \pm 0.29 \pm 0.01 \pm 0.03$	$2.68 \pm 0.28 \pm 0.01 \pm 0.03$
ω_{12}	0.05	0.05	0.05	0.05

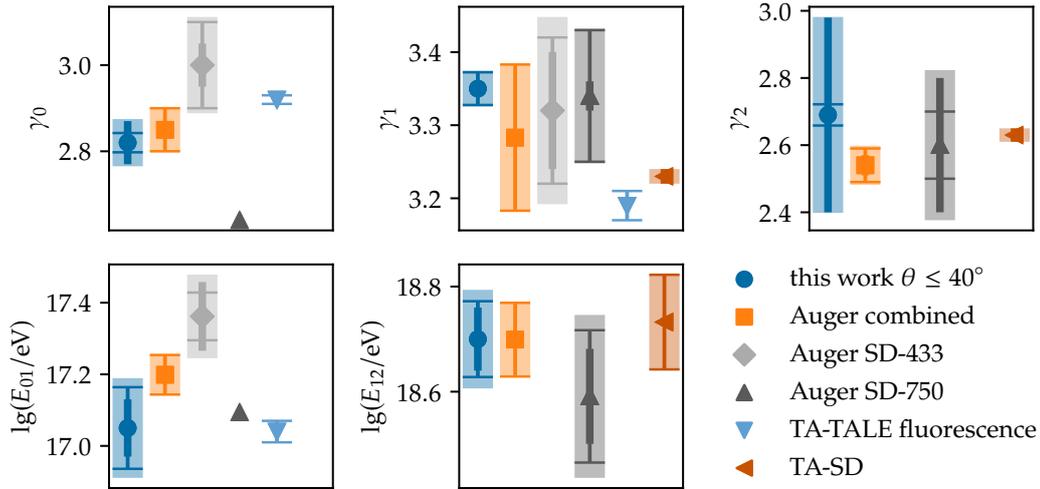


Figure 9.6: Parameters for the features with σ_{stat} (bold), $\sigma_{\text{sys}}^{\text{tot}}$ (thin with caps, includes σ_{bias}), and $\sigma_{\text{tot}}^2 = \sum_i \sigma_i^2$ (bold with transparency) shown for the publications given in Table 9.3 and the $\theta \leq 40^\circ$ spectrum presented in Section 9.4.

9.5 Comparing the Measurements with other Publications

A selection of measurements published by the Pierre Auger Collaboration (Auger) and the Telescope Array Collaboration (TA) is shown in Table 9.3. Fig. 9.6 shows that the values for the spectrum features presented in Section 9.4 are compatible with the values given in Ref. [21] and Ref. [49] within the uncertainties. For the latter, there is only tension for the spectral index γ_0 before the second knee. Since this value was fixed in Ref. [21] to a value incompatible with other experiments, this tension is not worrying. The position of the second knee and the ankle as well as the spectral index after the ankle as given in Ref. [50] and Ref. [51] are compatible with the measurements presented in this work within the uncertainties. Tensions arise for the spectral indices γ_0 and γ_1 . Since there is a known tension between the measurements of the Pierre Auger Observatory and the Telescope Array this is not unexpected.

More interesting is the tension of the second knee position and the spectral index γ_0 when comparing the results from Section 9.4 with the measurement from the SD-433 presented in Ref. [20]. The values are not compatible. However, since the energy scale of the SD-433 was cross-calibrated with an old SD-750 energy scale this tension might resolve when updating it with a more recent SD-750 energy calibration, e.g. the one presented in Chapter 8.

Table 9.3: Collection of different measurements of the spectrum.

parameter	Ref. [49]	Ref. [20]	value $\pm \sigma_{\text{stat}} \pm \sigma_{\text{sys}}$	Ref. [50] (no σ_{stat})	Ref. [51]
publication	Auger combined	Auger SD-433	Auger SD-750	TA-TALE fluorescence	TA-SD
experiment					
$\frac{J_0 \times 10^{13}}{\text{km}^{-2} \text{yr}^{-1} \text{sr}^{-1} \text{eV}^{-1}}$	-	$0.97 \pm 0.02 \pm 0.34$	$1.09 \pm 0.04 \pm 0.28$	-	-
$\text{Ig}(E_{01}/\text{eV})$	$17.20 \pm 0.01 \pm 0.06$	$17.36 \pm 0.10 \pm 0.07$	17.09	17.04 ± 0.04	-
γ_0	$2.85 \pm 0.01 \pm 0.05$	$3.00 \pm 0.05 \pm 0.10$	2.64	2.92 ± 0.01	-
γ_1	$3.28 \pm 0.00 \pm 0.10$	$3.32 \pm 0.08 \pm 0.10$	$3.34 \pm 0.02 \pm 0.09$	3.19 ± 0.02	3.23 ± 0.01
ω_{01}	0.25	0.25	$0.49 \pm 0.07 \pm 0.34$	sharp transition	-
$\text{Ig}(E_{12}/\text{eV})$	$18.70 \pm 0.01 \pm 0.07$	-	$18.59 \pm 0.09 \pm 0.13$	-	$18.73 \pm 0.01 \pm 0.09$
γ_2	$2.54 \pm 0.03 \pm 0.05$	-	$2.60 \pm 0.20 \pm 0.10$	-	2.63 ± 0.02
ω_{12}	0.05	-	0.05	-	sharp transition

CHAPTER 10

Conclusion and Outlook

In this thesis, I presented an updated reconstruction for the SD-750 data allowing the use of stations triggered only by the MoPS and TOTd triggers. This is essential when trying to reduce the energy threshold of the UHECR spectrum measured with the surface detector of the Pierre Auger Observatory. It is important to note that the updated reconstruction can be generalised for the other SD arrays.

Using simulations, I showed that the signals follow a log-normal distribution down to very small signals. To correctly use stations with such small signals in the reconstruction, I updated the likelihood including the station trigger probability. A new set of selection criteria for events where all parameters in the reconstruction can be fitted freely was developed for this thesis. Using these criteria and a new functional form of the LDF, I parameterised the LDF slope parameter to enable the reconstruction of low-energetic showers where only a few stations trigger. A correction for attenuation effects as well as an SD-FD hybrid energy calibration was performed to find the energy of a shower given the LDF reconstruction. In addition, I estimated the full-efficiency threshold of the SD-750 array in a data-driven way. This enabled investigations at the lower end of the energy spectrum obtained from the SD-750 data set.

By fitting the spectrum features for different zenith ranges and different thresholds of full efficiency, and showing their agreement with each other, I argued that the position of the second knee together with the change of the spectral index can be successfully measured with the SD-750. This is the first time that the second knee is measured with the SD-750 since in Ref. [21] fixed values of γ_0 and E_{01} were used because they “correspond to features below the measured energy region” [21, p. 12].

The measured position of the second knee at $E_{01} = 10^{17.05}$ eV and the change of the spectral index from $\gamma_0 = 2.82$ to $\gamma_1 = 3.35$, as presented in Chapter 9, are compatible with most of the other published measurements discussed in Section 9.5. The main uncertainties in this thesis arise from the lack of knowledge of the energy bias and the uncertainties of the fluorescence detector. All of these uncertainties can be reduced in future studies, allowing for a precise measurement of the second knee. For the SD-750, the uncertainties of the fit parameters for the ankle are dominated by low statistics, especially when restricting the data set to lower zenith angles.

A combined spectrum using the boxes shown in Fig. 8.2 might overcome these problems. In other words, the full-efficiency threshold can still be assumed to be at $E_{\text{full}} = 10^{16.9}$ eV since for small energies only data with $\theta \leq 20^\circ$ will be used. Since for higher energies the data from a much wider zenith range, e.g. $\theta \leq 55^\circ$, will be used, the ankle can be fitted more reliably due to better statistics. My new reconstruction, most importantly the updated likelihood described in Chapter 4, should be implemented into the standard reconstruction of Offline to allow the use of stations triggered only by MoPS and TOTd when reconstructing events. Further studies of the new lever-arm criteria, the new functional form of the LDF, and a detailed comparison of the old and the new reconstruction should lead the way to adopt my work presented in this thesis for the other surface arrays.

Investigations of the CIC, the array efficiency, the energy resolution, and the energy bias are necessary for further studies of the SD-750 energy spectrum. These studies should then be able to measure the spectrum features precisely. This is important when testing theoretical models of their physical origin.

Acknowledgments

Even though the work presented in this thesis was done (unless noted otherwise) and written by me, there are many people I want and have to thank.

I want to thank...

Prof. Dr. Ralph Engel for giving me the possibility to write this thesis, for his advice and for examining this thesis.

Priv.-Doz. Dr. Roger Wolf for kindly agreeing to be my second examiner.

Dr. Markus Roth for his expertise on the Pierre Auger Observatory and his input for my analysis.

Dr. Diego Ravnani for providing the energy-dependent CICs.

Dr. Valerio Verzi for the SD-FD energy calibration.

I want to thank all proofreaders, namely Gabriel Brichetto Orquera, Berenika Čermáková, Fiona Ellwanger, Dr. Steffen Hahn, Edyvania Emily Pereira Martins, Dr. Markus Roth, Dr. David Schmidt, Marc Schumann, Varada Varma Kizakke Covilakam, and Dr. Darko Veberič.

A special thanks goes to...

Dr. David Schmidt for his expertise on the experiment and its data analysis. In addition, I want to thank him for providing me with the necessary data I needed for my reconstruction.

Joaquín de Jesús for discussing the likelihood with me and suggesting the log-normal distribution.

Ezequiel Rodriguez for our productive discussions regarding the experiment, data analysis and statistics.

Gabriel Brichetto Orquera for his expertise in the energy spectrum and the unfolding procedure. I warmly thank him for providing me with his spectrum unfolding code.

I want to express my deep gratitude to Dr. Darko Veberič for his excellent knowledge and expertise regarding the experiment and the data analysis. His constant input and our (sometimes heated) fruitful discussions were a huge inspiration for me. I also want to thank him for his knowledge of Linux and – more importantly – movies.

I am very grateful to my colleagues in the IAP Auger group for the constant motivation and all the fun we had.

Last but not least, I want to deeply thank my family, namely my dad Karl-Heinz and brother Robin, and *mi amor* Ian for supporting me throughout my studies.

Bibliography

- [1] V. F. Hess, *Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten*, Phys. Z. **13** (1912), pp. 1084–1091.
- [2] P. Blasi, *The origin of galactic cosmic rays*, The Astronomy and Astrophysics Review **21** (2013), doi: [10.1007/s00159-013-0070-7](https://doi.org/10.1007/s00159-013-0070-7).
- [3] R. Aloisio, V. Berezhinsky, and A. Gazizov, *Transition from galactic to extragalactic cosmic rays*, Astroparticle Physics **39–40** (2012), pp. 129–143, doi: [10.1016/j.astropartphys.2012.09.007](https://doi.org/10.1016/j.astropartphys.2012.09.007).
- [4] V. Berezhinsky, *Extragalactic cosmic rays and their signatures*, Astroparticle Physics **53** (2014), pp. 120–129, doi: [10.1016/j.astropartphys.2013.04.001](https://doi.org/10.1016/j.astropartphys.2013.04.001).
- [5] S. Navas et al. (The Particle Data Group Collaboration), *Review of Particle Physics*, Phys. Rev. D **110** (2024), p. 030001, doi: [10.1103/PhysRevD.110.030001](https://doi.org/10.1103/PhysRevD.110.030001).
- [6] R. Engel and D. Schmidt, *Indirect Detection of Cosmic Rays*, Handbook of Particle Detection and Imaging, Springer International Publishing (2021), pp. 801–849, doi: [10.1007/978-3-319-93785-4_24](https://doi.org/10.1007/978-3-319-93785-4_24).
- [7] E. N. Parker, *The passage of energetic charged particles through interplanetary space*, Planetary and Space Science **13** (1965), pp. 9–49, doi: [10.1016/0032-0633\(65\)90131-5](https://doi.org/10.1016/0032-0633(65)90131-5).
- [8] A. E. Vladimirov et al., *Testing the Origin of High-energy Cosmic Rays*, ApJ **752**, 68 (2012), p. 68, doi: [10.1088/0004-637X/752/1/68](https://doi.org/10.1088/0004-637X/752/1/68).
- [9] A. Aab et al. (The Pierre Auger Collaboration), *Features of the Energy Spectrum of Cosmic Rays above 2.5×10^{18} eV Using the Pierre Auger Observatory*, Phys. Rev. Lett. **125** (2020), p. 121106, doi: [10.1103/PhysRevLett.125.121106](https://doi.org/10.1103/PhysRevLett.125.121106).
- [10] J. W. Mitchell and T. Hams, *Astrophysics and Space Instrumentation*, Handbook of Particle Detection and Imaging, Springer International Publishing (2021), pp. 761–799, doi: [10.1007/978-3-319-93785-4_23](https://doi.org/10.1007/978-3-319-93785-4_23).
- [11] W. Heitler, *The Quantum Theory of Radiation (2nd ed.)* Vol. 5, International Series of Monographs on Physics, Oxford: Oxford University Press (1944).
- [12] A. Schulz, *Measurement of the Energy Spectrum and Mass Composition of Ultra-high Energy Cosmic Rays*, PhD thesis, Karlsruher Institut für Technologie (2016).
- [13] A. Bridgeman, *Determining the Mass Composition of Ultra-high Energy Cosmic Rays Using Air Shower Universality*, PhD thesis, Karlsruher Institut für Technologie (2018), doi: [10.5445/IR/1000085504](https://doi.org/10.5445/IR/1000085504).
- [14] A. Haungs et al., *The KASCADE Cosmic-ray Data Centre KCDC: Granting Open Access to Astroparticle Physics Research Data*, Eur. Phys. J. C **78** (2018), doi: [10.1140/epjc/s10052-018-6221-2](https://doi.org/10.1140/epjc/s10052-018-6221-2).
- [15] The Pierre Auger Collaboration, *The Pierre Auger Cosmic Ray Observatory*, NIM-A **798** (2015), pp. 172–213, doi: [10.1016/j.nima.2015.06.058](https://doi.org/10.1016/j.nima.2015.06.058).

- [16] J. Stasielak (The Pierre Auger Collaboration), *AugerPrime - The upgrade of the Pierre Auger Observatory*, *International Journal of Modern Physics A* **37** (2022), p. 2240012, doi: [10.1142/S0217751X22400127](https://doi.org/10.1142/S0217751X22400127).
- [17] A. Aab et al. (The Pierre Auger Collaboration), *Observation of a large-scale anisotropy in the arrival directions of cosmic rays above 8×10^{18} eV*, *Science* **357** (2017), pp. 1266–1270, doi: [10.1126/science.aan4338](https://doi.org/10.1126/science.aan4338).
- [18] A. Aab et al. (The Pierre Auger Collaboration), *Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV*, *Phys. Rev. D* **90** (2014), p. 122005, doi: [10.1103/PhysRevD.90.122005](https://doi.org/10.1103/PhysRevD.90.122005).
- [19] G. Silli et al. (The Pierre Auger Collaboration), *Performance of the 433 m surface array of the Pierre Auger Observatory*, *PoS ICRC2021* (2021), p. 224, doi: [10.22323/1.395.0224](https://doi.org/10.22323/1.395.0224).
- [20] G. Bricchetto Orquera et al. (The Pierre Auger Collaboration), *The second knee in the cosmic ray spectrum observed with the surface detector of the Pierre Auger Observatory*, *PoS ICRC2023* (2023), p. 398, doi: [10.22323/1.444.0398](https://doi.org/10.22323/1.444.0398).
- [21] P. Abreu et al. (The Pierre Auger Collaboration), *The energy spectrum of cosmic rays beyond the turn-down around 10^{17} eV as measured with the surface detector of the Pierre Auger Observatory*, *Eur. Phys. J. C* **81** (2021), p. 966, doi: [10.1140/epjc/s10052-021-09700-w](https://doi.org/10.1140/epjc/s10052-021-09700-w).
- [22] A. Coleman, *Measurement of the Cosmic Ray Flux Above 100 PeV at the Pierre Auger Observatory*, PhD thesis, The Pennsylvania State University (2018).
- [23] A. Aab et al. (The Pierre Auger Collaboration), *Reconstruction of events recorded with the surface detector of the Pierre Auger Observatory*, *Journal of Instrumentation* **15** (2020), P10021, doi: [10.1088/1748-0221/15/10/P10021](https://doi.org/10.1088/1748-0221/15/10/P10021).
- [24] P. Meder, D. Schmidt, and D. Veberič, *Examination of WCD Signal Variance for Small Signals using Simulations*, Auger internal note GAP–2023–55 (2023).
- [25] R. Hiller and M. Roth, *An update on the signal accuracy using the infill array*, Auger internal note GAP–2012–12 (2012).
- [26] L. M. Bueno, P. Billoir, and I. C. Mariş, *Signal variance for the TOTd and MoPS triggers*, Auger internal note GAP–2014–35 (2014).
- [27] T. Schulz, *Enhanced UHECR Event Reconstruction by means of Sampling Lateral Distributions with Multiple Surface Sub-Detectors*, Bachelor’s Thesis, Karlsruher Institut für Technologie (2018).
- [28] P. Meder, D. Schmidt, and D. Veberič, *Revisiting LDF Likelihood*, Auger internal note GAP–2024–40 (2024).
- [29] G. Cowan, *Statistical Data Analysis*, Oxford University Press, USA (1998).
- [30] A. Coleman, P. Ghia, and M. Mostafá, *The Station Triggering Efficiency Using the New Triggers – A Parameterization Using the Doublets*, Auger internal note GAP–2018–17 (2018).
- [31] D. Veberič, private communication.
- [32] P. Meder et al., *New Lever-Arm Criteria to Fit the LDF Slope Parameters for the SD-750 Array*, Auger internal note GAP–2024–33 (2024).
- [33] J. Nishimura and K. Kamata, *The Lateral and Angular Distribution of Cascade Showers*, *Progress of Theoretical Physics* **5** (1950), pp. 899–901, doi: [10.1143/ptp/5.5.899](https://doi.org/10.1143/ptp/5.5.899).

-
- [34] J. Nishimura and K. Kamata, *On the Theory of Cascade Showers, I*, *Progress of Theoretical Physics* **7** (1952), pp. 185–192, doi: [10.1143/ptp/7.2.185](https://doi.org/10.1143/ptp/7.2.185).
- [35] K. Greisen, *Cosmic Ray Showers*, *Annual Review of Nuclear and Particle Science* **10** (1960), pp. 63–108, doi: [10.1146/annurev.ns.10.120160.000431](https://doi.org/10.1146/annurev.ns.10.120160.000431).
- [36] D. Newton, J. Knapp, and A. Watson, *The optimum distance at which to determine the size of a giant air shower*, *Astroparticle Physics* **26** (2007), pp. 414–419, doi: [10.1016/j.astropartphys.2006.08.003](https://doi.org/10.1016/j.astropartphys.2006.08.003).
- [37] P. Billoir, Pierre Auger Collaboration Mailing List, private communication.
- [38] W. H. Press et al., *Numerical recipes in C (2nd ed.): the art of scientific computing*, Cambridge University Press, USA (1992).
- [39] J. Albury, J. Bellido, and B. Dawson, *Exploring the energy threshold for full trigger efficiency of the Surface Detector with Hybrid events*, Auger internal note GAP–2018–38 (2018).
- [40] A. Schulz, *Measurement of the Energy Spectrum of Cosmic Rays between 0.1 EeV and 30 EeV with the Infill Extension of the Surface Detector of the Pierre Auger Observatory*, Diploma Thesis, Karlsruher Institut für Technologie (2012).
- [41] Q. Luce, *Lateral Distribution Function of SD-1500 m*, KA/BsAs group meeting, private communication (2020).
- [42] D. Veberič, *Lambert W function for applications in physics*, *Computer Physics Communications* **183** (2012), pp. 2622–2628, doi: [10.1016/j.cpc.2012.07.008](https://doi.org/10.1016/j.cpc.2012.07.008).
- [43] M. Stadlmaier, D. Veberič, and M. Roth, *Systematic Effects of the 5T5 Posterior Trigger*, Auger internal note GAP–2019–24 (2019).
- [44] D. Veberič et al., *Constant Intensity Cut: Unbinned Estimation of the Signal Attenuation Function*, Auger internal note GAP–2015–65 (2015).
- [45] D. Ravnani, private communication.
- [46] D. Veberič, private communication.
- [47] V. Verzi, private communication.
- [48] G. Bricchetto Orquera, private communication.
- [49] V. Novotný et al. (The Pierre Auger Collaboration), *Energy spectrum of cosmic rays measured using the Pierre Auger Observatory*, *PoS ICRC2021* (2021), p. 324, doi: [10.22323/1.395.0324](https://doi.org/10.22323/1.395.0324).
- [50] R. U. Abbasi et al. (The Telescope Array Collaboration), *The Cosmic Ray Energy Spectrum between 2 PeV and 2 EeV Observed with the TALE Detector in Monocular Mode*, *The Astrophysical Journal* **865** (2018), p. 74, doi: [10.3847/1538-4357/aada05](https://doi.org/10.3847/1538-4357/aada05).
- [51] Y. Tsunesada et al., *Joint analysis of the energy spectrum of ultra-high-energy cosmic rays as measured at the Pierre Auger Observatory and the Telescope Array*, *PoS ICRC2021* (2021), p. 337, doi: [10.22323/1.395.0337](https://doi.org/10.22323/1.395.0337).

APPENDIX A

Constant Intensity Cut (CIC) for 40°

The CICs presented in Chapter 7 used data with $\theta \leq 55^\circ$. Table A.1 shows the parameters for the CICs fitted to data with $\theta \leq 40^\circ$ [45, 46]. Future studies might benefit from this when limiting the spectrum studies to this zenith range. Figs. A.1 and A.2 show the S_{35} raw and intensity spectra, respectively. Deviations in the isotropy assumption for the bin containing $\theta = 25^\circ$ in Fig. A.1 are visible. However, the integral spectra in Fig. A.2 align well.

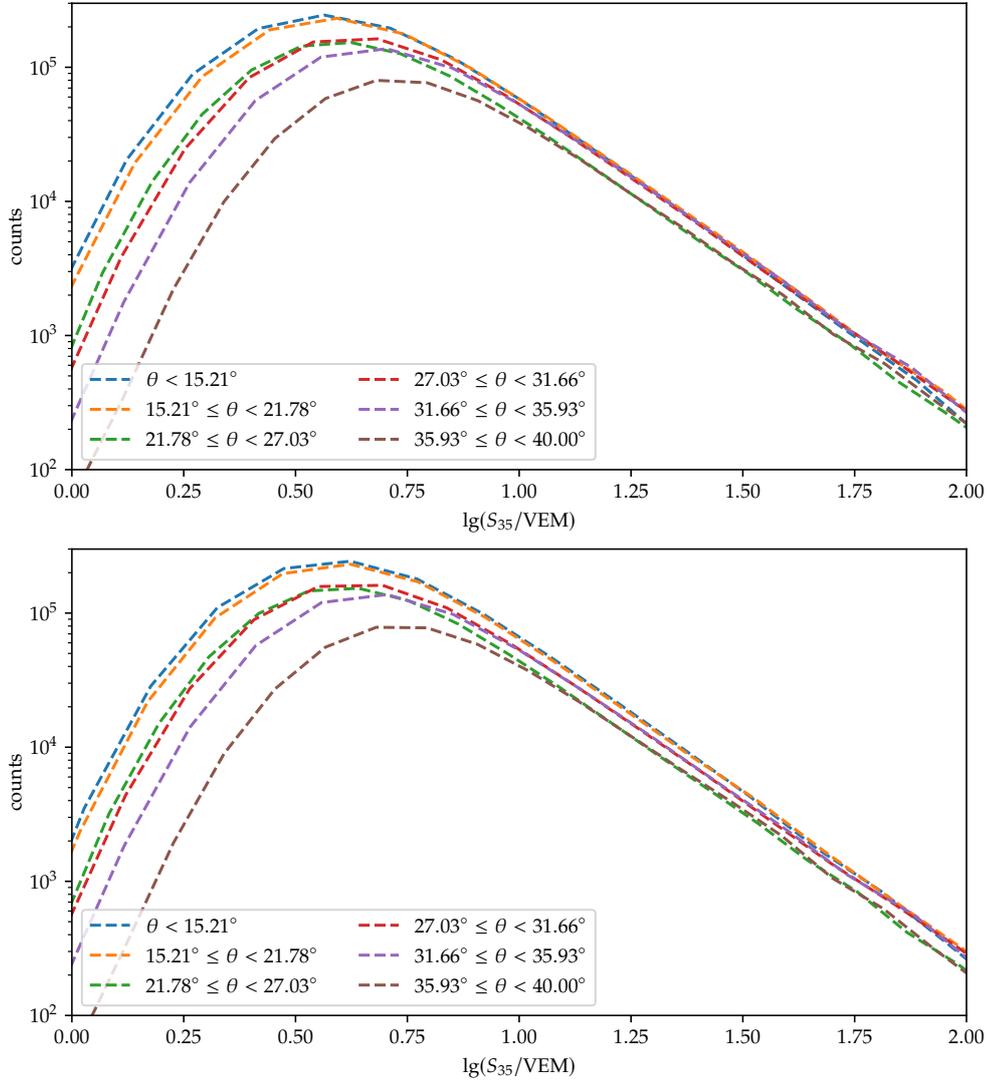
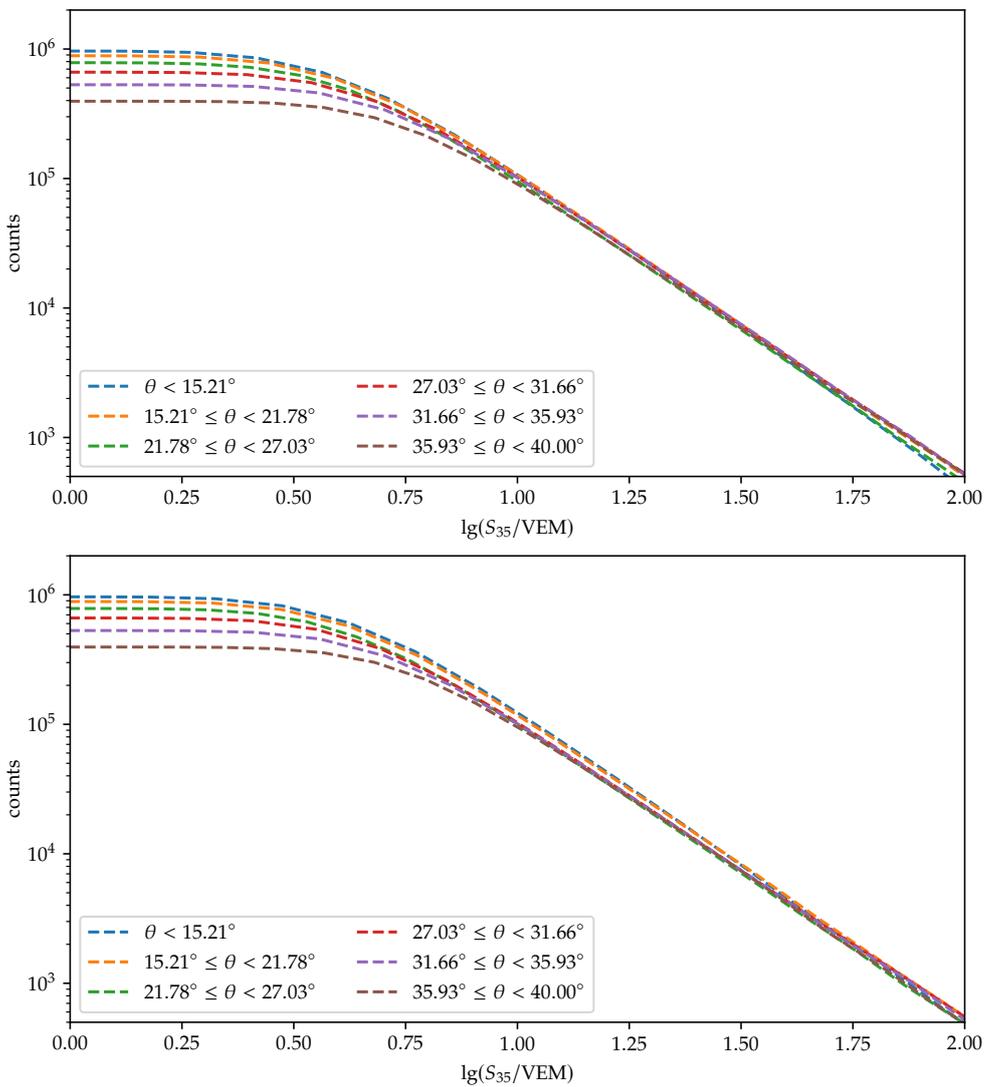


Figure A.1: Raw S_{35} spectra for bins equally distributed in $\sin^2 \theta$ using the CIC provided in Table A.1. *Top:* energy-independent CIC. *Bottom:* energy-dependent CIC.

Table A.1: Parameters for the attenuation functions in Eqs. (7.3) and (7.4) for $\theta_{\max} = 40^\circ$.

	energy-independent	energy-dependent		
		$k_{\alpha,0}$	$k_{\alpha,1}$	$k_{\alpha,2}$
a	1.752	1.600	-1.299	1.832
b	-1.580	-1.088	11.965	-24.583
c	0.033	-2.992	-27.658	50.791


Figure A.2: Integral raw S_{35} spectra for different zenith bins equally distributed in $\sin^2 \theta$ using the CIC provided in Table A.1 with $\theta_{\max} = 40^\circ$. *Top:* energy-independent CIC. *Bottom:* energy-dependent CIC.

APPENDIX B

Energy Spectrum

B.1 Calculating the Hexagon Area for the Exposure

With the distance d between the stations, the blue hexagon in Fig. 9.2 is defined by its height $h = d/2$. Using simple geometric relations, the distance between this hexagon's two neighbouring edges is $l = d/\sqrt{3}$. Therefore, the area of the hexagon is

$$A_{\text{hex}} = 6 A_{\text{triangle}} = 6 h \frac{l}{2} = 6 \frac{d}{2} \frac{1}{2} \frac{d}{\sqrt{3}} = \frac{\sqrt{3}}{2} d^2. \quad (\text{B.1})$$

B.2 Investigating the Influence of the Full-Efficiency Threshold

The analysis presented in Section 9.3 was also performed for other zenith ranges. The results are shown in Tables B.1 to B.5. Note that these tables also show fit values for full-efficiency thresholds below the ones obtained in Section 8.3.3.

B.3 Data Used for the Energy Spectrum

The raw data used to create the energy spectra Chapter 9 is provided in Table B.6. The fits were performed by using the procedure and detector properties presented in Chapters 8 and 9 with a tool provided by Ref. [48].

Table B.1: Fit parameters for the spectrum with $0 \leq \theta \leq 30^\circ$ for different thresholds $\lg(E_{\text{full}}/\text{eV})$ of full efficiency.

parameter $\lg(E_{\text{full}}/\text{eV})$	16.9	value $\pm \sigma_{\text{stat}} \pm \sigma_{\text{bias}}$ 17.0	17.1
$\frac{J_0 \times 10^{13}}{\text{km}^{-2} \text{yr}^{-1} \text{sr}^{-1} \text{eV}^{-1}}$	$1.17 \pm 0.03 \pm 0.15$	$1.17 \pm 0.03 \pm 0.18$	$1.16 \pm 0.00 \pm 0.16$
$\lg(E_{01}/\text{eV})$	$17.05 \pm 0.03 \pm 0.04$	$17.04 \pm 0.07 \pm 0.10$	$17.06 \pm 0.03 \pm 0.07$
γ_0	$2.82 \pm 0.06 \pm 0.12$	$2.82 \pm 0.12 \pm 0.01$	$2.85 \pm 0.04 \pm 0.08$
γ_1	$3.37 \pm 0.01 \pm 0.03$	$3.37 \pm 0.01 \pm 0.02$	$3.37 \pm 0.01 \pm 0.02$
ω_{01} (fixed)	0.25	0.25	0.25
$\lg(E_{12}/\text{eV})$	$18.71 \pm 0.09 \pm 0.03$	$18.71 \pm 0.09 \pm 0.05$	$18.71 \pm 0.08 \pm 0.02$
γ_2	$2.69 \pm 0.24 \pm 0.01$	$2.69 \pm 0.24 \pm 0.02$	$2.70 \pm 0.23 \pm 0.01$
ω_{12} (fixed)	0.05	0.05	0.05

Table B.2: Fit parameters for the spectrum with $0 \leq \theta \leq 40^\circ$ for different thresholds $\lg(E_{\text{full}}/\text{eV})$ of full efficiency.

parameter $\lg(E_{\text{full}}/\text{eV})$	16.9	value $\pm \sigma_{\text{stat}} \pm \sigma_{\text{bias}}$ 17.0	17.1
$\frac{J_0 \times 10^{13}}{\text{km}^{-2} \text{yr}^{-1} \text{sr}^{-1} \text{eV}^{-1}}$	$1.08 \pm 0.00 \pm 0.12$	$1.14 \pm 0.00 \pm 0.19$	$1.12 \pm 0.01 \pm 0.16$
$\lg(E_{01}/\text{eV})$	$17.20 \pm 0.03 \pm 0.07$	$17.03 \pm 0.05 \pm 0.24$	$17.05 \pm 0.08 \pm 0.09$
γ_0	$3.10 \pm 0.01 \pm 0.10$	$2.88 \pm 0.03 \pm 0.12$	$2.82 \pm 0.05 \pm 0.01$
γ_1	$3.35 \pm 0.01 \pm 0.01$	$3.34 \pm 0.01 \pm 0.00$	$3.35 \pm 0.01 \pm 0.02$
ω_{01} (fixed)	0.25	0.25	0.25
$\lg(E_{12}/\text{eV})$	$18.59 \pm 0.06 \pm 0.10$	$18.72 \pm 0.06 \pm 0.09$	$18.70 \pm 0.06 \pm 0.04$
γ_2	$2.91 \pm 0.26 \pm 0.22$	$2.70 \pm 0.27 \pm 0.18$	$2.69 \pm 0.29 \pm 0.01$
ω_{12} (fixed)	0.05	0.05	0.05

Table B.3: Fit parameters for the spectrum with $0 \leq \theta < 20^\circ$ for different thresholds $\lg(E_{\text{full}}/\text{eV})$ of full efficiency.

parameter $\lg(E_{\text{full}}/\text{eV})$	16.9	value $\pm \sigma_{\text{stat}} \pm \sigma_{\text{bias}}$ 17.0	17.1
$\frac{J_0 \times 10^{13}}{\text{km}^{-2} \text{yr}^{-1} \text{sr}^{-1} \text{eV}^{-1}}$	$1.21 \pm 0.04 \pm 0.14$	$1.20 \pm 0.03 \pm 0.17$	$1.20 \pm 0.01 \pm 0.16$
$\lg(E_{01}/\text{eV})$	$17.07 \pm 0.03 \pm 0.02$	$17.08 \pm 0.08 \pm 0.10$	$17.08 \pm 0.04 \pm 0.33$
γ_0	$2.83 \pm 0.07 \pm 0.17$	$2.87 \pm 0.13 \pm 0.02$	$2.85 \pm 0.03 \pm 0.28$
γ_1	$3.41 \pm 0.01 \pm 0.03$	$3.41 \pm 0.02 \pm 0.01$	$3.41 \pm 0.02 \pm 0.03$
ω_{01} (fixed)	0.25	0.25	0.25
$\lg(E_{12}/\text{eV})$	$18.66 \pm 0.08 \pm 0.02$	$18.65 \pm 0.08 \pm 0.01$	$18.65 \pm 0.08 \pm 0.05$
γ_2	$2.65 \pm 0.27 \pm 0.01$	$2.65 \pm 0.27 \pm 0.00$	$2.65 \pm 0.29 \pm 0.08$
ω_{12} (fixed)	0.05	0.05	0.05

Table B.4: Fit parameters for the spectrum with $20 \leq \theta < 30^\circ$ for different thresholds $\lg(E_{\text{full}}/\text{eV})$ of full efficiency.

parameter	value $\pm \sigma_{\text{stat}} \pm \sigma_{\text{bias}}$		
$\lg(E_{\text{full}}/\text{eV})$	16.9	17.0	17.1
$\frac{J_0 \times 10^{13}}{\text{km}^{-2} \text{yr}^{-1} \text{sr}^{-1} \text{eV}^{-1}}$	$1.23 \pm 0.11 \pm 0.22$	$1.10 \pm 0.07 \pm 0.15$	$1.11 \pm 0.01 \pm 0.18$
$\lg(E_{01}/\text{eV})$	$16.97 \pm 0.05 \pm 0.08$	$17.07 \pm 0.10 \pm 0.10$	$17.06 \pm 0.16 \pm 0.14$
γ_0	$2.59 \pm 0.16 \pm 0.01$	$2.90 \pm 0.20 \pm 0.05$	$2.87 \pm 0.10 \pm 0.07$
γ_1	$3.33 \pm 0.01 \pm 0.03$	$3.34 \pm 0.01 \pm 0.02$	$3.34 \pm 0.02 \pm 0.03$
ω_{01} (fixed)	0.25	0.25	0.25
$\lg(E_{12}/\text{eV})$	$18.77 \pm 0.14 \pm 0.07$	$18.76 \pm 0.14 \pm 0.02$	$18.77 \pm 0.52 \pm 0.27$
γ_2	$2.76 \pm 0.26 \pm 0.03$	$2.75 \pm 0.26 \pm 0.01$	$2.76 \pm 0.25 \pm 0.16$
ω_{12} (fixed)	0.05	0.05	0.05

Table B.5: Fit parameters for the spectrum with $30 \leq \theta < 40^\circ$ for different thresholds $\lg(E_{\text{full}}/\text{eV})$ of full efficiency.

parameter	value $\pm \sigma_{\text{stat}} \pm \sigma_{\text{bias}}$		
$\lg(E_{\text{full}}/\text{eV})$	16.9	17.0	17.1
$\frac{J_0 \times 10^{13}}{\text{km}^{-2} \text{yr}^{-1} \text{sr}^{-1} \text{eV}^{-1}}$	$1.11 \pm 0.06 \pm 0.20$	$1.01 \pm 0.02 \pm 0.12$	$1.01 \pm 0.01 \pm 0.12$
$\lg(E_{01}/\text{eV})$	$17.00 \pm 0.04 \pm 0.10$	$17.12 \pm 0.06 \pm 0.09$	$17.11 \pm 0.12 \pm 0.09$
γ_0	$2.54 \pm 0.12 \pm 0.11$	$2.88 \pm 0.10 \pm 0.03$	$2.86 \pm 0.17 \pm 0.02$
γ_1	$3.30 \pm 0.01 \pm 0.02$	$3.31 \pm 0.01 \pm 0.01$	$3.31 \pm 0.01 \pm 0.02$
ω_{01} (fixed)	0.25	0.25	0.25
$\lg(E_{12}/\text{eV})$	$18.67 \pm 0.12 \pm 0.04$	$18.68 \pm 0.12 \pm 0.00$	$18.68 \pm 0.12 \pm 0.02$
γ_2	$2.68 \pm 0.16 \pm 0.01$	$2.69 \pm 0.16 \pm 0.00$	$2.68 \pm 0.16 \pm 0.01$
ω_{12} (fixed)	0.05	0.05	0.05

Table B.6: Raw data for the spectra presented in Chapter 9. $\lg(E/\text{eV})$ is the midpoint of the bin.

$\lg(E/\text{eV})$	$n_{\text{events}}(\theta \leq 20^\circ)$	$n_{\text{events}}(\theta \leq 30^\circ)$	$n_{\text{events}}(\theta \leq 40^\circ)$	$n_{\text{events}}(\theta \leq 55^\circ)$
16.05	419	677	753	759
16.15	2935	4510	4825	4847
16.25	14688	23085	25041	25141
16.35	49719	81298	90291	90740
16.45	119717	205643	236599	238854
16.55	207314	375201	453774	461905
16.65	272136	515470	661127	684134
16.75	277251	542387	741077	788422
16.85	230740	460708	667851	743243
16.95	164528	332610	503590	595210
17.05	104157	213038	333307	423398
17.15	63530	130309	207003	281807
17.25	37488	77302	124107	179697
17.35	21757	45172	73050	111286
17.45	12689	26349	42933	67243
17.55	7283	15308	24976	39637
17.65	4271	8980	14636	23254
17.75	2319	5022	8305	13388
17.85	1324	2906	4751	7753
17.95	811	1691	2891	4672
18.05	451	959	1641	2714
18.15	230	533	894	1472
18.25	140	300	532	933
18.35	80	195	317	518
18.45	40	106	194	318
18.55	20	61	108	186
18.65	11	32	60	99
18.75	16	28	42	63
18.85	7	15	27	49
18.95	5	11	23	38
19.05	4	7	11	19
19.15	1	2	3	8
19.25	0	1	2	4
19.35	1	2	5	11
19.45	1	3	6	6
19.55	0	0	1	1
19.65	0	1	2	2
19.75	0	0	0	0
19.85	0	0	0	0