Proceedings of the 30th International Cosmic Ray Conference
Rogelio Caballero, Juan Carlos D'Olivo, Gustavo Medina-Tanco,
Lukas Nellen, Federico A. Sánchez, José F. Valdés-Galicia (eds.)
Universidad Nacional Autónoma de México,
Mexico City, Mexico, 2008

Vol. 4 (HE part 1), pages 641–644

30TH INTERNATIONAL COSMIC RAY CONFERENCE

Getting rid of artificial fluctuations in EAS simulated with thinning: estimate of artificial fluctuations and simple economical method to suppress them

D. S. GORBUNOV, G. I. RUBTSOV, S. V. TROITSKY.
Institute for Nuclear Research of the Russian Academy of Sciences, Moscow 117312, Russia
grisha@ms2.inr.ac.ru

Abstract: The most common way to simplify extensive Monte-Carlo simulations of air showers is the thinning approximation. We study its effect on the physical parameters reconstructed from simulated showers. To this end, we created a library of showers simulated without thinning with energies from 10^{17} eV to 10^{18} eV, different zenith angles and primaries. This library is publicly available. Various physically interesting applications of the showers simulated without thinning are discussed. Observables reconstructed from these showers are compared to those obtained with the thinning approximation. The amount of artificial fluctuations introduced by thinning is estimated. A simple method, multisampling, is suggested which results in controllable suppression of artificial fluctuation and simultaneously conserves computational resources as compared to the usual thinning.

Introduction

Experimental information about cosmic particles at very high energies is obtained through the study of atmospheric showers induced by these particles and is hence indirect. A necessary ingredient of these studies is therefore good understanding of a shower initiated by a primary particle with given parameters. Since the shower development is a complicated random process, the Monte-Carlo simulations are often used to model atmospheric showers[1]. Physical parameters are then reconstructed from the simulations and compared to real data.

At very high energies, however, the number of particles in a shower is so large that the simulations start to require unrealistic computer resources. Among several ways to simplify the problem and to reduce the computational time, the thinning approximation [2, 3] is currently the most popular one. Its key idea is to track only a representative set of particles; while very efficient in calculations and providing correct values of observables on average, this method introduces artificial fluctuations because the number of tracked particles is reduced by several orders of magnitude. These artificial fluctuations mix with natural ones and therefore reduce the precision of the determination of physical parameters.

Both the central value and the width of the distribution of the observed quantity are important for physical applications. The width of the distribution obtained in simulations arises from two sources: physical fluctuations and artificial fluctuations introduced by thinning.

The goal of the present work is to estimate the relative size of these artificial fluctuations and to develop an efficient resource-saving method to suppress them in realistic calculations.

All artificial showers used in present study are publicly available at *http://livni.inr.ac.ru* [4]. Library "Livni" contains artificial showers simulated with CORSIKA [5] without thinning.

Size of artificial fluctuations due to thinning

Having at hand a library of showers simulated without thinning, we may compare the observ-

1. A completely different approach [1] is to combine partial Monte-Carlo with analytical solutions of cascade equations and pre-simulated subshower libraries in the framework of *hybrid codes*.

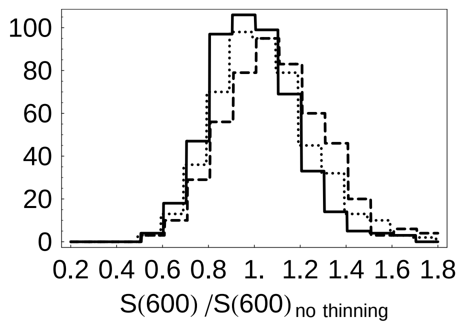

Figure 1: Distribution of $S(600)/S(600)_{\text{no thinning}}$, where $S(600)$ is reconstructed from 500 showers simulated with $\epsilon = 10^{-4}$ and the same random seed as the corresponding $\epsilon = 0$ shower, for three different artificial showers, induced by 10^{18} eV protons at the AGASA location.

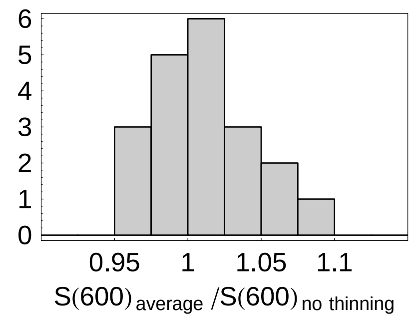

Figure 2: Distribution of $S(600)_{\text{average}}/S(600)_{\text{no thinning}}$, where $S(600)_{\text{average}}$ is the average of reconstructed $S(600)$ over a sample of 500 showers simulated with $\epsilon = 10^{-4}$ and the same random seed as the corresponding $\epsilon = 0$ shower, for 20 different random seeds. Showers are vertical, induced by 10^{17} eV protons at the Telescope Array location.

ables reconstructed from showers with and without thinning and estimate the effect of the approximation. To do that, for each shower without thinning ($\epsilon = 0$) we simulated a number of showers with different thinning levels ($\epsilon \neq 0$). All initial parameters (including the random seed numbers) were kept the same as in the $\epsilon = 0$ simulation, which enabled us to reproduce exactly the same first interaction in the entire set of showers. Three important observables — the signal density at 600 m from the shower axis $S(600)$, the muon density at 1000 m from the axis $\rho_\mu(1000)$, and the depth of the maximal shower development $X_{\max}$ — were reconstructed for each of the showers following the data-processing operation adopted by the AGASA experiment [6][2]. The detector response was calculated with the help of GEANT simulations in Ref. [8].

Figure 1 shows the distribution of the reconstructed $S(600)$ for showers with thinning simulated with the same initial random seed (and thus the same first interaction) as three representative $\epsilon = 0$ Livni showers. Though quite wide for $\epsilon = 10^{-4}$ thinning, the distributions of $S(600)/S(600)_{\text{no thinning}}$ are well centered at unity.

The distribution of the mean values of $S(600)/S(600)_{\text{no thinning}}$ for the ensembles of the thinned showers is presented in Fig. 2 for a uniform sample of twenty different $\epsilon = 0$ showers. For each of them, 500 showers with $\epsilon = 10^{-4}$ were simulated with the same first interaction as the corresponding $\epsilon = 0$ shower. The values of the observable averaged over 500 thinned showers approximate the "exact" $S(600)_{\text{no thinning}}$ with the accuracy of about 3%, which is consistent with the level of statistical fluctuations, $1/\sqrt{500} \sim 4\%$. We have found the same distributions for other observables considered, $\rho_\mu(1000)$ and $X_{\max}$.

To estimate the effect of thinning on the distribution of observables, we simulate samples of showers with fixed initial conditions but different random seeds for various thinning levels, including $\epsilon = 0$. We consider samples of $E = 10^{17}$ eV vertical proton-induced showers consisting of 20 showers with $\epsilon = 0$, 100 showers with $\epsilon = 10^{-5}$, 100 showers with $\epsilon = 10^{-5}$ and weight limitation, 100 showers with $\epsilon = 10^{-4}$ and 100 showers with $\epsilon = 10^{-4}$ and weight limitation.

Figure 3 illustrates the widths of the distributions obtained at different ϵ. Artificial fluctuations in $S(600)$ due to thinning are clearly seen by comparing $\epsilon = 10^{-4}$ case with others.

2. For the Telescope Array [7], we used the same procedure as for the AGASA experiment with straightforward modifications taking into account the thickness of scintillator detectors.

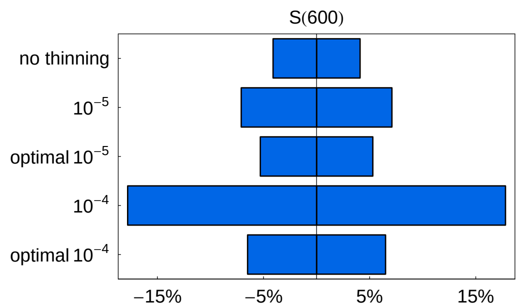

Figure 3: Width of the $S(600)$ distribution for 10^{17} eV vertical proton showers simulated with and without thinning for the Telescope Array observational conditions.

Multisampling: an economical method to suppress artificial fluctuations

We see from the results of the previous section that the use of thinning is well motivated when one is interested in the reconstruction of the central values of fluctuating observables. On the other hand, its use may limit the precision of composition studies, where the observed value of some quantity is compared to the simulated distributions of the same quantity for different primaries, and the width of these distributions is of crucial importance (see e.g. the proton–iron comparison in examples of Ref. [9]).

As it has been pointed out above, the effect of physical fluctuations on the distribution of an observable quantity should be in principle estimated by simulating a set of showers with the same physical parameters, with different random seeds and without thinning. As we may conclude from the previous section average of an observable over a sample of thinned showers with fixed initial random seed approximates the value of the same observable for an $\epsilon = 0$ shower with the same random seed with a good accuracy. The distribution of observables for $\epsilon = 0$ showers with different random seeds is then approximated by a distribution of these approximated observables calculated for samples with random seeds varying from one sample to another but fixed inside a sample. A practical way to do this is:

- instead of a single shower with $\epsilon = 0$, simulate N showers with some $\epsilon = \epsilon_0 \neq 0$ and fixed random seed;
- reconstruct the observable for each of N showers, average over these N realizations and keep this average value which approximates the result for a single shower without thinning;
- repeat the procedure M times for different random seeds to mimic a simulation of M showers without thinning and obtain the required distribution of the observable.

We will refer to this procedure as *multisampling* $(N \times \epsilon_0)$. Even for relatively large ϵ, averaging over sufficiently large number of showers (N) gives a good approximation to an $\epsilon_0 = 0$ value of an observable; the larger N the better the approximation. Required value of N may be estimated as follows. Consider the distribution of an observable reconstructed from showers simulated with the thinning level close to ϵ_0 for a given initial random seed. Assume that the distribution is Gaussian with the width σ (though the qualitative conclusions do not depend on the exact form of the distribution, we note that in practice it is indeed very close to Gaussian [10]); then one needs N measurements to know the mean value with the precision $\sim \sigma/\sqrt{N}$. Numerical results for the Livni showers demonstrate that $(N \times \epsilon_0)$ multisampling for $N \sim 15\ldots20$ and $\epsilon_0 \sim 10^{-4}$ results in the precision of $3 \div 4\%$ in reconstruction of $S(600)$, $\rho_\mu(1000)$ and $X_{\max}$ of the original $\epsilon = 0$ showers.

In Figure 4 we present the widths of the distributions obtained with the usual thinning and with multisampling for $E = 5 \cdot 10^{19}$ eV vertical proton-induced showers; the limited statistics (we used $n = 200$ showers) implies the statistical uncertainty of about $1/\sqrt{n} \sim 7\%$. For the case of $5 \cdot 10^{19}$ eV the multisampled distribution (which is expected to mimic the $\epsilon = 0$ distribution with a good accuracy) allows us to estimate the size of purely artificial fluctuations due to thinning which, for instance, for $\epsilon = 10^{-5}$ with weights limitations remain at the level of $\gtrsim 10\%$ for $S(600)$ and of $\gtrsim 12\%$ for $\rho_\mu(1000)$.

The CPU time is very sensitive to the choice of the hadronic interaction model: since thinning starts to

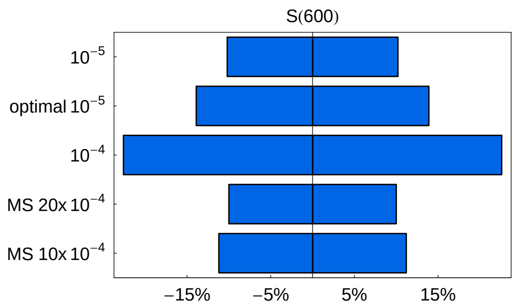

Figure 4: Width of the $S(600)$ distribution for 200 showers initiated by $5 \cdot 10^{19}$ eV vertical protons simulated with thinning and with multisampling for the Telescope Array observational conditions.

work when the number of particles is large enough, the first few interactions are simulated in full even for relatively large ϵ. If the high-energy model is slow, then the effect of multisampling on the computational time is not so pronounced. By variations of the hadronic interaction models, we estimated the average time consumed by QGSJET II, SYBILL, FLUKA and GHEISHA for simulations of showers at energies 10^{17} eV and $5 \cdot 10^{19}$ eV. For $5 \cdot 10^{19}$ eV vertical proton showers, (20×10^{-4}) multisampling is about 5 times faster than 10^{-5} thinning with weights limitation for SYBILL while for (very slow) QGSJET II, both take roughly the same time.

Discussion and conclusions

Artificial showers simulated without thinning were used for a quantitative direct study of the effect of thinning on the reconstruction of signal (S) and muon (ρ_μ) densities at the ground level as well as on the depth $X_{\max}$ of the maximal shower development. We demonstrated that thinning does not introduce systematic shifts into these observables, as was conjectured but never explicitly checked. We estimated the size of artificial fluctuations which appear due to the reduction of the number of particles in the framework of the thinning approximation; these unphysical fluctuations may affect the precision, e.g., of the composition studies. For instance, at the energies of $5 \cdot 10^{19}$ eV for vertical proton primaries, artificial fluctuations are about 10% in the signal density at 600 m and about 12% in the muon density at 1000 m for $\epsilon = 10^{-5}$ thinning with weight limitations. An effective method to suppress these artificial fluctuations, multisampling, is suggested and studied. The method does not invoke any changes in simulation codes but only in the parameters of, say, the CORSIKA input. Compared to the 10^{-5} thinning with weights limitations, it gives similar precision but allows one to gain an order of magnitude decrease in the required disk space. Gain in the CPU time depends on the speed of the high-energy interaction model: it is of order $5 \div 10$ for fast ones (SYBILL) and of order one for slow ones (QGSJET II).

We are indebted to T.I. Rashba and V.A. Rubakov for helpful discussions. This work was supported in part by the Russian Foundation of Basic Research grant 07-02-00820, by the grants of the President of the Russian Federation NS-7293.2006.2 and MK-2974.2006.2 (DG) and by the fellowship of the Russian Science Support Foundation (ST). Numerical part of the work was performed at the computer cluster of the Theory Division of INR RAS.

References

[1] L. G. Dedenko, Can. J. Phys. **46**, 178 (1968).
[2] A. M. Hillas, Nucl. Phys. Proc. Suppl. **52B**, 29 (1997).
[3] M. Kobal, Astropart. Phys. **15**, 259 (2001).
[4] G. I. Rubtsov, Talk given at the 3rd International Workshop on the highest energy cosmic rays and their sources.
[5] D. Heck, G. Schatz, T. Thouw, J. Knapp and J. N. Capdevielle, FZKA-6019.
[6] N. Chiba *et al.*, Nucl. Instrum. Meth. A **311**, 338 (1992).
[7] F. Kakimoto *et al.* [TA Collaboration], Proc. 28th ICRC, Tsukuba, 2003.
[8] N. Sakaki *et al.*, Proc. ICRC 2001, **1**, 333.
[9] D. S. Gorbunov, G. I. Rubtsov and S. V. Troitsky, arXiv:astro-ph/0606442.
[10] S. V. Troitsky, Talk given at the 3rd International Workshop on the highest energy cosmic rays and their sources.