Reconstruction of Vertical Events Recorded by the Surface Detector of the Pierre Auger Observatory

Daniela Mockler[*a] **for the Pierre Auger Collaboration**[†b]

[a]*Université Libre de Bruxelles, Brussels, Belgium*
[b]*Observatorio Pierre Auger, Av. San Martín Norte 304, 5613 Malargüe, Argentina*
 E-mail: auger_spokespersons@fnal.gov
 Full author list: http://www.auger.org/archive/authors_icrc_2019.html

At ground level, surface detector arrays sample the secondary particles from extensive air showers induced by primary cosmic rays with ultra-high energy interacting with the Earth's atmosphere. The recorded events are used to study the energy spectrum, arrival directions, and mass composition of the cosmic rays.

The surface detector of the Pierre Auger Observatory comprises 1660 water-Cherenkov stations deployed on a tessellated triangular grid. Each station is filled with 12 tons of ultra-purified water viewed by three 9-inch photomultipliers, which record the Cherenkov light produced mainly by electrons and muons entering the detector. Physics events are selected by exploring the temporal and spatial combination of the individual detector information in real time. The reconstruction algorithm determines the arrival direction and the energy of the primary particle, and will be described for selected events with zenith angles less than $60°$ (a dataset referred to as *vertical*). The presented reconstruction procedure has been successfully applied to measured and simulated showers.

---

[*]Speaker.

[†]for collaboration list see PoS(ICRC2019)1177

1. Introduction

The main objective of the Pierre Auger Observatory is the measurement of the energy spectrum, mass composition, and arrival direction of ultra-high energy cosmic rays. The manifold studies profit from the hybrid detector design, combining a large surface detector array with a fluorescence detector [1]. The flux of cosmic rays at the highest energies is as low as one particle per square-kilometer per steradian per century, and as such has to be measured by ground-based detectors. Upon entering the Earth's atmosphere, a primary cosmic ray initiates an extensive air shower. The fluorescence detector (FD) observes the longitudinal development of the shower by recording the fluorescence light emitted along the shower path. The surface detector (SD) measures the lateral distribution at the ground by sampling the secondary particles of the air shower with water-Cherenkov stations. Due to the tessellated array layout and a spacing of 1500 m between the stations, the threshold for full efficiency is reached for primary energies greater than 3×10^{18} eV. A smaller, nested array in the north-west of the Observatory with half of the original spacing extends the threshold of full efficiency down to 3×10^{17} eV [2, 3]. The whole array spans an area of 3000 km^2. The duty cycle of the surface detector is nearly 100%, while the uptime of the fluorescence detector is limited to clear and almost moonless nights. Thus, the bulk of data is provided by the SD.

In this contribution we present the reconstruction chain for the data measured with the SD, including the data-taking in individual stations, the trigger system deployed to select air shower events, and the reconstruction of these events for zenith angles up to 60° with the corresponding reconstruction uncertainties.

2. Surface detector

The SD array encompasses 1600 water-Cherenkov stations. Each station has a base of 10 m^2 and is filled with 12 tons of purified water up to a height of 1.2 m. The water volume, which is contained inside a diffusively-reflective liner, is viewed from the top by three 9-inch photomultiplier tubes which record the Cherenkov light produced by relativistic charged particles traversing the water. Each station is equipped with a solar photovoltaic system which provides power for the electronics and communication system, thus making the station self-contained. Due to the limited total bandwidth of only 1200 bits/s for data transmission from the stations to the central data acquisition system (CDAS), the calibration of each station is done automatically by the local electronics. The calibration is based on recording the average charge observed by a PMT for through-going muons from the atmospheric background. Each PMT provides two raw ADC traces for which an individual baseline has to be determined. One signal is directly taken from the anode, the other one is taken from the last dynode after an amplification and inversion within the PMT base electronics to a total nominal signal of 32 times the anode signal. This readout strategy allows for a broad dynamic range. An example of such a signal trace is shown in Fig. 1a. In every recorded trace, a reasonable signal range needs to be selected by defining a start and stop time for each signal trace, as indicated by the dashed vertical lines. For the calibration, a 61 s worth sample of background signals is stored together with the event data. From this, a histogram of integrated charges is created, as depicted in Fig. 1b. The position of the second maximum corresponds to the signal
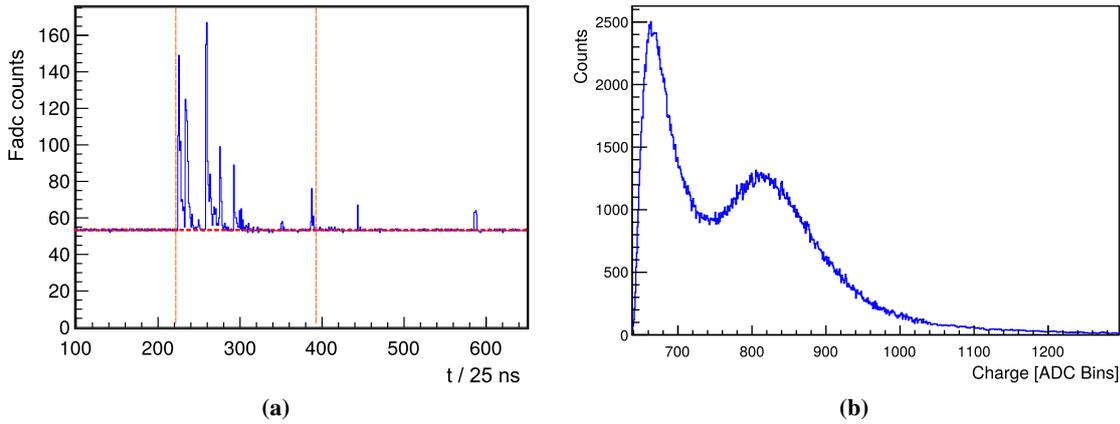
Figure 1: (a) Example of a raw ADC trace. The baseline is shown as horizontal dashed line, while the start and stop times of the signal are indicated by the dashed vertical lines. (b) Each ADC signal is calibrated into units of VEM. The calibration factor is extracted from the position of the maximum in the charge histogram (here at ∼800 ADC counts).

of background muons and is used to convert the ADC signals into the reference units of a Vertical Equivalent Muon (VEM), which is used as common reference level for the station signals [4]. After subtracting the baseline and calibrating the signal pieces into VEM units, the total station signal is then obtained as the average of the individual PMT signals. When a shower core falls close to a station, the dynamic range of the electronics is not sufficient and the signal saturates. If only the dynode channel saturates, the signal is recovered by reading out the anode channel. For a saturation of the anode channel, a signal-recovery algorithm is applied [5]. Based on the spatial and temporal combination of the first-level station triggers, an array trigger is formed by CDAS and all data satisfying this trigger are permanently stored [6].

3. Data selection and event building

The ingredients for the subsequent shower reconstruction are the start time and the size of the station signals with their respective uncertainties. The uncertainties on the signal start time and the total station signal are determined by analyzing data recorded with doublet stations, a subset of SD stations equipped with partner stations. Being only ∼11 m apart, the two stations of a doublet sample basically the same part of the air shower. The signal uncertainty can be described by a Poisson-like parameterization, including a dependence on the zenith angle [7]. The station signals and their uncertainties are used in the reconstruction of the lateral distribution function at the ground. The uncertainty of the signal start time affects the angular resolution. The main uncertainty stems from the uncertainty in the arrival time of the shower particles due to the thickness of the shower front. The particle density in the shower front decreases with increasing distance to the shower core and thus fewer particles enter the stations, resulting in a delayed start time of the signal.

At this stage, the events stored at CDAS not only contain real shower events but also accidental events. To select only the events due to air showers, an additional trigger level is applied offline. It requires at least three adjacent stations, whose signal start-times are compatible with a planar

shower-front moving at the speed of light. In next steps, the list of possible event stations has to be cleared from accidental stations, whose signal times coincide by chance. For this, a seed of three neighboring stations with largest signals is required to determine a first approximation of the development of the shower plane,

$$c(t_i - t_1) = -\hat{a}(\vec{x}_i - \vec{x}_1),$$
(3.1)

where $\vec{x}_1, t_1$ are the location and trigger time of the first seed station, respectively. With this, a provisional shower axis $\hat{a}$ is found. Next, all stations with trigger times not matching this shower front are discarded from the event. In addition, triggered stations which have no triggered neighbor station within 1.8 km or two triggered neighbors within 5 km are removed from the station list. As a last step, all active stations which exhibit no trigger but are otherwise functional are flagged as zero-signal stations and are kept for the subsequent event reconstruction as they provide additional constraints for the reconstruction.

4. Event reconstruction

In the following, the list of stations belonging to an event is used for the reconstruction. The shower geometry is approximated by fitting a plane for the shower front, using the information of the event stations:

$$\chi^2 = \sum_i \frac{(t_i - t_{\text{sh}}(\vec{x}_i))^2}{\sigma_{t_i}^2},$$
(4.1)

with t_{sh} as the expected time the shower front passes the station location $\vec{x}_i$ and σ_{t_i} being the start-time uncertainty of the signal in station i. A more realistic shower model is obtained when allowing for a curvature of the shower front. The true curvature of the shower front is unknown and hence the reconstruction utilizes two slightly different implementations of the shower curvature, a parallel-parabolic and a concentric spherical front-model. Both models are shown in Fig. 2a in addition to the start times of the triggered stations. The signal times are given relative to the planar shower front. The two station sets shown arise due to different modes used to select and calibrate the individual signals.

The reconstruction of the lateral profile of a shower depends on the recorded signal sizes. With increasing distance to the shower axis, the measured signals decrease, which can be described by a lateral distribution function (LDF) as

$$S(r) = S(r_{\text{opt}}) f_{\text{LDF}}(r),$$
(4.2)

with f_{LDF} being normalized such that $f_{\text{LDF}}(r_{\text{opt}}) \equiv 1$. The optimal distance r_{opt} is 1000 m for the surface array [8]. In general, this distance depends on the grid spacing and describes the optimal distance for which the LDF has the least systematic impact on the shower size estimator $S_{1000} \equiv S(r_{\text{opt}})$. As the true LDF is unknown, the functional form of f_{LDF} was chosen empirically. Similar to the implementation of the curvature model, two realizations of the LDF are available in the two algorithms used for analysis of the data,

$$f_{\mathrm{LDF}}^{\mathrm{NKG}} = \left(\frac{r}{r_{\mathrm{opt}}}\right)^{\beta} \left(\frac{r+r_1}{r_{\mathrm{opt}}+r_1}\right)^{\beta+\gamma}, \tag{4.3}$$

$$f_{\mathrm{LDF}}^{\mathrm{LLP}} = \begin{cases} \exp\left(\beta\rho + \gamma\rho^2\right), & r > r_c \\ \exp\left(\beta\rho + \gamma(2\rho - \rho_c)\rho_c\right), & r \leq r_c, \end{cases} \tag{4.4}$$

with $r_1 = 700\,\mathrm{m}$ and $r_c = 300\,\mathrm{m}$, $\rho = \ln\left(r/r_{\mathrm{opt}}\right)$. In Eq. (4.3), the LDF is described by a modified NKG function [9, 10], while Eq. (4.4) uses a log-log parabola. In both realizations, the LDF slope parameters are given as β and γ. The properties of the primary cosmic ray (mass, energy) as well as the shower development in the atmosphere affect the slopes of the LDF. To account for this, β and γ are parameterized respectively for both LDF realizations as function of zenith angle and shower size. Given the measured signals in the stations and their distances to the shower axis obtained in the previous geometry fit, the fit of the LDF is then maximizing the probability that the shower had a certain core position $\vec{x}_c$ and shower size S_{1000} written as likelihood

$$\log\mathscr{L} = \sum_i \log P\left(S_{1000}, \vec{x}_c | S_i, \vec{r}_i\right). \tag{4.5}$$

The probability P includes contributions of both triggered stations with signal and zero-signal stations, whose signals were not sufficient to cause a trigger. The resulting shower size estimator S_{1000} is affected by daily and seasonal atmospheric modulations which change the effective slant depth a shower traversed [11]. In addition, the shower size depends on the azimuth angle of the incident primary particle with respect to the geomagnetic field as charged particles are deflected by the field [12]. To obtain an unbiased shower size estimator both effects are corrected for. In the last step, S_{1000} has to be corrected for attenuation effects in the atmosphere. With increasing zenith angle, the electromagnetic shower component gets more and more attenuated as the amount of traversed atmosphere increases, resulting in a smaller observed shower size at the ground. Using a constant intensity cut method [13], the shower size estimator is converted to the shower size which would have been measured had the shower arrived at an angle of $38°$. This shower size is also referred to as S_{38} and is the unbiased energy estimate of the SD. The SD energy is then obtained by calibrating S_{38} to the energy measured by the FD using a subset of showers that can be reconstructed independently by both detectors [14, 15].

5. Reconstruction uncertainties

5.1 Angular resolution

The angular accuracy depends on the start times of the individual stations, which in turn depend on the particle distribution within the shower front as well as on the timing precision of the GPS. The angular resolution (AR) is investigated with full Monte-Carlo simulations for the two primary masses, proton and iron, using EPOS-LHC as hadronic interaction model. Each shower is distributed several times on the array using random core positions and the reconstructed shower directions $\hat{a}_{\mathrm{rec}}$ are then compared to the true shower axis $\hat{a}$ by calculating the space-angle η between the arrival directions. The distribution of η can be described by a Rayleigh distribution
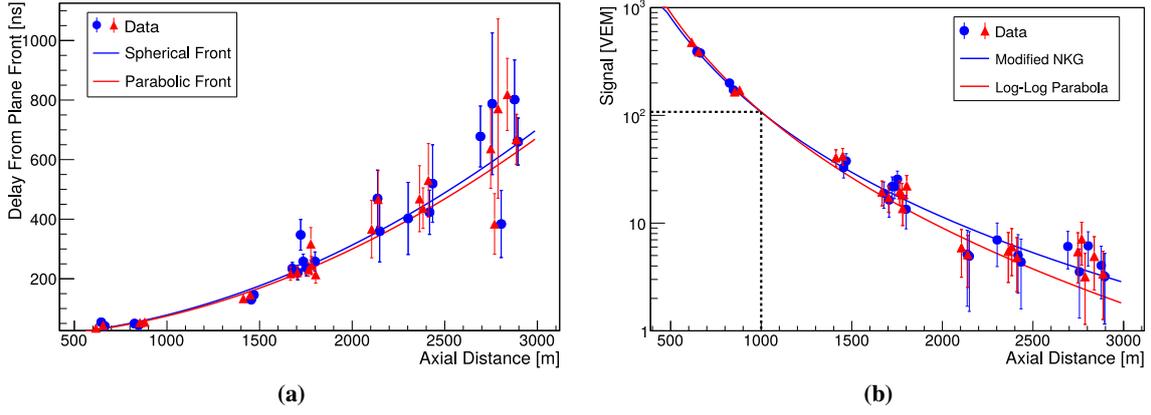
(a)   **(b)**

Figure 2: (a) The start times in individual stations as function of the distance to the reconstructed shower axis. The markers show the relative time difference to a planar shower front. Superimposed are the predicted times for the two implemented shower plane models. (b) The individual station signals are shown as a function of the distance to the shower axis. The dashed lines correspond to the signal extracted at 1000 m, which serves as shower-size estimator. The two functions used to fit the lateral distribution of the signals are given as solid lines.
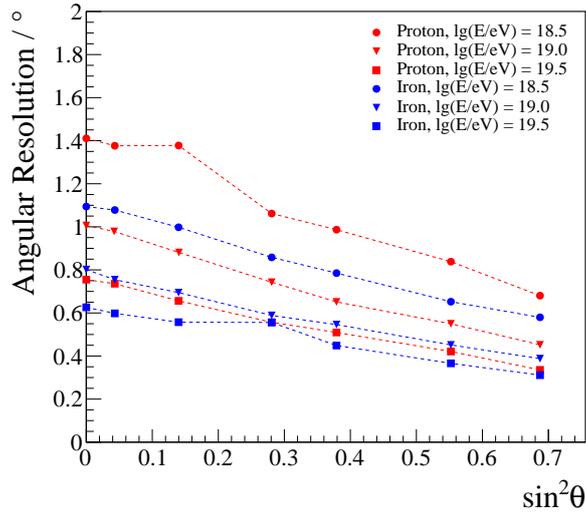

Figure 3: Angular resolution (AR) for the SD as a function of $\sin^2 \theta$. The AR is shown for different energies and two primary masses.

$(\eta/\sigma^2) e^{-\eta^2/2\sigma^2}$. The angular resolution is then defined as the angular radius that would contain 68% of the showers, so $AR = 1.5\sigma(\eta)$. As shown in Fig. 3, at the highest energies the angular resolution is better than $1°$ for all angles and both simulated masses.

5.2 Shower size estimator

The uncertainty of the shower-size estimator can be derived from Monte-Carlo simulations as well as from data. The total uncertainty on S_{1000} from simulations is extracted by comparing the signals in a ring of simulated stations at 1000 m to the reconstructed S_{1000} value given by the LDF

fit. For each simulated energy and zenith angle, the distribution of $\ln\left(S_{1000}^{\text{rec}}/S_{1000}^{\text{true}}\right)$ is fitted to a Gaussian function. The extracted sigma is then the quadratical combination of the statistical and systematic uncertainty of S_{1000}. For the data-driven method, the statistical uncertainty is directly given by the fitting error in S_{1000} arising from the uncertainty of the signals measured in each station. The systematic uncertainty is due to the unknown true LDF shape for a specific shower. For most events, the slope parameter β is parameterized as function of S_{1000} and zenith angle. Only for a subset of showers with many triggered stations, this parameter is a free parameter in the LDF fit. For this subset of showers, the fitted β_{free} is compared to the fixed value of β_{fix} obtained from the aforementioned parametrization. The dispersion σ_β of $\beta_{\text{fix}} - \beta_{\text{free}}$ can be parameterized as function of S_{1000}. Each shower is then reconstructed several times with a value for β sampled from a normal distribution centered at β_{fix} and with a standard deviation of σ_β. The systematic uncertainty on S_{1000} is given by the standard deviation of the resulting S_{1000} values. The statistical and systematic uncertainties are shown in Fig. 4a for two data sets, one set with saturated stations and one without. For both data sets, the statistical uncertainty is the dominant error on S_{1000} for values up to 50 VEM. In Fig. 4b, the S_{1000} values with β_{free} are compared to the ones obtained when using the parameterization for β. The difference is less than 3% and hence well within the systematic uncertainties.

As can be seen from the comparison of the combined statistical and systematic uncertainty of S_{1000} shown in Fig. 4c, both methods to derive the uncertainty provide compatible results. The uncertainty decreases from 15% at a shower size of 10 VEM, where the array is fully efficient for more inclined showers, to 5% at the highest shower sizes.

6. Conclusions

We have presented the reconstruction chain for the data measured with the SD. Due to the lack of knowledge of the true curvature of the shower front and the true LDF shape, the Observatory exploits two functional forms in the reconstruction to describe these shower properties. The agreement of both the shower geometry and the shower-size estimator is excellent for the two reconstruction modes. The angular resolution was investigated with simulations and was found to be better than 1% for the highest energies for both simulated primary masses. The same simulations were used to determine the uncertainty of the shower-size estimator. The uncertainty was found to be better than 10% at high energies for events without saturated stations, which is in agreement with the uncertainty derived from data.

References

[1] A. Aab [Pierre Auger Collaboration], Nucl. Instrum. Meth. A **798** (2015) 172.

[2] D. Ravignani [Pierre Auger Collaboration], Proc. 33rd ICRC.

[3] A. Coleman [Pierre Auger Collaboration], PoS(ICRC2019)225.

[4] X. Bertou [Pierre Auger Collaboration], Nucl. Instrum. Meth. A **568** (2006) 839–846.

[5] D. Veberič [Pierre Auger Collaboration], Proc. 33rd ICRC.

[6] J. Abraham [Pierre Auger Collaboration], Nucl. Instrum. Meth. A **613** (2010) 29.

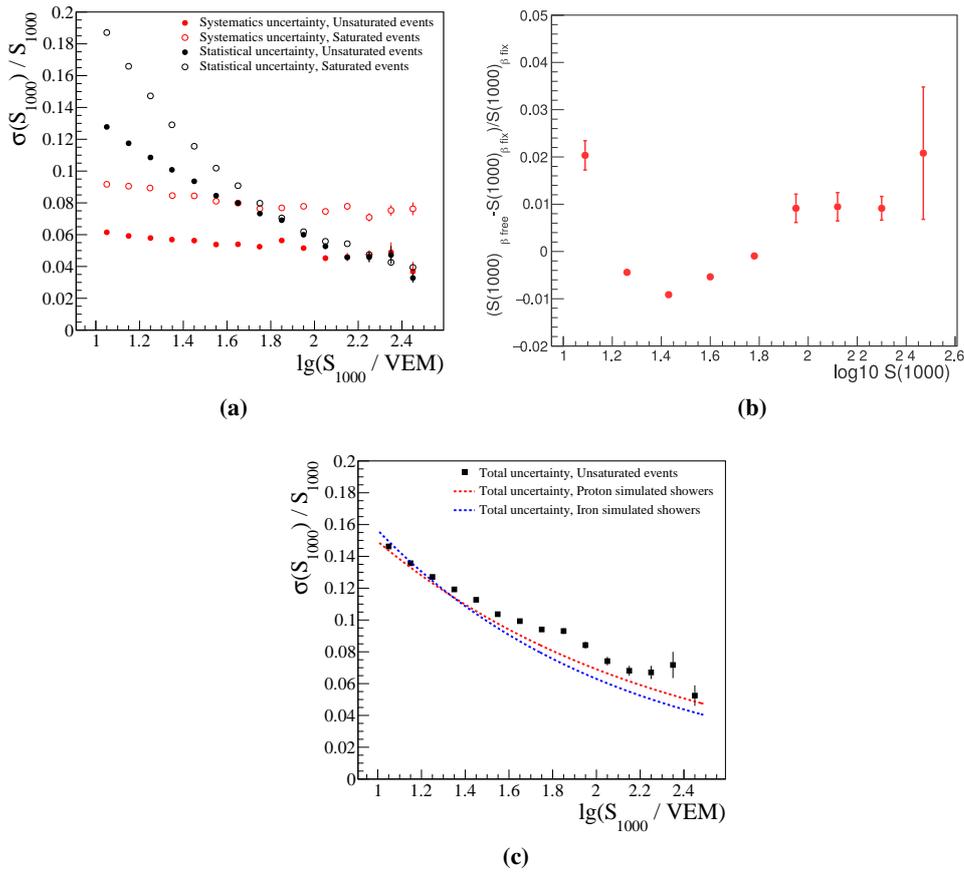[7] M. Ave [Pierre Auger Collaboration], Nucl. Instrum. Meth. A **578** (2007) 180.

PoS(ICRC2019)353

(a)

(b)

(c)

Figure 4: (a) The statistical and systematic uncertainties of the shower size S_{1000} as function of $\lg S_{1000}$. The data are divided into two subsets (with/without saturated stations). (b) Relative difference in S_{1000} for fitting β with respect to using the parameterization of it. (c) The combined statistical and systematic uncertainty derived from data for unsaturated events is compared to the uncertainty obtained from simulations.

[8] D. Newton and J. Knapp and A.A. Watson, Astropart. Phys. **26** (2007) 414–419.

[9] K. Kamata and J. Nishimura, Prog. Theor. Phys. Suppl. **6** (1958) 93.

[10] K. Greisen, Prog. Cosmic Ray Physics **3** (1956) 1.

[11] A. Aab [Pierre Auger Collaboration], JINST **12** (2017) P02006.

[12] P. Abreu [Pierre Auger Collaboration], JCAP **11** (2011) 022.

[13] J. Hersil et al., Phys. Rev. Lett. **6** (1961) 22.

[14] H.P. Dembinski, B. Kégl, I.C. Mariş, M. Roth, and D. Veberič, Astropart. Phys. **73** (2016) 44–51.

[15] V. Verzi [Pierre Auger Collaboration], PoS(ICRC2019)450.