PROCEEDINGS OF SCIENCE

Investigations of CORSIKA Thinning Levels Suitable for Studies of Photon-Hadron Discrimination at Ultra-High Energies

Fiona Ellwanger[a,*] **for the Pierre Auger Collaboration**[b]

[a] *Karlsruhe Institute of Technology, Institute for Astroparticle Physics, Karlsruhe, Germany*
[b] *Observatorio Pierre Auger, Av. San Martín Norte 304, 5613 Malargüe, Argentina*
Full author list: https://www.auger.org/archive/authors_2024_11.html

E-mail: spokespersons@auger.org

Cosmic ray detectors like the 3000 km^2 surface array of the Pierre Auger Observatory are capable of observing high-energy photons in the range of 10^{18} to 10^{20} eV if the flux is sufficiently high. However, no clear candidates for ultra-high-energy photons have been identified yet, so simulations must be used to study typical trigger patterns and observables for discriminating photons from hadrons, e.g., with neural networks. Thinning algorithms are applied to keep the computation time and file sizes in a manageable range since the simulation of ultra-high-energy particle showers is computationally expensive. In Corsika, particles with energies below a certain fraction of the primary energy, the thinning level, are exposed to thinning. In the case of thinning, only one of the particles emerging from an interaction is tracked. By assigning a corresponding weight, this particle then represents a number of its siblings. However, the weights of particles that originate from electromagnetic interactions can be 100 times larger than for hadronic interactions. In contrast to hadronic showers, where a major part of the signal in a surface detector is produced by muons, photon showers are almost purely electromagnetic. Using simulations of photon-induced showers with two different thinning levels, the influence on different observables used for photon-hadron discrimination is investigated. Effects deriving from both statistical sampling and detector simulations are considered. Possible influences on station-level as well as event-level observables are probed. With this study, we are reassured that the optimal thinning parameters determined for hadron-induced showers are also sufficient for photon-induced showers.

7th International Symposium on Ultra High Energy Cosmic Rays (UHECR2024)
17-21 November 2024
Malargüe, Mendoza, Argentina

[*] Speaker

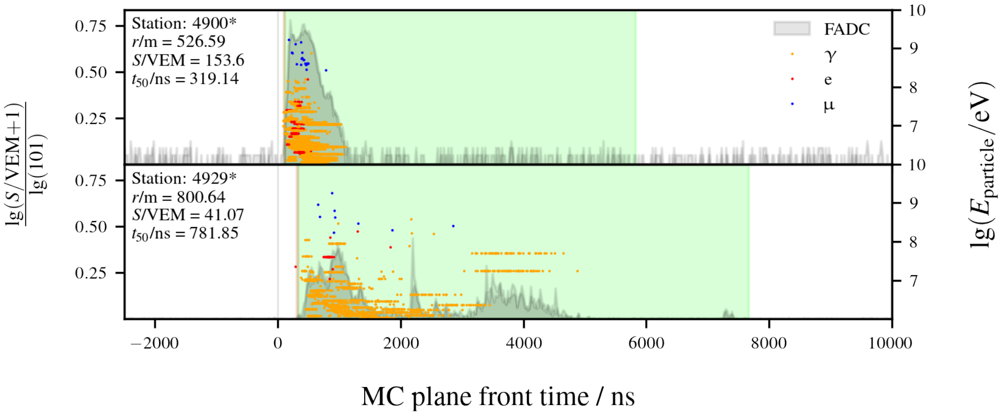

Figure 1: Traces of two stations together with the injected particles for an example event. Lines in the particle energy are clones that were smeared out in time according to a log-normal distribution. In the lower trace artifacts caused by a high number of clones can be observed.

1. Motivation

Thinning algorithms are applied to keep the computation time and file sizes in a manageable range since the simulation of ultra-high-energy particle showers is computationally expensive. In Corsika, particles with energies below a fraction of the primary energy E_prim, defined as the so-called *thinning level*, are exposed to thinning. Only one of these particles emerging from an interaction is tracked, and by assigning a corresponding weight w it then represents w of its siblings [1]. There are several parameters to adjust in CORSIKA that define how the thinning is done,

$f_\mathbf{thin}$**:** Fraction of the primary energy below which the particles are thinned. It is also called the *thinning level*.

$w_\mathbf{max}^\mathbf{Hadr}$**:** Maximum weight for particles from hadronic interactions.

$w_\mathbf{max}^\mathbf{EM}$**:** Maximum weight for particles from electromagnetic interactions.

In the detector simulation, the weights of the particles have to be taken into account by cloning. To avoid artifacts, the resampling algorithm is applied [2]. Each particle with weight w is interpreted as a flux ϕ_w of w such particles through a defined sampling area,

$$A_\mathrm{sam} = 4\,R^2\,\frac{\delta R}{R}\,\delta\phi\,\frac{\cos\theta_\mathrm{p}}{\cos\theta_\mathrm{sh}}, \tag{1}$$

where the sampling area is projected to the plane perpendicular to the particle direction, R is the distance to the shower axis, θ_p, and θ_sh are the zenith angles of the particle and the shower respectively. The size of this sampling area is chosen such that the flux through the effective detector area,

$$A_\mathrm{eff} = \pi\,r_\mathrm{s}^2\,\cos\theta_\mathrm{p} + 2\,r_\mathrm{s}\,h_\mathrm{s}\sin\theta_\mathrm{p}, \tag{2}$$

comprises one particle, where the water-tank radius is $r_\mathrm{s} = 1.8\,\mathrm{m}$ and height is $h_\mathrm{s} = 1.2\,\mathrm{m}$. The particle will then be injected into those detectors overlapping with the sampling area. That way, particles with high weights are more smeared out in the observation level compared to low-weighted particles. The steep lateral distribution of the particles can introduce biases as more weighted particles are going to be smeared away from the shower axis than towards it. To reduce this bias, the sampling area is limited to the ring section ($\delta R/R = 0.05$, $\delta\phi = 0.15 \approx 8.6°$). If the required sampling area exceeds this limit, the flux through the effective detector area is given by $\phi_{w_\mathrm{res}} = w_\mathrm{res}/A_\mathrm{eff}$. The resampled weight is given by,

$$w_\mathrm{res} = \frac{w\,A_\mathrm{eff}}{A_\mathrm{sam}^\mathrm{limit}} > 1, \tag{3}$$

which results in injecting clones. As they are clones, they all have the same type, energy, and zenith angle. To avoid unphysical spikes in the detector signal, the arrival time and position are smeared out for the clones. They are equally distributed over the effective detector area. The arrival time is smeared according to a lognormal distribution. The width depends on the delay of the original particle compared to the shower plane front.

Artifacts caused by clones, see, e.g. Fig. 1, previously had been observed to be rare in proton simulations. Artifacts can occur for very inclined particles when the effective area becomes larger than the sampling area or for particles with very high Corsika weights. In that case, the resampling algorithm does not suffice to reduce the weight.

In Corsika, a limit w_max is set so that the weights of particles can not exceed this number. This limit was found to reduce artificial fluctuations while keeping the computation time low [3]. However, $w_\mathrm{max}^\mathrm{EM} = 100\,w_\mathrm{max}^\mathrm{Hadr}$, which means that the weights of particles that originate from electromagnetic interactions can be 100 times higher than for hadronic interactions. In contrast to hadron-induced showers (HIS), where a significant part of the signal is produced by muons, photon-induced showers (PIS) are almost purely electromagnetic. In the following, we study possible side effects of the thinning level on photon-hadron discrimination.

2. Corsika simulations

Throughout this analysis, a discrete Corsika (version 7.7420) library is used.We use energies $\lg(E/\mathrm{eV}) \in \{18.0, 18.5, 19.0, 19.5, 20.0, 20.2\}$ and zenith angles $\theta \in \{0°, 38°, 65°\}$, where 30 PIS had been produced for each configuration. In order to understand the effects introduced by thinning, two thinning levels will be compared. Namely, the thinning levels $f_\mathrm{thin} = 10^{-6}$ (standard) and $f_\mathrm{thin} = 10^{-8}$ (better) are available. The other thinning parameters then follow according to the optimal thinning relation [3],

$$\begin{aligned} w_\mathrm{max}^\mathrm{EM} &= f_\mathrm{thin}(E_\mathrm{prim}/\mathrm{GeV}), \\ w_\mathrm{max}^\mathrm{Hadr} &= w_\mathrm{max}^\mathrm{EM}/100, \\ E_\mathrm{thin}^\mathrm{Hadr} &= E_\mathrm{thin}^\mathrm{EM} = f_\mathrm{thin}E_\mathrm{prim}. \end{aligned} \tag{4}$$

Please note that the file sizes for the “better” thinning can become bigger by almost a factor of 100 compared to the “standard” thinning level, therefore more aggressive thinning is necessary for larger simulation libraries.

2.1 Fluctuations in the number of particles

It is generally assumed that the number of particles N injected into a detector follows the Poisson statistics. However, this statement is only valid for the number of weighted Corsika particles N_{thin}. The statistics and, therefore, the uncertainty changes when clones are produced. It can be seen in Fig. 2 that the Poisson error bars approximately describe the fluctuations in N_{thin}. However, the fluctuations in N_{dethin} are larger than expected from Poisson statistics. The number of injected particles is given by $N = \sum_n^{N_{\text{thin}}} w_n$. Using the law of total variance, it follows,

$$\begin{aligned}\mathrm{Var}[N] &= \mathrm{E}[\mathrm{Var}[N|N_{\text{thin}}]] \\ &\quad + \mathrm{Var}[\mathrm{E}[N|N_{\text{thin}}]] \\ &= \mathrm{Var}[w]\,\mathrm{E}[N_{\text{thin}}] \\ &\quad + \mathrm{E}[w]^2\,\mathrm{Var}[N_{\text{thin}}],\end{aligned} \tag{5}$$

as N_{thin} is assumed to be Poissonian,

$$\begin{aligned}\mathrm{Var}[N_{\text{thin}}] &= \mathrm{E}[N_{\text{thin}}] \\ \Rightarrow \mathrm{Var}[N] &= \mathrm{E}[N_{\text{thin}}](\mathrm{Var}[w] + \mathrm{E}[w]^2).\end{aligned} \tag{6}$$

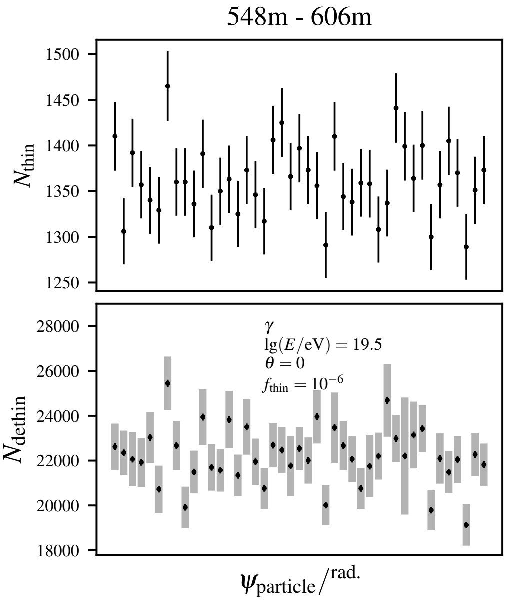

Figure 2: *Top:* Number of thinned particles in a ring of 548 m $< r <$ 606 m around the shower core. The error bars show the Poissonian errors $\sqrt{N_{\text{thin}}}$. *Bottom:* De-thinned number of particles injected into ring sections. The error bars represent Poisson errors $\sqrt{N_{\text{dethin}}}$. The shaded bar represents the error $\sqrt{\mathrm{Var}[N_{\text{dethin}}]}$ derived from the weight distribution in the respective ring section. The simulated vertical PIS has an energy of $10^{19.5}$ eV and a thinning level of 10^{-6}.

This means that artificial fluctuations are increased for large expectation values for the weights $\mathrm{E}[w]$, but also when the weight distribution is very wide [4].

Dividing the shower footprint into bins of the size of the maximum sampling area $A_{\text{sam}}^{\text{limit}}$ starting from $R = 150$ m to 1000 m, the resampled weight can be calculated for each particle in the Corsika file. Moreover, the number of thinned particles N_{thin}, the number of de-thinned particles N_{dethin}, the mean and the standard deviation of the weights can be computed. Using the above calculation for the variance, it follows for the uncertainty,

$$\sigma_w^2 = \mathrm{Var}[N_{\text{dethin}}], \qquad \sigma_{\text{Poisson}}^2 = N_{\text{dethin}}. \tag{7}$$

The relative difference between the uncertainties describes the “non-Poissonness” of the number of de-thinned particles, see Fig. 3. We see that in some bins $\mathrm{Var}[N_{\text{dethin}}]$ can be significantly larger than the Poisson uncertainty, especially for the electromagnetic component.

2.2 Estimation of the signal produced by clones

Now, however, we are not interested in the number of injected particles, but the signal they produce. For example, muons are low in number but contribute a major part of the signal in

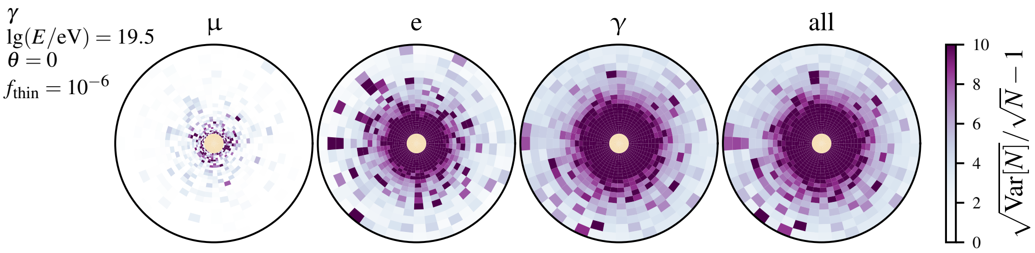

Figure 3: "Non-Poissonness" of the number of de-thinned particles in bins of azimuth and distance to the shower core (150 to 1000 m) for a vertical shower. Where the weights are small, N_dethin behaves more Poissonian-like.

HIS due to their higher energy. To estimate the signal each of the particles would produce in the detector without running the full detector simulation, a simple model for the detector response function (DRF) is used, see Fig. 4.

Using this, we can calculate the fraction of the signal that would be contributed by clones in the simulation. If w_i is the resampled weight of a particle i in a respective section, the total signal will be,

$$S_\mathrm{tot} = \sum_i^{N_\mathrm{thin}} w_i \, \mathrm{DRF}(E_{\mathrm{kin},i}). \tag{8}$$

For the clones this is,

$$S_\mathrm{clones} = \sum_i^{N_\mathrm{thin}} \max(w_i - 1, 0) \, \mathrm{DRF}(E_{\mathrm{kin},i}). \tag{9}$$

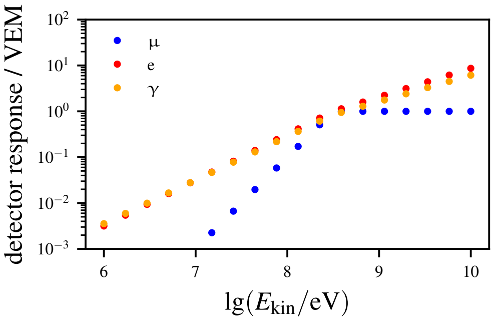

Figure 4: Simplified model of the Water-Cherenkov detector (WCD) response function (DRF). Taken from Ref. [5], Fig. 3.1.

The fraction of the signal contributed by clones is shown in Fig. 5. We see that the impact of clones is larger for the electromagnetic particles compared to muons. Moreover, as muons are more dominant in HIS, the signal contribution is smaller than that of PIS. We also see that for the "better" thinning level, the signal contribution of clones is negligible.

3. Detector Simulation

The size of the available test library at this point is rather small, so the shower-to-shower fluctuations can affect the comparison of distributions of the two thinning levels. Given the longitudinal profile of the showers, we observed that for some energies and zenith angles, the distributions of the depth of the first interaction X_1 differ. Therefore, we only select pairs of showers from both libraries that have an X_1 at least within 20 g/cm^2. That way, the two libraries still have the same size while having similar distributions in X_1.

Now, we consider the full Offline [12] detector simulation with the resampling algorithm. Each of the showers of the two libraries is simulated with the surface detector (SD-1500) of the Pierre Auger Observatory and additional rings of 24 detector stations at different distances to the shower

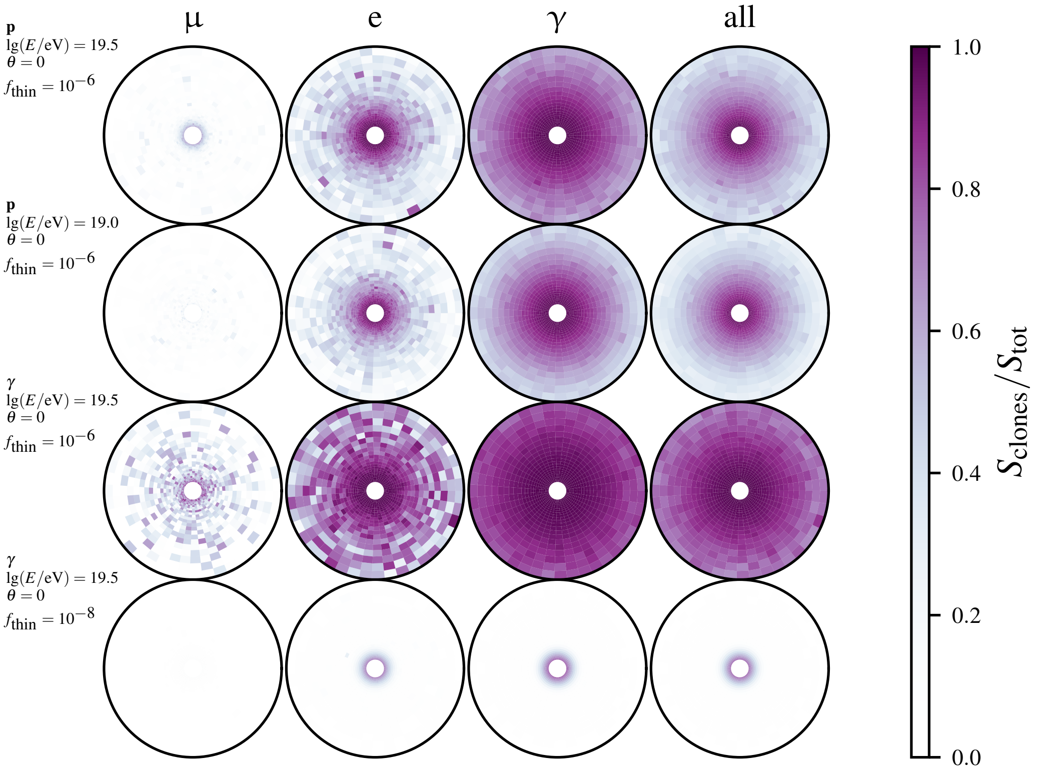

Figure 5: Signal fraction contributed by clones in bins of azimuth and distance to the shower core (150 to 1000 m) for a vertical shower.

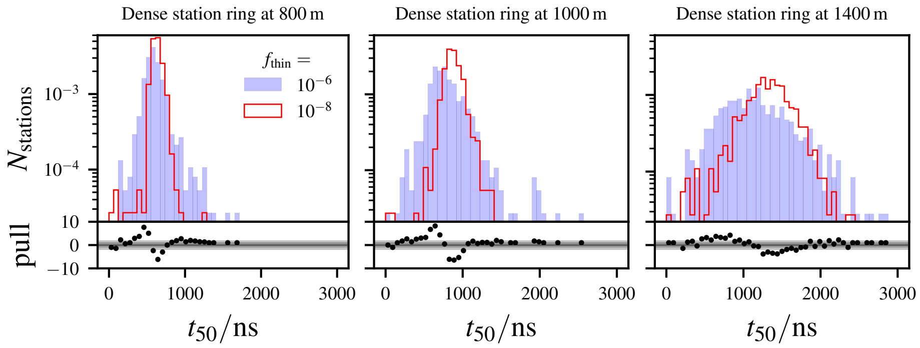

Figure 6: Distribution of t_{50} for the dense rings of 24 stations at 800, 1000, 1400 m distance to the shower axis for vertical showers at 10^{20} eV. A shift in the peak of the distribution is observed when comparing the two thinning levels.

axis. We observe that the detection and reconstruction efficiency for the showers of both thinning levels are compatible.

3.1 Risetime t_{50}

The risetime of a trace is sensitive to the muon content at the respective station and therefore used for photon-hadron discrimination. We investigate the connected observable t_{50}, which is the time after which 50% of the total signal is accumulated in the trace. The distribution of this

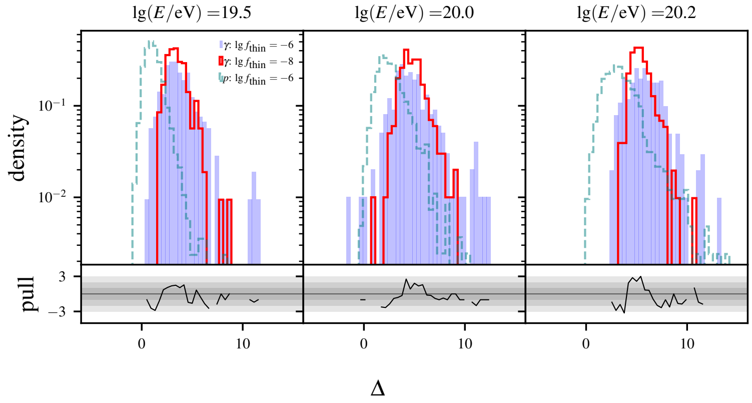

Figure 7: Distribution of Δ for the photon simulations with different thinning levels for different energies for a reconstructed inclination of 38°. Note that in a photon search only angles between 30° and 60° would be considered. For comparison also the distributions for proton simulations with zenith angles within 5° and energies within 0.1 in $\lg(E/\mathrm{eV})$ are shown (green dashed). In the lower plots the pull of the distributions with the two different thinning levels 10^{-6} (standard) and 10^{-8} (better) is shown.

observable is shown for three different detector rings at distances 800, 1000, and 1400 m from the shower axis. We see that the distribution for the standard thinning level is broader compared to the lower thinning level. Moreover, the peak of the distribution seems to be at higher t_{50} for the lower thinning level. We observe that this behavior appears mainly at the higher energies, where the weights are usually higher. This can be understood, considering that t_{50} reflects the width of the trace, while the trace itself is close to a time distribution of arriving particles. A higher thinning level means that fewer particles from this time distribution are sampled, which have higher weights in the distribution. If fewer particles are sampled, the probability of choosing a late particle decreases, while the integral is conserved by the weights. As a consequence, the trace becomes more narrow on average[1].

3.2 Photon-hadron discriminators

3.2.1 Observable Δ

An often used photon-hadron discriminating observable is Δ [11],

$$\Delta = \sum_i \frac{t_{12}^{(i)} - t_{\mathrm{b}}}{\sigma_{12}(\theta_{\mathrm{rec}})}, \tag{10}$$

where t_{12} is the asymmetry corrected risetime of the station, σ_{12} is a parameterization of the variance of t_{12} depending on the zenith angle θ_{rec}, and t_{b} is a benchmark parameterization for t_{12} derived from data depending on the distance to the shower axis and the gain of the station. In Fig. 7, it can be observed that no significant shift is introduced between the thinning levels. Please note that to obtain a better separation power, Δ has to be combined with other observables in a Fisher analysis, see [11].

[1] For a Gaussian distribution, the expectation value of the standard deviation depending on the sample size can be analytically found [9].

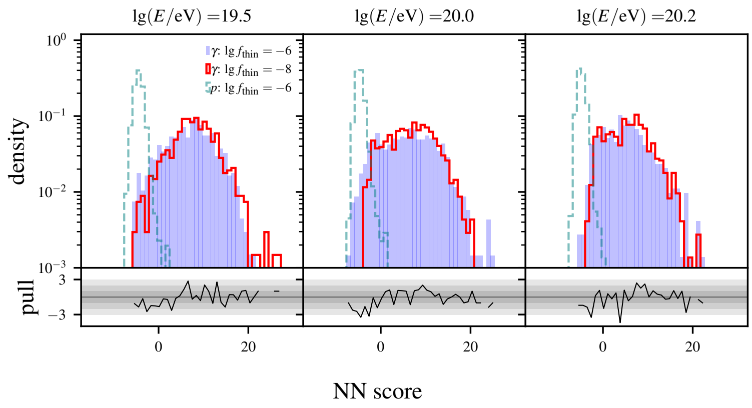

Figure 8: Same as Fig. 7, but for the neural network score and for all three zenith angles.

3.2.2 Neural network discriminator

Deep neural networks that analyze the shapes of the time traces of the surface detector have proven to be powerful tools in various reconstruction tasks at Auger. A neural network[2] was trained on photon-proton discrimination.

In Fig. 8, we see that the distributions of the predicted score for both thinning levels are compatible. We conclude that the changes in the trace introduced by thinning are not what is picked up by the network as a photon-hadron discriminating feature in this case.

4. Conclusion

We observe negligible effects caused by the standard thinning level for photon-hadron discrimination with classical or machine-learning approaches, where the measurements of several stations in an event are combined. However, special attention to the thinning level is advised if distributions for station-level observables like the risetime are considered.

References

[1] D. Heck and T. Pierog, Version 7.7500, 10 October 2023.

[2] P. Billoir, Astropart. Phys. **30** (2008) 270–285.

[3] M. Kobal, Astropart. Phys. **15** (2001) 259–273.

[4] P.M. Hansen *et al.*, Astropart. Phys. **34** (2011) 503—512.

[5] M. Pothast, PhD thesis, Radboud University Nijmegen (2023).

[6] Pierre Auger Coll., Phys. Rev. D **96** (2017) 122003.

[7] P. Sanchez Lucas, PhD thesis, Universidad de Granada (2017).

[8] Pierre Auger Coll., PoS **ICRC2023** (2023) 275.

[9] M. Roesslein *et al.*, ACCREDIT QUAL ASSUR **12** (2007) 495–496.

[10] R. Engel *et al.*, Comput. Softw. Big Sci. **3** (2018) 2.

[11] Pierre Auger Coll., JCAP **5** (2023) 021.

[12] S. Argirò *et al.*, Nucl. Instrum. Meth. A **580** (2007) 1485-1496.

[2] Same architecture as in [8] with slight modifications for the task of binary-classification.