Investigating signal fluctuations of the surface detector array of the Pierre Auger Observatory using pair tanks

Masterarbeit

zur Erlangung des akademischen Grades

Master of Science
(M.Sc.)

dem Fachbereich Physik der
Universität Siegen

vorgelegt von

Marcus Rammes

September 2008

Contents

List of Tables

List of Figures

List of Acronyms

ADC	**A**nalogue to **D**igital **C**onverter
AGN	**A**ctive **G**alactic **N**ucleus
APF	**A**erosole **P**hase **F**unction
CDAS	**C**entral **D**ata **A**cquisition **S**ystem
CLF	**C**entral **L**aser **F**acility
CMB	**C**osmic **M**icrowave **B**ackground
EA	**E**ngineering **A**rray
EAS	**E**xtensive **A**ir **S**hower
EM	**E**lectro-**M**agnetic
FADC	**F**lash **A**nalogue to **D**igital **C**onverter
FD	**F**luorescence (Telescope) **D**etector
FE	**F**ront **E**nd
FPGA	**F**ield **P**rogrammable **G**ate **A**rray
GAP	**G**iant **A**irshower **P**roject
HV	**H**igh **V**oltage
LDF	**L**ateral **D**istribution **F**unction
LDPE	**L**ow **D**ensity **P**oly**E**thylen
LIDAR	**LI**ght **D**etection **A**nd **R**anging
LsId	**L**ocal **s**tation **Id**entifier
MC	**M**onte **C**arlo (simulation)
PAO	**P**ierre **A**uger **O**bservatory
PMT	**P**hoto**M**ultiplier **T**ube
SD	**S**urface **D**etector
SN	**S**uper**N**ova
ToT	**T**ime **o**ver **T**hreshold
UHECR	**U**ltra **H**igh **E**nergy **C**osmic **R**ay
VEM	**V**ertical **E**quivalent **M**uon

1 Introduction

Man has started to observe the sky and the constellation of stars from the very beginning of his evolution, trying to get a systematic coherence into the structures of the universe.

With the invention of the first optical telescope by Johannes Kepler in 1608, the possible distances of observation were suddenly expanded. This unrevealed many details of planetary movement and heralded the beginning of modern astronomy.

During the last century, the observable range of wavelengths has been expanded rapidly; towards microwaves and radio signals as well as into the X-ray and gamma ray spectrum.

But only a small fraction of the matter existing in the universe is believed to be luminous. The hypothetic existance of dark matter will only be able to be detected by very rare decays, emitting only single photons or neutrinos of highest energy.
Besides, light is not deflected by magnetic fields. The measurement of intra- and extragalactic magnetic fields can only be performed by detecting the arrival of charged particles. The investigation of the chemical composition of particles expelled by active galactic nuclei or supernovae necessitates hadronic particles emitted by these objects.

The measurement of high and ultra-high energetic cosmic rays is a relatively young field of research. As the differential flux of cosmic particles is strongly downsizing with energy, experiments of steadily increasing dimensions have been necessary to extend the upper limit of observable energies. That is the reason why cosmic rays of highest energies cannot be measured directly, they have to be detected indirectly via their fragments they produce when traversing the atmosphere of the Earth down to the ground level.
A limited number of ground detector stations has to suffice to reconstruct the energy of the primary particle, exploiting the atmosphere as native detector volume.

Therefore, the determination of signal fluctuations of ground detector stations is important to improve the estimation of the energy spectrum of cosmic rays.

1.1 Topics of this thesis

In this thesis, signal fluctuations of the surface detector (SD) array of the Pierre Auger Observatory (PAO) will be examined using *pair tanks*, i.e. ground detector stations that are located at nearly the same position.

The thesis has five main parts of analysis:

- Studying the consistence of signal strength induced signal fluctuations with real data
- Modelization of a fluctuation model with additional zenith angle dependence
- Proposal of a preliminary model of signal fluctuations depending on the distance to the shower axis
- Cross-checking of models
- Estimation of the impact of signal fluctuations on primary energy reconstruction

The main objective of the thesis is to confirm and/or improve formerly published results regarding signal fluctuation analysis and, as far as possible, to propose theoretical models that predict the results of the analysis and to check if the possible sources of signal fluctuations are physically well understood and consistent.

The methods of analysis described in my thesis may differ from those of the originally performed publications, but will lead to similar results. The main progresses of this thesis compared to existing publications are

- Theoretical prediction and analytical description of the trigger threshold effect
- Theoretical model of signal fluctuations caused by the LDF effect
- Repetition of already performed analyses with an increased amount of statistics
- Confirmation and improvement of existing parameterizations and models

The analyses and all plots have been generated by the commonly used analysis software package ROOT (`http://root.cern.ch`) together with the usage of C++. Fits were performed using the MIGRAD technology implemented in ROOT.
The experimental data has been processed by the Auger Offline framework of version v2r4p1.

Part I
Theory

2 Extensive Air Showers (EAS)

All radiative objects in the universe (Sun, AGNs[1], SNs[2] and many more) are emitting relativistic particles. Their energy range is varying over magnitudes, from several keV up to approx. 10^{21} eV, depending on the kind of acceleration and source they are coming from.

In 1912, Viktor Hess flew in a balloon to altitudes up to 5 km and discovered that the ionization of the air strongly increases with height. Up to that time, the only source of radiation was believed to be radioactivity aroused from substances in the Earth's crust [1, p4]. Hess noted in the conclusion of a publication [2]:

> *Da der Emanationsgehalt der Luft in mehr als 1000 m Höhe doch wohl im Mittel nicht größer sein kann als an der Erdoberfläche, so ist zu schließen, daß die RaC-Teilchen der Luft nur etwa 1/20 der in der Höhe von 1000-2000 m herrschenden durchdringenden Strahlung erzeugen [...] Alle diese Tatsachen deuten darauf hin, daß ein sehr großer Teil der durchdringenden Strahlung nicht von den bekannten radioaktiven Substanzen in der Erde und der Atmosphäre herrührt.*

The only logical conclusion for him was that the increasing ionization had to stem from charged particles penetrating the atmosphere from outer space, so-called *cosmic rays*. For the discovery of cosmic rays, Hess received (together with C. D. Anderson) the Nobel Prize in 1936 [3].

In 1939, Pierre Auger measured cosmic rays the first time on ground level at two different places simultaneously, using two distant Wilson cloud chambers, once at the Jungfraujoch (elevation 3500 m) and once at Pic du Midi (elevation 2900 m) [4].

Cosmic radiation is distinguished in *primary cosmic radiation* and *secondary cosmic radiation*. Secondary cosmic radiation is just emerged by interactions of primary cosmic radiation with the atmosphere (see next section).

[1] Active galactic nuclei (AGNs) are high energetic areas around a black hole in the center of galaxies.

[2] Supernovae (SNs) are the luminous remnants of an exploded star of a mass high enough to collapse to a neutron star.

Primary cosmic radiation consists of 98% of nuclei and 2% of electrons whereas the nuclei are split into 87% of protons, 12% of α-particles and 1% of heavier nuclei [5, p231].

2.1 Interactions of primaries with the atmosphere

When nuclei of the primary cosmic radiation intrude the Earth's atmosphere, they will begin to interact with nuclei of the atmosphere via the strong interaction. In the first interaction levels, mainly hadrons will be produced as secondary particles. In inelastic $p-p$, $p-\alpha$ and $\alpha-p$ collisions, mainly charged and neutral pions are generated (a and b are integer numbers) [6, p114]:

$$
\begin{aligned}
p+p &\rightarrow p+p+a(\pi^+ + \pi^-) + b\pi^0\,, && (1)\\
p+p &\rightarrow p+n+\pi^+ + a(\pi^+ + \pi^-) + b\pi^0\,, && (2)\\
p+p &\rightarrow n+n+2\pi^+ + a(\pi^+ + \pi^-) + b\pi^0\,, && (3)\\
p+\alpha &\rightarrow p+\alpha + a(\pi^+ + \pi^-) + b\pi^0\,, && (4)\\
p+\alpha &\rightarrow p+n+\mathrm{He}^3 + a(\pi^+ + \pi^-) + b\pi^0\,, && (5)\\
p+\alpha &\rightarrow p+p+n+\mathrm{H}^2 + a(\pi^+ + \pi^-) + b\pi^0\,, && (6)\\
p+\alpha &\rightarrow 4p+n+\pi^- + a(\pi^+ + \pi^-) + b\pi^0\,, && (7)\\
p+\alpha &\rightarrow 3p+2n + a(\pi^+ + \pi^-) + b\pi^0\,, && (8)\\
p+\alpha &\rightarrow 2p+3n + a(\pi^+ + \pi^-) + b\pi^0 \quad \text{and} && (9)\\
p+\alpha &\rightarrow p+4n+2\pi^+ + a(\pi^+ + \pi^-) + b\pi^0\,. && (10)
\end{aligned}
$$

Charged pions decay mostly by

$$
\begin{aligned}
\pi^+ &\rightarrow \mu^+ + \nu_\mu \quad \text{and} && (11)\\
\pi^- &\rightarrow \mu^- + \bar{\nu}_\mu\,. && (12)
\end{aligned}
$$

with a mean lifetime of $\tau_{\pi^\pm} = (2.6033 \pm 0.0005) \cdot 10^{-8}\,\mathrm{s}$. The neutral pions immediately decay into photons (branching ratio $\Gamma_{\gamma\gamma}/\Gamma = (98.798 \pm 0.032)\%$ [7]):

$$
\pi^0 \rightarrow \gamma + \gamma\,, \qquad (13)
$$

with $\tau_{\pi^0} = (8.4 \pm 0.6) \cdot 10^{-17}\,\mathrm{s}$ [7]. The secondary muons decay via the processes

$$
\begin{aligned}
\mu^+ &\rightarrow e^+ + \nu_e + \bar{\nu}_\mu \quad \text{and} && (14)\\
\mu^- &\rightarrow e^- + \bar{\nu}_e + \nu_\mu\,. && (15)
\end{aligned}
$$

e^+/e^- pairs annihilate via the electromagnetic interaction into a photon pair, the same process may occur in the reverse direction, i.e.

$$e^+ + e^- \rightarrow \gamma + \gamma \quad \text{or} \tag{16}$$
$$\gamma + \gamma \rightarrow e^+ + e^- . \tag{17}$$

Due to their high energies, most of the muons survive down to ground level in spite of their short lifetime of $\tau_{\mu^\pm} = (2.19703 \pm 0.00004)\,\mu$s [7] by time dilatation

$$\tau = \gamma\tau_0 \tag{18}$$

with

$$\gamma = \frac{E}{m_0 c^2} = \frac{1}{\sqrt{1-\beta^2}}\,, \tag{19}$$

$\beta = v/c$ and v being the speed of the particle and c the speed of light.

All decay processes are statistical processes and therefore are subject to the decay law

$$N(t) = N_0 \cdot e^{-\lambda t} \tag{20}$$

with the decay constant $\lambda = 1/\tau$.

Besides this time dependent behavior of particle decay, there is an additional effect if the particle is not flying through vacuum but rather through a material in which additional collision processes play an important role.
The mean free path length λ for a spontaneous decay in matter can be written as [8, p11]

$$\frac{1}{\lambda} = \frac{1}{\beta\gamma c\tau_0 \varrho} \tag{21}$$

with $\varrho\,[\mathrm{g\,cm^{-3}}]$ being the matter density of the medium the particle is traversing. After a piece of slant depth dX,

$$dN = \frac{N_0}{\lambda} dX \tag{22}$$

particles of an original population N_0 will have decayed. As X is not necessarily constant, an integral form remains as solution of this equation:

$$N(\Delta X) = N_0 \exp\left(-\int \frac{dX}{\lambda}\right) . \tag{23}$$

2.2 Energy losses and scattering processes

Besides decays, all charged particles suffer energy losses when traversing a medium. The dominating process is dependent on the energy of the particle:

- **Ionization and excitation**
  The particle frequently hits atoms/nuclei of the medium and loses an amount of energy dE after crossing the medium by a distance dx (expressed in terms of *column density*; $dx = [\mathrm{g\,cm^{-2}}]$). The energy loss is described by the *Bethe-Bloch formula* [8, p16]. It is

$$-\left(\frac{dE}{dx}\right) = 4\pi N_A {r_e}^2 m_e c^2 z^2 \frac{Z}{A}\frac{1}{\beta^2}\left[\ln\left(\frac{2m_e c^2\gamma^2\beta^2}{I}\right) - \beta^2 - \frac{\delta}{2}\right] \quad (24)$$

  with

m_e	rest mass of the electron
r_e	classical electron radius
N_A	Avogadro's number
I	*Ionization constant* (approx. $I = 16 \cdot Z^{0.9\,\mathrm{eV}}$ for $Z > 1$)
δ	*density effect factor*

- **Bremsstrahlung and pair production**

  For higher energies, charged particles will additionally become subject to energy losses by *bremsstrahlung* (b_{bs}), *pair production* (b_{pp}) and *nuclear interactions* (b_{ni}) via photo-nuclear processes [8, p16].
  All these effects are contributing additively and are depending on the energy of the particle:

$$b(E) = b_{\mathrm{bs}}(E) + b_{\mathrm{pp}}(E) + b_{\mathrm{ni}}(E) \quad (25)$$

  and are assumed to be effects proportional to energy.

  Together with ionization losses ($a_{\mathrm{ion}}(E)$), we obtain an overall formula for all interaction processes mentioned above:

$$\frac{dE}{dx} = a_{\mathrm{ion}}(E) + b(E)E\,. \quad (26)$$

  The process of energy loss will continue until a minimal energy, the *critical energy* E_c, is reached. For electrons, E_c can be approximated by

$$E_c = \frac{800}{Z + 1.2} \tag{27}$$

where Z is the atomic charge of the medium traversed [8, p18]. For muons, the critical energy is $E_c \simeq 500\,\mathrm{GeV}$ [1, p139].

2.3 Sources of cosmic rays and acceleration mechanisms

There are several possible processes that can bring matter to high energies. All of these processes are responsible for accelaration of particles within a certain energy range, creating the spectrum of high energy cosmic rays.

2.3.1 Solar cosmic radiation

Our Sun emits a lot of particles, mainly photons, electrons/positrons, protons and neutrinos. The source of the highest particle flux (about $5 \cdot 10^8\,\mathrm{cm^{-2}\,s^{-1}}$ protons arrive at the Earth) is the solar wind [9, p176]. The proton energy is only a few keV so the solar wind does not play a role for the energy spectrum of high energy cosmic rays.

Much more interesting for the shape of the low end of the energy spectrum is the generation of high energy particles by solar flares. Solar flares are sudden eruptions of the Sun's chromosphere [9, p176]. Particles emitted by a solar flare can carry energies up to 50 GeV [9, p183].

2.3.2 Fermi acceleration

One theory of accelaration of particles when emitted from SN remnants is the *Fermi accelaration* [1, p63]. Particles expelled from the SN collide with interstellar clouds with inherent turbulent magnetic field. If the cloud has the relative velocity $\beta_{\mathrm{cl}} = v_{\mathrm{cl}}/c$, the energy of the particle in the coordinate system of the cloud is

$$E_0' = \gamma_{\mathrm{cl}}(E_0 + \beta_{\mathrm{cl}} p_0)\,. \tag{28}$$

with E_0 being the particle energy in the lab frame. The interaction of the particle with the magnetic field of the cloud is assumed to be totally elastic, i.e. particle energy and the amount of momentum will not change, only its direction will be reversed in this simple model. Leaving the cloud again, the particle will carry the energy

$$E_1 = \gamma_{\mathrm{cl}}(E_0' + \beta_{\mathrm{cl}} p_0') = E_0 {\gamma_{\mathrm{cl}}}^2 (1 + \beta_{\mathrm{cl}})^2\,. \tag{29}$$

The energy gain for one of such a process is then

$$\frac{\Delta E}{E} = \frac{E_1 - E_0}{E_0} = {\gamma_{\mathrm{cl}}}^2(1+\beta_{\mathrm{cl}})^2 - 1\,. \tag{30}$$

Considering the angles under which the particle enters (θ_1) and leaves (θ_2) the cloud, the average energy gain is $\frac{\Delta E}{E} \simeq 4/3{\beta_{\mathrm{cl}}}^2$.
After a number of n acceration steps, the particle energy will amount to

$$E_n = E_0\left(1+\frac{\Delta E}{E}\right)^n\,. \tag{31}$$

2.3.3 Acceleration by shock waves

Another acceleration mechanism is the acceleration of particles of the SN remnant by shock waves with velocity v_S emitted by the SN [1, p65].
The velocity of the shock wave is $v_S \simeq \frac{4}{3}v_R$ with v_R being the speed of the particle expelled from the SN remnant. If the radius of the shock wave is much larger than the gyroradius $r_g = \frac{pc}{ZeB}$ of the particle, the shock wave can be regarded as a hard plane where particles are reflected at. Transforming v_R and v_S (lab frame) into the frame of reference of the shock wave, the particles gain energy by a factor of

$$\xi \propto \frac{4}{3}\beta_S\,, \tag{32}$$

where β_S is the relative velocity $\beta = (u_1 - u_2)/c$ of the shock wave in the shock frame.

Acceleration by shock waves is much faster than the Fermi acceleration [1, p64].

2.3.4 Galactic confinement

The Milky Way has an inherent galactic magnetic field of approximately $B = 3\,\mu$G. Charged particles will be bent in the magnetic field with their Larmor radius dependent of the momentum of the particles.

For highly relativistic particles, there is a maximal energy for the particles being confined by the magnetic field, given by $E_{\max} \approx c \cdot p_{\max} \leq eBR$. In units appropriate for astroparticle physics, the maximum energy can be calculated by

$$E_{\max} = 10^5\,\mathrm{TeV}\frac{B}{3\,\mu\mathrm{G}}\frac{R}{50\,\mathrm{pc}} \tag{33}$$

with R being the size of the source. Assuming a gyroradius of $R = 5\,\mathrm{kpc}$, one obtains $E_{max} = 10^{19}\,\mathrm{eV}$. Particles with energies beyond this energy are most probably originating from extragalactic sources [10, p164].

2.4 Energy spectrum

The flux of cosmic particles arriving at the Earth has been subject to a lot of research and experiments established within the last decades. The basic quantity that has to be measured is the *differential energy spectrum*

$$j(E) = \frac{dN(E)}{dA\,d\Omega\,dE\,dt}\,[\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}\,\mathrm{GeV}^{-1}]\,. \tag{34}$$

For low energy particles ($E_0 < 15\,\mathrm{GeV}$ for protons), the geomagnetic field is an obstacle for particles reaching the surface of the Earth. The trajectories of charged particles will be bent, forcing them to move parallel to the magnetic field lines and thus parallel to the ground plane. An analytical theory for a dipole field approximation had been developed by Størmer [1, p99]. The solution expresses the particle motion in units of the *Størmer radius* $r_S = \sqrt{\mu_0 M/4\pi R}$ where M is the dipole moment of the Earth ($M \simeq 8.1 \cdot 10^{25}\,\mathrm{G\,cm}^3$). The rigidity is

$$R_S(r, \lambda_B, \theta, \varphi_B) = \left(\frac{M}{2r^2}\right)\left\{\frac{\cos^4\lambda_B}{\left[1 + (1 - \cos^3\lambda_B \sin\theta \sin\varphi_B)^{1/2}\right]^2}\right\} \tag{35}$$

where θ is the zenith angle of the particle, φ_B the azimuthal angle, measured clockwise from the direction of the magnetic south pole and λ_B the magnetic latitude.
The deflection of charged particles at energies below $15 \cdot 10^9\,\mathrm{eV}$ away from the Earth explains the flattening of the energy spectrum in Fig. 1 at the very left part of the scale.

For energies higher than $\approx 50\,\mathrm{GeV}$, charged particles are not influenced by the geomagnetic field anymore. From that point, the differential spectrum decreases with a power law [5, p233]

$$\frac{dN}{dE} \propto E^{-\gamma} \tag{36}$$

with the *spectral index* γ. The spectral index depends on the type of particle [1, p105]. The chemical composition of cosmic rays is quite well-known up to an energy of $10^5\,\mathrm{GeV}$ from balloon and satellite experiments. Beyond this energy, different particle types cannot be distinguished directly anymore due to their very low flux.

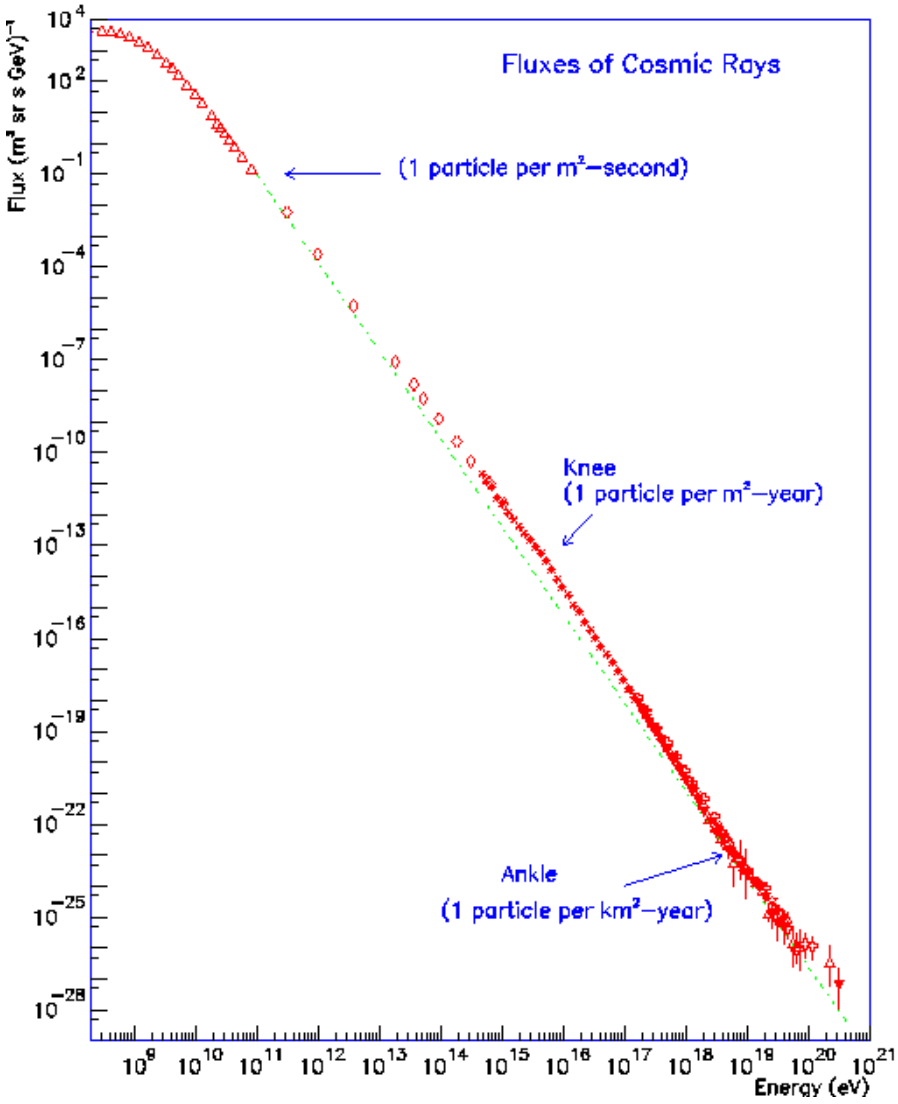


Figure 1: The energy spectrum of cosmic rays [11].

Up to an energy of approximately 10^6 GeV, cosmic rays are believed to originate from SN remnants [1, p11], the spectral index within that range is $\gamma \simeq 2.7$ constantly (see Fig. 2). Between $10^6 < E < 10^7$ GeV, the spectrum becomes steeper until it amounts to $\gamma \simeq 3$ [5, p233]. This region is known as the *knee* of the cosmic ray spectrum. The sources of cosmic rays with energies larger than those of the knee are assumed to be of extragalactic origin.

At energies of 10^{18} eV, the spectrum flattens again (called the *ankle*). For energies larger than 10^{20} eV, Greisen, Zatsepin and Kuzmin predicted a sharp cutoff in the spectrum due to interactions of the cosmic ray particles

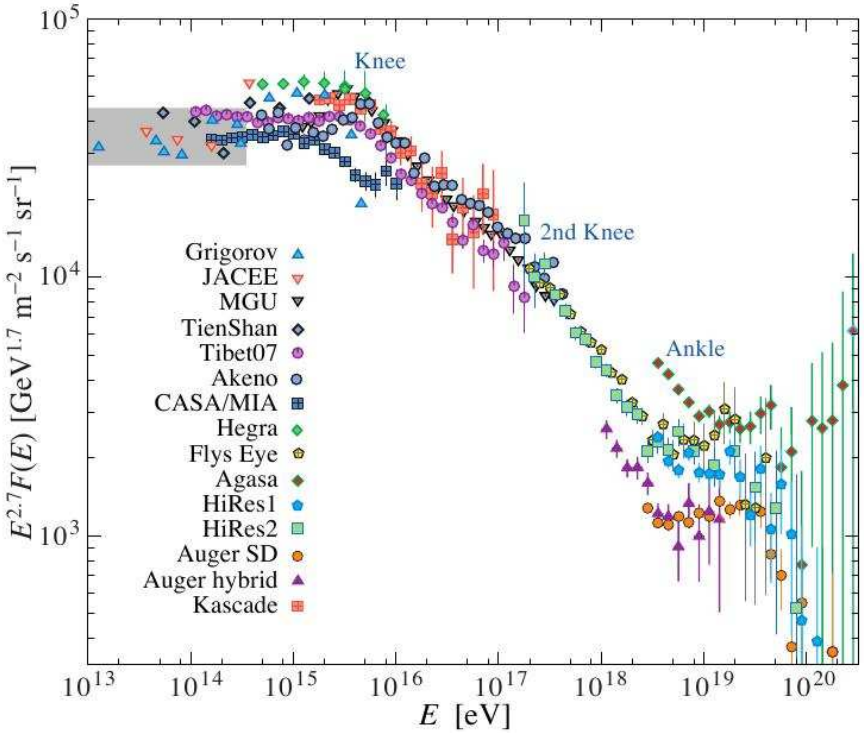


Figure 2: Energy spectrum of high and ultra-high energy cosmic rays measured by different experiments [7].

with the cosmic microwave background (CMB) by the reaction chains

$$p + \gamma_{\mathrm{CMB}} \rightarrow \Delta^+ \rightarrow p' + \pi^0 \tag{37}$$

or

$$p + \gamma_{\mathrm{CMB}} \rightarrow \Delta^+ \rightarrow n + \pi^+ \,. \tag{38}$$

Particles with energies $E > 10^{20}\,\mathrm{eV}$ would steadily lose energy by such processes until the threshold energy of $E_{\mathrm{GZK}} \approx 10^{20}\,\mathrm{eV}$ is reached [12]. This effect is known as the *GZK cutoff*.

Fig. 2 shows the energy spectrum of cosmic rays normalized to $E^{-2.7}$, measured by different EAS experiments.

2.5 Development of EAS in the atmosphere

An EAS starts with the first (hadronic) interaction (see (1)) in the higher regions of the atmosphere. The secondarily produced particles will create higher order generations of particles, this is called a *cascade*. We can seperate

an EAS into two kinds of cascades qualitatively: A *hadronic cascade* and an *electromagnetic cascade*. Besides, there is a *muonic component*. The development of an EAS is demonstrated schematically in Fig. 3.

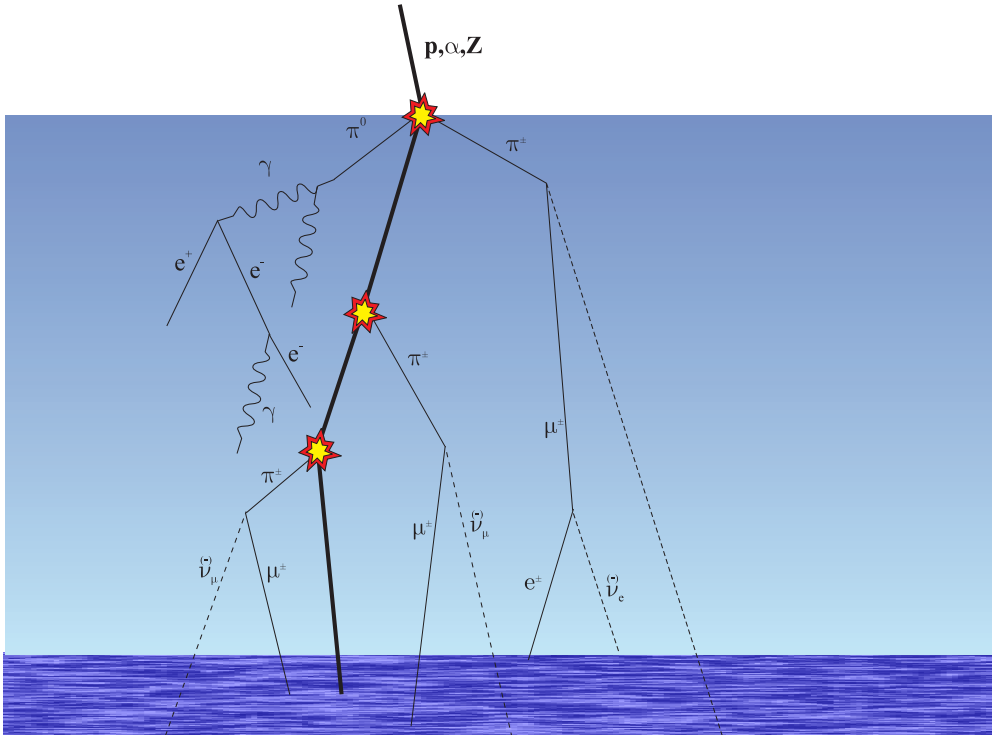


Figure 3: Schematic visualization of the development of an EAS in the atmosphere.

2.5.1 Electromagnetic cascade

The electromagnetic cascade of an EAS consists of photons, electrons and positrons. The longitudinal development of the electromagnetic cascade can be described by the *Heitler model* [1, p174]. One particle (according to the number of particle generation considered) decays after a length λ into two subparticles; both carrying half of the primary energy. This process keeps ongoing recursively, happening every radiation length $\Delta X = \lambda$, until the produced particles undershoot the *critical energy* E_c where further interactions can occur. For that reason, the maximum number of particles produced in the electromagnetic cascade is

$$N_{\max} = \frac{E_0}{E_c}\,. \tag{39}$$

$N_{\max}$ is reached after a maximum depth of

$$X_{\max} = \lambda \ln(E_0/E_c)\,. \tag{40}$$

The Heitler model is an oversimplification of the physics of an electromagnetic cascade.

2.5.2 Hadronic cascade

When the first generation of pions (according to (1)) has been generated, a pair of charged pions will produce on average two thirds of charged and one third of neutral pions in the reaction

$$\pi^+ + \pi^- \rightarrow \pi^+ + \pi^- + \pi^0\,. \tag{41}$$

The neutral pion will immediately decay and generate an electromagnetic cascade according to (13). The charged pions have a much longer lifetime and can therefore either decay or re-interact; pions of highest energy will preferably re-interact due to their large time dilatation. The decay/interaction competition of all charged mesons determines the shape of development of the hadronic cascade.

Although the process of hadronic cascades is much more complicated than that of electromagnetic ones, the Heitler model still can be used to describe the development of the hadronic cascades, too.

2.5.3 Muonic component

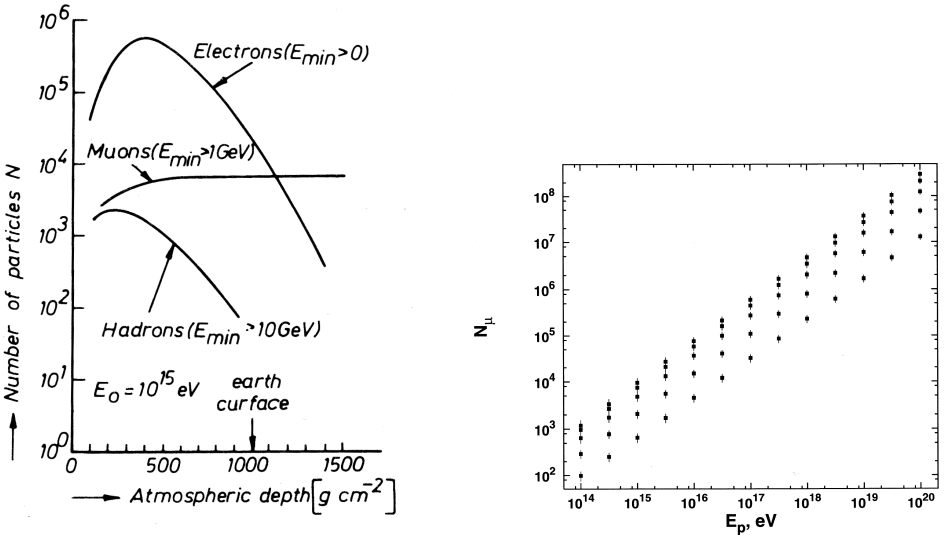
The charged pions of an EAS decay into a muon and a muon neutrino. Most of these muons have a momentum high enough to survive down to the ground level. For that reason, muons do not generate a cascade. After a short distance, the maximum number of muons is reached and kept for the remaining development of the air shower.

The number of muons can be estimated by

$$N_\mu^A = A\,[(E_0/A)/\epsilon_\pi]^\beta = A^{1-\beta} N_\mu^p \tag{42}$$

for a primary nucleus of mass A and energy E_0. N_μ^p is the number of muons for a corresponding EAS with a proton primary. β is set to $\beta = 0.85$.
By measuring the ratio N_e/N_μ at ground level, there is the possibility to reconstruct the the type of the primary nucleus. N_e can be obtained from (45) [1, pp187].

Fig. 4(a) demonstrates that the number of muons is constant for nearly the whole development of an EAS. Fig. 4(b) shows the number of muons in dependence of the primary energy.



(a) Comparison of the number of different types of particles of an EAS in dependence of the slant depth X [9, p151].

(b) Distribution of the number of muons for different primary energies. From top to the bottom the muon energy thresholds are 0.3, 1, 3, 10 and 30 GeV [1, p189].

Figure 4: The muonic component of an EAS.

2.5.4 Longitudinal shower development

The most important parameters of the longitudinal development of an EAS are the depth of shower maximum $X_{\max}$ and the number of electrons at that place $N_e^{\max}$. With a modified Heitler model for electromagnetic and hadronic component combined [1, p185], $X_{\max}$ can be described as

$$X_{\max} = X_0 \ln\left(\frac{2(1-K_{el})E_0}{(\langle m\rangle/3)\epsilon_0}\right) + \lambda_n(E_0) \tag{43}$$

if $\langle m\rangle = 12$ secondary pions are produced. The number of electrons at the shower maximum is

$$N_e^{\max} = \frac{1}{2}\frac{\langle m\rangle}{3}\frac{(1-K_{\mathrm{el}})E_0}{\epsilon_0}\,, \tag{44}$$

where E_0 is the primary energy of the nucleon and $\lambda_N = 80\,\mathrm{g\,cm^{-2}}$ the interaction length. ϵ_0 is the effective energy of the neutral pions. The factor $(1-K_{\mathrm{el}}) = 0.5$ describes the fraction of energy loss. With this choice of

values, one obtains $X_{\max} = 500\,\mathrm{g\,cm^{-2}}$ and $N_e^{\max} = 8 \cdot 10^4$ for a $10^5\,\mathrm{GeV}$ proton.

A continuous estimation of N_e dependent to X has been proposed by GAISSER [1, p186]:

$$N_e(X) = N_e^{\max} \left(\frac{X - X_1}{X_{\max} - \lambda} \right)^{\frac{X_{\max} - \lambda}{\lambda}} \exp\left(-\frac{X - X_1}{\lambda} \right) \tag{45}$$

with X_1 as depth of first interaction. This formula is well known as *Gaisser-Hillas formula*. Only electrons exhibit a $X_{\max}$, the number of muons remains more or less constant until the ground level is reached. Fig. 5 shows the shape of $N_e(X)$ for ten different air showers.

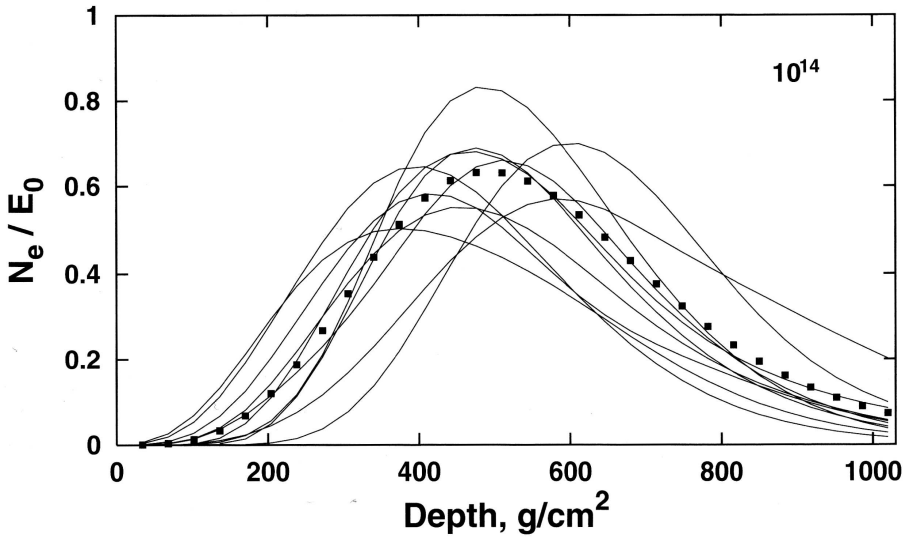


Figure 5: Longitudinal shower development for ten different EAS [1, p186].

2.5.5 Lateral shower distribution

Due to bremsstrahlung and pair production, secondary particles of an EAS are not emitted into the direction of the primary one. In addition, electrons change their direction by Coloumb scattering [1, p178]. The average angular deviation for multiple Coulomb scattering is given by

$$\left\langle \delta\theta^2 \right\rangle = \left(\frac{E_s}{E} \right)^2 \delta X \tag{46}$$

with the *effective energy* $E_s = m_e c^2 \sqrt{\frac{4\pi}{\alpha}} = 21\,\mathrm{MeV}$ [1, p179].

A theory for the lateral spread of shower particles had been developed by Greisen, Kamata and Nishimura [1, p179], therefore called NKG-like function. The particle density $\rho_e(r, X)$ in dependence of the distance r to the shower axis and the shower depth X is

$$\rho_e(r, X) = N_e(X) \frac{C(s)}{{r_M}^2} \left(\frac{r}{r_M}\right)^{s-2} \left(1 + \frac{r}{r_M}\right)^{s-9/2} \tag{47}$$

with the *Molière radius*

$$r_M = \left(\frac{E_s}{E_c}\right) X \,. \tag{48}$$

The Molière radius is a characteristic value that indicates the distance from the shower axis where the radiation has attenuated to approximately 1/4 of the intensity that is reached after one radiation length. s is a parameter that describes the "age" of the shower. $C(s)$ is a normalization coefficient derived from the definition of (47):

$$\frac{2\pi}{N_e(X)} \int_0^\infty r\rho(r)\, dr = 1 \,. \tag{49}$$

(47) is the most general form of the NKG-like function. For specific air shower experiments, a parameterization for different shower ages is meaningless as the lateral distribution of the particles of an EAS are always measured on ground level. For this purpose, (47) is adapted to the expected signals at the altitude of the ground detector stations.

3 The Pierre Auger Observatory

The Pierre Auger Observatory (PAO), located in the Pampa Amarilla (yellow prairie) near Malargüe, Western Argentina, is one of the latest and largest astroparticle experiments in the world.

The main goal of the PAO is the examination of primary cosmic particles with energies above $10^{19}\,\mathrm{eV}$ [13, p55]. The observatory reaches full acceptance at approx. $E > 3 \cdot 10^{18}\,\mathrm{eV}$ [14]. Beyond energies of $10^{19}\,\mathrm{eV}$, there are many open questions to physics concerning particle accelaration and sources of UHECRs as well as chemical composition of cosmic rays, energy spectrum and anisotropies of point sources[3]. Even exotic models such as topological defects or the existence of magnetic monopoles are subject to the investigation of UHECRs by the PAO [13, pp32].
As the cosmic particle flux (in units of $\mathrm{sr}^{-1}\,\mathrm{km}^{-2}\,\mathrm{yr}^{-1}$) decreases with a power law, ultra-high energy cosmic rays (UHECRs) cannot be measured directly by i.e. satellite or balloon experiments. UHECRs must be detected indirectly via extended air showers (EAS)[4].

The PAO consists of two independent detector systems: Four fluorescence detector (FD) telescope sites on the one hand and a surface detector (SD), containing 1600 water Čerenkov tanks located at a distance of 1.5 km to each other in a strict hexagonal shape on the other hand. The SD tanks are filled with $12\,\mathrm{m}^3$ highly purified water and equipped with 3 photomultipliers (PMTs), covering an area of approx. $3000\,\mathrm{km}^2$. This is about the size of the Saarland or Rhode Island. The layout of the PAO is shown in Fig. 6.
An online survey webpage ("Little brother") displays the status of the SD array (see Fig. 7).

The data of all SD stations and FD telescopes is transmitted to the central data acquisition system (CDAS) storage cluster situated in the central PAO office building. Fig. 10 shows a photo of the office building.

3.1 The fluorescence detector

The secondary particles of cosmic radiation excite the atoms of the atmosphere, mainly nitrogen. After the de-excitation, the atoms emit fluorescence

[3] In 2007, the PAO caused quite a stir when the collaboration published a study onto anisotropies. A general anisotropy of sources of UHECRs could be affirmed with more than 99% confidence level; moreover, a correlation with AGNs being point sources could be demonstrated for 20 out of 27 events [16].

[4] A rather new technique measuring EAS is the detection of radio signals from synchrotron radiation generated by e^+/e^- pairs produced via pair production.

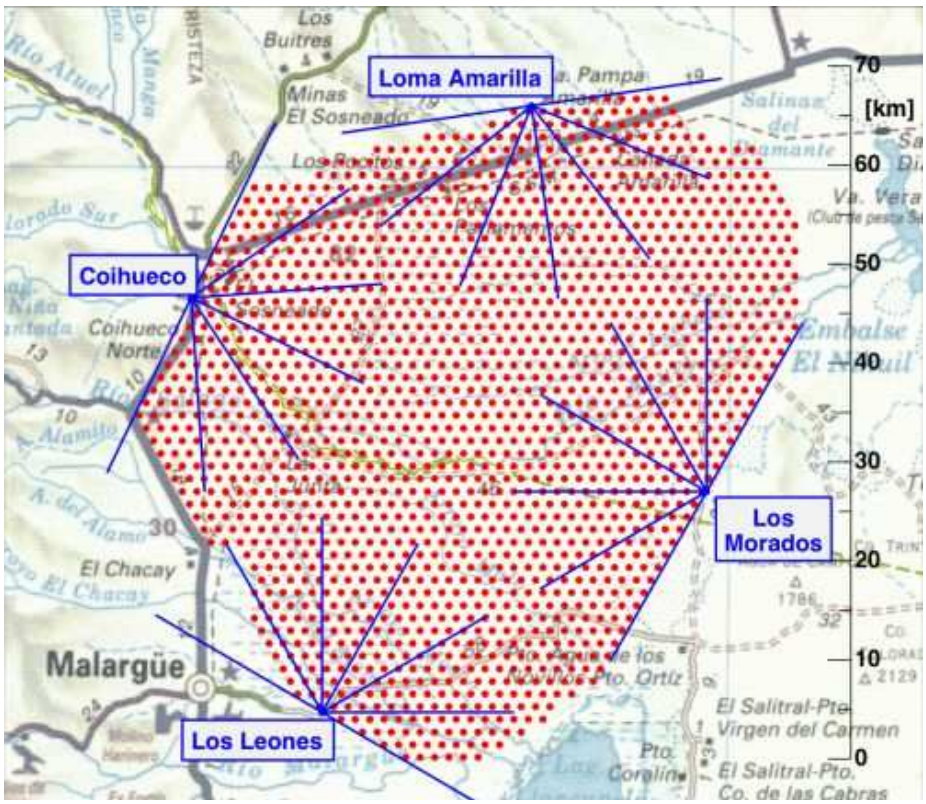


Figure 6: Map of the Southern site of the Pierre Auger Obervatory in the Pampa Amarilla, Argentina. The red dots indicate the positions of the SD stations, the blue lines are the viewing directions of the four FD eyes [15].

light. The fluorescence light can be detected by very sensitive telescopes. The fluorescence detector consists of an 20x22 array of hexagonal PMTs, the *camera*, arranged on a quasi-spherical support, located on the focal surface of a segmented mirror [17]. The PMTs are sensitive for detecting photons in the wavelength range between 300 nm and 400 nm [18]. Six telescopes are building up one *eye*. The PAO has four eyes: Los Leones, Coihueco, Los Morados and Loma Amarilla. Fig. 8 shows Los Leones with open bays. Each eye is equipped with a LIDAR system (see Sec. 3.3) that provides atmospheric and weather calibration data. Together with SD data, events can be reconstructed including SD and FD data (*hybrid events*). Hybrid events have a better angular resolution than mere SD events. The fluorescence detector can only be operated in clear, rainless, moonless nights.

3.2 The surface detector

The surface detector of the PAO is made up of about 1660 water Čerenkov tanks. Ultra-relativistic particles will produce Čerenkov light when traversing

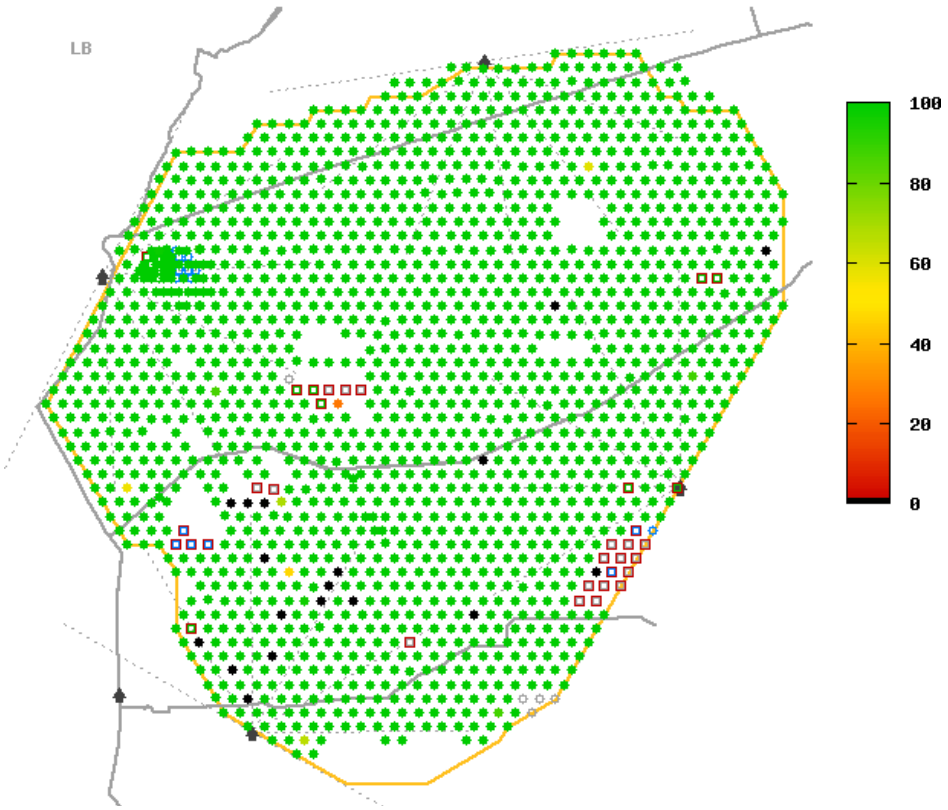


Figure 7: Status (uptime during day) of all deployed SD stations on the 21st of August in 2008 [19].

the tank. Three PMTs, mounted in a triangular placement on the top of each tank, are looking into the interior of the detector station and detect the Čerenkov photons (mainly scattered light). Each PMT has a low and a high gain output to handle weak signals as well as large ones. The PMT signals are digitalized by 10-bit-FADCs (Flash analogue-to-digital converter), so there is a range from 0 to 1023 channels available. Fig. 9 shows the pair tanks "Carmen" and "Miranda".

3.3 LIDAR

LIDAR stands for LIght Detection And Ranging. The FD telescopes of the PAO are intended to detect UV excitation radiation of nitrogen. The emitted light has to bridge a long distance through the atmosphere and is scattered by aerosoles. Furthermore, also Čerenkov photons generated by the particles of an EAS can in principle be detected.
All those background effects have to be taken into acount and are not constant in time. For this purpose, the LIDAR system periodically checks the

state of the atmosphere by scanning it with a laser system once per hour [18].

3.4 CLF

For calibration purposes, the Central Laser Facility (CLF) provides a laser system whose emitted wavelength is near to the excitation wavelength of nitrogen. The CLF shoots 50 pulses of 500 ms at a wavelength of 355 nm duration every 15 minutes vertically into the sky. Their power is 7 mJ, which corresponds to an EAS of a primary energy of $\approx 10^{20}$ eV [20], which is near the predicted limit of the GZK cutoff.
The laser shots can be seen in the FD data files. Depending on the state of the atmosphere (clouds, fog, dust, aerosoles, etc.) the beam is backscattered or diffused.

3.5 APF light source

Besides weather effects in the higher atmosphere, one has to consider dust and aerosoles on ground level due to Mie and Rayleigh scattering [21]. Objective of this light source is to determine an Aerosole Phase Function (APF).
The APF light source of the PAO is located ≈ 1.3 km southwestern of the Coihueco FD site, containing 3 Xenon flash lamps with wavelengths of 330, 360 and 390 nm to cover the full detection range of the FD PMTs.

Figure 8: The FD telescope site of "Los Leones" with open bays [22].

Figure 9: The Surface Detector stations "Carmen" and "Miranda", the first SD stations working in pair tank configuration. The FD site of Los Leones can be seen in the background [22].

Figure 10: Office building of the PAO with the CDAS transmission antenna on the left (photograph taken by myself during my FD shift stay in October 2007).

4 The surface detector tank

Each water Čerenkov tank of the PAO SD array has a diameter of 3.6 m and a height of 1.20 m. Inside the tank a sealed liner is embedded, filled with 12 000 l of highly purified water [23, p3]. A picture of a SD tank is shown in Fig. 11.

The liners of the stations consist of one outer layer of Tyvek[5], coated with a thin layer of TiO_2 pigmented low-density polyethylene (LDPE), followed by coats of clear LDPE, black LDPE and a terminal layer of clear LDPE.

The stations are equipped with three nine-inch-diameter PMTs of type XP 1805 [24], assembled in a regular triangular shape with a side length of 1.20 m, facing downwards perpendicularly into the water. The PMTs are protected by a transparent windows of polyethylene.
The tanks are designed for a lifetime of about 20 years, exposed to a temperature range from −15 °C up to +50 °C [23, p6].

The detector stations are self-containing with a solar module providing an average power of 10 W. A rechargeable battery stores charge for operation of the station at night.

An injection-molded housing outside the tank contains the tank electronics, including a front-end data processing unit, a GPS receiver, a radio transceiver and a power controller. One half of the front-end electronics had been tested at the Universities of Siegen and Wuppertal before being shipped to Argentina and mounted in the Čerenkov tanks [25].

4.1 Čerenkov radiation and its detection

When charged particles traverse a medium with a refraction index $n > 1$ with speed greater than speed of light in the medium $v > c_n$, they will produce photons emitted in a light cone with an aperture angle of

$$\cos\theta_C = \frac{1}{n\beta} \tag{50}$$

with $\beta = \frac{v}{c}$ where c is the speed of light in vacuum. The amount of energy (which is proportional to the number of Čerenkov photons) emitted by Čerenkov radiation is given by

$$\frac{dE}{dx} = \frac{(ze)^2}{c^2} \int_{\epsilon\omega > (1/\beta)^2} \omega \left(1 - \frac{1}{\beta^2 \epsilon(\omega)}\right) d\omega \tag{51}$$

[5] Tyvek 1025-BL is a strong and very diffuse reflective polyolefin of the Dupont company.

where $\epsilon(\omega)$ is the wavelength dependent dielectric coefficient [26, p737] and z the charge number of the particle. The functional relation of the dielectric coefficient is depending on the theoretical model considered (oscillator model [26, p358], anomolous dispersion and resonance absorption [26, ibid] and a model for high frequency limits [26, p362]).
For a very narrow frequency range (which is a good approximation due to the limited wavelength efficiency range of the PMTs), $\epsilon(\omega)$ can be considered as a constant. (51) can then be expressed by the constant index of refraction by substituting $n = \sqrt{\frac{\mu}{\mu_0}\frac{\epsilon}{\epsilon_0}}$ in terms of number of Čerenkov photons per unit of distance [10, p55]:

$$\frac{dN}{dx} = 2\pi\alpha z^2 \frac{\lambda_2 - \lambda_1}{\lambda_1 \lambda_2} \sin^2 \theta_C \tag{52}$$

with λ_1, λ_2 being the boundaries of the considered wavelength range.

The photons are detected by the PMTs with a certain probability depending on the wavelength, so we obtain

$$\frac{dN}{d\lambda} = 2\pi\alpha(1 - \frac{1}{{n_{\mathrm{H_2O}}}^2})\frac{\epsilon_{\mathrm{PMT}}(\lambda)}{\lambda^2} \tag{53}$$

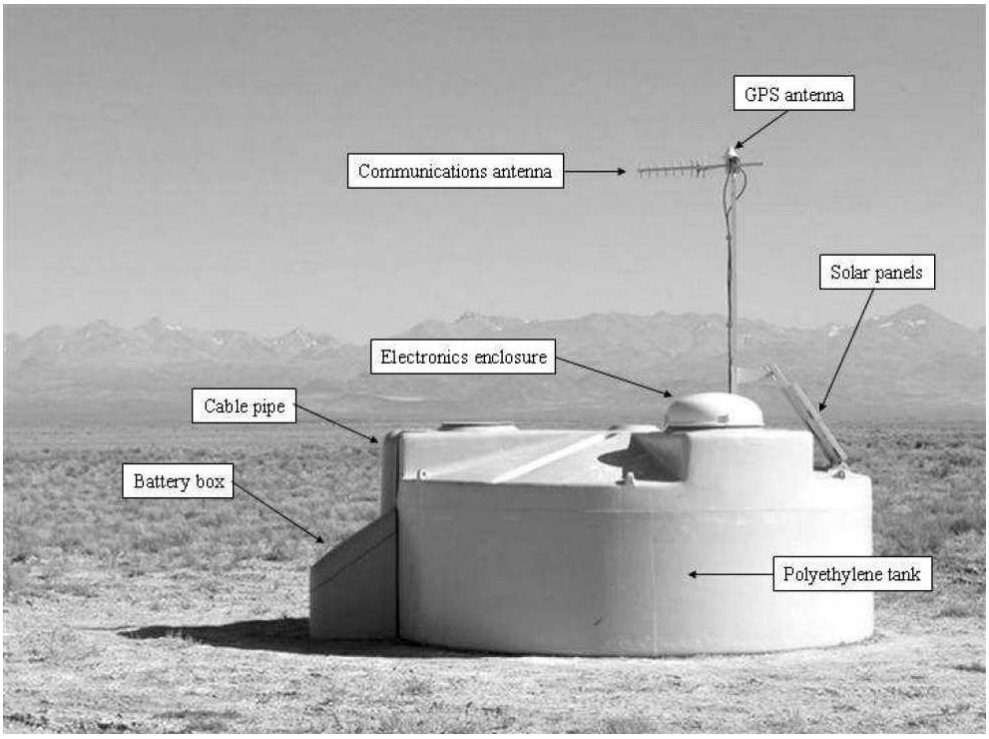


Figure 11: SD station and its most important components [23].

with the quantum efficiency $\epsilon_{\mathrm{PMT}}(\lambda)$ [27]. This is only an upper bound for estimating the number of Čerenkov photons detected in the stations. Within one absorption length L_{abs}, a certain number $dN_{\mathrm{H_2O}}$ of photons will be absorbed in the water by

$$dN_{\mathrm{H_2O}} = -\frac{c_{\mathrm{H_2O}}}{L_{\mathrm{abs}}} N(t) dt \,. \tag{54}$$

$c_{\mathrm{H_2O}}$ is the speed of light in water. Additional photons will be absorbed by the inner surface of the liner (area A) with an absorption probability C_{abs}. The liner envelopes a volume V of water. In this case, the number of absorbed photons dN_{wall} is angular dependent:

$$dN_{\mathrm{wall}} = -\frac{C_{\mathrm{abs}} A\, c_{\mathrm{H_2O}}}{2V} N(t) dt \int_0^1 \cos\alpha \, d(\cos\alpha) = -\frac{C_{\mathrm{abs}} A\, c_{\mathrm{H_2O}}}{4V} N(t) dt \,. \tag{55}$$

Taking into account the fraction of absorption of Čerenkov photons by the three PMTs, each of them covering an area of A_{PMT}, the overall absorption rate becomes

$$\frac{1}{N(t)} \frac{dN}{dt} = -c_{\mathrm{H_2O}} \left(\frac{1}{L_{\mathrm{abs}}} + \frac{C_{\mathrm{abs}} A + 3 A_{\mathrm{PMT}}}{4V} \right) . \tag{56}$$

4.2 VEM calibration

When particles produce Čerenkov light in a SD tank, the three PMTs will register a time-dependent signal. The read-out electronics is clocked with 40 MHz, so the signal can be divided into time bins with a width of 25 ns. 768 time bins are stored in a ring buffer.

The analogue signals of the PMTs are converted to channel numbers by the DACs of the FE (front-end) board according to the charge deposit in the PMTs of the (10bit-FADCs → 1024 channels numbered from 0 . . . 1023) [28, p840].
To make the integrated signals of the stations comparable to each other, the detector tanks have to be calibrated frequently to one common reference unit. This unit was chosen to be 1 VEM (Vertical Equivalent Muon). 1 VEM should be the signal measured for one centered muon traversing the tank vertically without being stopped. Actually, the unit VEM can either describe a peak of a current $I_{\mathrm{VEM}}^{\mathrm{peak}}$ induced in the PMT or the overall charge Q_{VEM} deposited by one muon.

As the signals in the PMTs are spread in time, one has to distinguish between the peak height $I_{\mathrm{VEM}}^{\mathrm{peak}}$ a VEM generates in one of the 768 recorded time

bins and the overall charge Q_{VEM} that is deposited in the PMTs. Therefore, there is a slight difference when plotting the charge (Q_{VEM}) and pulse height (I_{VEM}) histograms (Fig. 12).

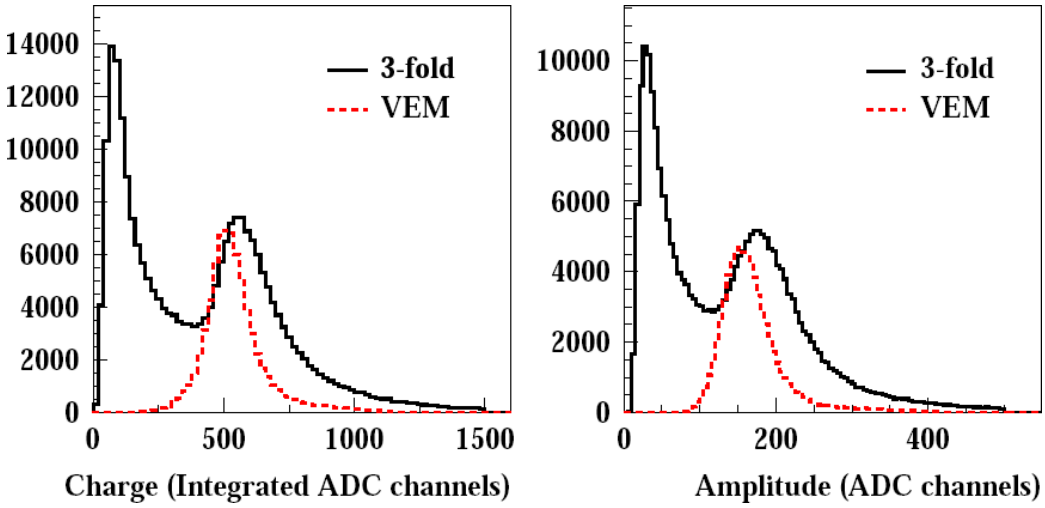


Figure 12: Charge and pulse height histogram used for VEM calibration [28].

The VEM charge Q_{VEM} for one PMT is calculated by integrating over 625 ns (corresponds to 25 time bins) from the point where its current peaks at the threshold of $I_{\mathrm{VEM}} = 1.75\,\mathrm{VEM}$ and then dividing by 1.75 [29, p813].

In Fig. 12, it can be seen that the position of the maximum measured by a reference SD tank (solid, black line) is not exactly the position of the ADC channel equivalent to 1 VEM because the angular muon distribution is convoluted with the mean muon track length (which is depending on the zenith angle either). By overlaying a muon distribution observed by an external muon telescope (dashed, red line) whose peak corresponds to exactly 1 VEM by selecting only vertical tracks, a correction factor between the peak observed by the SD tank and the actual VEM peak position can be obtained[6]. The VEM calibration procedure is performed in three steps [28, p841]:

1. The end-to-end gains of the 3 PMTs are adjusted by regulating their high voltages until the charge histograms of the PMTs match a dedicated point of the charge histogram of the reference tank. By this procedure, $I_{\mathrm{VEM}}^{\mathrm{peak}}$ is set to ADC channel no. 50 approximately[7].

[6] In [33], this correction factor is declared as 1.1, i.e. the peak height divided by 1.1 corresponds to 1 VEM.

[7] End-to-end gain means that the high voltage of the 3 PMTs is adjusted so that the three seperate PMT charge histograms agree.

2. Continually perform a local calibration to determine $I_{\mathrm{VEM}}^{\mathrm{peak}}$ in units of ADC channels to adjust the electronics trigger level.

3. Determine the value of $Q_{\mathrm{VEM}}^{\mathrm{peak}}$ using charge histograms and apply the conversion from $Q_{\mathrm{VEM}}^{\mathrm{peak}}$ to 1 VEM to obtain a conversion from the integrated PMT signals to VEM units.

Besides, a frequent online calibration is performed. The high voltage is not changed, only the channel number that corresponds to $I_{\mathrm{VEM}}^{\mathrm{peak}}$ is adjusted by a σ-δ compensation as long as the re-calibrated channel number shows no deviation larger than 20 channels of the nominal value of 50 channels.

4.3 The trigger system

The trigger system of the surface detector of the PAO consists of 5 sequential steps [30]. The T1 and T2 trigger levels are implemented harware sided in the tank electronics (programmed in the FPGA (Field Programmable Gate Array) of the front-end card). The T3 level (first offline level) decides whether the event is written to the CDAS database. T4 and T5 are pure offline triggers.

- **T1 trigger:** The T1 trigger has two different modes. The first mode is called *Time over Threshold* (ToT) and requires at least 13 bins in a 120 bins window (equivalent to a time window of 3 μs) with signal above a threshold of 0.2 $I_{\mathrm{VEM}}^{\mathrm{est}}$ with a coincidence in two of the three PMTs installed in a water Čerenkov tank. This trigger layer works at a rate of about 1.6 Hz.
  The second mode of the T1 trigger only requires one bin with a current above a threshold of 1.75 $I_{\mathrm{VEM}}^{\mathrm{est}}$ but a 3-fold coincidence of the PMTs. This mode is more noisy than the ToT trigger mode (trigger rate about 100 Hz) but allows to examine very fast events (200 ns), needed for the detection of horizontal muonic components.

- **T2 trigger:** The T2 trigger level directly promotes all events that have passed the T1 ToT trigger. If the event had triggered the T1 threshold mode, the T2 level checks for a 3-fold coincidence with a minimum current of 3.2 $I_{\mathrm{VEM}}^{\mathrm{est}}$.

- **T3 trigger:** The T3 trigger level has two different trigger conditions. The first one requires a coincidence of 3 tanks which have passed the T2 ToT trigger condition (3ToT). Since the T2 ToT already has a very low background, the 3ToT condition selects mostly physical events. The event rate of this T3 trigger is about 600 per day.
  The second T3 trigger condition requires a four-fold coincidence of any

T2 with a "moderate" compactness requirement (among the 4 fired tanks, one can be up to 6 km away from the others if compatible to the $\Delta t = d/c$ timing condition). This trigger is especially needed to detect very inclined showers that generate fast signals and have a wide topological pattern. This T3 selects about 400 events per day but only 2% of them are real showers.

- **T4 trigger (physical trigger):** The T4 physical trigger is implemented to let pass only real physical air showers from T3 data. In a first step, events with an estimated zenith angle of $\theta < 60°$ are selected. The T4 trigger is dual and requires either a compact 3ToT or a compact configuration of any locally triggered tanks where at least one fired station has 3 triggered tanks out of its first six neighbors, called 4C1 configuration. The tanks satisfying the 3ToT or 4C1 condition must have their trigger time compatible with speed of light (with a tolerance of $\Delta t = \pm 200$ ns to keep very inclined showers). With the 3ToT condition of the T4 trigger, less than 5% of all showers below 60° are lost. The 4C1 configuration rescues the 5% of showers below 60° that would get lost by the 3ToT level and also selects low energy events above 60°. 99% of the events that pass the two T4 triggers can be reconstructed.
  Fig. 13 shows the geometrical layout of the 3ToT and the 4C1 configuration.

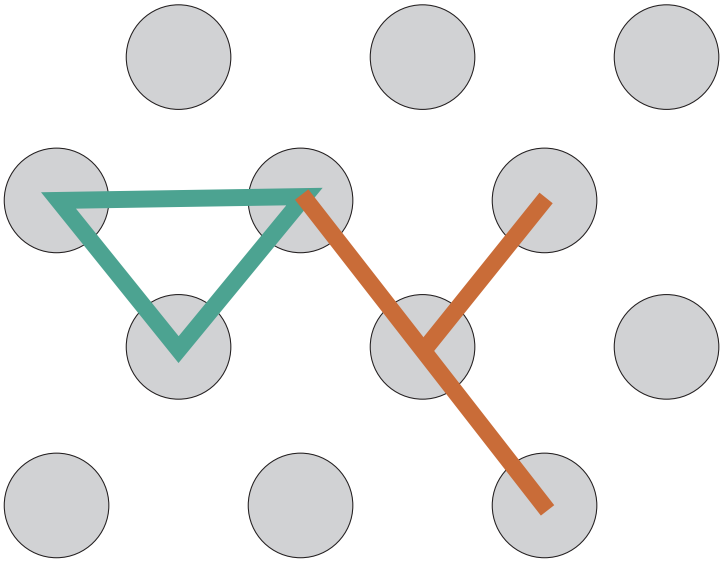


Figure 13: 3ToT (green) and 4C1 (red) T3 trigger configuration.

- **T5 trigger (quality trigger):** The T5 quality trigger ensures that only those events that can be reconstructed with a known energy and

sufficient angular accuracy to improve the quality of the measured Auger energy spectrum.
One objective of the T5 event selection is to filter events that have their shower core near the border of the SD array. In those cases parts of the shower are probably missing and the real core position could be located outside the array. This would also lead to wrong primary energy estimations.
The adopted T5 trigger requires that the tank with highest signal (*hottest tank*) must have at least 5 working tanks among its 6 closest neighbors at the time of the event detection and that the reconstructed shower core must be inside an equilateral triangle of working stations.

5 The Auger Offline framework

This section describes the most important features of the Auger Offline framework on the one hand and the official procedure of event reconstruction on the other hand.

The framework is mainly a large collection of C++ classes that represent all important features of the components of the detector systems of the PAO (FD and SD), for real data as well as for MonteCarlo simulations including full detector modelization using Geant4. The framework is based on the originally programmed CDAS code.

The execution of runnable programs is organized in seperately compiled modules, the chronology of processed modules is controlled by a configuration XML file. The particular modules can be steered by XML files, too. This allows the scientist to adjust the most important parameters of the algorithms of the modules without recompiling or modifying any source code.

Events are read by the `EventFileReader` module. It is capable to handle six different file formats [31]:

`CDAS`	Native format of T3 SD data stored in the CDAS system
`FDAS`	FD data
`Offline`	In principle CDAS data with additional information like results of event reconstruction and simulation data stored
`IoAuger`	Internal Offline format
`AIRES`	AIRES simulation output file of particles at ground (`grdpcles` file)
`CORSIKA`	CORSIKA simulation output file

The Auger Offline framework provides predefined modules that perform the official event reconstruction, detector calibration and offline triggers [32]. The most important modules for SD event reconstruction are explained in the following sections.

5.1 The `SdEventSelector` module

The `SdEventSelector` module checks stations that have been flagged as candidates and discards stations that are members of the EA. The list of stations (LsId, position, name, date of commission/decommission) can either be obtain from the `SStationList.xml` file or from a database. For pair

station treatment, the station with the higher LsId is removed by default (The removement can be exchanged against flagging the station as accidental). *Isolated stations*, i.e. stations that have no neighbor within a distance of 1800 m or only one within 500 m are flagged as accidental.

In a first step, the station data is checked for series of oscillations that could have been generated by lightnings. It can be chosen whether those events are removed from the analysis or not.

One purpose of the `SdEventSelector` is to calculate a *seed*, i.e. a tentative estimation of a shower axis $\vec{a}$ and a position of the shower core. For calculation of the seed, the three stations with the highest sum of signals are drawn on. The estimated shower plane must be compatible with speed of light, hence we have the equations

$$c(t_i - t_1) = -\vec{a} \cdot (\vec{x}_i - \vec{x_1}) \quad \text{for } i = 1 \ldots 3\,, \tag{57}$$

setting station 1 as local origin. The case $i = 1$ is trivial, the remaining conditions lead to

$$\begin{aligned} \vec{a} \cdot \vec{x}_{21} &= ct_{12} \quad \text{and} && (58)\\ \vec{a} \cdot \vec{x}_{31} &= ct_{13}\,. && (59)\\ & && (60) \end{aligned}$$

The shower axis can additionally be expressed by

$$\vec{a} = \alpha\vec{x}_{21} + \beta\vec{x}_{31} + \gamma(\vec{x}_{21} \times \vec{x}_{31})\,. \tag{61}$$

Together with (57), we obtain the linear equation system

$$\begin{pmatrix} \vec{x}_{21}^2 & \vec{x}_{21} \cdot \vec{x}_{31} \\ \vec{x}_{21} \cdot \vec{x}_{31} & \vec{x}_{31}^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = c \begin{pmatrix} t_{12} \\ t_{13} \end{pmatrix} \tag{62}$$

with the solution

$$\begin{aligned} \alpha &= c\frac{(t_{12}\vec{x}_{31}^2 - t_{13}\vec{x}_{21} \cdot \vec{x}_{31})}{D}\,, && (63)\\ \beta &= c\frac{(t_{13}\vec{x_{21}}^2 - t_{12}\vec{x_{21}} \cdot \vec{x_{31}})}{D} \quad \text{and} && (64)\\ D &= \vec{x}_{21}^2\vec{x}_{31}^2 - (\vec{x}_{21}^2 \cdot \vec{x}_{31}^2)^2\,. && (65) \end{aligned}$$

As $\vec{a}$ has to be normalized ($\vec{a}^2 \stackrel{!}{=} 1$), we must have

$$\gamma^2 = \frac{1 - |\alpha\vec{x}_{21} + \beta\vec{x}_{31}|}{|\vec{x}_{21} \times \vec{x}_{31}|^2} \,. \tag{66}$$

After estimating $\vec{a}$, the station start times are checked for time compatibility with a planar shower front, i.e. the predicted shower time t_{sh} at a position $\vec{x}$ has to be compared to the measured time of a station i with respect to its distance to $\vec{x}$:

$$t_{sh}(\vec{x}) = t_i - \vec{a} \cdot (\vec{x} - \vec{x}_i)/c \,. \tag{67}$$

For each station i, the time difference $\Delta t_i = t_i - t_{sh}$ must be in the range of

$$-1000\,\text{ns} < \Delta t_i < 2000\,\text{ns} \,. \tag{68}$$

Stations outside this time interval are flagged as accidentals.

The T4 and T5 trigger selection algorithms of the `SdEventSelector` module are described in Sec. 4.3.

5.2 The `SdPlaneFit` module

This module determines the shower plane of an event by χ^2 minimization of the time differences between the measured signal start time and the arrival time predicted by

$$ct(\vec{x}_i) = ct_0 - (\vec{x}_i - \vec{b}) \cdot \vec{a} \,, \tag{69}$$

where t_0 is the time offset, $\vec{a}$ the shower axis, t_i the measured start time and $\vec{x}_i$ the position of station i. With a time measurement uncertainty of σ_t, one has to minimize

$$\chi^2 = \frac{1}{{\sigma_t}^2} \sum_i \left[t_i - t(\vec{x}_i)\right]^2 = \frac{1}{c^2{\sigma_t}^2} \sum_i \left[ct_i - ct_0 + \vec{x}_i \cdot \vec{a}\right]^2 \,, \tag{70}$$

which can be rewritten as

$$\chi^2 = \frac{1}{{\sigma_t}^2} \sum_i \left[ct_i - ct_0 + x_i u + y_i v + z_i w\right]^2 \tag{71}$$

using the direction cosines $\vec{a} = (u, v, w)$. u, v, w are constrained by $u^2 + v^2 + w^2 - 1 = 0 \Rightarrow w = \sqrt{1 - u^2 + v^2}$. (71) is not linear, with the simplification $z_i \ll x_i,\, y_i$, the z component can be neglected an a linear χ^2 can be achieved:

$$\chi^2 = \frac{1}{{\sigma_t}^2} \sum_i \left[ct_i - ct_0 + x_i u + y_i v\right]^2 \,. \tag{72}$$

After having obtained u and v by minimization of the linear χ^2, w can be evaluated exploiting $w = \sqrt{1 - u^2 + v^2}$.

5.3 The LDFFinder module

Once the plane fit of the event has been performed, the LDFFinder module can be invoked. Amongst others this module calculates the primary energy of an event.

Generally, the predicted lateral signal is described by

$$S(r) = S_{1000} \cdot f_{LDF}(r) \tag{73}$$

with a normalized LDF f_{LDF}, i.e. $f_{LDF}(1000\,\mathrm{m}) = 1$.
The LDF fit is performed by executing sequential stages [31, p18]:

- **Stage 1: Estimation of a shower plane**
  The shower geometry estimated by the SdPlaneFit module is regarded as Stage 1
- **Stage 2: First estimation of S_{1000}**
  The signal of the station closest to a distance of 1000 m to the barycenter is taken es preliminary S_{1000} value.
- **Stage 3: Fit for S_{1000} and core position**
  In this step, S_{1000} and core location is fitted, assuming that the core is situated on a plane tangent to the reference ellipsoid containing the barycenter.
- **Stage 3.$\beta\gamma$: Fit for S_{1000}, core position, β and γ**
  Depending on the number of candidate stations, β and γ are gradually included as variable parameters subjected to the minimization process.
- **Stage 4.$\beta\gamma$: Treatment of zero-signal stations**
  This step is basically the same as Stage 3.$\beta\gamma$, but now including the estimated signal of zero-station signals.

The user can choose between two kinds of LDFs:

Modified power law

$$f_{LDF}(r) = \begin{cases} \left(\frac{r}{1000\,\mathrm{m}}\right)^{\beta+\gamma\ln(r/1000\,\mathrm{m})} & (r \geq 300\,\mathrm{m}) \\ \left(\frac{r}{1000\,\mathrm{m}}\right)^{\beta+\gamma\ln(300\,\mathrm{m}/1000\,\mathrm{m})} & (r < 300\,\mathrm{m}) \end{cases} \tag{74}$$

with the initial fit parameters

$$\beta_0 = 0.7 \arctan(6 \cdot (0.65 - \cos\theta)) - 3 \quad \text{and} \tag{75}$$
$$\gamma_0 = 0.05 \sin(8 \cdot (\cos\theta - 0.6)) - 0.5\,. \tag{76}$$

NKG-like function A slightly modified NKG function according to (47)

$$f_{LDF}(r) = \left(\frac{r}{1000\,\mathrm{m}}\right)^{\beta} \left(\frac{r + 700\,\mathrm{m}}{1700\,\mathrm{m}}\right)^{\beta+\gamma} \tag{77}$$

is assumed with initial fit parameters

$$\beta_0(\theta) = 0.9 \sec\theta - 3.3 \quad \text{and} \tag{78}$$
$$\gamma_0 = 0\,. \tag{79}$$

The default setting of using the NKG-like function is retained in this thesis.

In Fig. 14 the NKG-like LDF for a SD event (event ID: 4801049) with a reconstructed primary energy of $E_0 = (3.00 \pm 0.10) \cdot 10^{19}\,\mathrm{eV}$ and 24 candidate stations is shown as an example.
The black data points indicate the signals of the candidate stations, the red ones the signals of accidental stations. The small, blue triangles are saturated stations (lower limits of the recovered signals); the reverse, blue triangles are the upper bounds of the signals of silent stations.

Besides, the module calculates the impact point $\tilde{\vec{c}}$ of the shower from the shower axis $\vec{a}$ and the barycenter $\vec{b}$ by

$$\tilde{\vec{c}} = \vec{c} + \frac{\vec{n} \cdot (\vec{b} - \vec{c})}{\vec{n} \cdot \vec{a}} \vec{a}\,. \tag{80}$$

Two different types of minimization methods for fitting the LDF can be chosen in the `LDFFinder` module. For the minimization procedure, the module uses Minuit of the ROOT package.

5.3.1 χ^2 minimization

For χ^2 minimization, the algorithm calculates a weighted χ^2 (with weight factor σ_{S_i}, assuming that the signal fluctuates with $\sigma_{S_i} = \sqrt{S(\rho_i)}$) which leads to

$$\chi^2 = \sum_i \frac{[S_i - S(\rho_i)]^2}{{\sigma_{S_i}}^2} \tag{81}$$

with $\rho_i = |\vec{a} \times (\vec{x}_i - \vec{c})|$.

5.3.2 The maximum likelihood method

The main advantage of the maximum likelihood method compared to χ^2 minimization is the possibility of combining different signal distributions to one likelihood function which is defined as

$$L = \prod_i f_P(n_i, \mu_i) \prod_i f_G(n_i, \mu_i) \prod_i F_{\text{sat}}(n_i, \mu_i) \prod_i F_{\text{zero}}(n_i, \mu_i) \,. \tag{82}$$

From this formula we obtain the log-likelihood function

$$\log L = \sum_i \ln f_P(n_i, \mu_i) + \sum_i \ln f_G(n_i, \mu_i) + \sum_i \ln F_{\text{sat}}(n_i, \mu_i) + \sum_i F_{\text{zero}}(n_i, \mu_i) \,. \tag{83}$$

Small signals correspond to a small number of particles, thus we can assume Poissonian statistics:

$$f_P(n_i, \mu_i) = \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!} \,. \tag{84}$$

For a large number of particles, the Poissonian distribution converges towards a Gaussian one:

$$f_G(n_i, \mu_i) = \frac{1}{\sqrt{2\pi {\sigma_i}^2}} \exp\left(-\frac{(n_i - \mu_i)^2}{2{\sigma_i}^2}\right) \,. \tag{85}$$

Saturated signals represent a lower limit for the actual signal. The probability of detecting a signal larger than n_i can be obtained by integrating over all possible values larger than n_i:

$$F_{\text{sat}} = \int_{n_i}^{\infty} f_G(n, \mu_i)\, dn \,. \tag{86}$$

The signal of stations that do not trigger can be estimated by exploiting Poissonian distribution for a number of particles $n_i < n_{tresh} = 3$ (which corresponds to a signal $S_i \approx 3\,\text{VEM}$):

$$F_{\text{zero}}(n_{\text{thresh}}, \mu_i) = \sum_{n=0}^{n_{\text{thresh}}} f_P(n, \mu_i) \,. \tag{87}$$

Another important task of the `LDFFinder` module is the determination of a curved shower front by either an analytical approach or an exact fit.

5.4 The `SdRecPlotter` module

This module visualizes the most important results of the reconstructed data of a particular event such as station signals, shower axis and shower core within the station array; furthermore station timing, the LDF fit and a text box containing important statistics such as primary energy, zenith and azimuth angle are displayed. Fig. 15 gives an example of the graphical output of the module.

The upper left panel shows the shower axis and important signal information within the SD array geometry. The radii of the circles indicate the square roots of their signals, the colors the station timing, circles with border lines saturated stations; semicircles are twin stations. Red crosses mark accidental stations, hollow circles are silent stations. The red, sliced circles depict the shower core, the red line is the shower axis.

The timing information of the stations is drawn in the upper right panel. The black data points are the start times of the candidate stations, the red ones those of accidentals. The light blue points above indicate the rise times $t_{50\%}$, with the lower error bar being $t_{50\%} - t_{10\%}$ and the higher error bar $t_{90\%} - t_{50\%}$.
The light blue, dashed curves indicate the limits of station timings, the green circles the fit of the shower front curvature.

The lower left panel is the LDF fit and displays the same kind of graphics as shown in Fig. 14. The meanings of its entries have already been described in Sec. 5.3.

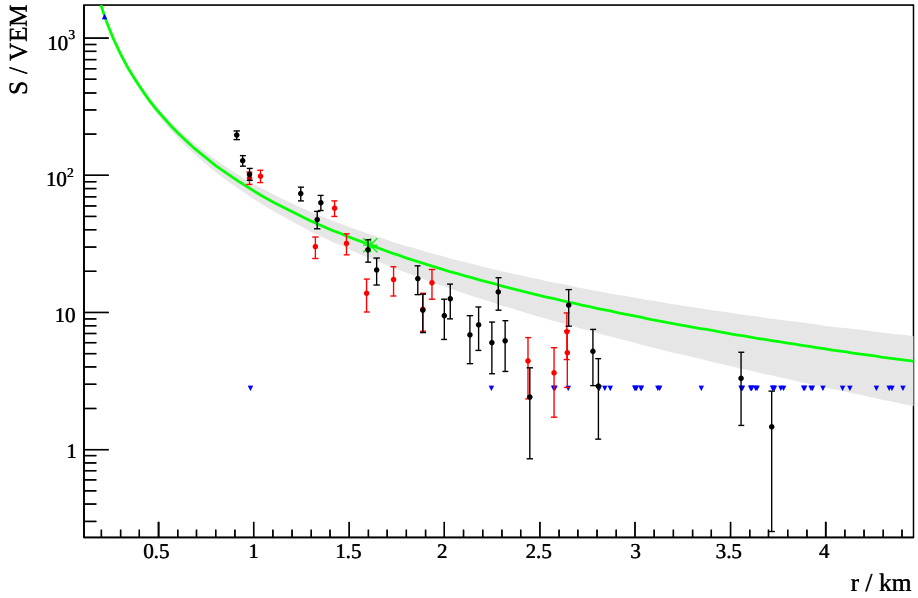


Figure 14: LDF fit (event ID: 4801049) obtained from the `LDFFinder` module.

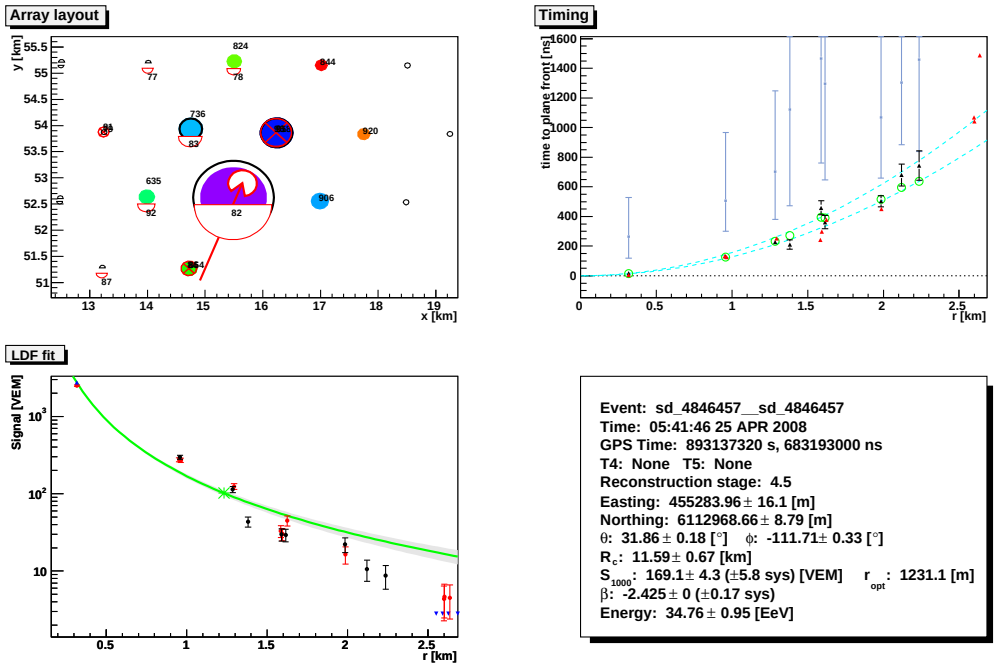


Figure 15: Example for a visualization of a reconstructed SD event generated by the `SdRecPlotter` module.

6 Measuring signal fluctuations

One of the most important parameters of an EAS that has to be obtained from SD reconstruction is the estimated primary energy of the event. Only with the reliable knowledge of the signal uncertainties, which have a significant influence onto the LDF fit (see Sec. 12), it is possible to combine the results to a differential energy spectrum. According to (81) and (83), the uncertainty of the reconstructed S_{1000} is a direct parameter of the uncertainty of the primary energy.

S_{1000} can be obtained from a LDF fit to the signals of the candidate tanks once the shower core and the shower axis have been reconstructed. For the fit procedure it is necessary to estimate the uncertainties of the particular tank signals.

In order to investigate size and origins of signal fluctuations, it lends itself to measure the particles of one EAS with two identical detector tanks at nearly the same location.
These *twin tanks* (or *pair tanks*) are positioned at a distance of 11 m to each other. Due to this short distance, the signals induced by the particles of an EAS can initially be assumed to be almost the same, thus the following *Null hypothesis* can be established:

Without fluctuations, corresponding pair tanks would measure the same signal of one particular EAS.

Of course, due to statistical and electronic fluctuations, both tanks will return slightly different signals. The reasons are [33]:

- **Tank fluctuations:** Fluctuations inside the tank due to calibration uncertainties, temporal detector instability (e.g. temperature deviations). They can be parameterized as follows:

$$\left(\frac{\sigma}{\mu}\right)^2_{\text{exp}} = \left(\frac{\sigma}{\mu}\right)^2_{\text{light}} + \left(\frac{\sigma}{\mu}\right)^2_{\#\text{pe}} + \left(\frac{\sigma}{\mu}\right)^2_{\text{gain}} \tag{88}$$

  where $\left(\frac{\sigma}{\mu}\right)^2_{\text{exp}}$ are the overall tank fluctuations, consisting of contributions stemming from the fluctuations of the number of Čerenkov photons $\left(\frac{\sigma}{\mu}\right)^2_{\text{light}}$, the generation of photoelectrons $\left(\frac{\sigma}{\mu}\right)^2_{\#\text{pe}}$ and the uncertainty of the gain amplification in the PMT anode $\left(\frac{\sigma}{\mu}\right)^2_{\text{gain}}$ [34].

- **Sampling fluctuations:** Poissonian fluctuations of the number of particles that hit the tank.
- **Muon track length:** Muons are capable to traverse the tank volume completely due to their high energy and small energy loss dE/dx (see Sec. 2.2) whereas $e^{\pm}$ and photons will be stopped/absorbed immediately. While propagating through the water, one muon generates a certain number of Čerenkov photons that is proportional to the distance the muon covered in the detector tank. An increased signal also increases signal fluctuations. This effect is depending on the zenith angle of the particle.
- **LDF effect:** The particle density at each tank can change as they are 11 m apart. The LDF is a smooth approximation to the real lateral shower distribution; a larger fluctuation of the number of particles can be expected the nearer a tank is located to the shower axis.
- **Azimuthal effect:** When not all 3 PMTs in a detector are providing a signal (e.g. one PMT is removed from CDAS because it is masked as "Raining PMT"), the measured, total signal will depend slightly on the azimuthal angle ϕ[8].

The usage of pair tanks for signal fluctuation analysis was first proposed by Alan Watson in 2001, the first analysis of this type was the measurement of coincidence rates in pair tanks [35].
The analysis of signal fluctuations began with the EA (Engineering Array) in 2003 [36] with the pair stations "Carmen" (LsId 49) and "Miranda" (LsId 64). Both stations were commissioned on 2000-01-01 and decommissioned on 2005-06-28 [37] [38].

The tank with the lower LsId is the "main" tank, i.e. its signal is considered for event reconstruction. The other one is tagged as accidental for fluctuation analysis (by default, it is discarded). The signal of the main tank will always be denoted by S_1, the signal of the partner station by S_2.
The following terminology will be used in the analysis of this thesis:

Pair tank event A pair tank event (or just pair event) is an EAS recorded by the SD CDAS with both stations of at least one tank pair having their integrated VEM traces availabe. The event and the tanks have to fulfil the specifications described in Sec. 7.

Pair tank sample A pair tank sample (or just sample) is one single tank pair (with signals available for both stations) of a pair tank event[9].

[8] The azimuthal effect will not be investigated in this thesis.
[9] One pair tank event may contain more than one sample.

Fig. 16 demonstrates the differences of signals measured in corresponding pair stations.

The red line of Fig. 16 indicates equality of the signals of the two corresponding stations of a tank pair and has an angle of 45° to the $S_1(S_2)$-axis. Hence the distance d of a data point in the plot to this bisector is $d = (S_2 - S_1) \cdot \cos 45° = \frac{S_2 - S_1}{\sqrt{2}}$. By normalizing d to the average signal by dividing it by the mean signal $\bar{S} := \frac{S_1 + S_2}{2}$, the *relative signal deviation* is obtained [39] [40] [33]:

$$\frac{\Delta S}{\bar{S}} := \sqrt{2} \cdot \frac{S_1 - S_2}{S_1 + S_2} \,. \tag{89}$$

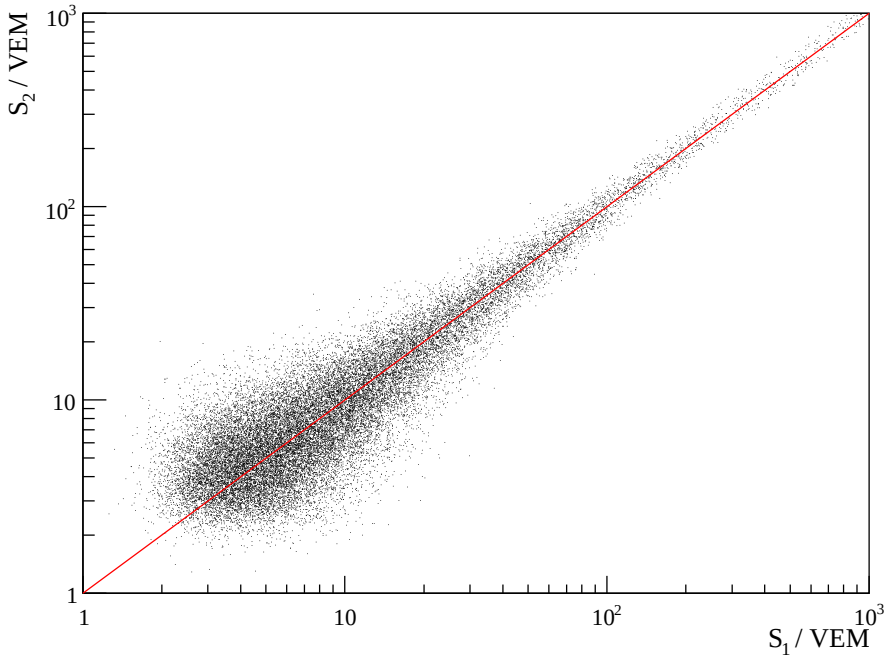


Figure 16: Distribution of integrated VEM signals in corresponding pair stations, plotted from the data set of Sec. 7 without any additional cuts applied.

Part II
Analysis

7 The data set

The analysis of signal fluctuations in this thesis is performed with Auger SD data (CDAS format) recorded from October 1st 2006 until April 30th 2008. The data is available as one file for each month and is stored in a server farm in Lyon. From this raw data set, events with at least one tank pair triggered were selected, regardless of having real air showers or not (i.e. no check on T4 or T5 trigger level) which resulted into 214083 T3 events with at least one tank pair triggered[10]. In Tab. 2 all pair stations considered for the creation of this data set are listed.
If both pair stations had been active at the event time, they were checked for availability of their integrated VEM traces. If both pair stations exhibited a signal (VEM trace), the following conditions were requested:

- Both stations had to be tagged as candidates
- Trigger data/information had to be available
- General data (such as integrated signal, rise time, etc.) must have had been recorded/reconstructed
- Rise time had to be greater than 50 ns

These event preselections were copied to the local server cluster of the High Energy Physics Department of the University of Siegen. There, non-T4 events were removed from the preselections. The T4 selection criteria were chosen as follows:

- At least three candidate stations at the beginning of the seed finding
- Usage of Bottom-Up-Selection
- No T5 trigger used
- At least 6 active stations required at the event time
- Rejection of bad periods, i.e. periods where the SD did not work properly

[10]For pair tank analysis, the signal of the partner station has to be kept by tagging the station as “accidental” in the Offline framework (by default, the station is discarded).

- Removing of lightning events, i.e. signals with certain oscillation patterns that are originated from lightnings

This T4 trigger configuration implements the shower cuts proposed in [40] on the one hand, on the other hand it makes sure that only real air shower events are taken into account for the analysis of signal fluctuations.

With these selection criteria, 16425 pair events remained for pair tank analysis providing 29218 samples (corresponds to a yield of approximately $1.8\,\frac{\text{samples}}{\text{pair event}}$).

The official cuts for event selection are separated into *shower cuts* and *tank cuts*. Tank cuts refer to the signal induced in the tank, shower cuts to the physics requirements of timing or geometry. The following cuts had been proposed in [40, p4]:

- *Tank cuts:*
  1. All three PMTs have to be reported as OK in both tanks
  2. The signal amplitude has to be smaller than $500\,\text{VEM}_{\text{peak}}$ ($S_{\text{peak}} < 500\,\text{VEM}_{\text{peak}}$)
- *Shower cuts:*
  3. The event has to be successfully reconstructed with three or more tanks exhibiting a signal (after the removal of random and isolated stations)
  4. The reconstructed core position has to be beyond 200 m from both tanks to avoid the LDF effect ($r_c > 200\,\text{m}$)
  5. The time difference between the signals in the two tanks has to be less than 200 ns ($\Delta t < 200\,\text{ns}$)
  6. Three (or more) first neighboring tanks must have their trigger times compatible to a shower front

The cuts 2, 4 and 5 were implemented into the analysis of this thesis manually. Furthermore, a 3 VEM threshold cut was applied additionally, i.e. all samples where at least one tank of the pair exhibits a signal $S < 3\,\text{VEM}$ were discarded. This cut avoids the bias that the T1 and the T2 triggers of the tanks have not exactly the same trigger threshold and ensures a clean data set also for very low signals.

The 3 VEM threshold cut together with the cuts 2, 4 and 5 are called *additional cuts* in this thesis.

All relevant values needed for the analysis were written into an ASCII file[11].

Many SD events exhibit stations with saturated signals, especially the tanks that are located near to the the shower core. Up to now, a reliable recovery procedure for saturated signals is still in progress. Therefore, saturated stations are excluded (besides a few exceptions that will be mentioned explicitely).

For distances lower than 200 m to the shower axis, deviations due to the LDF play a significant role. The $r_c < 200\,\mathrm{m}$ cut reduces this influence onto signal fluctuations.

To be able to compare the influences of all those effects, five different sub data sets were created; three ones with one single additional cut plus the 3 VEM threshold cut, one with none of the cuts and one dataset with the combination of all (see Tab. 1).
Most of the analyses in this thesis are performed using the data set with all additional cuts applied.

Additional cuts applied	Remaining samples
No cuts	29218
$\Delta t < 200\,\mathrm{ns}$ (incl. 3 VEM threshold cut)	25129
$r_c > 200\,\mathrm{m}$ (incl. 3 VEM threshold cut)	25515
$S_{peak} < 500\,\mathrm{VEM}_{peak}$ (incl. 3 VEM threshold cut)	25485
All cuts	24908

Table 1: Remaining pair signal samples of the T4 selection after applying additional cuts.

Fig. 17 indicates the positions of all tank pairs considered for the creation of the data set, noted in UTM coordinates[12], related to the reference ellipsoid WGS84 in which the PAO is located.
Fig. 18 is the upper left cutout of Fig. 17 and shows the *infill array* where the grid points of the SD array are located at shorter distances (750 m) to each other. Some stations of the infill array have a third partner tank, building up *triplets*. Due to their low interspace, it is difficult to distinguish the specific stations of the pairs/triplets in the drawing.

[11] Reading out an ASCII file for the analysis turned out to be extensively faster than parsing the corresponding ROOT data set created by the `EventFileExporter` module.

[12] UTM is a coordinate system that divides the Earth's surface into slices parallel to the axis of the Earth. Their projections onto cylinders perpendicular to the axis result in ellipsoids. Within such an *reference ellipsoid*, points are determined by planar coordinates [41].

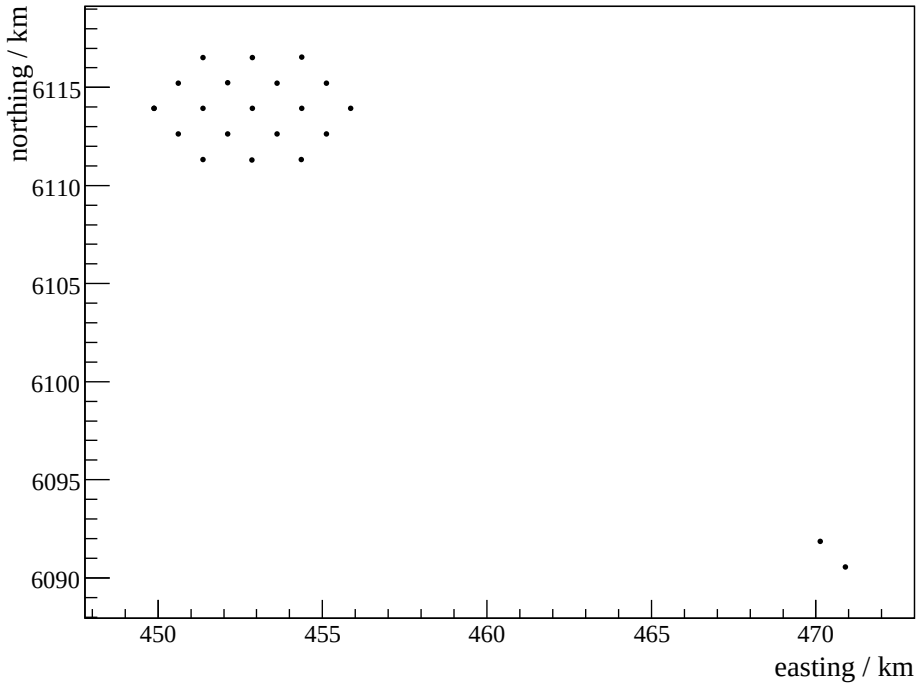


Figure 17: Positions of tank pairs and triplets considered in this analysis.

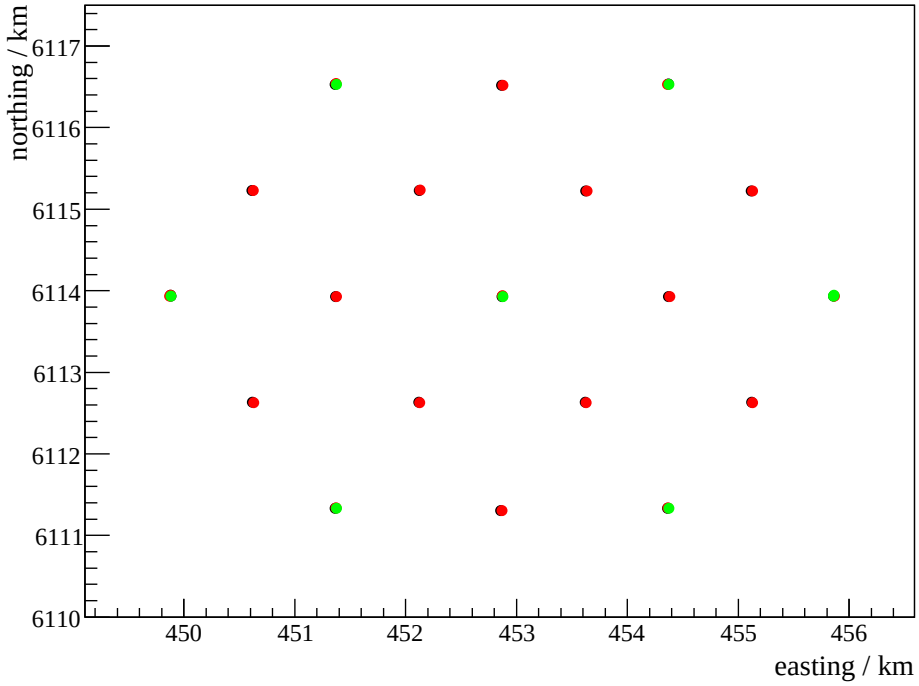


Figure 18: Infill array with twin (red) and triplet (green) stations.

pair no.	LsId	name	northing [m]	easting [m]	altitude [m]	commissioned
1	72	France	6115227.7	450615.3	1568.47	2006-09-22
	688	Toune	6115227.1	450626.13	1568.47	2004-12-07
2	73	Italy	6116519.3	452867.0	1548.15	2006-08-23
	695	Rosalia	6116518.6	452877.97	1548.15	2004-12-09
3	77	Chichino	6115225.6	453625.2	1539.56	2006-08-23
	707	Abelardo	6115225	453635.81	1539.56	2004-12-07
4	78	Frias	6115224.9	455115.93	1522.95	2006-10-06
	824	Tatanza	6115225.1	455127.27	1522.95	2005-02-18
5	80	Romeri	6115231.6	452122.2	1554.13	2006-08-23
	669	Toribio	6115232.2	452132.87	1554.13	2004-12-07
6	82	La Plata	6112630.8	455122.2	1514.22	2006-09-27
	657	Guadalupe	6112626.1	455130.42	1514.39	2004-12-21
7	83	Louisiana	6113929.6	454370.75	1528.95	2006-11-02
	736	Feche	6113929.6	454381.66	1528.95	2004-12-07
8	87	Isidorito	6111306.8	452856.65	1528.07	2006-11-02
	663	Tromen	6111305.8	452867.84	1528.07	2004-10-12
9	89	Santa Fe	6113929	451367.26	1555.97	2006-09-14
	734	Constanza	6113928.6	451378.12	1555.97	2004-12-07
10	90	Cordoba	6112631	450619.1	1553.46	2006-11-01
	651	La Salinilla	6112626.9	450629.67	1553.07	2004-10-12
11	91	San Juan	6112631.9	452118.7	1541.04	2006-11-02
	643	Borbaran	6112628.1	452128.87	1541.04	2004-10-12
12	92	Chubut	6112630.8	453618.3	1528.38	2006-11-02
	635	Hilda	6112628.4	453628.69	1528.38	2004-10-12
13	139	Dia	6091848.8	470131.53	1394.02	2004-01-01
	186	Noche	6091846.2	470121.29	1394.02	2004-01-01
14	140	Moulin	6090547.3	470893.43	1391.66	2004-01-01
	185	Rouge	6090550	470882.79	1391.66	2004-01-01
The following tank pairs are members of triplets						
15	71	Czech Republic	6116537.5	454368.9	1533.49	2006-09-01
	74	UCLA	6116528.2	454363	1533.49	2006-09-01
16	74	UCLA	6116528.2	454363	1533.49	2006-09-01
	713	Stavros	6116527.6	454373.92	1533.49	2004-12-09
17	75	Pipi	6116528.8	451364.8	1563.69	2006-11-01
	76	Lety	6116537.9	451371.1	1563.69	2006-11-02
18	76	Lety	6116537.9	451371.1	1563.69	2006-11-02
	698	Pea	6116528.2	451376.03	1563.69	2004-12-09
19	79	Alpataco	6113927.5	452868.44	1542.87	2006-09-14
	81	Ruso	6113937	452873.72	1542.87	2006-09-14
20	81	Ruso	6113937	452873.72	1542.87	2006-09-14
	119	Piuquen	6091851.6	471633.08	1391.56	2004-01-01
21	84	CSU	6111331.5	454362.06	1515.45	2006-11-02
	85	Michigan	6111341.2	454367.36	1515.45	2006-09-27
22	85	Michigan	6111341.2	454367.36	1515.45	2006-09-27
	664	El Clarin	6111331.4	454372.56	1515.45	2004-10-12
23	86	Malbec	6111331	451365.63	1539.77	2006-11-02
	88	Emy	6111340.2	451372.04	1539.77	2006-11-02
24	88	Emy	6111340.2	451372.04	1539.77	2006-11-02
	660	El Cenizo	6111330.4	451377.01	1539.77	2004-10-12
25	93	Corrientes	6113943.1	449879.54	1568.58	2006-11-02
	94	Tierra del Fuego	6113933.9	449873.65	1568.58	2006-11-02
26	94	Tierra del Fuego	6113933.9	449873.65	1568.58	2006-11-02
	710	Oye	6113933.4	449884.73	1568.58	2004-12-06
27	95	Santa Cruz	6113947.1	455860.97	1512.41	2006-10-20
	96	Neuquen	6113935.3	455856.6	1511.51	2006-10-20
28	96	Neuquen	6113935.3	455856.6	1511.51	2006-10-20
	918	Alma	6113935.2	455867.61	1511.51	2005-04-26

Table 2: Pair stations included in the analysis of this thesis.

8 Signal strength dependency

Objective of this section is to obtain a reliable model that predicts signal fluctuations in dependence of the signal itself.

8.1 Link between Poissonian and Gaussian statistics

If we want to obtain the probability of detecting a certain number of particles that hit a SD tank, we assume that this number is Poissonian distributed. A Poissonian distribution

$$P(k) = \frac{\mu^k \cdot e^{-\mu}}{k!} \tag{90}$$

returns the probability for obtaining a positive, integer value k for a given expectation value μ whereas a Gaussian distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{91}$$

is a probability density valid for arbitrary real numbers.

The standard deviation σ of the Poissonian distribution is determined by its expectation value μ by $\sigma = \sqrt{\mu}$. The main differences between the two distributions is the fact that (90) can return only positive integer numbers. Especially for small expectation values, there is an asymmetry of the Poissonian distribution because its median[13] is not equal to its expectation value.

For an increasing μ, the Poissonian distribution becomes more and more symmetric and approximates a Gaussian distribution:

$$p(k) \simeq \frac{1}{\sqrt{2\pi\mu}} \exp\left(-\frac{(k-\mu)^2}{2\mu}\right) . \tag{92}$$

Although this formula is a Gaussian distribution from the mathematical point of view, it is a one-parameter distribution like (90) with $\sigma = \sqrt{\mu}$. In its general form, μ and σ of a Gaussian distribution are independent from each other.

8.2 Zeroth order model

The signal measured in a SD station is caused by a certain integer number of particles. Therefore, we can expect a Poissonian-like behavior of the signal

[13]The median is the point of a distribution where exactly one half of the entries is situated below and the other half above that value.

uncertainties with $\sigma \approx \sqrt{S}$.
For an increasing number of particles, the Poissonian distribution becomes a Gaussian distribution in good approximation.

The average signal deviation can be calculated from the signals of the two pair tanks by[14]

$$\langle\sigma\rangle = \frac{1}{2}(\sigma_1 + \sigma_2) = \frac{1}{2}(\sqrt{S_1} + \sqrt{S_2})\,. \tag{93}$$

Defining a "relative sigma" σ_{rel} by

$$\sigma_{\mathrm{rel}} = \frac{S_1 - S_2}{\sqrt{2}\,\langle\sigma\rangle}\,, \tag{94}$$

one would expect a Gaussian distribution of signal differences normalized by their mean deviation, so the width of the Gaussian should be near to 1. The factor $\sqrt{2}$ in the denominator stems from $\Delta S = \frac{S_2 - S_1}{\sqrt{2}}$.

Furthermore, the Null Hypothesis implies $S_1 - S_2 \approx 0$, so we expect a mean of $\left\langle \frac{\Delta S}{\langle\sigma\rangle} \right\rangle \simeq 0$.

Fig. 19 shows the result of this zeroth order model. The Gaussian fit returns a mean of $\langle\sigma_{\mathrm{rel}}\rangle = (-3.3 \pm 0.5) \cdot 10^{-2}$ and a spread of $\sigma = 0.852 \pm 0.005$ ($\chi^2/dof = 214.5/34$). This implies the assumption that a mere Poissonian fluctuation model without any additional parameters overestimates the signal uncertainties with respect to the observed fluctuations. The reason is that 1 VEM does not always correspond to exactly one particle. In average, we have approximately 1 VEM $\widehat{=}$ 0.64 particles [43, p13]. This conversion factor from the number of particles to the corresponding signal value measured in units of VEM will be called p_σ from now on. In the following sections, more sophisticated models for signal fluctuations will be established to determine p_σ.

8.3 Method

Assuming a special statistical distribution of relative signal deviations would not take into account any systematics that are disregarded by the considered distribution. This would lead to a bias.
Therefore, the detector accuracy is defined as the RMS of a distribution of the relative signal deviations (89) [39] [40] [33]:

[14] The average of the two uncertainties is chosen rather than $\sigma^2 = {\sigma_1}^2 + {\sigma_2}^2$ in this case because $\sigma_{1,2}$ are no statistical fluctuations but functional values of a model.

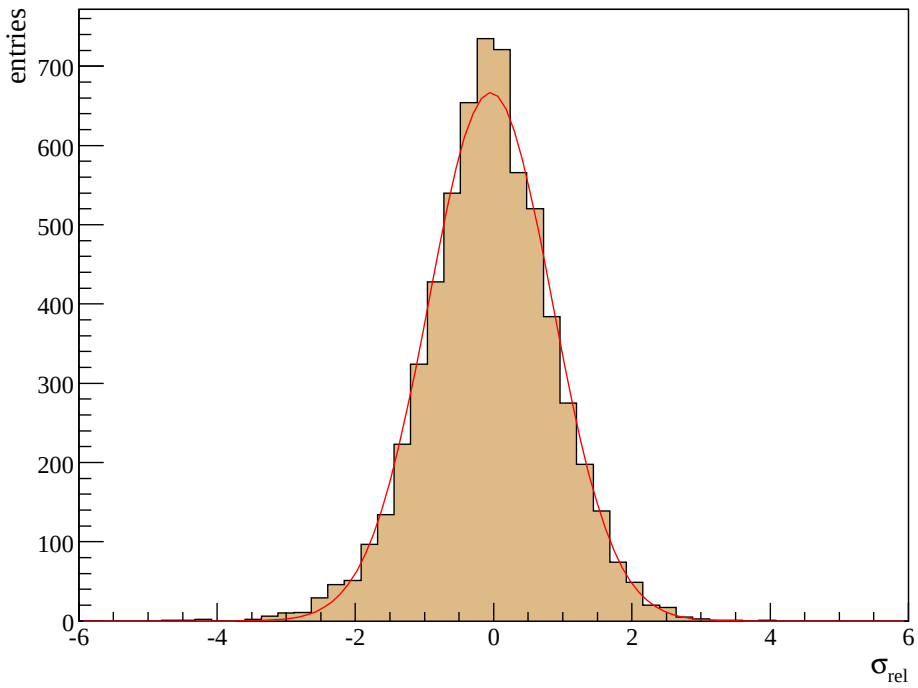


Figure 19: The zeroth order model: histogram of σ_{rel} with Gaussian fit.

$$RMS = \left[\left(\sum_{i=1}^{M} \left(\frac{\Delta S}{\bar{S}} \right)_i^2 \right) / M - \left(\left(\sum_{i=1}^{M} \left(\frac{\Delta S}{\bar{S}} \right)_i \right) / M \right)^2 \right] \cdot \frac{M}{M-1} \quad (95)$$

with M being the number of entries in that histogram. The RMS has an uncertainty, too, its variance is given by

$$V\left[\left(\frac{\sigma}{\bar{S}} \right)^2 \right] = \frac{1}{M} \left(m_4 - \frac{M-3}{M-1} m_2^2 \right) \quad (96)$$

where m_i is the i-th central moment ($m_i = \sum \left[\left(\frac{\Delta S}{\bar{S}} \right) - \left\langle \left(\frac{\Delta S}{\bar{S}} \right) \right\rangle \right]^i$) [40].

In this thesis, signal fluctuations are examined using pair tanks in a similar way as described in [36] and [40].
A 2-D histogram is filled with the relative signal deviations of all pair samples, setting the x-axis as $\bar{S} = (S_1 + S_2)/2$ and the y-axis as $\Delta S/\bar{S}$. Then this histogram is divided into N logarithmic slices of $\bar{S}$ (i.e. the N slices have thickness of $\Delta \log \bar{S}$). For each of those N slices, the projection along the $\frac{\Delta S}{\bar{S}}$-axis is built. From this 1-D histogram the RMS is calculated. The RMS is considered to be the overall *relative signal fluctuation* $\frac{\sigma}{\bar{S}}$ for that slice:

$$\left(\frac{\sigma}{\bar{S}}\right)_i := RMS\left\{\left(\frac{\Delta S}{\bar{S}}\right)_{ij}\right\}_{j=1}^{M_i}, \quad i = 1 \ldots N\,. \tag{97}$$

Fig. 20 demonstrates the principle of the procedure for obtaining $\left(\frac{\sigma}{\bar{S}}\right)$ described above.

By calculating the mean of the projection of one slice along the x-axis (that provides the values of all $\bar{S}$ within that bin[15]), the $\bar{S}$ value corresponding to the signal fluctuation value in the slice can be obtained.

Finally, the squares of the RMS values of the y-projections ($\left(\frac{\sigma}{\bar{S}}\right)^2$) and their corresponding means of the x-projections ($\bar{S}$) can be drawn into a diagram.

Fig. 20 visualizes the procedure described above. In Fig. 21, an example for a signal fluctuation plot is shown together with the y-projection histograms of the individual slices[16]. For the example, all additonal cuts have been removed. The means $\left\langle\frac{\Delta S}{\bar{S}}\right\rangle$, the number of entries and the RMS values of the projections are listed in Tab. 3. The mean values are ≈ 0 in all distributions as they have been predicted by the zeroth error model.

Fig. 22 shows the result for $\left(\frac{\sigma}{\bar{S}}\right)^2$ vs. $\bar{S}$ of the whole data set with the cuts of Sec. 7 applied individually (together with the 3 VEM threshold cut, except Fig. 22(a) and Fig. 22(f)). The signal range $\bar{S}$ has been divided into 19 logarithmic bins along the $\bar{S}$-axis in the range of $1.0 \ldots 4601$ VEM.

Fig. 22 and 21 are the only exceptions in this section where not all additional cuts have been implemented.

8.4 Poissonian-like fluctuation model

Assuming a Poissonian-like behavior ($\sigma \propto \sqrt{\bar{S}}$) of the squared signal fluctuations $\left(\frac{\sigma}{\bar{S}}\right)^2$, the signal fluctuation model

$$\left(\frac{\sigma}{\bar{S}}\right)^2 = \frac{{p_\sigma}^2}{\bar{S}} \tag{98}$$

can be fitted to the data points obtained as described in Sec. 8.3. This model will be denoted as *Poissonian-like model* or *Poisson fit* henceforward.

[15] The usage of $\bar{S}$ once as average of the signals of one sample $\bar{S} = (S_1 + S_2)/2$ and once as mean of all signals in a histogram is ambiguous but the correct meaning should turn out clearly from the current context.

[16] The signal fluctuations are always plotted as squared terms to intensify their slope vs. $\bar{S}$ for better visibleness.

Fig. 23 shows the signal fluctuations with all official cuts applied together with a Poisson fit in the range of 10.0 VEM $\leq \bar{S} \leq$ 80.0 VEM. We obtain

$$\left(\frac{\sigma}{\bar{S}}\right)^2 = \frac{(0.903 \pm 0.013)^2}{\bar{S}} \quad (\chi^2/dof = 13.7/4)\,. \tag{99}$$

Although the numerator σ and the demoninator $\bar{S}$ of (97) are not independent values that can be handled separately, they are regarded as individual variables. We often find the equivalent notation

$$\sigma^2 = {p_\sigma}^2 \cdot \bar{S} \quad \text{or even} \quad \sigma = p_\sigma \cdot \sqrt{\bar{S}} \tag{100}$$

instead of (98) [48] [40] [46] [39].

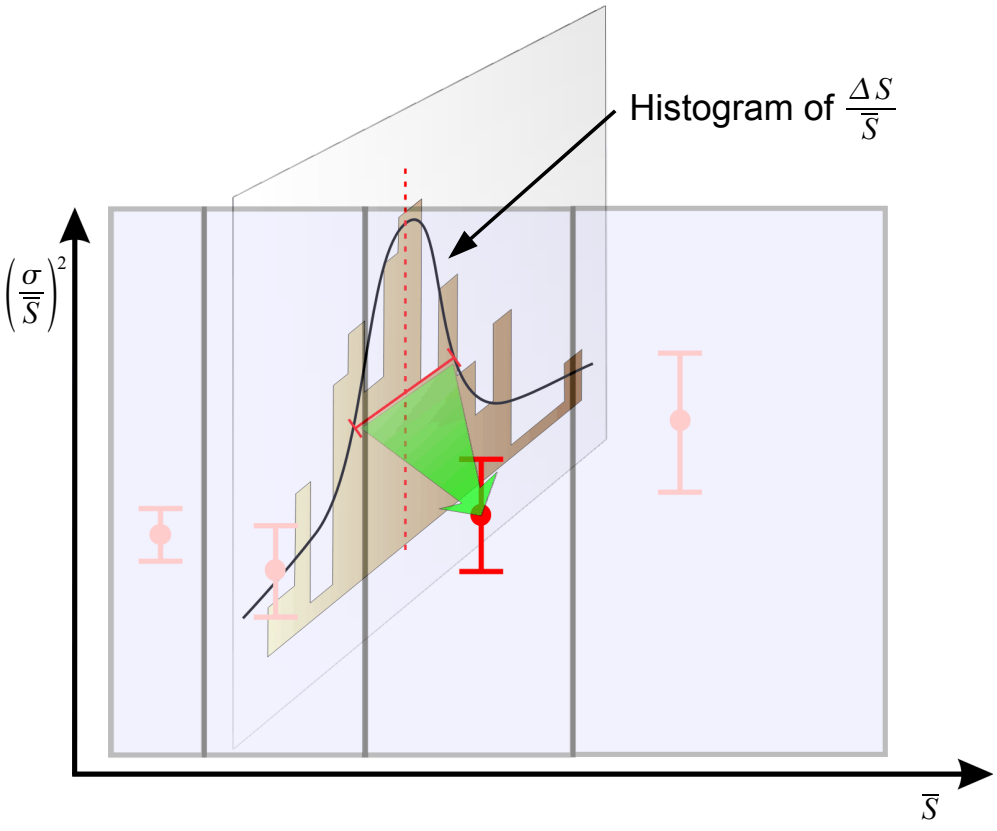
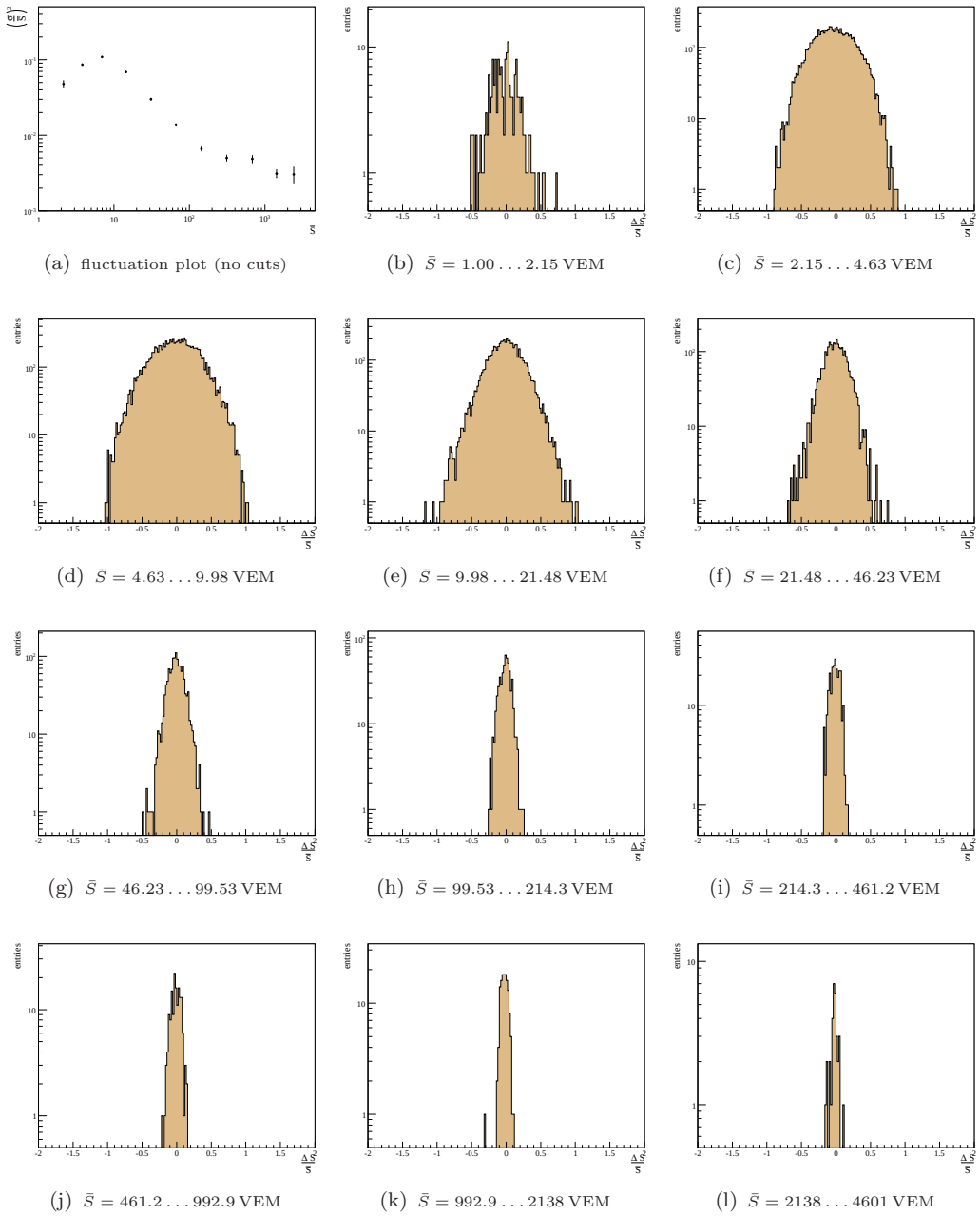


Figure 20: Stereoscopic visualization of the calculation of relative signal fluctuations. The $\bar{S}$-axis is divided into slices with equal widths of $\Delta \log \bar{S}$, each slice contains one histogram of relative deviations $\frac{\Delta S}{\bar{S}}$. Here, only one of these histograms is shown. The square of the RMS of the histograms is projected onto the $\left(\frac{\sigma}{\bar{S}}\right)^2$-axis together with the RMS error.



(a) fluctuation plot (no cuts) (b) $\bar{S} = 1.00 \ldots 2.15$ VEM (c) $\bar{S} = 2.15 \ldots 4.63$ VEM

(d) $\bar{S} = 4.63 \ldots 9.98$ VEM (e) $\bar{S} = 9.98 \ldots 21.48$ VEM (f) $\bar{S} = 21.48 \ldots 46.23$ VEM

(g) $\bar{S} = 46.23 \ldots 99.53$ VEM (h) $\bar{S} = 99.53 \ldots 214.3$ VEM (i) $\bar{S} = 214.3 \ldots 461.2$ VEM

(j) $\bar{S} = 461.2 \ldots 992.9$ VEM (k) $\bar{S} = 992.9 \ldots 2138$ VEM (l) $\bar{S} = 2138 \ldots 4601$ VEM

Figure 21: Distributions of $\frac{\Delta S}{S}$ projections for 11 bins in $\bar{S}$. The resulting signal fluctuation plot can be seen in the upper left panel. For this example, all additional cuts have been removed.

For large signals ($\bar{S} \gtrsim 300$ VEM), the fluctuations seem to reach a constant level. In the official guide-line for SD analysis [43], the fluctuation model (98) is noted together with an additional "noise constant" p_N as

Bin #	Entries	bounds [VEM]	$\left\langle \frac{\Delta S}{S} \right\rangle$	$RMS\left\{\left(\frac{\Delta S}{S}\right)_{ij}\right\}_{j=1}^{M_i}$
1	173	1.00 . . . 2.15	$(-2.3 \pm 1.7) \cdot 10^{-2}$	0.218 ± 0.012
2	7187	2.15 . . . 4.63	$(-3.0 \pm 0.3) \cdot 10^{-2}$	0.293 ± 0.002
3	10583	4.63 . . . 9.98	$(-1.4 \pm 0.3) \cdot 10^{-2}$	0.330 ± 0.002
4	5678	9.98 . . . 21.48	$(-6.6 \pm 3.5) \cdot 10^{-3}$	0.262 ± 0.002
5	2665	21.48 . . . 46.23	$(-7.9 \pm 3.4) \cdot 10^{-3}$	0.173 ± 0.003
6	1295	46.23 . . . 99.53	$(-7.2 \pm 3.3) \cdot 10^{-3}$	0.117 ± 0.002
7	540	99.53 . . . 214.3	$(-1.2 \pm 0.3) \cdot 10^{-2}$	0.081 ± 0.002
8	249	214.3 . . . 461.2	$(-1.3 \pm 0.4) \cdot 10^{-2}$	0.071 ± 0.003
9	153	461.2 . . . 992.9	$(-1.2 \pm 0.6) \cdot 10^{-2}$	0.070 ± 0.004
10	135	992.9 . . . 2138	$(-2.6 \pm 0.5) \cdot 10^{-2}$	0.056 ± 0.003
11	32	2138 . . . 4601	$(-2.4 \pm 1.0) \cdot 10^{-2}$	0.055 ± 0.007

Table 3: Means and RMS values of the $\frac{\Delta S}{S}$ distributions of Fig. 21.

$$\left(\frac{\sigma}{\bar{S}}\right)^2 = \frac{{p_\sigma}^2}{\bar{S}} + {p_N}^2 \,. \tag{101}$$

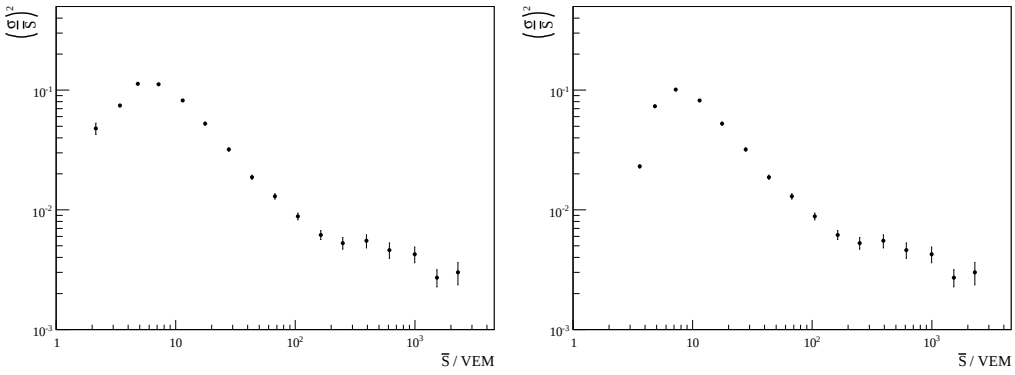
The fit of (101) yields (see Fig. 25)

$$\left(\frac{\sigma}{\bar{S}}\right)^2 = \frac{(0.934 \pm 0.008)^2}{\bar{S}} + (0.034 \pm 0.004)^2 \quad (\chi^2/dof = 42.4/9) \,. \tag{102}$$

As we have large signals in that region of constant signal flucutation level, this effect is more likely originating from saturation and recovery effects than from noise. Additional analysis would have to be performed to investigate this behavior for high signals.
Nevertheless, p_N makes fitting of the Poissonian-like model possible also for ranges of large signals.

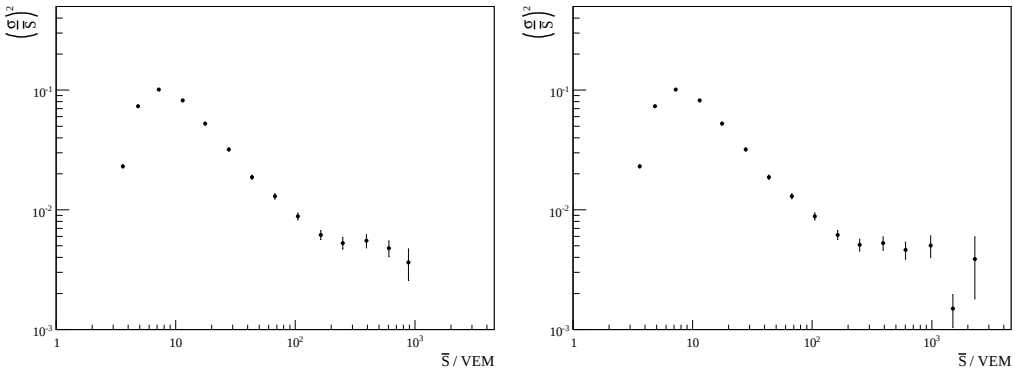
8.5 The threshold effect

For fluctuation analyses using pair tanks, both of the corresponding tanks must have been triggered to enable the calculation of $\frac{\Delta S}{S}$.
The trigger levels T1 and T2 of the SD array are installed hardware sided in the tank electronics. Only signals passing the T2 level are submitted to the CDAS. Each PMT and tank electronics of each tank can be considered to react slightly different to the same composition, momentum and number of particles entering the tank. This bias can be avoided by setting a threshold artificially which has to be chosen a little larger than the trigger threshold of the tank electronics.



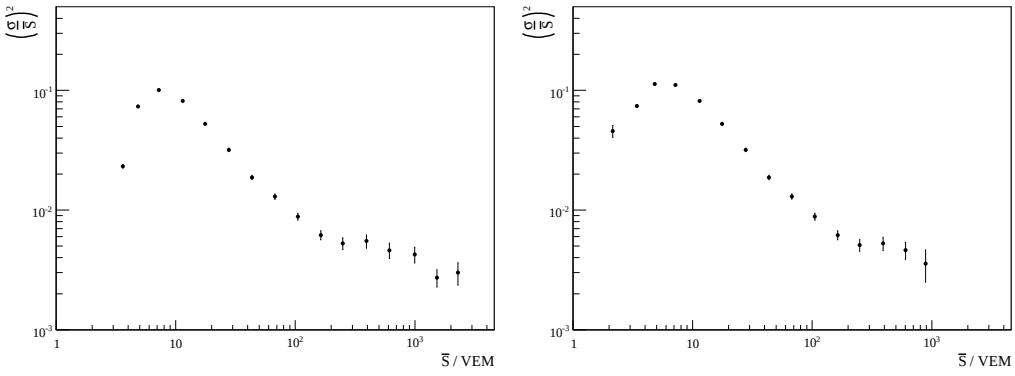
(a) No cuts

(b) 3 VEM threshold cut



(c) $S < 500\,\mathrm{VEM_{peak}}$ saturation cut

(d) $r_c > 200\,\mathrm{m}$ distance to shower axis cut



(e) $\Delta t < 200\,\mathrm{ns}$ timing cut

(f) all cuts except 3 VEM threshold cut

Figure 22: Signal fluctuation plots for different combinations of additional cuts, drawn in the range of $1.0\,\mathrm{VEM} \leq \bar{S} \leq 4601\,\mathrm{VEM}$.

A cutoff level of $S_{\mathrm{th}} = 3.0\,\mathrm{VEM}$ seems to be suitable to obtain a clean set of pair tank samples.

As the measured signals in the SD stations fluctuate according to a Poissonian-

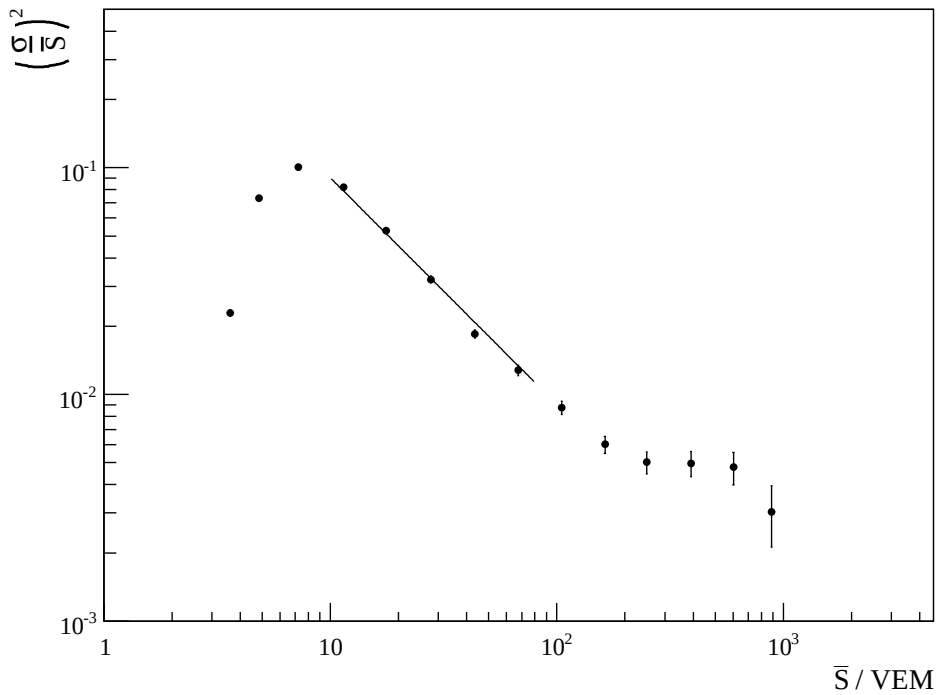


Figure 23: Signal fluctuation plot with Poisson fit (98) in the range 10.0 VEM $\leq \bar{S} \leq$ 80.0 VEM.

like behavior, there is the probability that at least one of the two pair tanks does not trigger although the average signal is larger than 3.0 VEM. Those samples are rejected and not available for the analysis. Fig. 24 illustrates the threshold effect.

This effect is purely signal strength dependent (assuming $\sigma \propto \sqrt{\bar{S}}$). On the one hand, the spread of the Gaussian distribution increases with higher signals by $\sigma = p_\sigma \cdot \sqrt{S}$, on the other hand, the probability of undershooting the threshold level increases with $\bar{S}$ approaching the 3.0 VEM limit.

8.5.1 Theoretical model of the threshold effect

The probability of undershooting the threshold level for one single tank is

$$P(\bar{S} < S_{\text{th}}) = \frac{1}{\sqrt{2\pi {p_\sigma}^2 \bar{S}}} \int_{-\infty}^{S_{\text{th}}} \exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2 \bar{S}}\right) dx\,. \qquad (103)$$

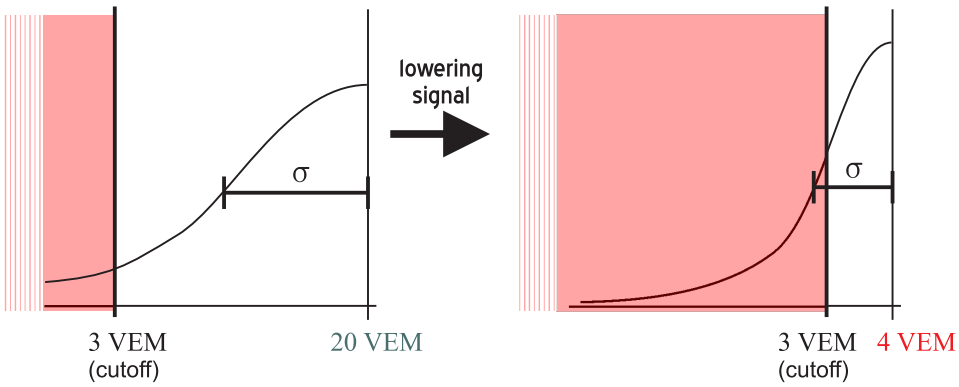


Figure 24: Illustration of the 3 VEM threshold effect. For a large signal, the probability of undershooting the 3 VEM threshold (light red area) is small (left panel). When lowering the signal, this probability increases by shifting the expectation value more and more towards the exclusion zone (right panel). The spread σ of the Gaussian increases with higher signals.

with $\sigma = p_\sigma \cdot \sqrt{\bar{S}}$. Because we have to consider the response of two independent tanks (the sample is rejected, if either the first tank is below the threshold or the other one), the overall probability of losing a sample is twice the error function[17]. Finally, the probability of not losing a sample is

$$\bar{P}(\bar{S}, \sigma) = 1 - 2P(\bar{S} < S_{\mathrm{th}})\,. \tag{104}$$

This effect leads to an apparent decrease of signal fluctuations when diminishing the signal down to S_{th} because the possible range for fluctuations will become more and more restricted with lower $\bar{S}$. This effect can be seen clearly in Fig. 23 or Fig. 22.

The threshold effect can now be evaluated analytically by

$$\left(\frac{\sigma}{\bar{S}}\right)^2 = \frac{{p_\sigma}^2}{\bar{S}} \cdot \bar{P}^2(\bar{S}, \sigma)\,. \tag{105}$$

Using (98), one can obtain $\sigma = p_\sigma \cdot \sqrt{\bar{S}}$ from a fit to experimental data. Therefore, the notation $\bar{P}(\bar{S}, p_\sigma)$ instead of $\bar{P}(\bar{S}, \sigma)$ will be used in future references.
Fig. 25 demonstrates the good agreement of this theoretical model with experimental data. In this case, (105) was fitted with an additional noise constant

[17] The case that both tanks exhibit signals below the threshold has not to be considered for $\bar{S} = \frac{S_1+S_2}{2}$ is a boundary condition that prevents both signals being either above or below $\bar{S}$ simultaneously.

$$\left(\frac{\sigma}{\bar{S}}\right)^2 = \frac{{p_\sigma}^2}{\bar{S}} \cdot \bar{P}^2(\bar{S}, p_\sigma) + {p_N}^2 \tag{106}$$

in the range of $3.0\,\mathrm{VEM} \leq \bar{S} \leq 4601\,\mathrm{VEM}$ with the fit result

$$\left(\frac{\sigma}{\bar{S}}\right)^2 = \frac{(0.933 \pm 0.007)^2}{\bar{S}} \cdot \bar{P}^2(\bar{S}, 0.933) + (0.037 \pm 0.004)^2 \tag{107}$$
$$(\chi^2/dof = 58.2/12)\,.$$

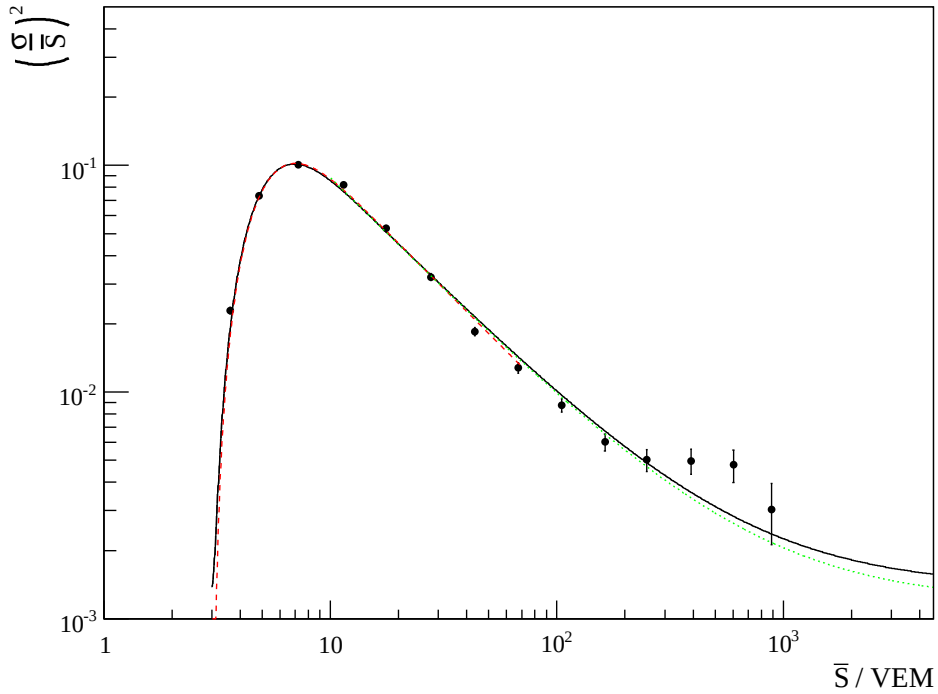


Figure 25: Fit of the fluctuation model including theoretical prediction of the threshold effect with (black curve) and without (red dashed curve) noise constant and a Poisson fit with noise constant (green dashed curve).

In Fig. 25, also the model with threshold prediction without noise constant (105) (red dashed curve, fit range $3.0\,\mathrm{VEM} \leq \bar{S} \leq 80.0\,\mathrm{VEM}$) is drawn. The fit result is

$$\left(\frac{\sigma}{\bar{S}}\right)^2 = \frac{(0.952 \pm 0.007)^2}{\bar{S}} \cdot \bar{P}^2(\bar{S}, 0.952) \tag{108}$$
$$(\chi^2/dof = 59.6/7)\,.$$

8.6 Impact of the fit range onto p_σ and p_N

When the data set is subdivided, e.g. into bins of zenith angle or distance to the shower core, it might happen that there are no data points in bins at the lower or the higher end of the signal range available anymore. For this reason, the influence of varying the fit range (which is equivalent to the exclusion of data points) on the Poissonian-like model (101) and the model including threshold prediction (106) is investigated in this section.

The signal fluctuation plots in the range of $1.0\,\mathrm{VEM} \leq \bar{S} \leq 4601\,\mathrm{VEM}$ are divided into 19 bins in $\bar{S}$. Due to the 3.0 VEM threshold cut, the first two bins do not contain data points. The fit range is set to $S_0 \leq \bar{S} \leq 4601\,\mathrm{VEM}$ and data points are removed from the fit by gradually increasing S_0 from an initial value of $S_0 = 2.64\,\mathrm{VEM}$.
The variation of the fit parameters p_σ and p_N dependent of the number of data points excluded is given in Tab. 4.

The vanishing of data points in the high end of the signal spectrum can be compensated by discarding the noise constant p_N.

		Poissonian model		Model with thresh. prediction	
NPR	S_0 [VEM]	p_σ	p_N	p_σ	p_N
0	2.64	—	—	0.933 ± 0.007	0.037 ± 0.004
1	4.12	—	—	0.936 ± 0.007	0.034 ± 0.004
2	6.42	—	—	0.935 ± 0.007	0.034 ± 0.004
3	10.0	0.934 ± 0.008	0.034 ± 0.004	0.936 ± 0.008	0.034 ± 0.004
4	15.6	0.912 ± 0.011	0.039 ± 0.004	0.912 ± 0.011	0.039 ± 0.004
5	24.3	0.872 ± 0.015	0.045 ± 0.004	0.872 ± 0.015	0.045 ± 0.004
6	37.9	0.82 ± 0.02	0.051 ± 0.004	0.82 ± 0.02	0.051 ± 0.004

Table 4: Deviation of fit parameters of signal fluctuation models by removing data points from the lower end (NPR = number of points removed).

Both models return stable fit results when increasing the lower bound of the fit range. For signal values up to $\bar{S} \approx 40\,\mathrm{VEM}$, they lead to nearly equal parameters p_σ and p_N with equal fit errors.

It can be concluded that for $10.0\,\mathrm{VEM} \lesssim \bar{S} \lesssim 40\,\mathrm{VEM}$ both models are suitable. For extending the fit range below $\approx 10.0\,\mathrm{VEM}$, the model with threshold prediction has to be used anyway because the Poissonian-like model is not able to follow the threshold clipping.

8.7 Correction of signal fluctuations

Despite the fact that signal fluctuations are described sufficiently well with (105) or (106), there is the possibility of correcting the measured signal fluctuations to the expected level they would have without threshold effect. This correction can be performed by two different approaches:

1. Deriving a correction function $R(\bar{S})$ from (105)
2. A Toy MC procedure that simulates the impact of the threshold effect and the Gaussian-like behavior of the tank response. This approach had been introduced in [36] and is repeated with higher statistics in this thesis.

The correction function $R(\bar{S})$ can be obtained by calculating the ratio of the theoretical models for signal fluctuations with and without the 3.0 VEM threshold cut:

$$R(\bar{S}) = \frac{\left(\sigma/\bar{S}\right)^2_{\text{no thresh.}}}{\left(\sigma/\bar{S}\right)^2_{\text{3 VEM thresh.}}} = \frac{\frac{{p_\sigma}^2}{\bar{S}}}{\frac{{p_\sigma}^2}{\bar{S}} \cdot \bar{P}^2(\bar{S},\sigma)} = \bar{P}^{-2}(\bar{S},\sigma)\,. \tag{109}$$

The corrected data points have uncertainties that are given by

$${\sigma_R}^2 = {\sigma_{\Delta y}}^2 + {\sigma_{\Delta R}}^2 + {\sigma_{\Delta p_\sigma}}^2\,. \tag{110}$$

For details on (110), see App. A.

Fig. 26 shows the result of the correction procedure using $R(\bar{S})$ (blue squares).

The Poissonian-like fluctuation model (98) is fitted to the corrected data within an extended fit range down to $\bar{S}_{min} = 3.0\,\text{VEM}$ (blue line of Fig. 26). The result of the fit is

$$\left(\frac{\sigma}{\bar{S}}\right)^2 = \frac{(0.933 \pm 0.005)^2}{\bar{S}} \quad (\chi^2/dof = 14.2/7)\,. \tag{111}$$

The green diamonds/dashed line in Fig. 26 are the data points corrected via the Toy MC procedure and the corresponding Poisson fit (see next section).

8.7.1 Correction via Toy MC

The impact of the threshold effect onto the signal fluctuations also can be corrected by a Toy MC simulation [36] [40]. The procedure is as follows:

1. Plot the signal spectrum of the measured pair tank samples
2. Fit a power law to the spectrum

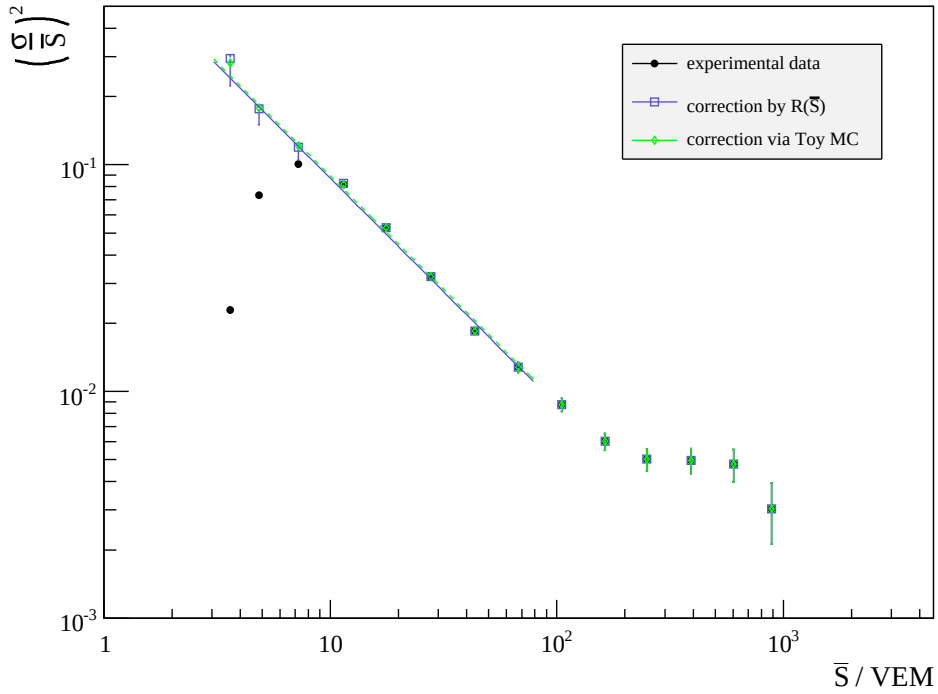


Figure 26: Correction of signal fluctuations using $R(\bar{S})$ (blue squares) and via Toy MC procedure (green diamonds) together with Poisson fits (98) in the range $3.0\,\mathrm{VEM} \leq \bar{S} \leq 80\,\mathrm{VEM}$. The original data is marked by black points.

3. Take this power law as a probability function and create a random set of simulated signal values

4. Take each entry of the random set and fluctuate it twice using a Gaussian distribution. The spread of the Gaussian is taken from $\sigma = p_\sigma \cdot \sqrt{\bar{S}}$ with p_σ being the fit result of (98). This step simulates the response of the two pair tanks

5. Create two simulated signal spectra from the fluctuated values of the last step; one with and one without the threshold cut

6. From those two data sets, create two plots $\bar{S}$ vs. $\left(\frac{\sigma}{\bar{S}}\right)^2$

7. Calculate a correction factor for each $\bar{S}$ bin

8. Apply the correction factors to the real data fluctuations

The power law $\Phi(S) = \Phi_0 \cdot {S_1}^{\alpha}$ is fitted in the range $20.0\,\mathrm{VEM} \leq S_1 \leq 45.0\,\mathrm{VEM}$[18] to the experimental signal spectrum of the main tanks (see Fig. 27). The fit returns

[18] The fit range had to be chosen so small because the fit procedure failed to converge for a larger range.

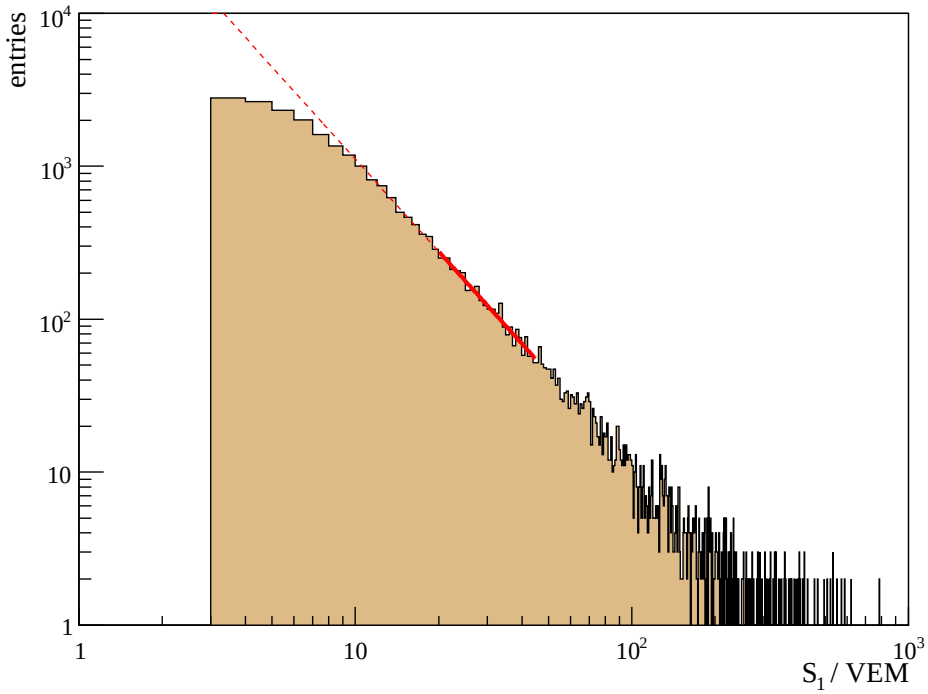


Figure 27: Signal spectrum of all pair tank samples. Only the signal of the main tank (S_1) is considered. The thick red line is the fit of a power law in the range 20.0 VEM $\leq S_1 \leq$ 45.0 VEM, the dashed thin one indicates the extrapolation to low signals.

$$\Phi(S) = (11 \pm 3) \cdot 10^4 \cdot S^{-2.00 \pm 0.08} \,. \tag{112}$$

From $\Phi(S)$, a new histogram containing 1000 bins is filled in the range of 1 VEM $\leq S_{\text{sim}} \leq$ 1000 VEM with 30,000 random values generated according to $\Phi(S)$. This procedure simulates the expected signal spectrum.
By taking a value out of the simulated spectrum by random, the detector response is emulated by fluctuating the obtained signal of the spectrum twice via a Gaussian distribution. The Gaussian function has been defined for $\pm 5\sigma$ with $\sigma = p_\sigma \cdot \sqrt{S}$. p_σ is obtained from the Poisson fit (99). From the fluctuated values, the averages $\bar{S}_{\text{sim}} = (S_{1,\text{sim}} + S_{2,\text{sim}})/2$ for each pair are stored and two samples are created, one containing all 30,000 events and a second one in which both signals have to be above the threshold of 3 VEM. In this sample with threshold cut applied, 8606 entries remain. Fig. 28 shows the distributions of the two samples.

From the simulated signal spectrum, two relative signal fluctuation plots are created, one with and the other one without threshold cut (see Fig. 29). The blue points indicate the relative signal fluctuations without threshold cut, the red ones those with 3 VEM threshold cut.

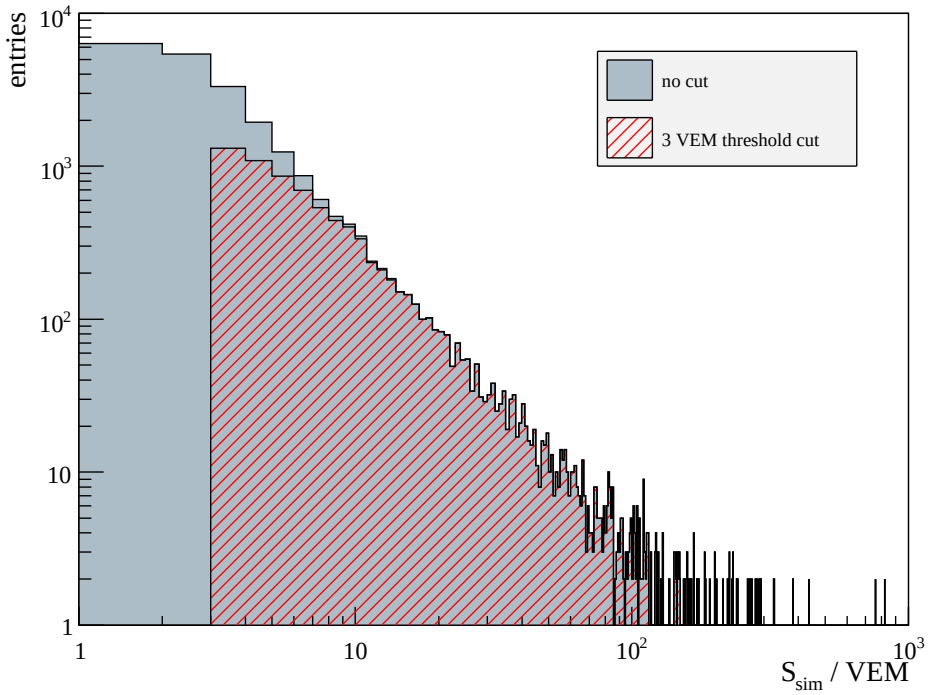


Figure 28: Simulated signal spectrum with (red shaded) and without (blue gray) 3 VEM threshold cutoff.

The correction factors are calculated by dividing the simulated signal fluctuations without threshold cut by those with threshold cut, evaluated separately for each signal bin. The uncertainties of the corrected data points are put together by scaling the uncertainties of the data point i with the corresponding correction factor R_i and by multiplying the data point with the uncertainty of the correction factor:

$$\left(\sigma_i^{\mathrm{MC}}\right)^2 = {R_i}^2 \left[\Delta\left(\frac{\sigma}{\bar{S}}\right)_i^2\right]^2 + \left(\frac{\sigma}{\bar{S}}\right)_i^2 (\Delta R_i)^2 \,. \tag{113}$$

ΔR_i is given by the statistical errors of the simulated data points of Fig. 29:

$$(\Delta R_i)^2 = [\Delta\left(\frac{\sigma}{\bar{S}}\right)^2_{\mathrm{sim}}]^2 + [\Delta\left(\frac{\sigma}{\bar{S}}\right)^2_{\mathrm{sim\ (3\ VEM\ cut)}}]^2 \,. \tag{114}$$

In Tab. 5, the correction factors obtained by the described Toy MC procedure and by the correction via $R(\bar{S})$ are listed together with their uncertainties.

The fit result of (98) to the corrected data points (green) via the Toy MC procedure in Fig. 26 is

$$\left(\frac{\sigma}{\bar{S}}\right)^2 = \frac{(0.944 \pm 0.005)^2}{\bar{S}} \quad (\chi^2/dof = 31.0/7) \,. \tag{115}$$

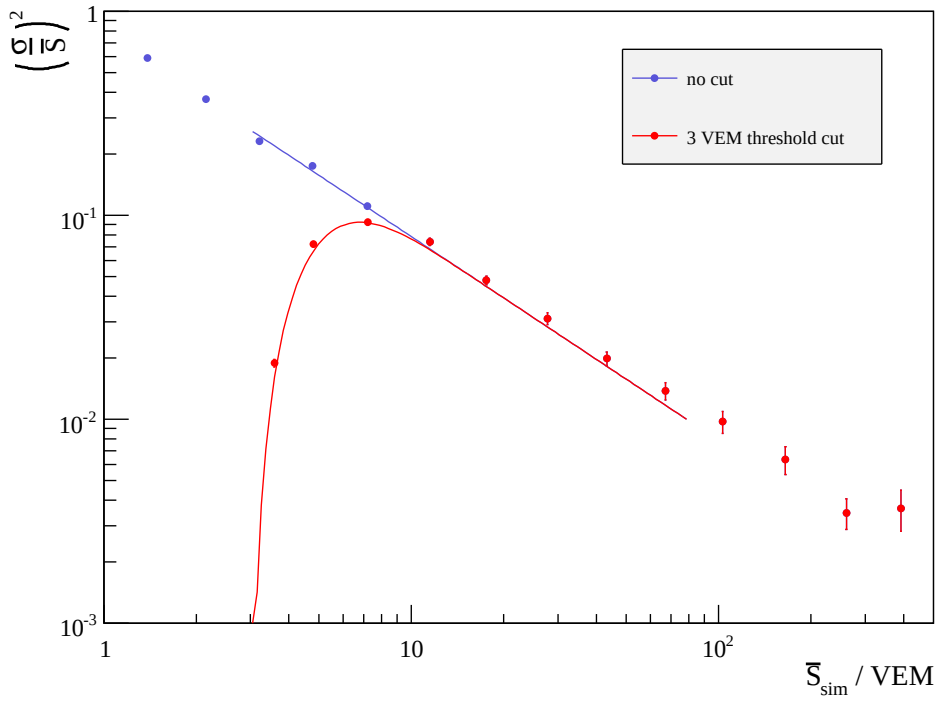


Figure 29: Signal fluctuation plot obtained from the simulated data set with (red) and without (blue) threshold cut. The red curve is a fit of the model including threshold prediction (105), the blue one the Poissonian-like model (98), both applied in the range 3.0 VEM $\leq \bar{S}_{\text{sim}} \leq$ 80 VEM. In this plot only the relevant range of $1 \leq \bar{S}_{\text{sim}} \leq 500$ VEM is displayed. The binning corresponds to that of Fig. 23.

Bin #	$R(\bar{S})$	Toy MC
1	no data points	no data points
2	no data points	no data points
3	12.85 ± 0.12	12.248 ± 0.004
4	2.410 ± 0.003	2.428 ± 0.006
5	1.18706 ± 0.00012	1.200 ± 0.004
6	$1.011494 \pm 2 \cdot 10^{-6}$	1.000 ± 0.004
7	$1.000228 \pm 4 \cdot 10^{-9}$	1.0000 ± 0.003
8 . . . 19	1.0000 ± 10^{-14}	1.0000 ± 0.002

Table 5: Correction factors for $\left(\frac{\sigma}{\bar{S}}\right)^2$ for the method using the correction function $R(\bar{S})$ and for the Toy MC method. The uncertainties of the factors for the bins 8 . . . 19 are the maximum errors within that range.

The two methods of correcting the signal fluctuations lead to a very similar result, the fit parameters of the Poissonian-like fluctuation model are compatible to each other within their statistical uncertainties. $R(\bar{S})$ has larger

uncertainties for large correction factors than those of the corresponding factors obtained from the Toy MC method, but the $R(\bar{S})$ method estimates the uncertainties of the corrected data points in a more trustable way.

In order to investigate the compatibility of the fluctuation models with the Toy MC data, the fluctuation model with threshold prediction (105) (red) is fitted to the red points, the mere Poissonian-like model (98) (blue) to the blue points, both within a fit range of $3.0\,\mathrm{VEM} \leq \bar{S} \leq 80\,\mathrm{VEM}$. The fit results are

$$p_\sigma^{\text{no thresh.}} = 0.887 \pm 0.006 \tag{116}$$

for the blue line and

$$p_\sigma^{3\text{ VEM thresh.}} = 0.903 \pm 0.013 \tag{117}$$

for the red curve.

The p_σ obtained from the fits are a bit lower than the corresponding fit results (99) and (108) to real data.

8.8 Conclusions

In this section, four different methods of signal fluctuation analysis have been presented:

1. Direct fit of a Poissonian-like model
2. Poissonian-like model of signal fluctuations compensated for low signals using a correction function $R(\bar{S})$
3. Poissonian-like model of signal fluctuations compensated for low signals using a Toy MC correction procedure
4. Analytical model that decribes the threshold effect based on the Poissonian-like model

All four models can be used together with an additional noise constant p_N.

The Toy MC method has two disadvantages:

- The individual correction factors are constant within a whole bin of $\bar{S}$. As $\bar{S}$ of the data point will deviate from the $\bar{S}$ of the simulated point within one bin, the correction cannot be exact.

- It is rather difficult to estimate the uncertainties of the Toy MC correction reliably.

Generally, the correction of signal fluctuations does not bring any advantages compared to a direct fit of (105) or (106) to the uncorrected data points. For these reasons, both correction methods will not be followed in this thesis anymore.
Nevertheless, the fit of (105) in Fig. 29 is in good agreement with the simulated data points with 3 VEM threshold cut.

The best method is the model including the theoretical prediction of the threshold effect because it expands the possible fit range down to signals of $\bar{S}_{\min} \geq 3\,\text{VEM}$ and will therefore be used preferably in this thesis.
In cases where not enough data points for fits down to $\bar{S} = 3\,\text{VEM}$ are available, the simple Poissonian-like model will be used. Then, the fit range has to be restricted to $\bar{S}_{\min} \gtrsim 10\,\text{VEM}$. Anyway, additional data points for higher signals can be taken into account by adding the noise constant, either as a fixed value (e.g. $p_N = 0.037 \pm 0.004$ from (107)) or as a free fit parameter.

Besides, it can be concluded that each of the four methods of signal fluctuation analysis might in principle be used as they lead to very similar results with comparatively low fit uncertainties.

The fit results obtained for the four models presented in this section are listed in Tab. 6. The features of the two models that avoid a correction of signal fluctuations and will be used in the following sections of this thesis are recapitulated in Tab. 7.

Method	p_σ	p_N
Poisson fit (98)	0.903 ± 0.013	—
Poisson fit with noise constant (101)	0.934 ± 0.008	0.034 ± 0.004
Poisson fit to corrected data ($R(\bar{S})$ method)	0.933 ± 0.005	—
Poisson fit to corrected data (Toy MC method)	0.944 ± 0.005	—
Model incl. threshold prediction (no noise constant)	0.952 ± 0.007	—
Model incl. threshold prediction and noise constant	0.933 ± 0.007	0.037 ± 0.004

Table 6: Fit results of the signal fluctuation models presented in Sec. 8.

Method	defined by	advantages	disadvantages
Poisson fit (98)	$\frac{{p_\sigma}^2}{\bar{S}}$	—	Limited fit range
Poisson fit with noise constant (101)	$\frac{{p_\sigma}^2}{\bar{S}} + {p_N}^2$	Extended fit range to higher signals	Lower bound of fit range restricted
Model incl. threshold prediction (105)	$\frac{{p_\sigma}^2}{\bar{S}} \cdot \bar{P}^2(\bar{S}, p_\sigma)$	Extended fit range to lower signals	Upper bound of fit range restricted
Model incl. threshold prediction and noise constant (106)	$\frac{{p_\sigma}^2}{\bar{S}} \cdot \bar{P}^2(\bar{S}, p_\sigma) + {p_N}^2$	Widest fit range	—

Table 7: Comparison of the signal fluctuation models presented in Sec. 8.

9 Zenith angle dependency

A second order effect of signal fluctuations is the dependency of the zenith angle of an EAS. This effect is hidden when only considering signal strength induced fluctuations. The zenith angle dependency will be deconvoluted from the signal strength induced fluctuations in this section.

9.1 Basics

As described in Sec. 8, the main origin of signal fluctuations is the Poissonian-like behavior of the number of particles hitting the pair tanks. But actually, we do not measure single particles but rather continuous signals from PMTs generated by particles, expressed in units of VEM.

Particles entering the Čerenkov tanks at ground level are mainly muons and EM particles (electrons/positrons and photons). While EM particles have a typical kinetic energy of some MeV, muons have energies in the GeV range. Besides, muons have a smaller energy loss dE/dx than EM particles. For these reasons, muons will traverse the whole tank, regardless of the length of passage through the water they have to bridge. The length of passage L_μ is strongly depending on the zenith angle of the particle.

The correlation between the deposited charge in the PMTs of a Čerenkov tank and the track length had been investigated by positioning two scintillators at opposite sides of the tank, either one scintillator above the lid and one below the bottom or at opposite sides of the outer shell [44].
Frome these measurements, a linear dependence between the deposited charge and the muon track length could be confirmed as well as a linear dependence of the uncertainty of the charge to its square root [45]. The experimental results had been verified via Toy MC and a semi-analytical model that describes the tank geometry exactly [45] [53].

In a simple model, the diameter of the tank can be expanded to infinity. Then the cylindrical geometry of the tanks does not play a role anymore. We obtain

$$L_\mu = \frac{h_0}{\cos\theta} = h_0 \cdot \sec\theta \tag{118}$$

for a tank with infinite diameter and height h_0. In a more elaborate model including tank geometry, the track length of muons will depend on the place where the muon enters the tank and its zenith angle. Regarding the cylindrical tank geometry, four different cases can be distinguished:

1. Muons that enter through the top lid and leave through the bottom lid

2. Muons that enter through the top lid and leave through the outer shell

3. Muons that enter through the outer shell and leave through the bottom lid (equivalent to case 2.)

4. Muons that enter and leave through the outer shell

Comparing the impact of these four cases onto the mean muon track length, it can be seen that the first case makes the biggest contribution (see Fig. 30). This case is described exactly by (118). Thus, a simple model for zenith angle dependency of signal fluctuations based upon (118) is proposed:

$$p_\sigma(\theta) = a + b \cdot \sec\theta\,. \tag{119}$$

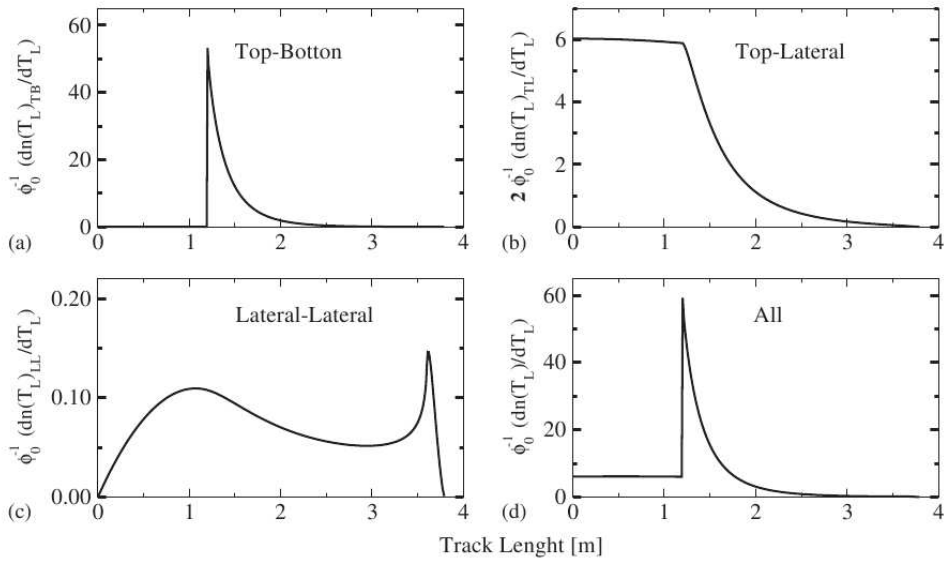


Figure 30: Distribution of muon fluxes vs. track length L_μ for the four different cases of tracks, obtained from the semi-analytical model [45]. It can be seen that case (a) makes the dominant contribution to the flux.

9.2 Analysis

To obtain the parameters a and b of (119), the data set is divided into 4 logarithmic bins of $\sec\theta$ in the range of $1.0 \leq \sec\theta \leq 2.2$. Fig. 31 shows the signal fluctuation plots for each bin. Due to the reduced ammount of entries in each $\sec\theta$ bin, the number of bins in $\bar{S}$ in each plot is reduced. The signal range is restricted to $3.0\,\mathrm{VEM} \leq \bar{S} \leq 700\,\mathrm{VEM}$. Then, the signal fluctuation model with threshold prediction without noise constant (105) is

Color	$\sec\theta$ range	p_σ	$\langle\sec\theta\rangle$
black	$1.000 \ldots 1.218$	0.845 ± 0.007	1.11 ± 0.06
red	$1.218 \ldots 1.483$	0.943 ± 0.010	1.33 ± 0.08
green	$1.483 \ldots 1.806$	1.09 ± 0.02	1.61 ± 0.09
blue	$1.806 \ldots 2.200$	1.211 ± 0.04	1.97 ± 0.11

Table 8: Fit results of $p_\sigma(\theta)$ and calculated mean and RMS of $\sec\theta$. The colors correspond to those of Fig. 31.

fitted to each plot.

Besides, the mean and the RMS of all $\sec\theta$ entries are calculated for each bin separately (see Tab. 8).
Fig. 32 shows the obtained values of p_σ plotted against $\langle\sec\theta\rangle$ and a fit according to (119). The fit returns

$$p_\sigma = (0.32 \pm 0.04) + (0.47 \pm 0.03) \cdot \sec\theta \qquad (\chi^2/dof = 2.11/2)\,. \tag{120}$$

The parameterization of signal fluctuations of the form $\sigma = (a + b \cdot \sec\theta)\sqrt{\bar{S}}$ (without noise constant) has been implemented officially in the `LDFFinder` module of the Auger Offline framework, taking the official values of $a = 0.32 \pm 0.09$ and $b = 0.42 \pm 0.07$ [48, p6].

9.3 Results

Combining the analyses of Sec. 8 and 9, the central results of this thesis can be compressed in one formula by

$$\left(\frac{\sigma}{\bar{S}}\right)^2 = \bar{P}^2(\bar{S}, p_\sigma(\theta))\,\frac{{p_\sigma}^2(\theta)}{\bar{S}} + (0.039 \pm 0.003)^2 \tag{121}$$

with

$$p_\sigma(\theta) = (0.32 \pm 0.04) + (0.47 \pm 0.03) \cdot \sec\theta\,. \tag{122}$$

The results obtained in this analysis are in good agreement with the values used for the official event reconstruction. In this thesis, are larger data set has been considered, reducing the statistical fit uncertainties of a and b of (119) by a factor of 2.

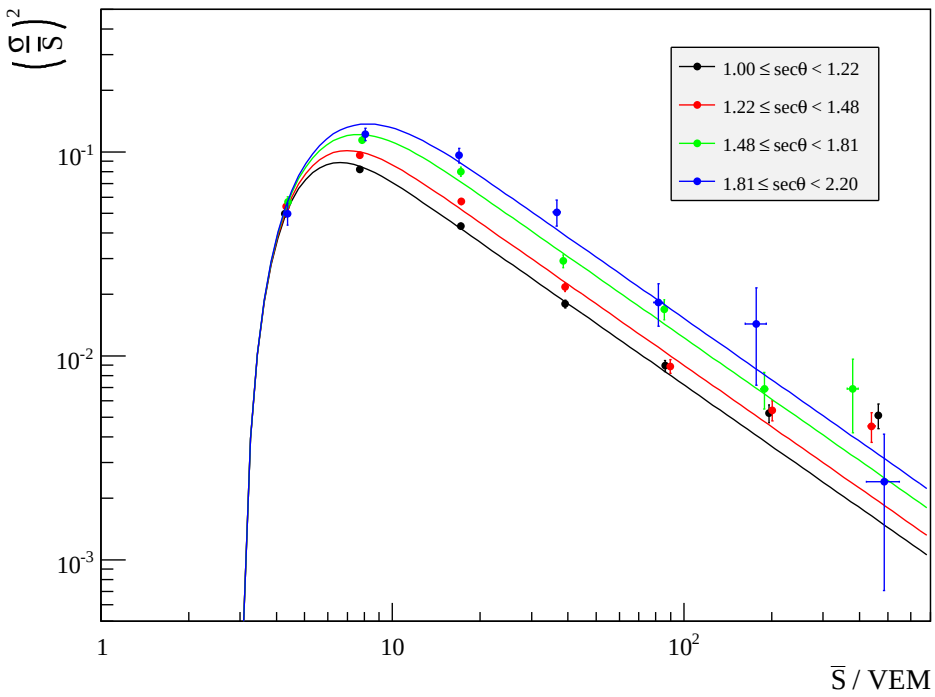


Figure 31: Signal fluctuation plots of four $\sec\theta$ bins with analytical signal fluctuation model (105) in the range of $3.0\,\mathrm{VEM} \leq \bar{S} \leq 700\,\mathrm{VEM}$.

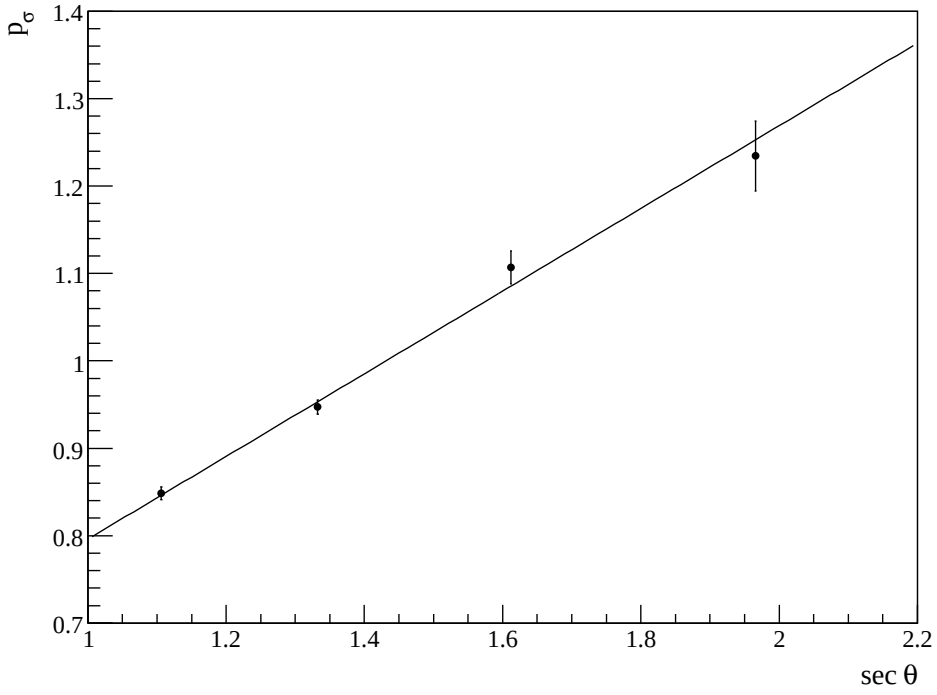


Figure 32: Zenith angle dependency of p_σ with fit of (119) in the range of $1.0 \leq \sec\theta \leq 2.2$.

10 Dependency on the distance to the shower axis

As mentioned in the beginning of Sec. 6, the steep slope of the LDF near the shower core leads to significant differences of the particle density in the pair tanks besides differences due the Poissonian fluctuations. This effect is investigated in this section.

10.1 Basics

The Auger Offline framework does not provide a method for obtaining the distance to the shower axis, only the distance to the shower core. This had to be implemented manually.
For a given shower axis $\hat{a}$ ($\hat{a}$ is a unity vector), the given position of the shower core $\vec{c}$ and the position of a station $\vec{s}$, the distance to the shower axis r_c can be evaluated. First, the segment between the station and the shower core by can be expressed by

$$\vec{r}_1 = \vec{s} - \vec{c}. \tag{123}$$

Then, there is a segment from the shower core into the direction of the shower axis that is unknown in length, so it has to be parameterized as straight line

$$\vec{r}_2(\lambda) = \lambda \hat{a}\,. \tag{124}$$

The distance vector $\vec{d} = \vec{r}_2 - \vec{r}_1$ is perpendicular to $\hat{a}$, hence

$$\begin{aligned}
\vec{d}(\lambda) \cdot \hat{a} &= (\vec{r}_2 - \vec{r}_1) \cdot \hat{a} \stackrel{!}{=} 0 && (125)\\
&\Rightarrow (\vec{s} - \vec{c} - \lambda\hat{a}) \cdot \hat{a} && (126)\\
&= (\vec{s} - \vec{c}) \cdot \hat{a} - \lambda = 0 && (127)\\
&\Rightarrow \lambda = (\vec{s} - \vec{c}) \cdot \hat{a}\,. && (128)
\end{aligned}$$

With $r_c = |\vec{d}|$ and the solution for λ, we obtain

$$r_c = \left|\left[(\vec{s} - \vec{c}) \cdot \hat{a}\right] \hat{a} - \vec{s} + \vec{c}\right|\,. \tag{129}$$

As the range of r_c is investigated down to 0 m, the $r_c < 200\,\mathrm{m}$ cut has to be removed. There are many saturated stations for low distances to the shower core, so the $S_{\mathrm{peak}} < 500\,\mathrm{VEM_{peak}}$ saturation cut is removed as well to keep as much statistics as possible. Without saturation cut, there are 25097 samples in the r_c range of $0\,\mathrm{m} \leq r_c \leq 1980\,\mathrm{m}$ considered; including the cut, 24902

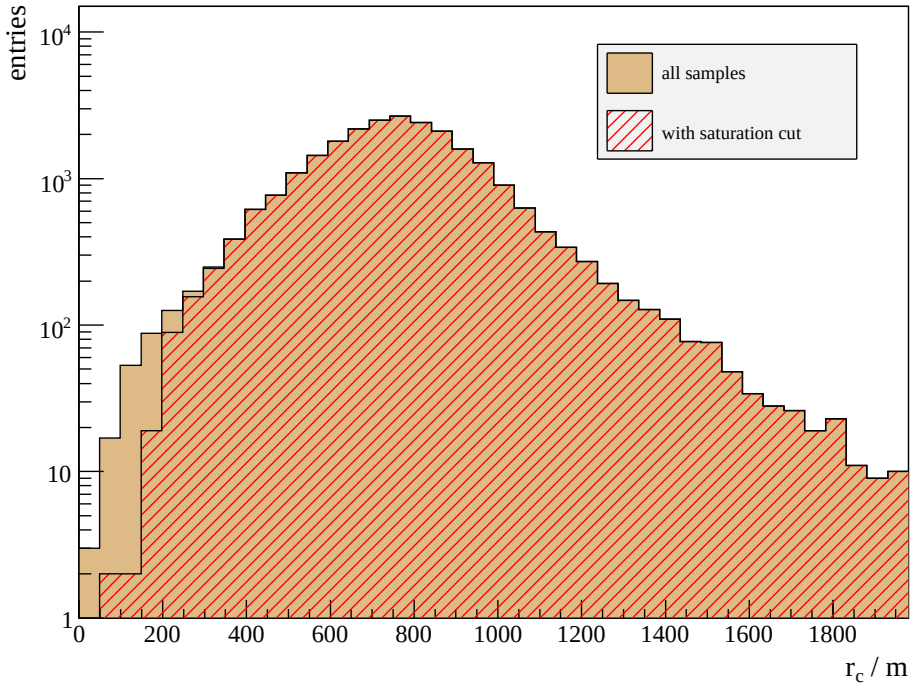


Figure 33: Distribution of samples vs. r_c without (fawn) and with (red shaded) $500\,\mathrm{VEM_{peak}}$ saturation cut.

samples remain. The distribution of entries vs. r_c is shown in Fig. 33, indicating that nearly all samples with $r_c < 200\,\mathrm{m}$ exhibit a saturated signal.

A theoretical model for a signal fluctuation dependency on the distance to the shower axis can be derived from the LDF.
The assumption is that the overall signal fluctuation is the square sum of the statistical uncertainty $\sigma_{\mathrm{VEM}} = p_0 \cdot \sqrt{S}$ and an additional uncertainty that is proportional to the difference of particles due to the influence of the LDF, the signal uncertainty can be expressed by

$$\sigma_{\mathrm{LDF}} = \Delta N(r_c) = p_0 \cdot \Delta S(r_c)\,. \tag{130}$$

$p_0 = p_\sigma$ from Sec. 8 is assumed to be the correct conversion factor from the number of particles to the corresponding signal in units of VEM. The signal $S(r_c)$ is then proportional to $\left(\frac{r_c}{1000\,\mathrm{m}} \cdot \frac{r_c+700\,\mathrm{m}}{1700\,\mathrm{m}}\right)^\beta$, the difference of the signal $\Delta S(r_c)$ within a small interval of distance Δr_c is then

$$\Delta S(r_c) = \frac{\partial S(r_c)}{\partial r_c} \Delta r_c \tag{131}$$

with

$$\frac{\partial S}{\partial r_c} \propto \beta \cdot \left(\frac{2r_c + 700\,\mathrm{m}}{1000\,\mathrm{m} \cdot 1700\,\mathrm{m}} \right) \left(\frac{r_c}{1000\,\mathrm{m}} \cdot \frac{r_c + 700\,\mathrm{m}}{1700\,\mathrm{m}} \right)^{\beta-1} . \tag{132}$$

Hence we obtain

$$\begin{aligned} \sigma^2 &= {\sigma_{\mathrm{VEM}}}^2 + {\sigma_{\mathrm{LDF}}}^2 \\ &= {p_0}^2 \cdot S + {p_0}^2 \left(\frac{\partial S(r_c)}{\partial r_c} \right)^2 (\Delta r_c)^2 . \end{aligned} \tag{133}$$

From this formula we can identify the relative signal fluctuations

$$\left(\frac{\sigma}{S} \right)^2 = \frac{{p_0}^2}{S} \left(1 + \frac{1}{S} \left(\frac{\partial S(r_c)}{\partial r_c} \right)^2 (\Delta r_c)^2 \right) \tag{134}$$

and finally $p_\sigma(r_c)$:

$$p_\sigma(r_c) = p_0 \sqrt{1 + \frac{1}{S} \left(\frac{\partial S(r_c)}{\partial r_c} \right)^2 (\Delta r_c)^2} . \tag{135}$$

Δr_c is assumed to be isotropic in each bin of r_c, so it is arguable to set $\Delta r_c = \mathrm{const}$. The means of the absolute values of Δr_c are plotted in Fig. 34 for confirmation, the fit of a constant returns $\langle |\Delta r_c| \rangle = (6.24 \pm 0.02)\,\mathrm{m}$.

Hence the prefactor β of (132) and Δr_c can be merged to a normalization parameter S_0. Setting $S = S(r_c)$ as the signal obtained from the LDF, we can simplify

$$\begin{aligned} p_\sigma(r_c) &= p_0 \sqrt{1 + S_0 \left(\frac{\partial f(r_c)}{\partial r_c} \right)^2 \frac{f(r_c)^{2\tilde{\beta}-2}}{f(r_c)^{\tilde{\beta}}}} \\ &= p_0 \sqrt{1 + S_0 \left(\frac{\partial f(r_c)}{\partial r_c} \right)^2 f(r_c)^{\tilde{\beta}-2}} \end{aligned} \tag{136}$$

with $f(r_c) = \frac{r_c}{1000\,\mathrm{m}} \frac{r_c + 700\,\mathrm{m}}{1700\,\mathrm{m}}$ and $\frac{\partial f(r_c)}{\partial r_c} = \frac{2r_c + 700\,\mathrm{m}}{100\,\mathrm{m} \cdot 1700\,\mathrm{m}}$.

p_0 is a constant level of signal fluctuation that is approximated for an increasing r_c when the influence of the LDF effect does not play a role anymore. $\tilde{\beta}$ is expected to be approximately the β that is obtained when inserting the average $\langle \sec\theta \rangle$ for $\sec\theta$ in (78)[19].

[19] This can be done because the `LDFFinder` module is being used with fixed β.

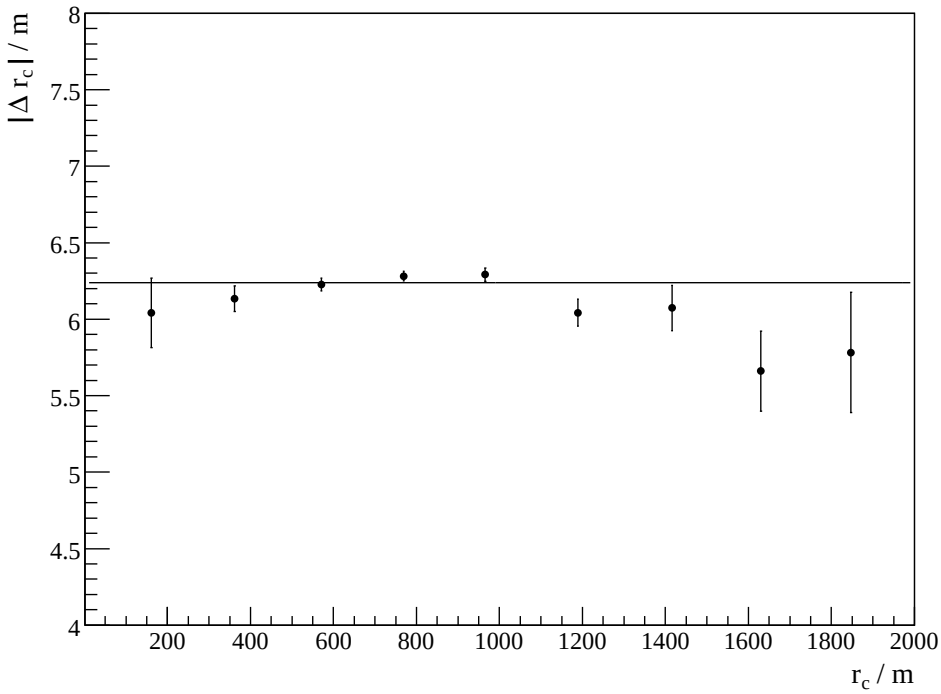


Figure 34: $|\Delta r_c|$ of the 9 bins in r_c with fit of a constant.

10.2 Analysis

In order to study the model (136), the dataset is split into 9 equidistant bins of r_c in the range of $0\,\mathrm{m} \leq r_c \leq 1980\,\mathrm{m}$, taking the distance of the first station of the tank pair as r_c. For each bin, a signal fluctuation plot (range $1.0 \ldots 4601\,\mathrm{VEM}$) containing 10 bins in $\bar{S}$ is created. Then the Poissonian-like model (98) with additional noise constant

$$\left(\frac{\sigma}{\bar{S}}\right)^2 = \frac{p_\sigma(r_c)^2}{\bar{S}} + {p_N}^2$$

is fitted in the range $10.0\,\mathrm{VEM} \leq \bar{S} \leq 4601\,\mathrm{VEM}$ to the first five signal fluctuations plots. For the remaining four plots, the analytical model with threshold prediction incl. noise constant (106) is fitted in the range $3.0\,\mathrm{VEM} \leq \bar{S} \leq 4601\,\mathrm{VEM}$ because for high distances to the shower axis, smaller signal values can be expected so the impact of the threshold cutoff cannot be neglected anymore. Besides, for large distances r_c, there are no data points in the range where (98) could be fitted. The result is shown in Fig. 35.
p_N is fixed to $p_N = 0.037$ from (107). The resulting fit values for the $p_\sigma(r_c)$ plotted against $\langle r_c \rangle$ are shown in Fig. 36, the corresponding fit values are given in Tab. 9.

Color	r_c [m]	$\langle r_c \rangle$ [m]	$p_\sigma(r_c)$
black	0 . . . 220	161	1.38 ± 0.12
red	220 . . . 440	362	0.97 ± 0.02
bright green	440 . . . 660	571	0.906 ± 0.010
bright blue	660 . . . 880	769	0.943 ± 0.015
yellow	880 . . . 1100	966	0.95 ± 0.03
pink	1100 . . . 1320	1189	1.07 ± 0.04
aqua	1320 . . . 1540	1416	1.01 ± 0.06
green	1540 . . . 1760	1630	0.98 ± 0.10
blue	1760 . . . 1980	1847	0.73 ± 0.12

Table 9: Binning, $\langle r_c \rangle$ and fit results for 9 equidistant bins in r_c. The colors correspond to those in Fig. 35.

Then, (136) is fitted to Fig. 36, yielding

$$p_0 = 0.925 \pm 0.009\,, \tag{137}$$
$$S_0 = (210 \pm 500)\,\mathrm{VEM} \quad \text{and} \tag{138}$$
$$\tilde{\beta} = -1.9 \pm 0.9 \tag{139}$$

with $\chi^2/dof = 31.6/7$. A constant level of p_0 is reached asymptotically, the value agrees with $p_\sigma = 0.934 \pm 0.008$ of (102) as well as with $p_\sigma = 0.933 \pm 0.007$ of (107).

From (78) and $\langle \sec\theta \rangle = 1.3 \pm 0.3$ (see Sec. 11.1), one obtains

$$\langle \beta \rangle = 0.9 \cdot (1.3 \pm 0.3) - 3.3 = -2.1 \pm 0.3\,, \tag{140}$$

which agrees with (139).

Although the results of this section are compatible with the assumed model (136) and the formerly obtained fit results for p_σ, it should be mentioned that fitting the p_σ needed for (136) is not robust yet, neither against varying the number of bins in r_c, nor when changing the number of bins of $\bar{S}$ in each of the signal fluctuation plots (Fig. 35). It took a lot of trials to find a combination of binning and an appropriate range of r_c where the χ^2 minimization of the fit procedure converged for each p_σ fit.
In the range of low $r_c \lesssim 400\,\mathrm{m}$, at least one additional p_σ would be necessary to stabilize the fit result of Fig. 36. A larger amount of data is required to make the uncertainty model in dependence on the distance to shower axis more confidential.

The statistical error of the determination of the shower core is $\delta r_c \approx 50\,\mathrm{m}$ [49]. The parameterization of signal fluctuations depending on the distance to the shower axis is only significant for $r_c \lesssim 400\,\mathrm{m}$. Therefore, using the model only makes sense for very low distances when the LDF effect becomes dominant. In the most cases, the model with zenith angle dependency is to be chosen preferably.

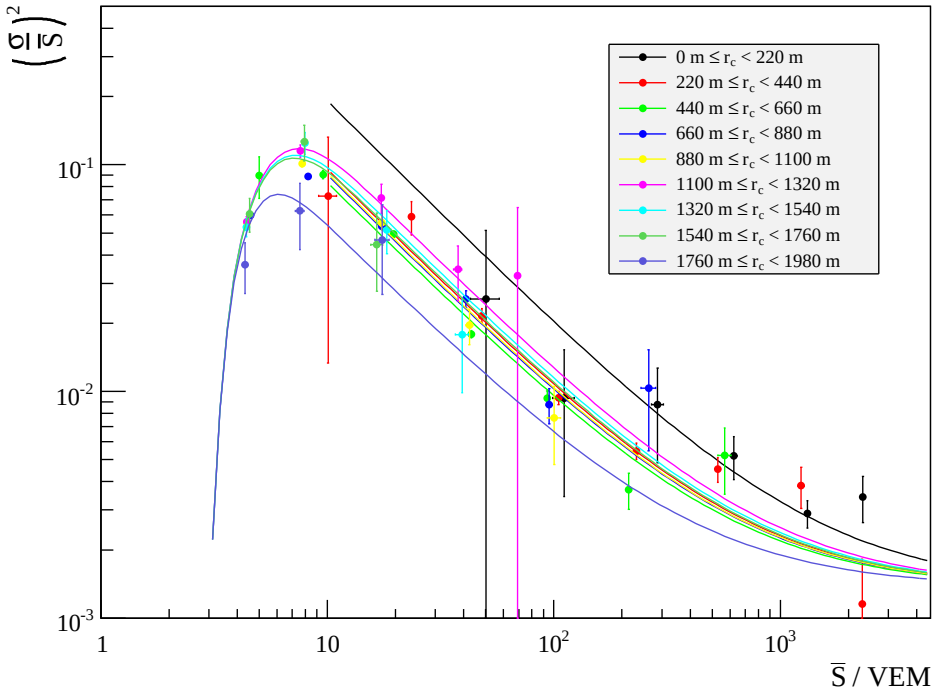


Figure 35: Signal fluctuation plots for 9 bins in r_c.

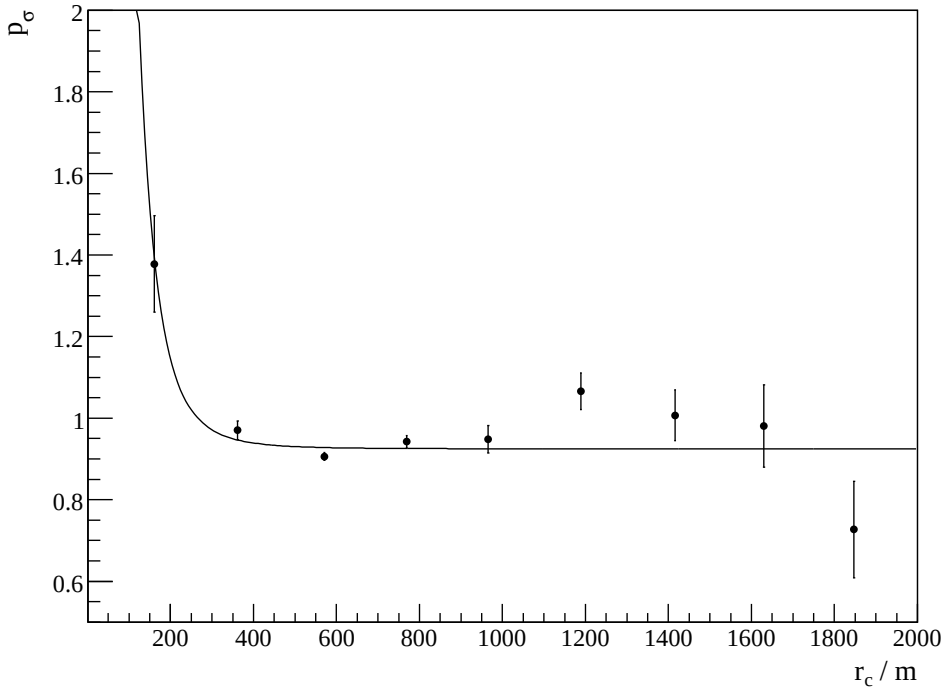


Figure 36: Fit of (136) to the $p_\sigma(r_c)$ obtained from Fig. 35.

11 Model checks

In this section, the signal fluctuation model including signal strength and zenith angle dependency will be checked against the less sophisticated models (105) and the "zeroth order" model. Furthermore, the model including zenith angle dependency will be tested for its significance compared to the uncertainties of signal fluctuations.

11.1 Model with and without zenith angle dependency

The zenith angle model of signal fluctuations of Sec. 9 can be checked against the signal strengh dependency model that has been obtained in Sec. 8. In principle, the fit values for p_σ in (101) or (106) represent the average $\langle p_\sigma \rangle$ over the whole range of occuring zenith angles. Determining the mean value of the $\sec\theta$ distribution of all pair samples and inserting this mean value into (120) should lead to a similar result as in (102) and (107).

Fig. 37 shows the distribution of samples vs. $\sec\theta = 1.0 \ldots 5.0$ (corresponding to zenith angles of 0° to 78.5°). The mean value is $\langle \sec\theta \rangle = 1.3 \pm 0.3$. Inserting that mean value into (120), one obtains $p_\sigma(\langle \sec\theta \rangle) = 0.94 \pm 0.14$ which is in good agreement with the results of (102) and (107).

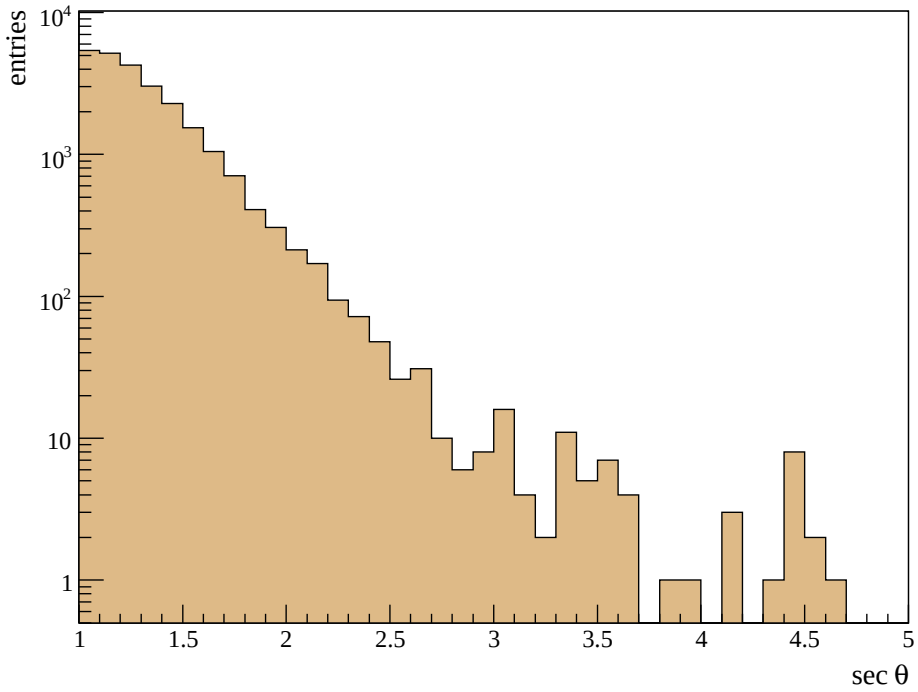


Figure 37: Distribution of all pair tank samples vs. $\sec\theta$ (all additional cuts applied).

11.2 Zeroth order model

With the parameterization (121), the zeroth order model that introduced Sec. 8 as motivation can be performed again. The result is shown in Fig. 38. One obtains a mean of $(-4.5 \pm 0.7) \cdot 10^{-2}$ and a spread of $\sigma_{\text{rel}} = 1.083 \pm 0.005$ for the Gaussian fit.

Despite the fact that, with complete fluctuation parametrization, a σ_{rel} very close to 1 is now obtained, the value still deviates more than 15 standard deviations with respect to σ. This implies that there is either still some hidden dependency of the signal fluctuation model or the deviations are not perfectly Gaussian distributed.

Comparing the fit results with the RMS and mean of the histogram ($\left\langle \frac{\Delta S}{S} \right\rangle = (-4.6 \pm 0.7) \cdot 10^{-2}$, $RMS = 1.087 \pm 0.005$), there is full compatibility of the corresponding values.

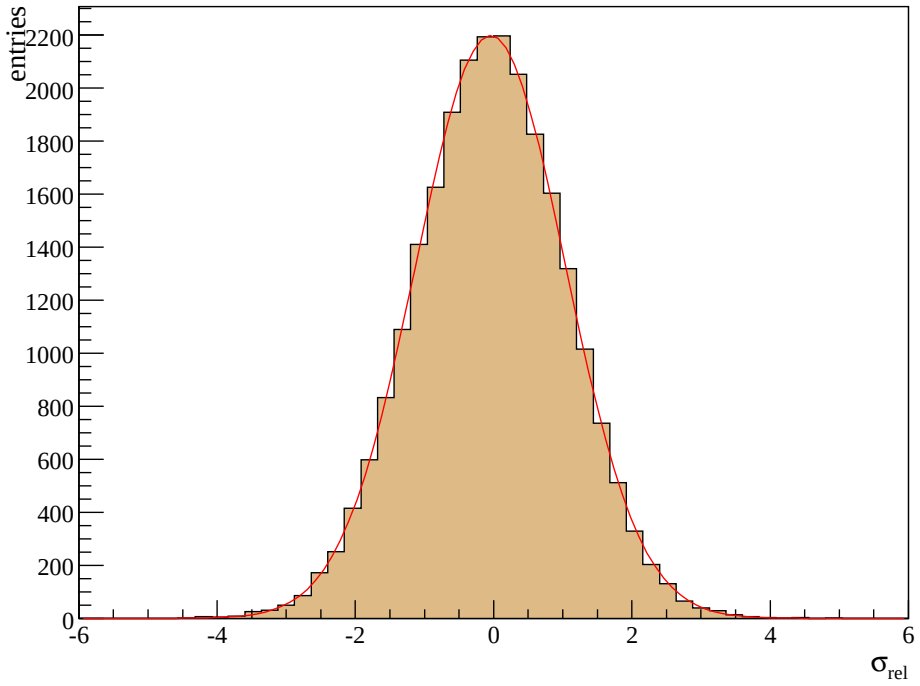


Figure 38: Distribution of the corrected zeroth error model σ_{rel} with Gaussian fit.

11.3 Significance of the model including threshold prediction

Within their fit errors, the fit parameters p_σ and p_N of (121) should be compatible with the statistical errors of the data points.

$$\eta = \frac{\left(\frac{\sigma}{S}\right)^2_{\text{theo}} - \left(\frac{\sigma}{S}\right)^2_{\text{exp}}}{{\sigma_{\text{theo}}}^2 + {\sigma_{\text{exp}}}^2} \tag{141}$$

compares the difference between the theoretically expected $\left(\frac{\sigma}{S}\right)^2_{\text{theo}}$ and the real fluctuations $\left(\frac{\sigma}{S}\right)^2_{\text{exp}}$ to the statistical errors σ_{exp} and the fit errors σ_{theo} of both.

σ_{theo} is given by the uncertainty of p_σ which has been evaluated in (163) and (164) in Sec. 8.7 on the one hand and the fit error of p_N according to (101) on the other hand. As p_σ is depending on a, b and $\sec\theta$ (119), the overall fit uncertainty of (121) is obtained via error propagation:

$${\sigma_a}^2 = \left(\frac{\partial \left(\frac{\sigma}{S}\right)^2_{\text{theo}}}{\partial a}\right)^2 (\Delta a)^2 = \left(\frac{\partial \left(\frac{\sigma}{S}\right)^2_{\text{theo}}}{\partial p_\sigma}\frac{\partial p_\sigma}{\partial a}\right)^2 (\Delta a)^2\,, \tag{142}$$

$${\sigma_b}^2 = \left(\frac{\partial \left(\frac{\sigma}{S}\right)^2_{\text{theo}}}{\partial b}\right)^2 (\Delta b)^2 = \left(\frac{\partial \left(\frac{\sigma}{S}\right)^2_{\text{theo}}}{\partial p_\sigma}\frac{\partial p_\sigma}{\partial b}\right)^2 (\Delta b)^2 \tag{143}$$

and

$${\sigma_\theta}^2 = \left(\frac{\partial \left(\frac{\sigma}{S}\right)^2_{\text{theo}}}{\partial(\sec\theta)}\right)^2 (\Delta\sec\theta)^2 = \left(\frac{\partial \left(\frac{\sigma}{S}\right)^2_{\text{theo}}}{\partial p_\sigma}\frac{\partial p_\sigma}{\partial(\sec\theta)}\right)^2 (\Delta\sec\theta)^2\,. \tag{144}$$

The values for $\Delta(\sec\theta)$ have been taken from Tab. 8. The impact of Δp_N is just

$${\sigma_{p_N}}^2 = \left(\frac{\partial \left(\frac{\sigma}{S}\right)^2_{\text{theo}}}{\partial p_N}\right)^2 (\Delta p_N)^2 = (2\,p_N \Delta p_N)^2\,. \tag{145}$$

The overall error of the theoretical uncertainty thus becomes

$${\sigma_{\text{theo}}}^2 = {\sigma_a}^2 + {\sigma_b}^2 + {\sigma_\theta}^2 + {\sigma_{p_N}}^2\,. \tag{146}$$

Fig. 39 shows the result of the complete error model (121) in a 2-D plot. The shape of (106) along the x-axis can as well be seen as a slight slope due to the $\sec\theta$ dependency of p_σ along the y-axis.

Fig. 40 shows the measured signal fluctuations for the four bins of $\sec\theta$, each one divided into 9 bins of $\bar{S}$ (see Fig. 31). The resulting values of η are displayed in Fig. 41.

Fig. 42 combines the entries of Fig. 41 in a 1-D histogram. Only bins containing data points have been considered, so the distribution only has 28 entries instead of 36 being the number of bins ($4 \cdot 9 = 36$). For a perfect modelization of signal fluctuations, one expects a Gaussian-like behavior with mean of zero and a sigma of 1.
The Gaussian fit yields $\langle\eta\rangle = 0.2 \pm 0.3$ and $\sigma_\eta = 1.6 \pm 0.3$ ($\chi^2/dof = 2.85/2$). $\langle\eta\rangle$ is approximately zero within the fit uncertainties. $\sigma_\eta > 1$ indicates that some few data points do not fit to the theoretical model within their statistical errors. Besides that there is an asymmetry towards positive η.

Nevertheless, the fit error of σ_η is quite large so the theoretical model (121) is still absolutely acceptable. Besides, the Gaussian fit in Fig. 42 is performed to only a few entries with a large bin width. An additional unbinned fit should be performed to avoid this bias.

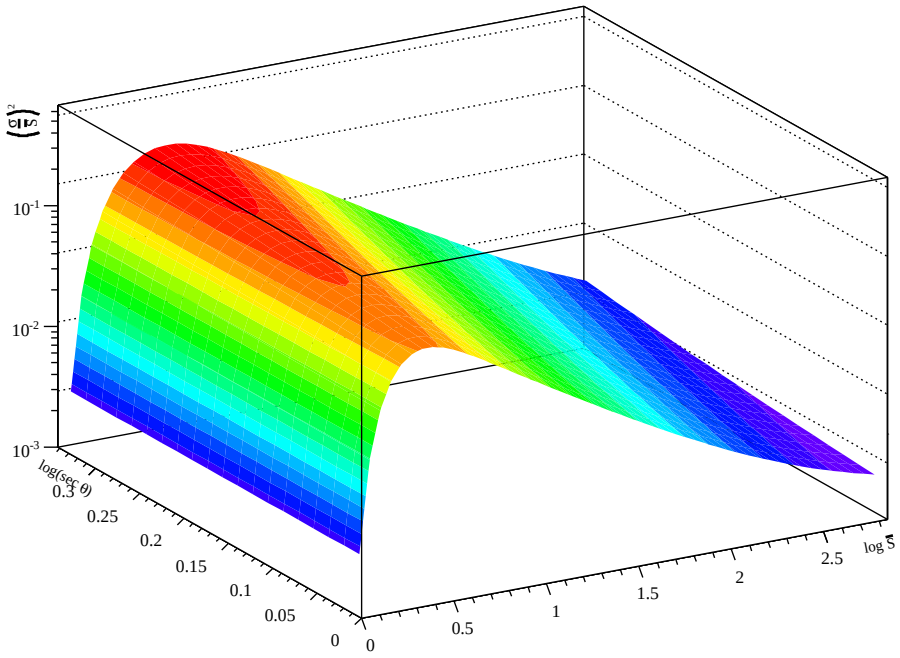


Figure 39: 2-D function plot of the signal fluctuation model (105) with theoretical prediction of the 3.0 VEM threshold.

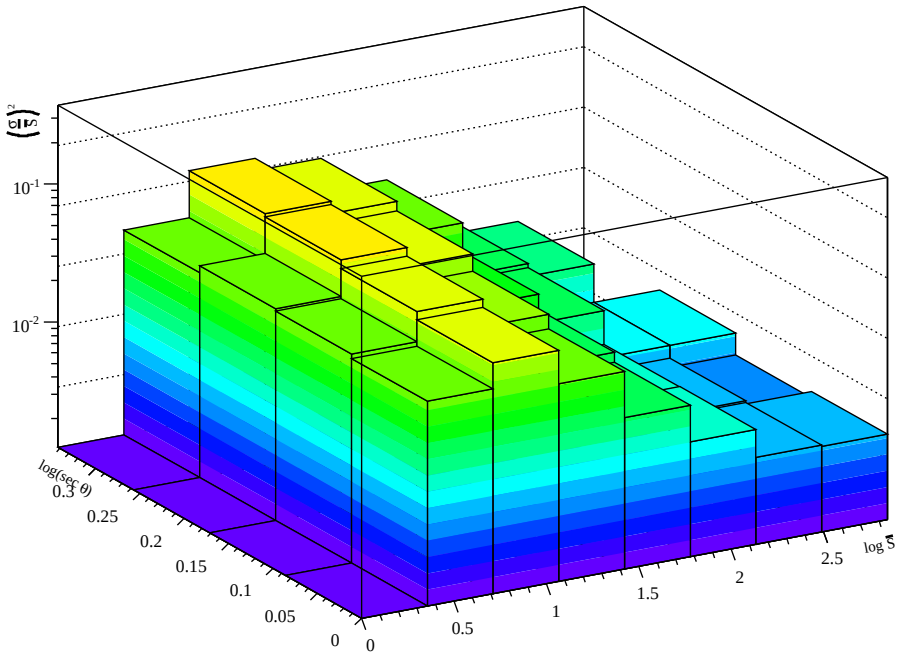


Figure 40: Measured signal fluctuations as a function of $\log \bar{S}$ and $\log(\sec\theta)$.

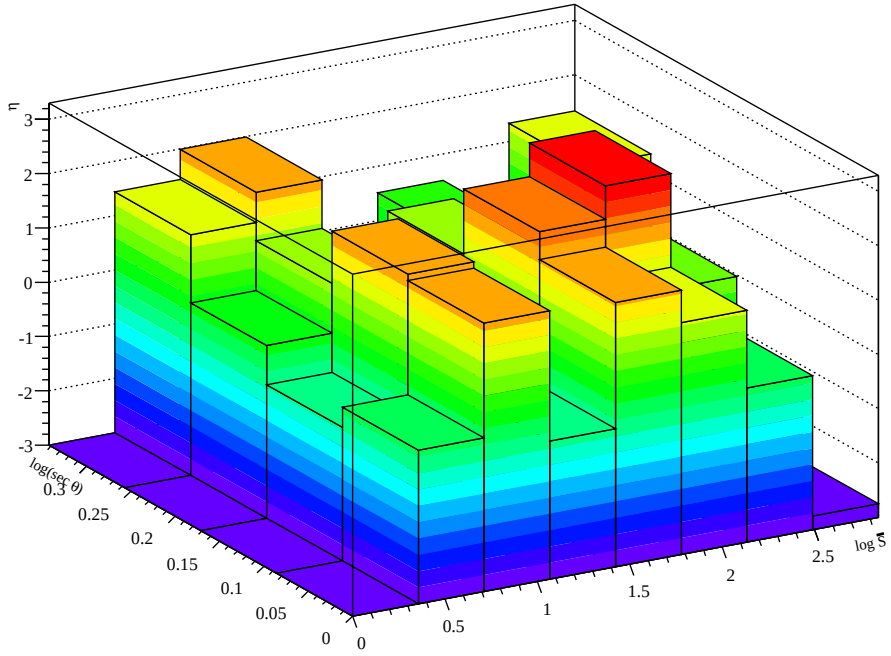


Figure 41: Significance η according to (141).

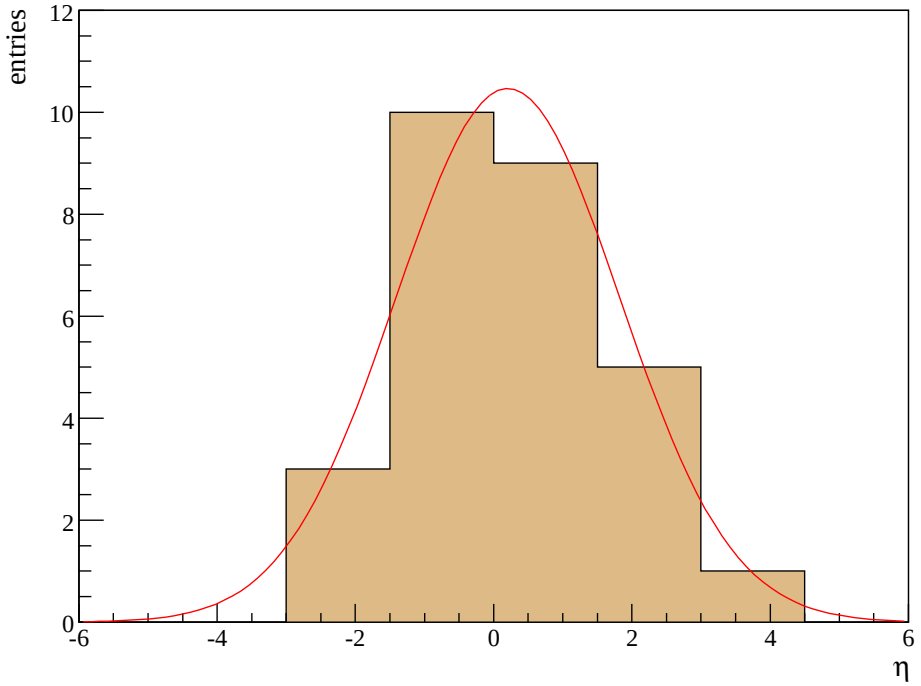


Figure 42: Histogram of all significance values cumulated from Fig. 41.

12 Impact of signal fluctuations onto energy estimation

The primary energy of the particle inducing an EAS can be estimated once the S_{1000} value is obtained from the LDF fit.

The estimation of the LDF will be influenced by signal uncertainties, regardless whether using the χ^2 fit procedure or the maximum likelihood method. In this section, the quantitative influence of signal uncertainties on the estimation of S_{1000} and the primary energy will be investigated.

First, all T5 events are extracted from the pair tank data set described in Sec. 7, yielding 8884 T5 events. Then the `LDFFinder` module is slightly modified to make it possible to control the uncertainties of the station signals and to choose one of the following parameterizations of signal fluctuations:

1. Original (official) parametrization:

$$p_\sigma = 0.32 + 0.42 \sec\theta \tag{147}$$

2. Parameterization using (120):

$$p_\sigma = 0.32 + 0.47 \sec\theta \tag{148}$$

3. Shifting the signals of the candidate stations upwards/downwards by $\left(\frac{\sigma}{S}\right)^2 = \frac{(0.32+0.47\cdot\sec\theta)^2}{S} + (0.039)^2$ which results in

$$S_+ = S\left[1 + \left(\frac{\sigma}{S}\right)\right] \text{ (upwards)} \quad \text{and} \quad S_- = S\left[1 - \left(\frac{\sigma}{S}\right)\right] \text{ (downwards)}\,. \tag{149}$$

The signal uncertainties of the shifted signals are assumed to be

$$\sigma_\pm = S_\pm \cdot \sqrt{\frac{{p_\sigma}^2}{S_\pm} + {p_N}^2}\,. \tag{150}$$

For the S_- case, negative signal values have to be intercepted (a 3 VEM threshold cut is not applied in the T5 event selections as we are not dealing with pair tank analysis here but rather with event reconstruction). The reason is that a Gaussian distribution is no more a good approximation for very low signals and allows the occurence of negative signal values. Instead, Poissonian statistics should have been used but this is not possible because the signal values are real numbers. Stations with $S_- < 0$ VEM are therefore set to 0 VEM (silent station).

The `SdEventSelector` module selected 4145 events from the T5 event data set. Four subsets containing all 4145 energies for the four cases mentioned above (E_{official}, E_{new}, E_+ and E_-) were created.

Fig. 43 shows the relative deviations

$$\left(\frac{\Delta E}{E}\right)_{\text{fluc}} = \frac{E_{\text{new}} - E_{\text{official}}}{E_{\text{official}}} \tag{151}$$

of the reconstruction of the primary energy with respect to the official/updated parameters of (120). The mean energy deviation amounts to $\left\langle \left(\frac{\Delta E}{E}\right)_{\text{fluc}} \right\rangle = (-0.200 \pm 0.008)\%$ with a RMS of $RMS\left(\frac{\Delta E}{E}\right)_{\text{fluc}} = (0.497 \pm 0.005)\%$,

It is remarkable that the reconstructed energy is shifted downwards by up to $\approx 4\,\%$ only due to slightly changed fluctuation parameters. This effect is a systematic influence and will appear by sure. Nevertheless, the maximal energy deviation caused by the new fluctuation parameters is still far below the overall systematic uncertainty of energy reconstruction of $\approx 22\,\%$ [50].

Fig. 45 and 47 show the relative deviations

$$\left(\frac{\Delta E}{E}\right)_{\pm} = \frac{E_{\pm} - E_{\text{new}}}{E_{\text{new}}}\,. \tag{152}$$

The impact onto energy reconstruction by shifting all station signals upwards or downwards is much more drastic. Lowering the signals in a systematic manner leads to a mean energy decrease of $\left\langle \left(\frac{\Delta E}{E}\right)_{-} \right\rangle = (-16.1 \pm 0.2)\%$ ($RMS\left(\frac{\Delta E}{E}\right)_{\text{fluc}} = (10.33 \pm 0.11)\%$). The energy is decreased down to $\approx -95\,\%$.

A similar behavior can be observed for a systematic increase $S \to S_+$ (Fig. 45). We obtain $\left\langle \left(\frac{\Delta E}{E}\right)_{+} \right\rangle = (+15.90 \pm 1.77)\%$ and $RMS\left(\frac{\Delta E}{E}\right)_{\text{fluc}} = (10.80 \pm 0.12)\%$. One event energy is reconstructed $\approx 300\,\%$ above the orginal value.

In a further step, eventual dependencies of deviations of energy reconstruction on primary energy itself are investigated. In Fig. 44 and 46, no clear dependency can be discovered (With correlation factors of $-5.6\,\%$ for $\left(\frac{\Delta E}{E}\right)_{\text{fluc}}$ and $-15.2\,\%$ for $\left(\frac{\Delta E}{E}\right)_{+}$).

Fig. 48 indicates a slight dependency of $\left(\frac{\Delta E}{E}\right)_{-}$ linear to primary energy (correlation factor $19.3\,\%$). This systematic (if really existing) might arise from setting the station signals to zero if $S_- < 0\,\text{VEM}$. Considering Poissonian statistics for very low signals instead of Gaussian statistics would avoid the

problem of signals shifted to the negative, so maybe this effect would then disappear.

Although the energy deviations obtained by shifting signals seem to be very significant, it has to be remembered that we are assuming a systematic behavior of all candidate stations into the same direction. From the statistical point of view, the probability that a particular station has its real signal outside one standard deviation is $1 - 0.683 = 0.317$ [42, p20]. This is approximately one third. The probability that the signal deviates towards one appointed direction is then $p_{\pm} \simeq 0.5 \cdot \frac{1}{3} = \frac{1}{6}$. Hence the overall probability p_{wc} for a "worst case" scenario that all N candidate station exhibit signals shifted into the same direction due to statistical effects is

$$p_{\mathrm{wc}} \approx \left(\frac{1}{6}\right)^N . \tag{153}$$

For a minimum number of $N = 3$ candidates, this worst case probability becomes $p_{\mathrm{wc}} \approx \frac{1}{108} = 0.9\,\%$. About one in hundred 3-fold events may therefore exhibit a very large energy deviation.

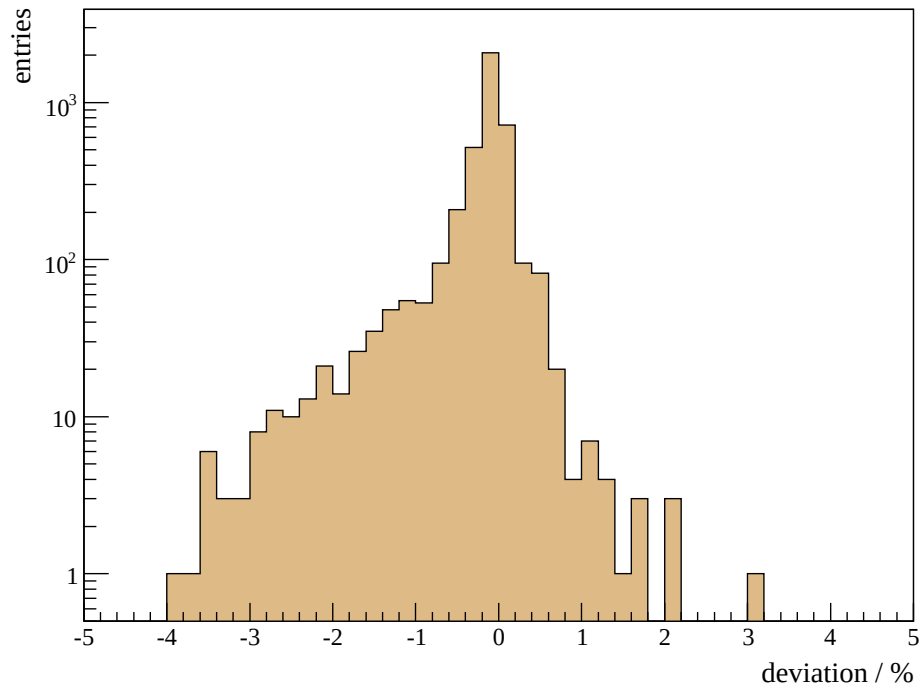


Figure 43: Relative deviations of the reconstructed primary energy with the official parameters of (120) compared to the parameters obtained in this thesis.

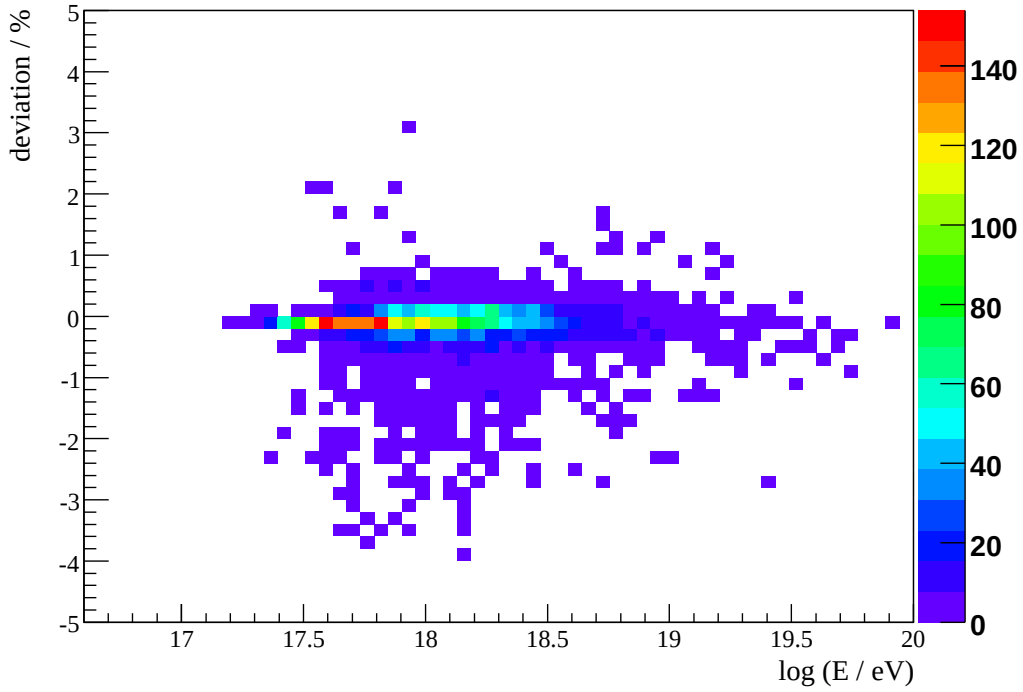


Figure 44: Relative energy deviations vs. energy for the new/official parameters of (120).

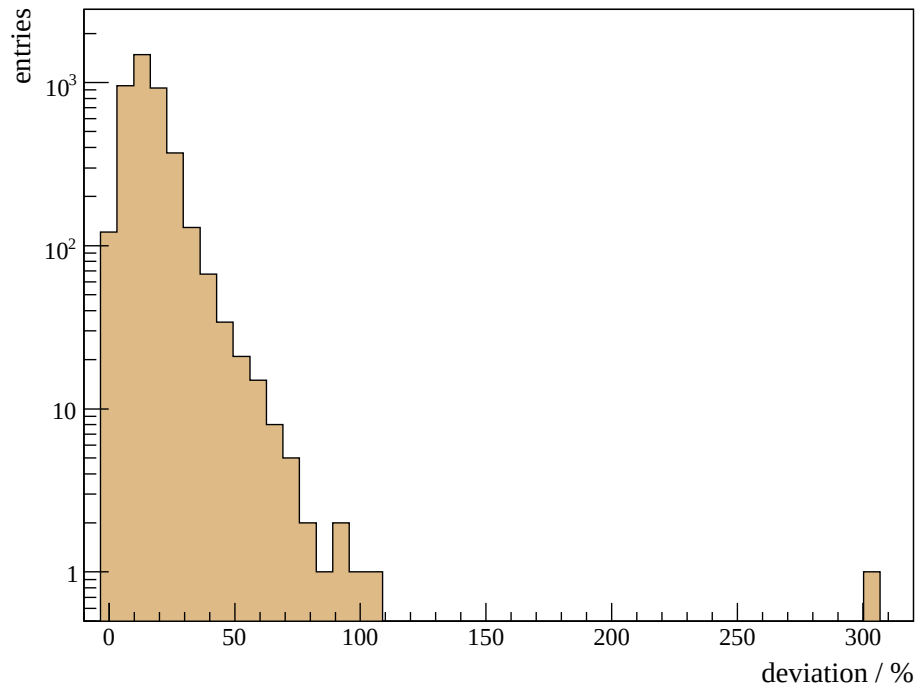


Figure 45: Relative deviations of the reconstructed primary energy after shifting the signals of the candidate stations upwards.

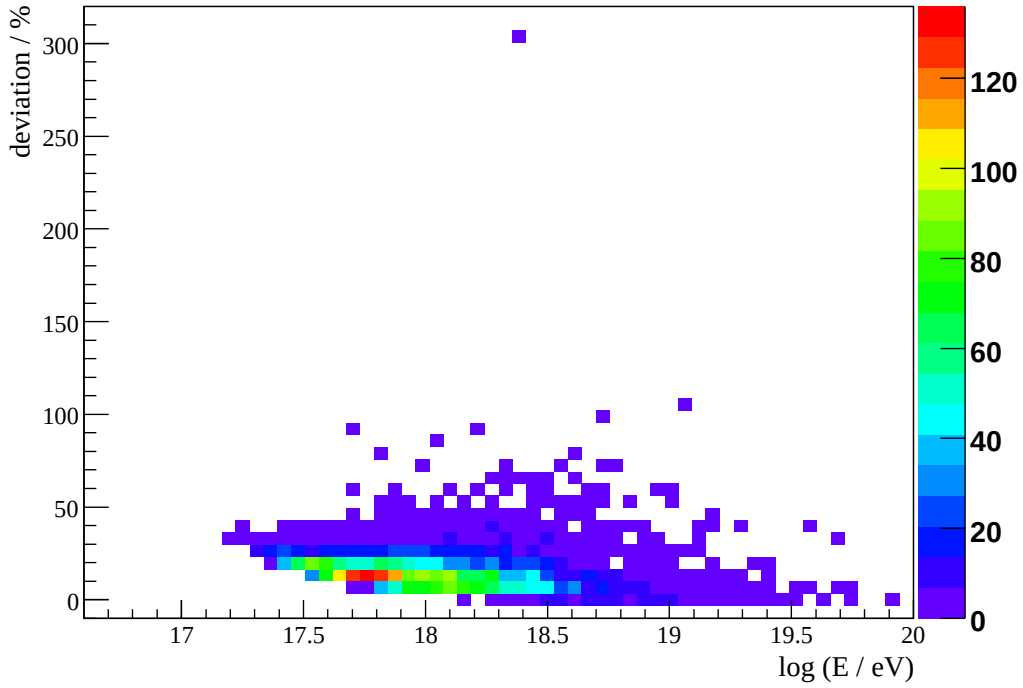


Figure 46: Distribution of the relative energy deviations vs. energy for signals shifted upwards.

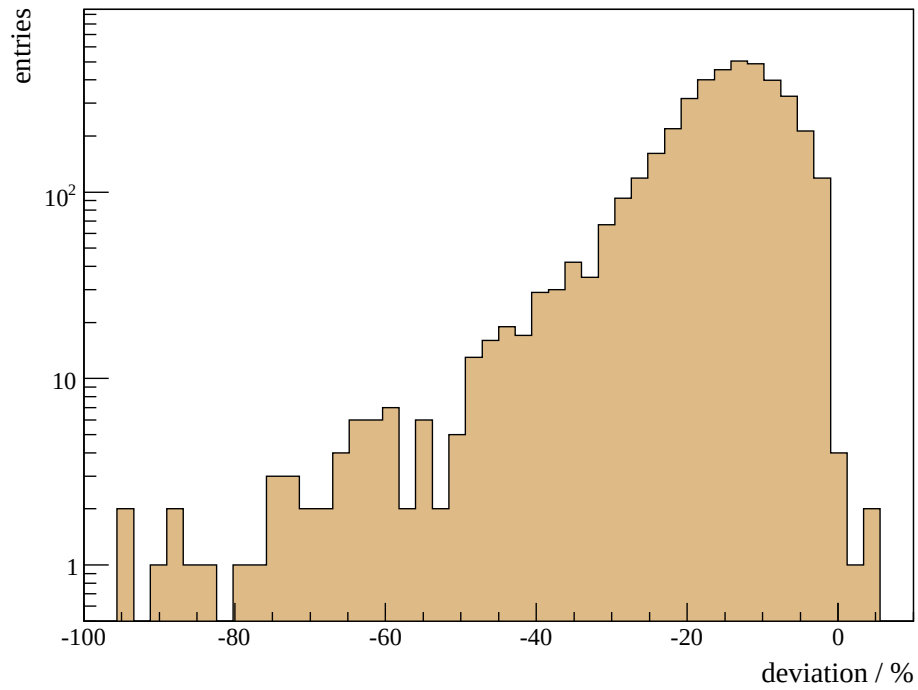


Figure 47: Relative deviations of the reconstructed primary energy after shifting the signals of the candidate stations downwards.

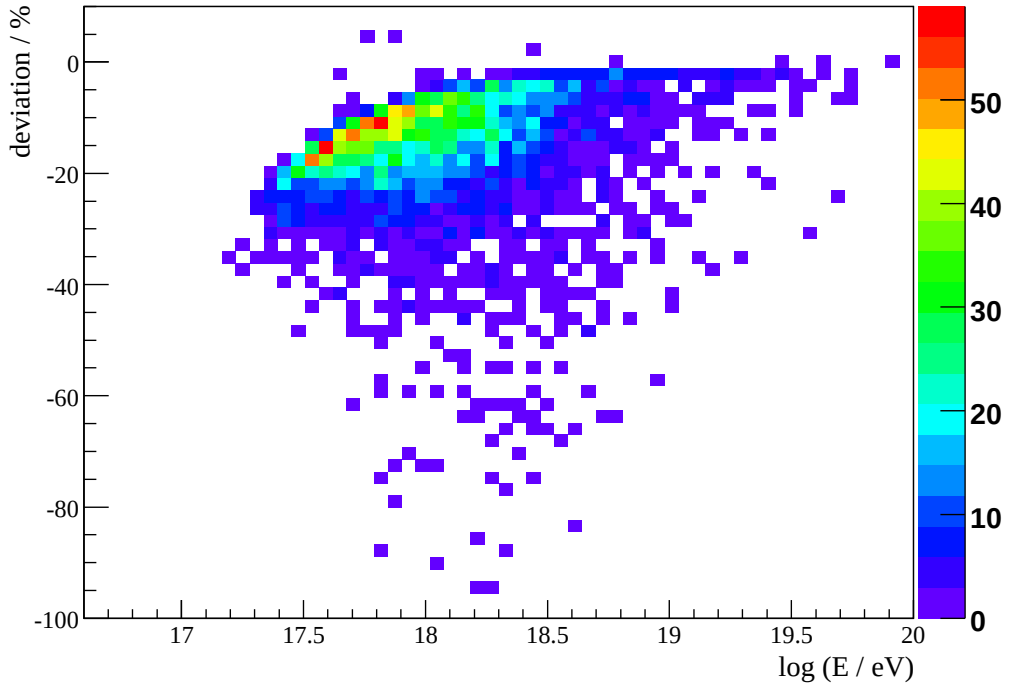


Figure 48: Distribution of the relative energy deviations vs. energy for signals shifted downwards.

13 Summary

In this thesis, signal fluctuations of the Pierre Auger surface detector array were investigated by the help of using signals from pair tanks in the range of $3.0\,\mathrm{VEM} \leq S \leq 4601.4\,\mathrm{VEM}$ and covering a range of zenith angles of $1.0 \leq \sec\theta \leq 2.2$.

A data set of 16425 T4 air shower events, yielding 29218 pair tank samples, was created containing Auger data from October 2006 until April 2008. The six official shower and tank cuts proposed in [40] and four further cuts were applied to the data.

In a first step, the zeroth order model was considered for the purpose of motivation, assuming that the signals of the tanks vary only due to Poissonian statistics. The zeroth order model demonstrated that a Poissonian statistics overestimates the measured fluctuations:

$$\sigma_{\mathrm{rel}} = 0.852 \pm 0.005 \quad (\chi^2/dof = 214.5/34)\,.$$

This result indicated that a conversion factor p_σ is required that converts the measured signal to the corresponding number of particles.
The signal fluctuations $\left(\frac{\sigma}{\bar{S}}\right)^2$ were measured as a function of the average signal $\bar{S}$. In the range of $10.0\,\mathrm{VEM} \leq \bar{S} \leq 80.0\,\mathrm{VEM}$, a model assuming a Poissonian-like behavior was fitted, yielding

$$\left(\frac{\sigma}{\bar{S}}\right)^2 = \frac{(0.903 \pm 0.013)^2}{\bar{S}} \quad (\chi^2/dof = 13.7/4)\,.$$

The signal fluctuations exhibit a steep clipping of towards low signals down to the applied signal threshold of $\bar{S}_{\mathrm{th}} = 3.0\,\mathrm{VEM}$. This clipping was predicted to stem from the threshold effect, assuming that the recorded signals fluctuate around the average signal in a Gaussian-like behavior. A model that describes the threshold effect (105) was developed in the framework of this thesis. A fit of this model to the data yielded

$$\begin{aligned}\left(\frac{\sigma}{\bar{S}}\right)^2 &= \frac{(0.952 \pm 0.007)^2}{\bar{S}} \cdot \bar{P}^2(\bar{S}, 0.952) \\ &\quad (\chi^2/dof = 59.6/7)\,.\end{aligned}$$

Both fluctuation models can be amended by an additional noise constant p_N. Including the noise constant, the fits of the models returned

- *Poissonian-like model:*

$$\begin{aligned} p_\sigma &= 0.934 \pm 0.008\,, && (154)\\ p_N &= 0.034 \pm 0.004\,, && (155)\\ \chi^2/dof &= 42.4/9\,. && (156) \end{aligned}$$

- *Model with prediction of the threshold effect:*

$$\begin{aligned} p_\sigma &= 0.933 \pm 0.007\,, && (157)\\ p_N &= 0.037 \pm 0.004\,, && (158)\\ \chi^2/dof &= 58.2/12\,. && (159) \end{aligned}$$

Both models are able to describe the data well, but only the model including the prediction of the threshold effect can take into account data points below a signal of $\bar{S} \lesssim 10\,\mathrm{VEM}$.

The behavior of the fits of both fluctuation models including the noise constant were investigated for the case that data points are excluded from the lower end of the fit range. Both models showed a consistent behavior.

To test the effect of the 3 VEM threshold cut, a Toy MC simulation was performed. The Toy MC shows the same behavior as the data, the influence of the threshold on the simulation can be well described by the model (105) developed in this thesis.

Furthermore, the dependence of the signal fluctuations of the distance to the shower core was analyzed. A theoretical model based on statistical considerations and exploiting the LDF was established. It turned out that the impact of the LDF effect onto the overall signal fluctuations can be neglected for distances of $r_c \gtrsim 400\,\mathrm{m}$.

A zenith angle dependency of the signal fluctuations proportional to $\sec\theta$ as proposed in [48] could be confirmed:

$$\begin{aligned} \left(\frac{\sigma}{\bar{S}}\right)^2 &= \bar{P}^2(\bar{S}, p_\sigma) \cdot \frac{{p_\sigma}^2}{\bar{S}} + (0.039 \pm 0.003)^2\,,\\ p_\sigma(\theta) &= (0.32 \pm 0.04) + (0.47 \pm 0.03) \cdot \sec\theta\,. \end{aligned}$$

The fit results for the parameters of $p_\sigma(\theta)$ are in good agreement with those published by the Pierre Auger Collaboration [48, p6]. This model is the central result of this thesis as it represents the official parameterization of signal

fluctuations for the SD event reconstruction.

The signal fluctation models were tested and cross-checked. A significance check was performed by comparing the differences between the theoretical model (121) of signal fluctuations with the experimental values. All tests showed that the models of signal fluctuations investigated in this thesis are consistent.

Finally, the influence of signal fluctuations on the estimation of the primary energy was examined. The new fit parameters for (121) gained in this thesis lead to an average drop of the estimation of the primary energy of $\left\langle \left(\frac{\Delta E}{E}\right)_{\text{fluc}} \right\rangle = (-0.200 \pm 0.008)\%$. A maximal energy deviation of $\approx -4\,\%$ could be observed.
By shifting all signals of the candidate stations systematically upwards or downwards by one σ according to (121), an average increase of (10.80±0.12)% and a decrease of $(-16.1 \pm 0.2)\%$ respectively could be observed.

13.1 Outlook

As mentioned in this thesis, one problem that still has to be resolved is the treatment of events with saturated signals in the PMTs. An algorithm to recover the signals of saturated stations reliably is still being developed. The latest work on this topic considers two different approaches: The undershoot method (Variation of the PMT baseline after the signal has attenuated) and the shape time distribution. The signal uncertainties of both methods had been analyzed using twin tanks [51]. Furthermore, it is planned to lower the high voltage (HV) level of the third tank of the triplets to avoid saturations for very high signals [52]. Diminishing the HV level would decrease the PMT gain so a higher number of particles could be detected by the station without being saturated.

By the years, enough statistics will be available to eventually deconvolute the zenith angle and distance to the shower axis dependencies. Up to now, one of the two parametrizations has to be chosen alternatively. As $p_\sigma(r_c)$ is dominating only for very low r_c, the fluctuation model depending on the zenith angle should be chosen in the most cases.

Part III
Appendix

A Calculation of the uncertainties of $R(\bar{S})$

The correction function $R(\bar{S})$ can be obtained by calculating the ratio of the theoretical models for signal fluctuations with and without 3.0 VEM threshold cut:

$$R(\bar{S}) = \frac{\left(\sigma/\bar{S}\right)^2_{\text{no thresh.}}}{\left(\sigma/\bar{S}\right)^2_{3\,\text{VEM thresh.}}} = \frac{\frac{{p_\sigma}^2}{\bar{S}}}{\frac{{p_\sigma}^2}{\bar{S}} \cdot \bar{P}^2(\bar{S},\sigma)} = \bar{P}^{-2}(\bar{S},\sigma)\,. \tag{160}$$

The correction function implies several uncertainties that have to be calculated (for easier readability it is $y := \left(\frac{\sigma}{\bar{S}}\right)^2$):

- The statistical uncertainties of the uncorrected signal fluctuations have to be scaled by $R(\bar{S})$ ($\rightarrow \sigma_{\Delta y}$):

$$\sigma_{\Delta y} = R(\bar{S}) \cdot \Delta y\,. \tag{161}$$

- The statistical uncertainties ΔS of $\bar{S}$ themselves lead to slightly different values of $R(\bar{S})$ when shifting $\bar{S}$ by $\pm\Delta S$ ($\rightarrow \sigma_{\Delta R}$). As $R(\bar{S})$ is strictly monotonic decreasing, the condition $R(\bar{S}+\Delta S) < R(\bar{S}) < R(\bar{S}-\Delta S)$ holds.
  Due to the difference of $R(\bar{S}+\Delta S) \neq R(\bar{S}-\Delta S)$, the uncertainties $\sigma_{\Delta R}$ are asymmetric. One gets

$$\sigma^{\pm}_{\Delta R} = R(\bar{S}) - R(\bar{S} \pm \Delta S) \cdot y\,. \tag{162}$$

  σ_{Δ_R} is also considered to be of statistical kind.

- The fit parameter p_σ has an uncertainty which has to be taken into account by error propagation of $R(\bar{S})$ (second order effect) ($\rightarrow \sigma_{\Delta p_\sigma}$). The impact of Δp_σ on the uncertainty is

$$\sigma_{\Delta p_\sigma} = \left| \frac{\partial R(\bar{S})}{\partial p_\sigma} \Delta p_\sigma \right| \tag{163}$$

  with

$$\frac{\partial R(\bar{S})}{\partial p_\sigma} = -4R(\bar{S})^{-3} \exp\left(-\frac{(S_{\text{th}}-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right)\left(\frac{S_{\text{th}}-\bar{S}}{\sqrt{2\pi{p_\sigma}^4\bar{S}}}\right)\,. \tag{164}$$

The evaluation of $\frac{\partial R(\bar{S})}{\partial p_\sigma}$ requires a longer calculation, see next section.

The overall uncertainties of the corrected signal fluctuations are

$$(\sigma_R)^2 = (\sigma_{\Delta y})^2 + (\sigma_{\Delta R})^2 + (\sigma_{\Delta p_\sigma})^2 \,. \tag{165}$$

A.1 Evaluation of $\frac{\partial R(\bar{S})}{\partial p_\sigma}$

In order to obtain $\sigma_{\Delta p_\sigma} = \frac{\partial R(\bar{S})}{\partial p_\sigma} \Delta p_\sigma$, the differentiation $\frac{\partial R(\bar{S})}{\partial p_\sigma}$ has to be calculated.

From $R(\bar{S}) = \big(1 - 2P(\bar{S}, p_\sigma)\big)^{-2}$, we initially obtain

$$\frac{\partial R(\bar{S})}{\partial p_\sigma} = 4R(\bar{S})^{-3} \frac{\partial P(\bar{S}, p_\sigma)}{\partial p_\sigma} \,. \tag{166}$$

We evaluate

$$\frac{\partial P(\bar{S}, p_\sigma)}{\partial p_\sigma} = \frac{\partial}{\partial p_\sigma} \left(\frac{1}{\sqrt{2\pi {p_\sigma}^2 \bar{S}}} \int_{-\infty}^{S_{\text{th}}} \exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2 \bar{S}} \right) dx \right) . \tag{167}$$

The differentiation of the square root term yields

$$\frac{\partial}{\partial p_\sigma} \left(\frac{1}{\sqrt{2\pi {p_\sigma}^2 \bar{S}}} \right) = -\frac{1}{{p_\sigma}^2 \sqrt{2\pi \bar{S}}} \,. \tag{168}$$

The differentiation of the integral leads to

$$\begin{aligned} & \frac{\partial}{\partial p_\sigma} \int_{-\infty}^{S_{\text{th}}} \exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2 \bar{S}} \right) dx & (169) \\ = & \int_{-\infty}^{S_{\text{th}}} \frac{\partial}{\partial p_\sigma} \exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2 \bar{S}} \right) dx & (170) \\ = & \int_{-\infty}^{S_{\text{th}}} \exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2 \bar{S}} \right) \frac{\partial}{\partial p_\sigma} \left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2 \bar{S}} \right) dx \,. & (171) \end{aligned}$$

The differentiation of the inner term is

$$\frac{\partial}{\partial p_\sigma} \left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2 \bar{S}} \right) = \frac{(x-\bar{S})^2}{{p_\sigma}^3 \bar{S}} \,. \tag{172}$$

The resulting integral $\int_{-\infty}^{S_{\text{th}}} \left(\frac{(x-\bar{S})^2}{{p_\sigma}^3 \bar{S}} \right) \exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2 \bar{S}} \right) dx$ can be simplified by substitution of variables. Let

$$\xi = \frac{x - \bar{S}}{\sqrt{2\bar{S}}p_\sigma} \Rightarrow \exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right) = \exp(-\xi^2)\,. \tag{173}$$

Then it is

$$\sqrt{2\bar{S}}p_\sigma\, d\xi = dx \quad \text{and} \quad \frac{(x-\bar{S})^2}{{p_\sigma}^3\bar{S}} = \frac{2}{p_\sigma}\xi^2\,. \tag{174}$$

Thus we obtain

$$\int \left(\frac{(x-\bar{S})^2}{{p_\sigma}^3\bar{S}}\right) \exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right)\, dx = 2\sqrt{2\bar{S}}\int \xi^2 \exp(-\xi^2)\, d\xi\,. \tag{175}$$

This integral can be solved exploiting integration by parts:

$$\begin{aligned}\int \xi^2 \exp(-\xi^2)\, d\xi &= \int \xi\cdot\xi\exp(-\xi^2)\, d\xi &(176)\\ &= \frac{1}{2}\left(\int \exp(-\xi^2)\, d\xi - \xi\exp(-\xi^2)\right)\,. &(177)\end{aligned}$$

We can now reinsert the definition for ξ:

$$\begin{aligned}&\frac{1}{2}\left(\int \exp(-\xi^2)\, d\xi - \xi\exp(-\xi^2)\right) &(178)\\ = \;&\frac{1}{2}\left(\frac{1}{\sqrt{2\bar{S}}p_\sigma}\int_{-\infty}^{S_{\text{th}}} \exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right)\, dx + \ldots\right.\\ &\left.\ldots - \left[\frac{x-\bar{S}}{\sqrt{2\bar{S}}p_\sigma}\exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right)\right]_{-\infty}^{S_{\text{th}}}\right)\,.\end{aligned}$$

It is

$$\lim_{x\to-\infty}\frac{x-\bar{S}}{\sqrt{2\bar{S}}p_\sigma}\exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right) = 0\,, \tag{179}$$

thus it remains

$$\left[\frac{x-\bar{S}}{\sqrt{2\bar{S}}p_\sigma}\exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right)\right]_{-\infty}^{S_{\text{th}}} = \frac{S_{\text{th}}-\bar{S}}{\sqrt{2\bar{S}}p_\sigma}\exp\left(-\frac{(S_{\text{th}}-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right)\,. \tag{180}$$

Inserting this into (169) together with (175) yields

$$\begin{aligned} & \frac{\partial}{\partial {p_\sigma}} \int_{-\infty}^{S_{\text{th}}} \exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right) dx \qquad (181) \\ = \; & \frac{1}{p_\sigma}\left[\int_{-\infty}^{S_{\text{th}}} \exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right) dx - \exp\left(-\frac{(S_{\text{th}}-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right)\left(S_{\text{th}}-\bar{S}\right)\right] \\ = \; & \frac{1}{p_\sigma}\left[\sqrt{2\pi{p_\sigma}^2\bar{S}}\cdot P(\bar{S},p_\sigma) - \exp\left(-\frac{(S_{\text{th}}-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right)\left(S_{\text{th}}-\bar{S}\right)\right]. \end{aligned}$$

With this result combined with (168), we can now evaluate (167). For better readability, let the square root term of (167) be u and the integral be v. Then we have

$$\frac{\partial(u\cdot v)}{\partial p_\sigma} = \frac{\partial u}{\partial p_\sigma}\cdot v + u\cdot\frac{\partial v}{\partial p_\sigma} \qquad (182)$$

with

$$\begin{aligned} \frac{\partial u}{\partial p_\sigma}\cdot v &= -\frac{1}{{p_\sigma}^2\sqrt{2\pi\bar{S}}}\int_{-\infty}^{S_{\text{th}}} \exp\left(-\frac{(x-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right) dx \qquad (183) \\ &= -\frac{\sqrt{2\pi{p_\sigma}^2\bar{S}}}{{p_\sigma}^2\sqrt{2\pi\bar{S}}}\cdot P(\bar{S},p_\sigma) = -\frac{P(\bar{S},p_\sigma)}{p_\sigma} \end{aligned}$$

and

$$\begin{aligned} u\cdot\frac{\partial v}{\partial p_\sigma} &= \frac{1}{\sqrt{2\pi{p_\sigma}^2\bar{S}}}\cdot\frac{1}{p_\sigma}\left[\sqrt{2\pi{p_\sigma}^2\bar{S}}\cdot P(\bar{S},p_\sigma) + \ldots\right. \qquad (184) \\ &\quad \left.\ldots - \exp\left(-\frac{(S_{\text{th}}-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right)\left(S_{\text{th}}-\bar{S}\right)\right] \\ &= \frac{P(\bar{S},p_\sigma)}{p_\sigma} - \exp\left(-\frac{(S_{\text{th}}-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right)\left(\frac{S_{\text{th}}-\bar{S}}{\sqrt{2\pi{p_\sigma}^4\bar{S}}}\right). \end{aligned}$$

Finally, we can write

$$\frac{\partial P(\bar{S},p_\sigma)}{\partial p_\sigma} = -\exp\left(-\frac{(S_{\text{th}}-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right)\left(\frac{S_{\text{th}}-\bar{S}}{\sqrt{2\pi{p_\sigma}^4\bar{S}}}\right) \qquad (185)$$

and hence

$$\frac{\partial R(\bar{S})}{\partial p_\sigma} = -4R(\bar{S})^{-3}\exp\left(-\frac{(S_{\text{th}}-\bar{S})^2}{2{p_\sigma}^2\bar{S}}\right)\left(\frac{S_{\text{th}}-\bar{S}}{\sqrt{2\pi{p_\sigma}^4\bar{S}}}\right). \qquad (186)$$

References

[1] T. Stanev
High Energy Cosmic Rays
Springer Press, 2^{ns} edition (2003)

[2] Viktor F. Hess
Physik. Zeitschr. 13, 1084 (1912)

[3] Official web page of the Nobel Foundation (2008)
`http://nobelprize.org`

[4] P. Auger
Extensive Cosmic-Ray Showers
Rev. Mod. Phys. 11, 288 - 291 (1939)

[5] H. V. Klapdor-Kleingrothaus, K. Zuber
Teilchenastrophysik
Teubner Press (1997)

[6] R. Schlickeiser
Cosmic Ray Astrophysics
Springer Press (2002)

[7] W.-M. Yao et al. (Particle Data Group)
G. Phys. G 33, 1. (2006)
`http://pdg.lbl.gov`

[8] Peter K. F. Grieder
Cosmic rays at Earth
Researcher's Reference Manual and Data Book
Elsevier Press (2001)

[9] O. C. Allkofer
Introduction to Cosmic Radiation
Karl Thiemig Press (1975)

[10] C. Grupen
Astroparticle physics
Springer Press (2005)

[11] *Le plus grand Détecteur de Rayons Cosmiques au Monde* (french website, 2008)
`http://lpnhe-auger.in2p3.fr/`

[12] K. Greisen
End to the Cosmic-Ray Spectrum?
Phys. Rev. Lett. 16, 748 (1966)

[13] The Auger Collaboration
The Pierre Auger Observatory Design Report
2nd edition (1997)

[14] M. G. Gomez, I. Allekotte, X. Bertou
The Auger Surface Detector acceptance from real data
GAP note 2004-092 (2004)

[15] German webpage of the PAO (2008)
`http://www.auger.de`

[16] The Pierre Auger Collaboration
Correlation of the highest-energy cosmic rays with thepositions of nearby active galactic nuclei
Astroparticle Physics 29 (2008), 188

[17] S. Argiro
Performance of the Pierre Auger Fluorescence detector and Analysis of well reconstructed events
28th International Ray Conference, 333 (2003)

[18] S. Y. Ben Zvi
The Lidar system of the Pierre Auger observatory
Nucl. Instr. Meth. A574 (2007) 171-184

[19] "Little brother" - Array status (2008)
`http://lb.auger.org.ar/SD/`

[20] B. Fick
The Central Laser Facility at the Pierre Auger Observatory
submitted to 29th ICRC Pune India (2005)

[21] John A.J. Matthews and Miguel A. Mostafa
APF light sources for the Auger Southern Observatory
28th International Cosmic Ray Conference (2003)

[22] International webpage of the PAO (2008)
`http://www.auger.org`

[23] I. Alekotte et al.
The Surface Detector Tank of the Pierre Auger Observatory
arXiv:0712.2832v1 (2007)

[24] D. Barnhill et al.
Testing of Photomultiplier Tubes for Use in the Surface Detector of the Pierre Auger Observatory
Subm. to Nucl. Instr. Meth. (2007)

[25] K.-H. Kampert et al.
The Front-End Cards of the Pierre Auger Surface Detectors: Test Results and Performance in the Field
Astron. Nachr. / AN 328 (2007), No. 7

[26] John D. Jackson
Klassische Elektrodynamik
de Gruyter Press, 4th edition (2006)

[27] P. Billoir
Analytic Evaluation of the Number of Photo-electrons per VEM and the Decay Time in Water Tanks
GAP-note 2007-030 (2007)

[28] X. Bertou et al.
Calibration of the surface array of the Pierre Auger Observatory
Nucl. Instr. Meth. A568 (2006) 839-846

[29] X. Bertou
Calibration and Monitoring of the Pierre Auger Surface Detectors
28th International Cosmic Ray Conference (2003) 00, 813-816

[30] D. Allard et al.
The trigger system of the Pierre Auger Surface Detector: operation, efficiency an stability
29th International Cosmic Ray Conference (2005) 00, 101-106

[31] S. Argiró et al.
The Offline Software Framework of the Pierre Auger Observatory
Subm. to Proc. CHEP 04, September 27 - October 1, 2004, Interlaken, Switzerland (2004)

[32] D. Veberič, M. Roth
SD Reconstruction (Offline Reference Manual)
GAP-note 2005-035 (2005)

[33] M. Ave, P. Bauleo, T. Yamamoto
Signal Fluctuation in the Auger Surface Detector Array
GAP-note 2003-030 (2003)

[34] A. Etchegoyen
Track Geometry and Smearing of the Bump Calibration
GAP-note 2002-078 (2002)

[35] J. W. Cronin
Carmen Miranda's Beautiful Song
GAP-note 2001-040 (2001)

[36] A. Castellina
Accuracy of signal measurements in the Auger Surface Detector
GAP-note 2003-031 (2003)

[37] XML Station list of the Auger Offline software v2.0
`share/auger-offline/config/SStationList.xml`

[38] XML Station list of the Auger Offline software v2.4p1
`share/auger-offline/config/SStationList.xml`

[39] M. Ave, A. S. Chou, T. Yamamoto
Signal Fluctuations in the Auger Surface Detector: New Result using the Preproduction Array
GAP-note 2004-025 (2004)

[40] P. Bauleo, A. Castellina, R. Knapik, G. Navarra, J. Harton
Auger Surface Detector Signal Accuracy from Production Tanks
GAP-note 2004-047 (2004)

[41] USGS (U. S. Geological Survey), Eastern Geographic Science Center (2008)
`http://erg.usgs.gov`

[42] T. Fließbach
Statistische Physik
Spektrum Akademischer Verlag, 3rd edition (1999)

[43] D. Allard, M. Ave, N. Busca et al.
A guide-line to the Auger-Surface-Detector Analysis
GAP-note 2006-024 (2006)

[44] C. B. Bonifazi et al.
Response of a Water Cherenkov Detector to Oblique and Quasi-Horizontal Muons
GAP-note 2001-018 (2001)

[45] A. Etchegoyen et al.
Muon-track studies in a water Cherenkov detector
Nucl. Instr. Meth. A545 (2005) 602-612

[46] Denis Allard et al.
A guide-line to Auger-Surface-Detector Analysis
GAP-note 2006-024 (2006)

[47] Internal e-mail conversation with T. Yamamoto and M. Ave
March $26^{\text{th}}/28^{\text{th}}$ 2008

[48] M. Ave, P. Bauleo, A. Castellina et al.
The accuracy of signal measurements with the water-Cerenkov detectors of the Auger Observatory
Subm. to Nucl. Instr. Meth. A (2006)

[49] I. Allekotte
The Pierre Auger Observatory: Present Status and Future Prospects
GAP-note 2003-099 (2003)

[50] B. R. Dawson
Hybrid Performance of the Pierre Auger Observatory
30^{th} International Cosmic Ray Conference (2007)

[51] M. Aglietta, G. Navarra et al.
Recovery of Saturated Signals of the Surface Detector
GAP-note 2008-030 (2008)

[52] Internal e-mail conversation with G. Navarra, P. Bauleo, A. Castellina et al.
June - August 2008

[53] D. Supanitsky, X. Bertou
Semi-Analytical Model of the Three Fold Charge Spectrum in a Water Čerenkov Tank
GAP-note 2003-113 (2003)

Acknowledgements

I would like to use the opportunity to thank the crowds of people devoid of their support the creation of this thesis would not have been possible, especially

Ivor Fleck for his time-consuming supervision

Thomas Bäcker, Rodica Tcaciuc, Peter Buchholz, Stefan Grebe and **Claus Grupen** for their scientific support

Hendrik Czirr, Rodica Tcaciuc and **Thomas Bäcker** for taking time for reading and commenting this thesis before submission

Uwe Fröhlich and **Hendrik Czirr** for the funny moments

Monika Wingerath, Philip Wingerath, Helena Wingerath, Irmgard Schmidt, Thurid Mannel, Sebastian Ochel, Thomas Dietershagen, Evi Schönhofen, Diana Hector, Katrin Jänen, Viola Sidenstein, Sandra Wächtler, Ralf Collmann, Claudia Kronz, Tobi (Ramone) Müller, Volker Krenzer, André Ortmann, Mario Franz, Michael Scharun for support, relief, relaxation, advise and motivation

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Masterarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate und Ergebnisse Anderer kenntlich gemacht habe.

(Ort, Datum) (Unterschrift)