

Computer time optimization in Extensive Air Shower simulations.

Juan de Dios Álvarez

Universidad Michoacana de San Nicolás de Hidalgo

E-mail: juandedios@ifm.umich.mx

Umberto Cotti

Universidad Michoacana de San Nicolás de Hidalgo

E-mail: ucotti@ifm.umich.mx

Cederik de León*

Universidad Autónoma de Puebla – Universidad Michoacana de San Nicolas de Hidalgo.

E-mail: cederik@gmail.com

The computation time used by the Extensive Air Showers (EAS) simulators is an important variable to be considered when generating shower simulated libraries.

The thinning method implemented in the EAS simulator codes has been a fundamental tool when the computational resources were limited, nevertheless at present the computers allows to perform more calculations in a shorter time. Computer tools are such that the use of thinning nowadays is an option that can be avoided and one can perform simulations without it, achieving the totality of the information generated along the EAS development. This, anyway, imply the use of the maximum calculation time, on the other hand, it is always more efficient to have the option that allows a less time consuming, useful for example in the first stages of an EAS simulation analysis. It thus arise, naturally, the question about what is the optimal thinning that should be applied to perform the required EAS simulation, minimizing computer error and fluctuations.

We present an analysis of the convergence of the fluctuations of the observable of interest as a function of the thinning used, which allows to find the suitable thinning value and the number of simulated EAS needed in order to get the required observables with a value that guarantee the minimization of its artificial fluctuations.

We prove that the computation time can be optimized in a controlled way that allows to know which are the minimum numbers of EAS to average, the thinning value to be considered, which will allow to obtain a convergent value of the desired physical observable related to the fluctuations between EAS.

keyword: CR Extensive Air Shower Simulation

35th International Cosmic Ray Conference — ICRC2017

10–20 July, 2017

Bexco, Busan, Korea

*Speaker.

1. Introduction and background

Two decades ago Hillas presented his well know thinning method [1], proposed with antelation [2], to solve the problem related to manage, track and store the huge amount of information from secondary particles produced by an extensive air shower.

The computation time required to perform an EAS simulation is a determining variable that needs to be considered when performing studies or generating shower simulated libraries, as it would rise proportionally to the primary energy, this will lead to a huge and sometime prohibitive amount of processing time.

The thinning method implemented in the EAS simulator codes has been a useful tool when the computational resources were limited, nevertheless at present the computers allows to perform more calculations in a shorter time. Computer tools are such that the use of thinning is an option that could be avoided and one can perform simulations without it, achieving the totality of the information generated along the EAS development. This anyway imply the use of the maximum calculation time, on the other hand, is always more efficient to have the option that allows the less time consuming as a function of the specific analysis of interest.

In this two decades improvements to the original thinning method have been made, as the statistical thinning [3], different procedures to unthinnig, dethinning or resampling [4, 5] all with the aim of replacing the information lost in thinning, and a few revisions have been written [6, 7]. The improvement of computing and storage capacities has made it possible to increase the maximum energy at which the full simulation of EAS without thinning [8, 9, 10] can be performed, eventually parallelizing it [11, 12] on different cpu, optimizing the storage processes and making possible the comparison of the results against the same simulations with thinning.

The application of the thinning procedure induce the production of artificial or unphysical fluctuations on the observables values of the simulates showers. This kind of fluctuations are different from (and can hamper) the physical or experimental ones (table 1), however it is possible to quantify how the artificial ones are hiding the information of the physicals [13, 14]. The aim of this work is to provide a simple way to obtain an answer to the following question: Given a specific computer hardware and a defined calculation time, what are the optimum thinning to use and the appropriate number of shower in order to obtain the optimum value which minimize the artificial fluctuation and the root mean square error (RMSE)?

Physical fluctuations	Experimental fluctuations	Artificial fluctuations
Depth of first interaction; Multiplicity, inelasticity, etc., (in first interaction); Secondary interactions.	Detector response; Sampling fluctuations.	Thinning; Un-thinning.

Table 1: Classification of the fluctuations in a simulated shower, arising from different sources, physical processes, measurement process, and the artificial ones due to the thinning algorithm.

2. The process and its computer related issues

The method consist in an approach based on the balance between the efficiency desired and the expected precision of the results which permits to avoids the characterization of the full set of variables that determines the performance of the computer hardware available and permits to optimize the computing time. This is achieved producing several sets of generated EAS with a decreasing thinning value, comparing the artificial fluctuations between them and monitoring the decreasing errors, the corresponding fluctuation of the observables and the computing time, until the reach of a set of values of thinning, fluctuation, computing time, that meet the desired requirements. By this way one can optimize quickly the efficiency of the computer hardware available, without the necessity of external resources, like performance tests on hardware, software benchmarks, hardware and/or software (operating system and its very own tools) fine-tuning.

The starting point, is getting the data to plot the value of a specific observable with a very fast thinning runs (which depends on the computer resources available) and a relative high number of EAS to simulate, until obtain a certain convergence. The following steps are to proceed to improve the thinning and reduce the number of EAS in different ways.

We will first illustrate the procedure through an application used by the authors as a part of a work previously done [15], and we will later analyze and generalize it. Using the Aires shower simulator [16] in a 8MB, Intel Core i7, 2.9GHZ, 4 cores computer machine, we generate showers produced from a proton, as a primary particle, of 10 EeV of energy and a zenith angle of 0 degrees.

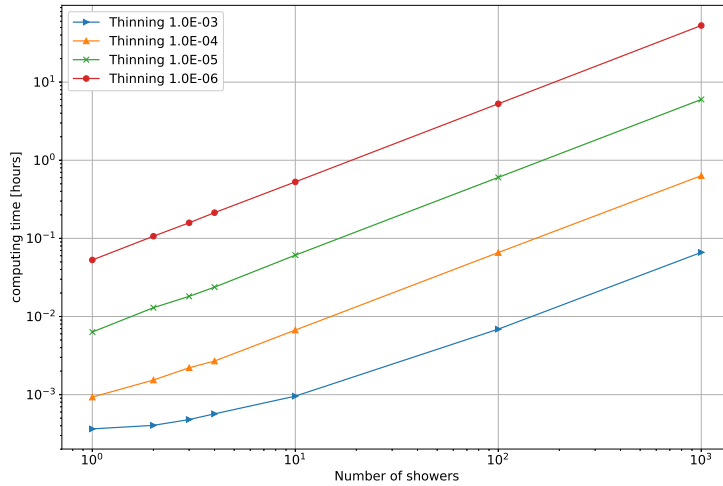


Figure 1: Computing efficiency characterization generated through the generation of 28 different libraries of the same event for different number of showers and thinning

In figure 1 we report the execution times needed to generate 28 different libraries of the same event, with 4 different thinning values (from 10^{-3} to 10^{-6}) and 7 distinct amounts of showers (from 1 to 1000). We will demonstrate later that only one point with an adequate thinning is enough to approximately generate the same plot.

By inspection is easy to realize that for a given computing time there may be more than one and up to four (the totality of the available sets) different thinning values, for which the tasks are performed in the same given time, obviously through the production of distinct amounts of showers. For a given computing time, the number of showers increase one order of magnitude in correspondence of a one order of magnitude decreasing of thinning. It is useful to establish a criterion that allows to select, for a given time, the most efficient combination of the numbers of showers and thinning according to the study of the desired observables. In a larger analysis [17] we have found that this criterion is not absolute but depends on the observable that are being studied, as in the example that will follow. Here we present a case related to the study of the lateral profile of the density of different particles for specific inputs.

The linear behavior (in logarithmic scale) of the graphs of figure 1, allows to easily extrapolate the full behavior, its computing times, of the hardware available for any value of the thinning and the number of showers, through the knowledge of only one point. Let m be the value of the slope of one of the lines plotted in 1 are described by an equation

$$Y(X) = mX + Y(0)$$

then these variables will correspond to the equivalent equation

$$\log y(x) = m \log x + Y(0)$$

which can be rewritten as

$$y(x) = ax^m \quad (2.1)$$

with $Y(0) = \log a$. Being the logarithms in base 10 in our case. The (logarithmic) slope, m is related to the load efficiency of the computing system, if the system response to load increments is linear then m will be equal to 1. This means that, for a given thinning if the computing time for a set is T , then the computing time for 2 sets will be $2T$. The above hypothesis is not always completely true [17], since it depends on the efficiency in which the software is managing the system. In this case we assume a value of $m = 1$.

Given the computing time T_{N_ϵ} required to generate a set of N_ϵ showers with a thinning ϵ , from equation (2.1) one obtains the expression for the approximate time necessary to generate $xN_{10^n\epsilon}$ showers with a thinning $10^n\epsilon$ is

$$T_n(x) \simeq 10^n T_{N_\epsilon} x. \quad (2.2)$$

The value 1 for the exponent of x is related, as already mentioned, to the linearity of the efficiency of the computing system. The fact that a variation in thinning, ϵ , by a factor of 10^n is reflected by the same factor 10^n in the computing time, is the consequence that this variation implies an increase (reduction) in the same proportion to the amount of particles that the simulator is tracking in the calculation of the shower.

Once the efficiency of the system is determined, is it possible to determine which combination is the most suitable (the efficiency) for the type of study being performed. The different combinations of numbers of showers and thinnings reflects respectively on the values of the RMSE and the artificial fluctuations between one set and another, but this will depend on the observables being studied.

To illustrate the above we will compare how two distinct observables, the lateral distribution of muons in figure 2 and the lateral distribution of all charged particles in figure 3, have different fluctuations and RMSE although they have been generated with the same 4 libraries with different numbers of showers and thinning, but under similar execution times.

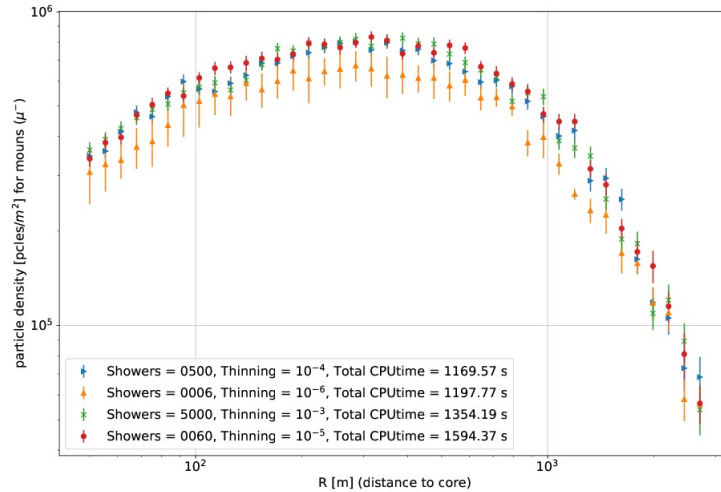


Figure 2: Lateral distribution of the density of muons for different numbers of showers and values of thinning but equivalent computing times

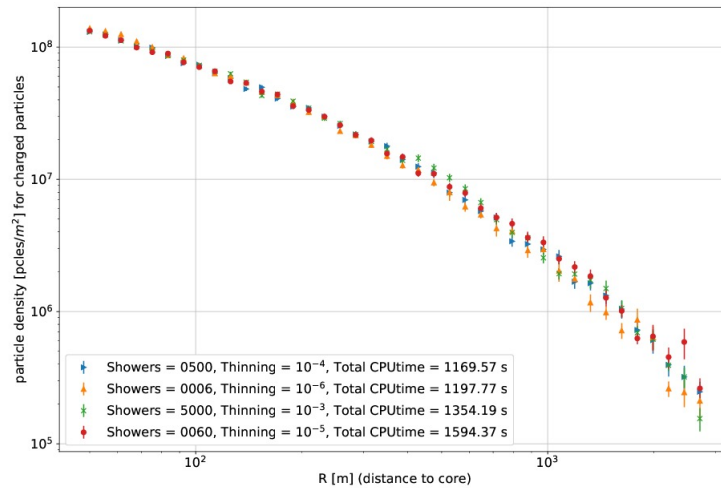


Figure 3: Lateral distribution of the density of charged particles for different numbers of showers and values of thinning but equivalent computing times

It can be observed that, specially in the case of muons, which have lower statistics, there are results that by the small number of showers present a higher RSME although with a better

thinning (value of ε smaller). From the inspection of figure 2 the choice of the value of $\varepsilon = 10^{-6}$ would be ruled out while if the study is oriented to all charged particles, any combination could be considered.

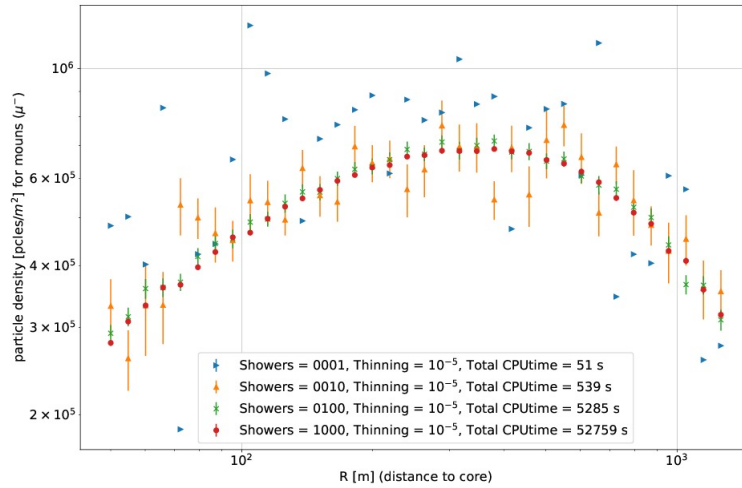


Figure 4: Lateral distribution of the density of muons for different numbers of showers same thinning minimizing the shower to shower artificial fluctuations

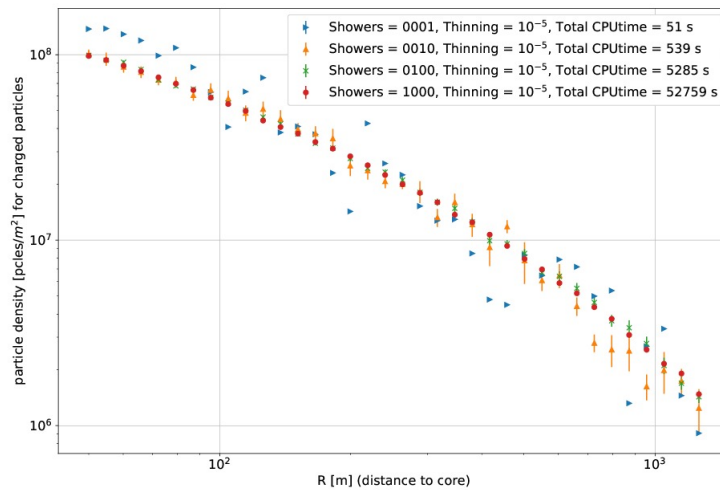


Figure 5: Lateral distribution of the density of charged particles for different numbers of showers same thinning minimizing the shower to shower artificial fluctuations

The second optimization shown in figures 4 and 5 relates processes, increasing the computing time, for a given thinning, varying the amount of showers to be averaged. The shower to shower

artificial fluctuation allows to determine for which values the convergence of the observable studied is sufficient and optimizes the resources of computation available, mainly the time.

Again if the study relates in particular to the lateral density of muons, it is immediately observed that from the production of 100 showers the artificial fluctuations and the RMSE are sufficiently reduced and no further investment of computation time is required to average one More rain, which in this case would involve a saving of a factor of 10 in the computation time that will be invested in the generation of other necessary libraries (for example with other zenith angles).

If the study refers to the lateral density of muons, from figure 4 one observes that starting from the production of 100 showers the artificial fluctuations and the RMSE are sufficiently reduced and no further investment of computation time is required to average an higher number of showers. In this case this would mean a saving, compared to the slower option, of 90% in the computation time.

A result similar to the previous one is obtained from the inspection of figure 5, in this case the computation time could be even lower since a number of showers greater than 10 but less than 100 will be sufficient. This represents a saving of time Of computation much greater than 90% compared to the slowest option studied.

3. Summary and conclusion

We presented a method that allows reduce the artificial fluctuations and the RMSE of the observables that are studied until the reach of an optimum value, minimizing at the same time the investment of computational time necessary.

The method consists of three steps, the first being to characterize the efficiency of the available computer equipment, for a given event, through the measurement of a process that can be of the duration of a few minutes. This allows through equation (2.2) the generation of any of the graphs of figure 1 and therefore the prediction of computation time of any magnitude as a function of the thinning and the number of showers to be generated.

The second step allows to discard values of the set of thinning and number of showers that are not appropriate for the study that is desired to be done. With the third step one directly minimize the artificial fluctuations through the recognition of the asymptotic behavior of the value of the desired observable, as a function of the number of showers generated.

The method and example presented shows how this allows drastically a decreasing in the computing time when compared to times that are normally used.

4. Acknowledgements

The authors thankfully acknowledge the computer resources, technical expertise and support provided by the Laboratorio Nacional de Supercómputo del Sureste de México, CONACYT Network of national laboratories, Red de Física de Altas Energías (Red FAE).

References

- [1] Hillas A.M. *Shower simulation. Lessons from MOCCA*, *Nucl.Phys.B.Proc.Suppl.* **1997** (52) 29
- [2] A.M. Hillas, *17th Int. Cosmic Ray Conference, Paris 8*, **1981** 193

- [3] Kobal M., *A thinning method using weight limitation for air-shower simulations*, *Astropart.Phys.* **2001** (15) 259
- [4] Billoir P., *A sampling procedure to regenerate particles in a ground detector from a “thinned” air shower simulation output*, *Astropart.Phys.* **2008** 30 270
- [5] Stokes B.T., Cady R., Ivanov D., Matthews, G.B. Thomson., *Dethinning extensive air shower simulations*, *Astropart.Phys.* **2012** (35) 759 [arXiv:1104.3182]
- [6] Allen J., Castellina A., Engel R., Kasahara K., Knurenko S., Pierog T., Sabourov A., Stokes B.T., Ulrich R., Sako T., Ostapchenko S., *Air shower simulation and hadronic interactions*, *Eur.Phys.J.W.o.C* **2013** 53 01007
- [7] Ostapchenko S., *Status of air shower simulations*, *Nucl.Instr.Meth.Phys.Res.A* **2012** (662) S168 [arXiv:1010.0137]
- [8] Kuzmin V.A., Rubtsov G.I., *No-thinning simulations of extensive air showers and small-scale fluctuations at the ground level*, *J.E.T.P.Lett.* **2007** (85) 535
- [9] Gorbunov D.S., Rubtsov G.I., Troitsky S.V., *Air-shower simulations with and without thinning. Artificial fluctuations and their suppression*, *Phys.Rev.D* **2007** (76) 043004
- [10] Gorbunov D.S., Rubtsov G.I., Troitsky S.V., *Getting rid of artificial fluctuations in EAS simulated with thinning. Estimate of artificial fluctuations and simple economical method to suppress them*, *ICRC — 2007* **2007** (4) 641
- [11] Pierog T., Engela R., Hecka D., Poghosyanb G., Oehlschlägera J., Veberic D., *Ultra-high energy air shower simulation without thinning in CORSIKA*, *ICRC — 2015*
- [12] Stokes B.T., Cady R., Ivanov D., Matthews J.N., Thomson G.B. *Using CORSIKA to quantify Telescope Array surface detector response*, *ICRC — 2011* **2011** (2) 254
- [13] Hansen P.M., Vazquez R.A., Alvarez-Muñiz J., *Studying shower to shower fluctuations with simulations*, *ICRC* (**2009**)
- [14] Hansen P.M., Alvarez-Muñiz J., Vázquez R.A., *A comprehensive study of shower to shower fluctuations*, *Astropart.Phys.* **2011** (34) 503 [arXiv:1004.3666]
- [15] Álvarez J.deD., Cotti U., de León C., *Search for signals of New Physics in Extensive Air Showers induced by Ultra High Energy Cosmic-Rays*, *ICRC* (**2017**)
- [16] Sciutto S.J., *AIRES, a system for air shower simulation and analysis*, *Proc. 27th ICRC* **2001** (1) 237
- [17] Cotti U., de León C. *work in progress*