Study of statistical thinning with fully-simulated air showers at ultra-high energies

R. BRUIJN[1], J. KNAPP[1], I. VALIÑO[2]
[1] *School of Physics and Astronomy, University of Leeds, Leeds, LS2 9JT, United Kingdom*
[2] *Dep. Física de Partículas & IGFAE, Universidad de Santiago de Compostela, 15782 Santiago, Spain*
r.bruijn@leeds.ac.uk
DOI: 10.7529/ICRC2011/V02/0228

Abstract: In the simulations of ultra-high energy (> 10^{18} eV) extensive air-showers, statistical thinning techniques are used to reduce the computational requirements. These techniques can introduce artificial fluctuations and systematic shifts in the air-shower observables. Using a modified version of the CORSIKA air-shower simulation program, a library of showers which are both "thinned" and "unthinned" has been produced to study the adverse effects on the observables at ground level. The library contains showers ranging in energy from 5×10^{18} to 2×10^{19} eV, zenith angles between 0° and 60° and proton, iron and photon primaries. Even though the detailed information on the shower front is lost, typical observables as amplitude and risetime show no significant biases or enhanced fluctuations with the standard version of thinning used. The arrival time of the shower front, however, may show a bias. Also, photon showers require a better-quality thinning.

Keywords: extensive air showers, numerical simulation, statistical thinning

1 Introduction

Simulations of ultra-high energy cosmic-ray showers are essential in the interpretation of the data of extensive air-shower arrays, such as the Pierre Auger Observatory [1]. At these ultra-high energies, the computational demands of fully simulating these showers are prohibitively large. The full simulation of a 10^{19} eV shower would require the tracking and storage of approximately 10^{10} individual particles. It would require a computation time of about 1.5 years on a single, standard (2 GHz) CPU. The storage requirement would amount to approximately 5 TB. As computation time scales with the energy of a primary particle, a 10^{20} eV shower would be practically impossible to simulate in this way.
To reduce the computational requirements to manageable levels, *statistical thinning* techniques [2, 3] are used. Only a representative subset of the particles is tracked, and the others are discarded. To account for the discarded particles, a weight w is assigned to the tracked particles. Parameters in the thinning algorithm are the fraction of the energy of the primary particle below which particles at a vertex are thinned, called the *thinning level* ε, and the maximum weight of an individual particle, $w_{\max}$. Larger thinning levels and maximum weights mean coarser thinning, less tracked particles and faster computing times. The reduced number of tracked particles implies enhanced artificial fluctuations and correlations due to single particles with high weights. Details of the shower front are lost and only a coarser description remains. From the subset of particles with weights arriving at ground level, a set of particles without weight is reconstructed before particles can be injected into the detector simulation. This re-sampling algorithm [4] considers all weighted particles in an area larger than the detector and involves randomisation and corrections for the arrival time of the particles, assuming a plane shower front and a log-normal smearing in time to avoid pile-up of particles originating from the same parent. Air shower arrays which aim at the highest energy cosmic rays, consist typically of a sparse array of detectors at ground. Therefore, the particles are in general detected at significant distances from the shower core. At these distances, the particle densities are small and thus especially sensitive to the detrimental effects of thinning.
In this paper we study the dependence of the fluctuations and biases as function of primary energy, zenith angle and primary species, and the effects of thinning and the subsequent re-sampling on the arrival time of the first particles and the detailed time structure of the signals as recorded by detectors of the Pierre Auger Observatory.

2 Simulation

A modified version of CORSIKA [5] has been used to produce a library of unthinned showers [6]. This version of CORSIKA allows a parallel simulation of showers by splitting them after the first few interactions into n sub-showers, each carrying approximately $1/n$ of the primary energy. By running these sub-showers on a cluster of CPUs running in parallel, a gain in time proportional to the number of CPUs is achieved (for our simulations typically about a factor of 100 i.e. 7.5 days in stead of 1.5 year). Disk space has been saved by not storing most of the particles very close to the shower core, where detectors anyway would be saturated. Typically the storage requirement was reduced by 85%.

CORSIKA has also been modified to apply different levels of thinning in parallel besides the unthinned version. This is essential for this study, as it allows comparison of the same shower at different thinning levels and separation of physical fluctuations from thinning fluctuations. During the simulation 10 different thinning algorithms are applied. In this study we consider 5 with thinning levels of 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5} and 10^{-4} for both the electro-magnetic and hadronic components of the shower. The weight limitation is set to the *optimum thinning* value [3] and thus a smaller maximum weight for hadronic particles compared to electro-magnetic particles. The re-sampling as described above has been used with a sampling area (defined by polar coordinates r,ϕ) of $\Delta r/r$ = 0.1 and $\Delta\phi$ = 0.15. This restricts the number of detectors at a given radius from the shower core to 20.

To ease analysis, the particles at ground are sorted into radial and azimuthal bins. The simulation parameters for the library of showers are shown in table 1. The depth of first interaction (X_0) has been set to an average value and the random number seeds are chosen to yield an average multiplicity in the first interaction. The hadronic interaction models used are QGSJETII-03 for high energies, (>100 GeV) and URQMD for low energies.

The simulation of the detector [7] response includes all relevant processes for photons, electrons, positrons and muons and Cherenkov light production for all charged particles. The response of the photo-multipliers and subsequent digitisation are also simulated. The particular detectors considered are the water-Cherenkov tanks as used in the Pierre Auger Observatory [1]. These are essentially cylindrical water-filled tanks with a height of 1.2 meters and a surface area of 10 m^2. Three photo-multipliers, facing downwards from the top, detect the Cherenkov light originating from the shower particles. The photo-multiplier signal is digitised with 40 MHz analog-to-digital converters (FADCs). In this way, *time traces* are obtained which provide a detailed measurement of the passing of the shower front within bins of 25 ns.

3 Analysis

Vertical showers are best suited to study potential biases and fluctuations in the observables. Their azimuthal symmetry allows the characterisation of the observables as function of core distance r and thinning level ε, as both the average values and the fluctuations between the detectors can be calculated. Besides the detailed time traces S_i, the quantities being investigated specifically are:

Primary	E [EeV]	θ [°]	X_0 [g/cm^2]	B_x, B_z [μT]
p	10	0	64.5	20.1, -14.2
p	10	0	40.8	0, 0
p	20	0	39.6	0, 0
p	5	0	44.0	0, 0
p	5	45	46.0	20.1, -14.2
p	10	45	29.0	0, 0
p	10	60	19.9	0, 0
Fe	10	0	10.8	0, 0
γ	10	0	50.3	0, 0

Table 1: Summary of the shower library used in this study.

Start time, t_0 The time at which the signal in the traces exceeds a threshold, which is associated with the arrival time of the earliest particles in the shower front. This quantity is used in the geometric reconstruction of the shower and the curvature of the shower front, which is composition sensitive.

Amplitude, S The total integrated signal in a detector. It is used to reconstruct the core position of the shower and to estimate its energy.

Risetime, $t_{1/2}$ The time it takes for the integrated signal to grow from 10% to 50% of the total signal in a detector. It is related to the longitudinal development of the shower and is composition sensitive.

Bin-to-bin "jump", ΔV_i = S_i = S_{i-1} Its distribution can be used to estimate the muon content of the detector signal.

3.1 Traces

The FADC traces contain detailed information on the history of the energy deposited by the shower particles. This detailed information, however, is lost in the thinning procedures. When re-sampling a thinned shower, there is no direct correspondence between the original shower particles and the ones injected into the detector simulation. This allows for different sets of injected particles (realisations), drawn from the same set of weighted particles. Figure 1 illustrates the effects of different thinning levels on the traces. The top row corresponds to the unthinned case. This means that the particles falling on the detector are the ones which are injected and thus represent the *true* population of particles. For multiple realisations for the same tank and shower, of which two are shown in the first row, only the interactions in the tank cause stochastic fluctuations. Thus, the two realisations show a good overall agreement with each other. When re-sampling is applied to the unthinned shower (second row), the resemblance with the unthinned (*true*) traces (top row) is significantly worse. Also, the two realisations look less similar. This is due to the fact that the particles injected into the detectors differ between the two realisations. When standard thinning and re-sampling are applied (third row), the resemblance with the unthinned (*true*) traces is even worse. Also, the two realisations of the

same tank and shower show clear differences in both the overall structure and the substructures of the electromagnetic and muonic contributions in the trace. It is evident that in the left trace, the muonic contribution is lagging behind compared to the true distribution and to the other realisation for the same thinning level. With more severe thinning (10^{-4}), the influence of a low number of high-weight particles dominate, and, even after re-sampling, make the different realisations start to resemble each other. Also, the total signal differs significantly (60%) from the total signal in the unthinned traces.
It is clear that the detailed structure of the traces is affected by the thinning and resampling, however, the derived quantities listed above are not necessarily affected as significantly.

3.2 Observables

The observables considered have been introduced above. Their mean and fluctuations are calculated from 20 detectors at the same core distance on a shower by shower basis and averaged over 50 realisations for each shower. For showers with proton and iron primaries, and the angles and energies considered (see table 1) the risetime and amplitude show no significant bias up to a thinning level of 10^5 in the distance range of 600 to 1400 m. The fluctuations, however, increase rapidly for $\varepsilon > 10^{-6}$. The increase is most severe at the highest energies where the weights per particle are larger and a smaller relative number of muons is tracked. At 10^{-5} the fluctuations increase with 20% for a proton primary at 5×10^{18} eV and 100% for 2×10^{19} eV. The increased fluctuations in risetime and amplitude also depend on the primary mass at a given energy, with iron being least affected. The primaries that are most affected by the thinning are the photons due to the smaller particle densities, higher weights and smaller muon contributions. Already at the standard thinning level of 10^{-6}, the amplitude and risetime show increased fluctuations of about 40 and 30%, respectively. Smaller thinning levels are thus recommended to avoid these fluctuations. Up to a thinning level of 10^{-6} for nuclear primaries, and up to 10^{-7} for photons, the amplitude and risetime can be used reliably. In contrast to the amplitude and risetime, the start time of the signal does show a bias. The bias is also present when there is no thinning, but the re-sampling algorithm is applied to the unthinned showers. The origin of this bias, which is of the order of 10% is thought to originate from the assumption of a plane shower front in the re-sampling algorithm. The time shift can have consequences for analyses which rely on the timing of the shower front, notably studies involving the radius of curvature of the shower front.
The bin-to-bin jumps have been studied for nuclear primaries. The spectrum is unaffected for all but the largest thinning levels. This suggests that muon-counting techniques are unbiased, despite the loss of the detailed time traces. While the distribution of peaks seems to be reproduced reasonably well, studies relying on the correlation of (muon) signals between tanks should avoid thinning and resampling.

4 Conclusion

A study on the effects of statistical thinning techniques in the simulation of extensive air-showers has been expanded by generating a library of unthinned showers. These showers have been simulated using a special version of CORSIKA, which does not only parallelise the simulation, but also allows for comparisons between unthinned and thinned versions of the same showers. The library covers energies, angles and primary types which are detected by the Pierre Auger Observatory. Even though the detailed information on the particle arrival structure of shower fronts is lost, important observables show no bias and enhanced fluctuations up to the standard thinning level of 10^{-6}. A notable exception is the measured arrival time of shower front, which is biased towards earlier times. Also, the simulation of photon primaries requires finer thinning, 10^{-7} or better, as the lower particle densities lead to larger thinning fluctuations.

References

[1] J. Abraham *et al.* [Pierre Auger Collaboration], Nucl. Instrum. Meth. A **523** (2004) 50
[2] A. M. Hillas, Proc. 17^{th} ICRC, Paris 8 (1981) 193
[3] M. Kobal, Astropart. Phys. **15** (2001) 259
[4] P. Billoir, Astropart. Phys. **30** (2008) 270
[5] D. Heck et al., Report FZKA 6019 (1998), Forschungszentrum Karlsruhe
[6] F.Schmidt, J. Knapp and M. Zha, *Results from the Simulation of an Unthinned Proton Shower at* 5×10^{18} *eV*, GAP note 2005-095
[7] S. Argiro *et al.*, Nucl. Instrum. Meth. A **580** (2007) 1485 [arXiv:0707.1652 [astro-ph]].

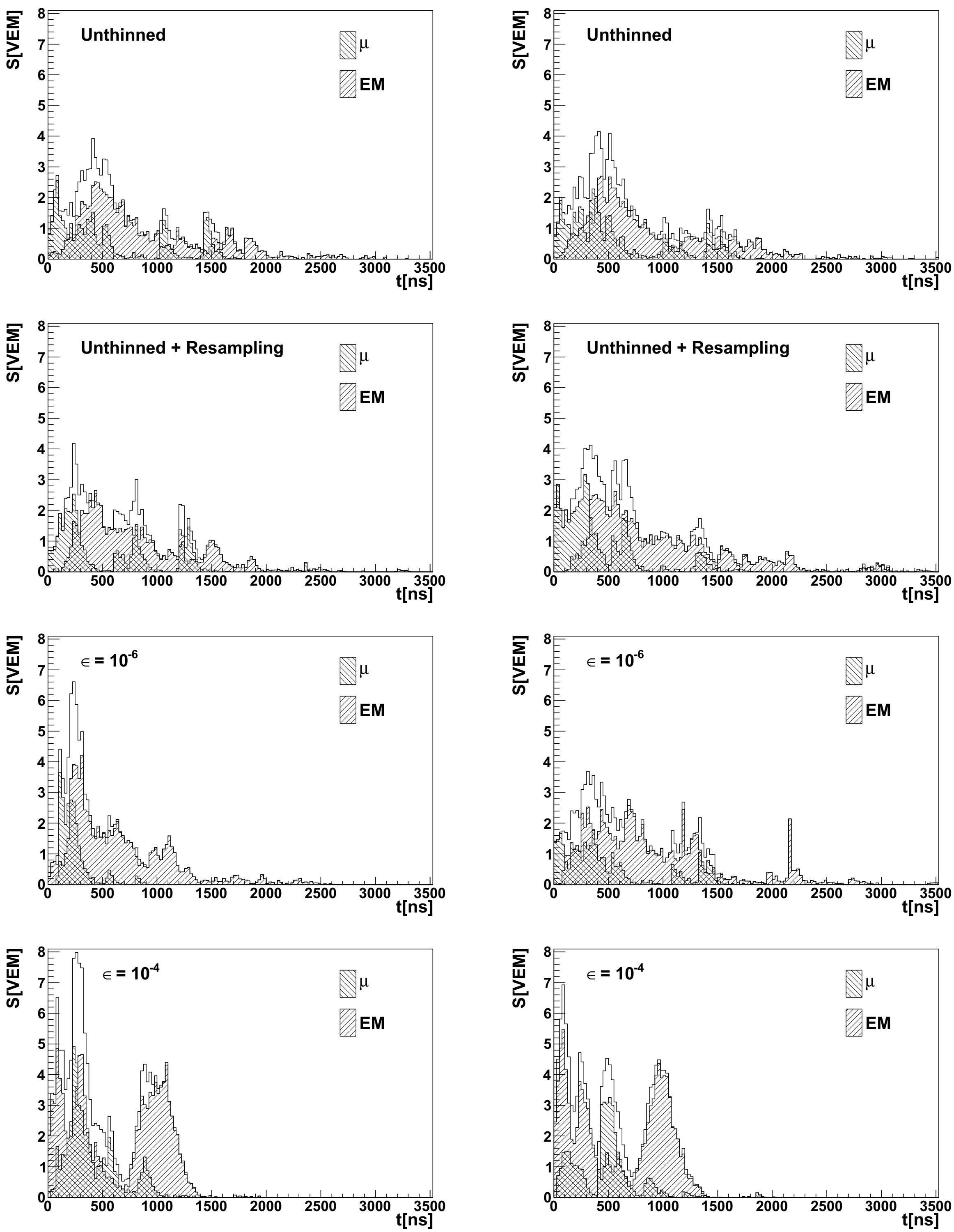

Figure 1: Simulated FADC traces for a single tank at 1000 m from the shower core of a vertical 10^{19} eV proton shower. All plots show the same detector for the same shower. The rows correspond to the different thinning levels and the two columns show different realisations of the simulation. The top row shows two unthinned realisations, the second row the unthinned realisations with re-sampling applied, the third and fourth row a thinning level of 10^{-6} and 10^{-4}, respectively.