Astroparticle Physics 30 (2008) 270–285

Contents lists available at ScienceDirect

Astroparticle Physics

journal homepage: www.elsevier.com/locate/astropart

A sampling procedure to regenerate particles in a ground detector from a "thinned" air shower simulation output

Pierre Billoir

L.P.N.H.E., CNRS/Université P. et M. Curie (Paris 6)/Université D. Diderot (Paris 7), 4 Place Jussieu, 75252 Paris CEDEX 05, France

ARTICLE INFO

Article history:
Received 8 April 2008
Received in revised form 1 October 2008
Accepted 3 October 2008
Available online 21 October 2008

PACS:
95.75.Pq
96.50.sd

Keywords:
Cosmic rays
Atmospheric showers
Thinning
Ground detector
Sampling
Regeneration
Statistical fluctuations

ABSTRACT

Atmospheric showers induced by ultra-high energy cosmic rays contain a huge number of particles. Simulation packages cannot generate, follow and store all of them within a reasonable time. The "thinning" procedure was invented to cope with this problem; it provides a file containing a limited sample of *weighted* particles. We describe here a method to regenerate a set of particles entering a ground detector from such a file, reproducing as far as possible the properties of the signal, without using tabulations or building smoothed densities in the parameter space of particles. We discuss the possible biases on the amplitude and the time distribution of the signal and practical ways to suppress them. We also evaluate the *artificial* fluctuations due to the reduced size of the thinned sample and the dispersion of the weights, compared to the *natural* fluctuations due to the discrete nature of the shower and to the finite size of the detector. We define practical criteria to control the statistical quality, that is, to maintain the artificial fluctuations at an acceptable level.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The cascade of interactions induced by an ultra-high energy particle in the upper atmosphere has a very large size, of the order of 10^{12} secondary particles for a primary energy of $E_{prim} = 10^{19}$ eV. A full simulation of such showers cannot be performed with usual computing resources. The "thinning" technique [1] was designed to obtain, within a reduced computing time and a limited storage, estimations of both the longitudinal profile and the distribution of secondary particles at the ground level. It provides a subsample of particles with statistical weights, that allows to obtain unbiased estimators of any quantity that can be expressed as a sum over particles: local flux through a given surface, time distribution, etc. The principle is to apply to each interaction in the cascade below a *thinning energy* $E_{thin} = \epsilon E_{prim}$ the following procedure: keep a daughter particle with a probability p and, if kept for further interactions, define its statistical weight w as w_m/p, where w_m is the weight of the mother particle (the primary having weight 1).

The weight of any particle kept in the cascade is exactly the inverse of the probability to have been kept, and for any additive quantity q the sum $\sum_{kept} w_i q_i$ (over the particles kept within a given region) has an expectation value equal to the sum $\sum_{all} q_k$ that would have been obtained by summing, within the same region, over *all* particles of the full cascade. The original method [1] consisted in keeping only one daughter particle per interaction, with a probability proportional to its energy, to ensure the exact conservation of the total weighted energy; however different algorithms may be used, provided they do not bias the expectation values. The principle of the method is illustrated in Fig. 1.

Such estimators cannot account for the detailed structure of the shower. The local densities are obtained as averages over larger or smaller distances, depending on the statistics of the thinned sample. Short range correlations such as small subshowers giving local overdensities cannot be adequately reproduced. Because of the reduced number of particles, the sums on thinned samples are affected by the so called *artificial* fluctuations: the relative variance of a sum is larger than what would be obtained for the same sum on a fully simulated shower.

Details of the thinning procedure include choices for the values of p at each step for each particle, as function of the nature and the energy of the particle. These details are optimized in terms of the best compromise between computing resources required and the resulting fluctuations of the signal in the detector. In particular, it is clear that, for a given level of thinning, better results are achieved if the distribution of the weights is homogeneous, and

E-mail address: billoir@lpnhe.in2p3.fr

0927-6505/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.astropartphys.2008.10.002

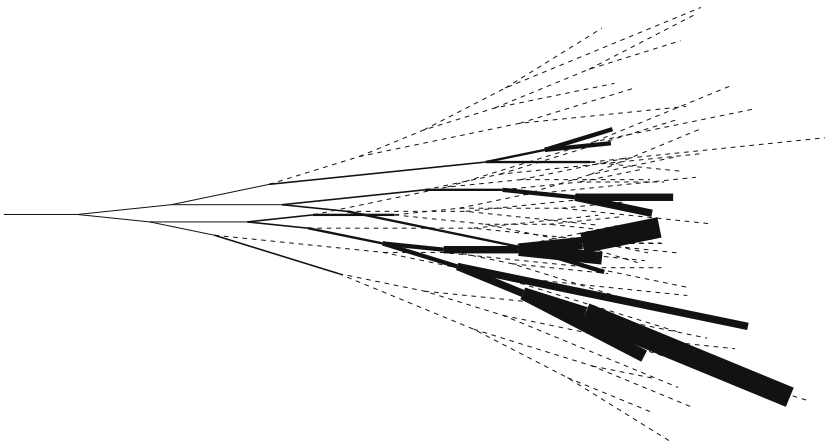

Fig. 1. Principle of the thinning procedure in a cascade of interactions: the dashed lines are particles not actually followed; the solid lines are the particles kept in the output file, with a width proportional to their statistical weight.

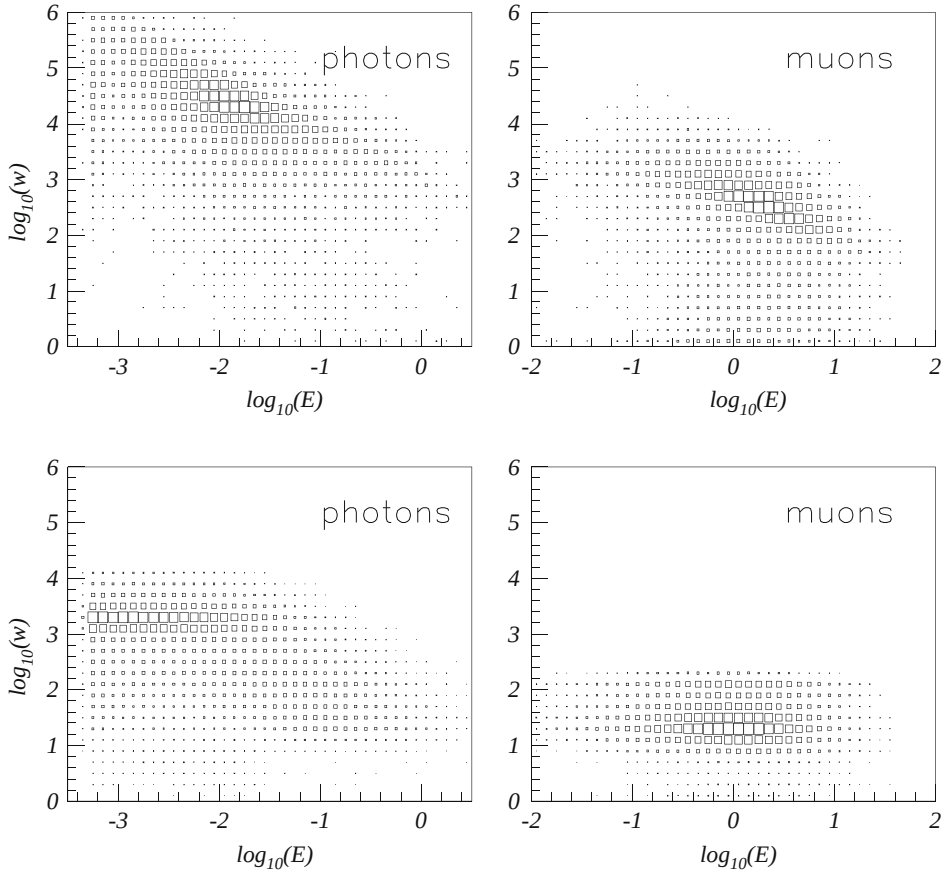

Fig. 2. Effect of the AIRES weight limitation algorithm for photons and muons at more than 500 m from the core, for a vertical proton shower at 10^{19} eV, with $\epsilon = 10^{-7}$ (electrons and positrons behave like photons). Distribution of weight vs energy (GeV): Top: without limitation: the weight is roughly proportional to $1/E$; Bottom: with limitation ($W_f = 1$ for photons, 1/88 for muons).

more precisely, if it has no tail at high values. Weight limitations procedures were defined to improve the thinning algorithm with respect to this condition, while keeping the computation time acceptable. Technical details may be found in [2], as well as a discussion on the impact of the dispersion of weights on the statistical quality of the thinned sample. Fig. 2 shows the effect of the weight limitation on a vertical proton shower using the simulation package AIRES [3], where the limitation is governed by a "thinning weight factor" W_f: the weight of photons, e^+ or e^- cannot be larger than $A_0\ E_{thin}\ W_f$ ($A_0 = 14\ \mathrm{GeV}^{-1}$). If the detector has an enhanced sensitivity of to muons, W_f is replaced for the hadronic particles by a much lower value (by default $W_f/88$), to reduce further the maximum weight of the muons.[1] As can be seen on Fig. 2, the mean weight is well below the allowed maximum.

[1] A small fraction of the muons come indirectly from the electromagnetic cascade, through photonuclear interactions. This produces a tail in the distribution of weights, above the nominal maximum for hadrons, with little impact on the artificial fluctuations.

The aim of the shower simulation is to provide estimators for signals in detectors. In this study we are interested to ground detectors, which are small compared to their spacing. In these conditions the simulation output file contains very few particles (or none at all) falling exactly onto a detector: they cannot be used to represent the actual flux of particles on this detector. The purpose of the *resampling* (or "unthinning") procedure is to use a wider sampling in the ground particle file, to obtain a fair simulation of the particles entering the detector.

The detector response to a single particle depends *a priori* on the type of particle and on its energy, impact point, and direction. A reliable simulation of the response to a flux of particles depends on the local density in a multidimensional *parameter space* of position, direction, time, and energy. This density cannot be described by a tabulation, nor by a factorization of densities, because the parameters are more or less correlated. Smoothing or parametrizing is difficult in an n-dimensional space; moreover, the choice of a pattern can distort a distribution. Here we want to obtain a representation without any prejudice, except the assumption that the density in parameter space depends *smoothly* on the position. Short range correlations, if any, are spoiled by the thinning algorithm, and difficult to restore without introducing large unwanted fluctuations. In this method they are ignored.

In Section 2 we describe the resampling procedure. In Section 3 we evaluate biases on some features of the signal in the detector, and we examine possible corrections. In Section 4, we evaluate the biases by a Monte Carlo simulation in the framework of the Surface Detector of the Pierre Auger Observatory [5]. In Section 5, we give analytical expressions for the artificial fluctuations due to the thinning-resampling procedure, and we propose some quantitative tools to control them; we discuss the impact of artificial fluctuations on the global analysis of a shower.

2. The local resampling procedure

2.1. Basic principle

We define a *sampling region* on the ground around the detector to be simulated. This region should be small enough to reproduce well the density in the parameter space (energy, direction, time) at the detector position, but large enough to keep the artificial fluctuations below an acceptable level, allowing to reproduce the natural fluctuations of the signal in this detector; finding a good compromise is the guiding line of the next Sections. The particles within this region are used, according to their weight, to regenerate particles with the same energy, direction and time, within the detector surface. The number of particles hitting the detector has an expectation value equal to the number of particles into the sampling region, scaled by the *sampling ratio* $R = \mathscr{A}_d/\mathscr{A}_{sr}$ where $\mathscr{A}_d$ is the area of the detector and $\mathscr{A}_{sr}$ the area of the sampling region. In other terms, each particle in the thinned file, with a weight w, is considered as being equivalent to a flux $w/\mathscr{A}_{sr}$ of particles of its nature with its energy, direction and arrival time.[2] This is illustrated on Fig. 3. The number of particles of this type falling into the detector is generated according to a Poisson law with a parameter equal to the *resampled weight* $w_r = w\mathscr{A}_d/\mathscr{A}_{sr}$. If $w_r \ll 1$, this procedure amounts to keep this particle with a probability w_r; else, it may generate several *clones*. In any case, it reproduces both the average and the Poisson fluctuations expected from an incident flux, supposed to be uniform at the scale of the detector.

For a more precise description, the detector area should be understood as an *effective* one if it is not a simple horizontal surface: $\mathscr{A}_d$ is the projection onto an horizontal plane, along the direction $\vec{n}_p$ of the *incident particle* (hatched area on Fig. 4); that is, $\mathscr{A}_d$ is *not the same* for all particles. For example, if we consider a vertical cylinder of radius R and height H, we have:

$$\mathscr{A}_d = \pi R^2 + 2RH\tan(\theta_p) \tag{1}$$

where θ_p is the angle between $\vec{n}_p$ and the vertical direction. More generally, the effective area is the sum of all elements $\mathrm{d}S$ of the detector surface hit by $\vec{n}_p$, multiplied by a projection factor depending on the orientation of the particle and of the surface element, as illustrated in Fig. 4:

$$\mathrm{d}\mathscr{A}_d = \cos(\alpha_p)/\cos(\theta_p)\,\mathrm{d}S \tag{2}$$

where α_p is the angle between $\vec{n}_p$ and the direction normal to $\mathrm{d}S$. To simulate an entry point according to the "uniform flux" hypothesis, a point is taken randomly within $\mathscr{A}_d$, and then projected onto the detector surface along $\vec{n}_p$.

2.2. Time assignment

The shower may be described in first approximation as a "pancake" just behind the *front plane*, defined as the plane perpendicular to the shower axis (unit vector $\vec{n}_s$[3]), moving at speed c. For an inclined shower, the difference of arrival time between different ground points within the sampling region may be large compared to the typical time scale of the signal in a given detector (the thickness of the shower divided by c). Then a time correction is needed, to keep invariant the *delay* $\tau = t_p - t_{fp}$ of the resampled particle w.r.t. the front plane, that is, to preserve the intrinsic structure of the shower, assuming that it does not change much within the sampling region:

$$t' = t_p + \vec{n}_s \cdot (\vec{r}_p - \vec{r}_d)/c \tag{3}$$

where $\vec{r}_p$ is the position of the ground particle and $\vec{r}_d$ the position of the detector. For a thick detector, a further delay accounts for the propagation from an horizontal plane of reference to the detector surface.

Another problem may appear if the resampled weight w_r is large: in that case, the ground particle has to be regenerated several times; in the real world, the arrivals occur at random times, while the procedure described above gives the same time for all clones of this particle. If the time structure of the signal is exploited for physics analysis such as detecting peaks in a signal trace over time, the artificial piling of particles is a flaw in the simulation. A smearing procedure may be applied to cure it, exploiting the fact that the shower has a typical overall structure: a fast rise behind the front plane, followed by a slow, roughly exponential decrease, with a time scale approximately proportional to the distance from shower axis. A possible implementation consists in applying a log-normal smearing to the delay:

$$\tau' = \tau\exp(\sigma G) \tag{4}$$

where G is a random number from a Gaussian distribution with mean 0 and variance 1, and σ a small number (typically 0.1–0.2). With this procedure, the spread of the clones is small in the early part of the signal and larger in the tail, so that the overall shape is practically unchanged; moreover the delay is never negative. The clones are prevented from piling up as soon as the delay is large compared to the width of the signal of a single particle; in the early part, they may overlap, but this would happen anyway in the real

[2] This flux *does not* represent in any way a subshower related to this particle; it has a pure statistical meaning: the sum of all these fluxes is an unbiased approximation of the actual flux in a real shower.

[3] Here we take the convention that $\vec{n}_s$ goes *upwards* (opposite to the actual movement of the shower).

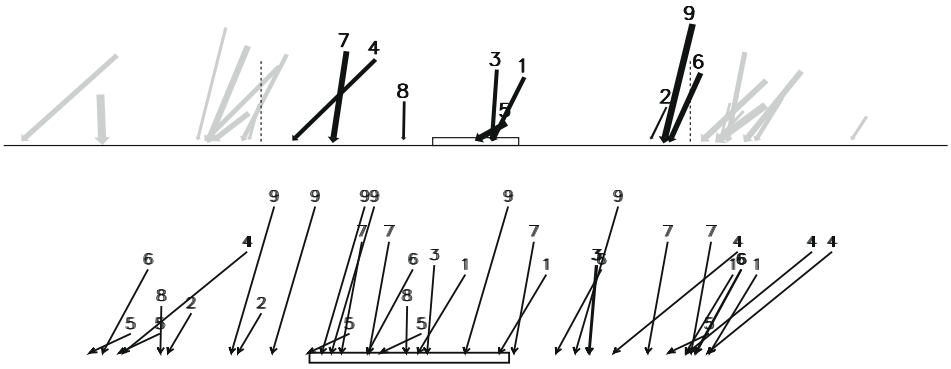

Fig. 3. Resampling from a weighted ground particle file. Top: each weighted particle is represented by an arrow (length proportional to energy, width proportional to weight); the sampling region around the detector (horizontal rectangle) is delimited by the dashed lines. Bottom: each arrow falling in the sampling region (in black on the top) is converted into a flux of particles with the same direction and energy, with an intensity proportional to the weight, falling on the ground. Similar arrows are placed at positions generated uniformly and independently, according to these fluxes; they give a fair representation of the particles hitting the detector.

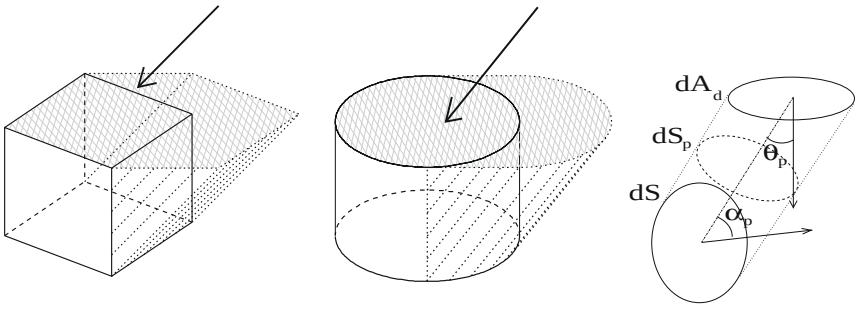

Fig. 4. Effective area (hatched) of a detector receiving a flux directed along the arrow: top area + projection of the lateral area onto an horizontal plane. Right: an elementary area dS on the detector surface corresponds to $dS_p = dS\cos(\alpha_p)$ in an "intrinsic" section (normal to the direction of the particle), and to $d\mathscr{A}_d = dS_p/\cos(\theta_p)$ when projected onto an horizontal surface.

world. The impact of this procedure depends on the time resolution of the detector; it will be examined using simulations in Section 4.4.

2.3. Definition of the sampling region

In a first approximation the shower has an azimuthal symmetry around its axis (the "core"). It is convenient to introduce the *shower frame* defined by $\vec{n}_s$ and two perpendicular axes in the front plane, for example an horizontal one $\vec{n}_h$ and the other one $\vec{n}_v$ into the vertical plane containing $\vec{n}_s$ (see Fig. 5).

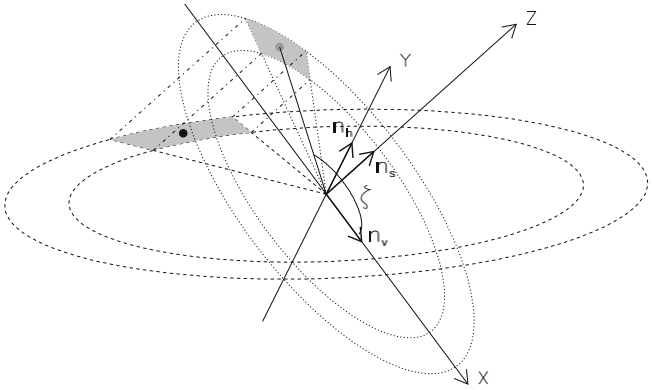

Fig. 5. Definition of the shower frame: Z is along the shower axis, X in the vertical plane going through the shower axis, and Y in the horizontal plane. The "standard" sampling region around the detector (black point) is in grey: it is the projection onto the ground, along Z-axis, of a region in the shower plane (X, Y) delimited by a circular crown and an angular sector.

Let (X,Y,Z) be the coordinates along the axes defined by $(\vec{n}_v,\ \vec{n}_h,\ \vec{n}_s)$. Z is the longitudinal coordinate along the shower axis, $r = \sqrt{X^2+Y^2}$ is the distance to the core, and ζ is the azimuthal angle of (X, Y) in the shower frame ($\zeta = 0$ for the upstream direction). The density in our parameter space depends mainly on r. The dependence on ζ accounts for the asymmetry, and the dependence on Z accounts for the longitudinal evolution of the shower. The dependence on Z also may be expressed as a dependence on $\cos(\zeta)$, if the ground is horizontal.

In these conditions, a natural choice for the sampling region around a detector at position $(r_d,\ \zeta_d)$ is (grey area in Fig. 5):

$$|r - r_d| < \delta|r_d|; \quad d(\zeta,\zeta_d) < \alpha \tag{5}$$

where $d(\zeta,\zeta_d) \leqslant \pi$ is the angular distance from the detector. The constants δ and α should be small enough to ensure the validity of the local averaging (see Section 3) and to avoid artificial correlations between neighbouring detectors due to an overlap of their sampling regions. However δ and α should be large enough to avoid large artificial fluctuations (see Section 5). Hereafter, this will be called a "standard" sampling region.

2.4. First approach to the artificial fluctuations

At this stage, we can consider that the resampling procedure is satisfactory if the resampled weights are well below 1, that is, for any category, the sample of particles entering the detector is extracted from a much larger set, where the fluctuations on any interesting physical quantity are negligible as compared to the fluctuations in the detector itself. The resampled weight w_r may be large, either because the original weight is large, or because the sampling region is small: for example, the area of a standard

sampling region scales as r_d^2, then w_r may be large at short distance from the core.

We will study here the distribution of w_r in the case of the water Cherenkov tanks of the Pierre Auger Observatory (cylinders of diameter 3.6 m, height 1.2 m). In such a thick detector the "hard" particles (giving the highest signals) are the muons and the photons/electrons of high energy. Typical running conditions ($E_{prim} = 10^{19}$ eV, $\epsilon = 10^{-7}$, $W_f = 1$) give the distribution shown in Fig. 6. The resampled weights for the hard particles are in general well below 1 at distances of the order of 1000 m or more, so no clones are generated. "Soft" particles (low energy photons/electrons), or any kind of particles at low distance, will generate clones, but the global signal in the detector will be generally reproduced in a satisfactory way:

- If many particles hit the detector within a short time, their individual signals cannot be distinguished and the superposition of the clones reproduces well the time dependence of the global signal. This point will be discussed later in more detail.
- At a short distance from the core, the signal is very large, then its relative statistical fluctuation is very small, possibly dominated by the artificial one. If this occurs in only one detector, the errors on the reconstructed parameters of the shower are dominated by the errors on the other detectors, so the overall impact of artificial fluctuations in this detector is negligible. If artificial fluctuations dominate in several detectors this is no longer true, but the systematic errors on the physical quantities of interest are often larger than the errors induced by the fluctuations in these detectors. In that case, the resampling procedure may be considered as acceptable as soon as the artificial fluctuations do not dominate the final errors, including the systematic ones.

To summarize, the quantitative criteria proposed in the following aim to be a guide for evaluating the impact of artificial fluctuations, but they need to be adapted to the practical context of an experimental setup and to the actual state of the systematic errors. For example, within the primary energy range covered by an experiment, the thinning ratio may be modulated as a function of E_{prim}: a constant thinning energy E_{thin} makes w_r independent of E_{prim}, but at highest energies this option has a enormous cost in CPU time and file storage, while the gain on the statistical quality at short distance may be useless. On the contrary, the usual practice of a constant thinning ratio ϵ is not optimal: if ϵ is well adapted to highest energies, the statistical quality is too good for the lowest ones.

3. Biases and possible corrections

3.1. General considerations

If the density dependence on the position is smooth, the average value in the sampling region is approximately equal to the density at its center of gravity, which is approximately the position of the detector. As the sampling region cannot be arbitrarily small, both these simplifications induce a bias. We want here to evaluate the deviations for some physically interesting quantities: the amplitude of the signal and some variables describing the distribution in time. For example, at second order in α and δ, the average of

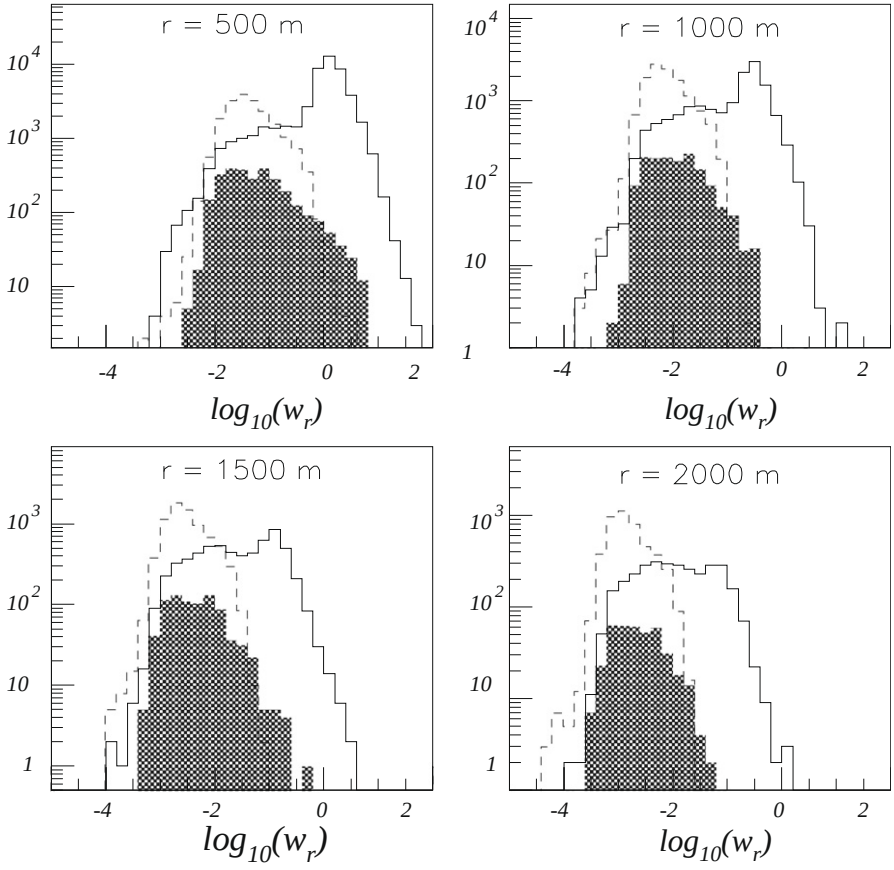

Fig. 6. Distribution of resampled weights in a vertical proton shower (AIRES, $E_{prim} = 10^{19}$ eV), with weight limitation ($E_{thin}/E_{prim} = 10^{-7}, W_f = 1$), into Auger Cherenkov tanks. Solid histogram: photons (black: above 100 MeV). Dashed: muons.

a density $f(r,\zeta)$ over a standard sampling region is given by an integration in polar coordinates:

$$\frac{1}{4\alpha\delta r_d^2}\int_{r_d-\delta r_d}^{r_d+\delta r_d}\int_{\zeta_d-\alpha}^{\zeta_d+\alpha} f(r,\zeta) r\,\mathrm{d}r\,\mathrm{d}\zeta = f(r_d,\zeta_d)+\frac{1}{6r_d}\left(\frac{\partial^2 g}{\partial r^2}(\delta r_d)^2+\frac{\partial^2 g}{\partial \zeta^2}\alpha^2\right) \quad (6)$$

where $g(r,\zeta)=rf(r,\zeta)$.

In the spirit of the resampling method proposed here, f is *a priori* unknown, and should not be used explicitly to estimate the response of the detector. However, we can reduce the dependence to an external model (e.g. an average parametrization of f) if we use it *only* to evaluate a *correction to the bias*. Alternatively, a correction may be applied to the weights or to the parameters of the ground particles; for example, an *ad hoc* dependence in r may compensate the non-linearity of g in Eq. 6. We propose hereafter corrections on weights, times and directions, using a reasonable external modelling, a general symmetry or a scaling hypothesis.

3.2. Bias and correction on the number of particles

The main source of bias is the dependence of the density f on r. The function $g(r)=rf(r)$ is usually concave, so the local density is overestimated (see Fig. 7. The relative excess is $\simeq g''(r)/g(r).(\delta r)^2/6$. This can be written as $\eta(\eta-1)\delta^2/6$ if $f(r)$ is locally proportional to $r^{-\eta}$. In typical showers with moderate zenith angle θ, η is of the order of 3–4, so the excess is typically δ^2 to $2\delta^2$. With $\delta\sim 0.1$, this is much less than the natural fluctuations; moreover, all detectors are roughly affected by the same factor, then the only noticeable effect is a slight overestimation of the energy of the shower.

The value of η depends mainly on r_d and θ. There may be variations due to the nature of the primary, the modelling of high energy interactions, and the shower-to-shower variations; all of them are well below 1. Then a correction may be reasonably applied to the weights, as a function of r_p, distance of the ground particle to the shower axis:

$$w' = w\left(\frac{r_p}{r_d}\right)^{\eta-1} \quad (7)$$

As an example, Fig. 7 shows the effect of such a correction on the total energy of photons and on the number of muons, at different distances from the core, for the superposition of ten vertical showers. The sampling zone is a wide circular crown ($\alpha = 180°$, $\delta = 0.2$); the "true" value is evaluated using $\delta = 0.05$. Although the uncorrected values (points at $\eta = 1$) have various biases, a correction with $\eta\sim 3.5$ gives good results in all cases, and the precise tuning of η is not crucial.

However, this may be wrong for particular configurations, for example for a shower with a late development, that hits the ground before reaching its maximal size: in that case the lateral distribution is steeper. The value of η may be tuned to account for the nature of the primary, the distance to the core, etc; in any case keeping $\delta \lesssim 0.1$ is a safe option.

The dependence in the longitudinal position Z mentioned above may be expressed as a function of the slant depth X (thickness of air traversed along the shower axis from the entrance in atmosphere), then the local density is biased if the longitudinal evolution is not linear. This evolution may be described by a profile function $G(X-X_{max})$ where X_{max} is the usual parameter defining the position of the maximal size, which may be precisely evaluated in simulated showers. Once parametrized, G may be used to correct the weights. It depends on the nature of the particle, and slightly on r.

3.3. Delay and time spread

As the distribution of arrival times depends on r the signal is artificially stretched by the extension in r of the sampling region (see Fig. 8). Both the average and the spread will be biased. The cascade of interactions propagating from the core produces a curved front. The simulations suggest an approximately spherical shape, or a paraboloid described by $\tau_f(r)=r^2/(2Rc)$ (R: radius of curvature), and a distribution of the delay after the front scaling approximately as r. Within this picture, the average delay of the particles in the sampling region is the sum of the average of $\tau_f(r)$ and the average of $\tilde{\tau}_d\cdot r/r_d$, where $\tilde{\tau}_d$ is the average delay after the front at $r=r_d$. Using the local parametrization $f(r)\propto r^{-\eta}$ and the same second order computation as in Section 3.2:

$$\int_{r_d-\delta r_d}^{r_d+\delta r_d} r^\alpha\,\mathrm{d}r = r_d^{\alpha+1}\int_{-\delta}^{+\delta}(1+u)^\alpha\,\mathrm{d}u \simeq r_d^{\alpha+1}\times\int_{-\delta}^{+\delta}(1+\alpha u+\alpha(\alpha-1)u^2/2)\mathrm{d}u = 2\delta r_d^{\alpha+1}(1+\alpha(\alpha-1)\delta^2/6)$$

the first term may be written as:

$$\frac{\int_{r_d-\delta r_d}^{r_d+\delta r_d}\tau_f(r)rf(r)\,\mathrm{d}r}{\int_{r_d-\delta r_d}^{r_d+\delta r_d} rf(r)\,\mathrm{d}r} = \frac{1}{2Rc}\frac{\int_{r_d-\delta r_d}^{r_d+\delta r_d} r^3 f(r)\,\mathrm{d}r}{\int_{r_d-\delta r_d}^{r_d+\delta r_d} rf(r)\,\mathrm{d}r} \simeq \frac{r_d^2}{2Rc}\left(1-\frac{2\eta-3}{3}\delta^2\right) \quad (8)$$

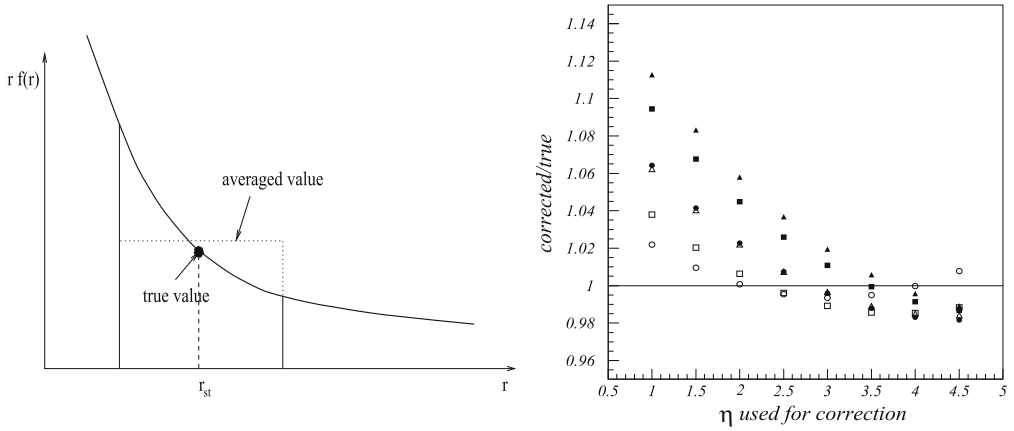

Fig. 7. Sampling biases on the amplitude of the signal. Left: origin of the overestimation. Right: effect of a r-dependent correction on the energy of photons (solid symbols) and the number of muons (open symbols). Circles, squares, triangles are for $r_d = 500, 100, 1500$ m, respectively.

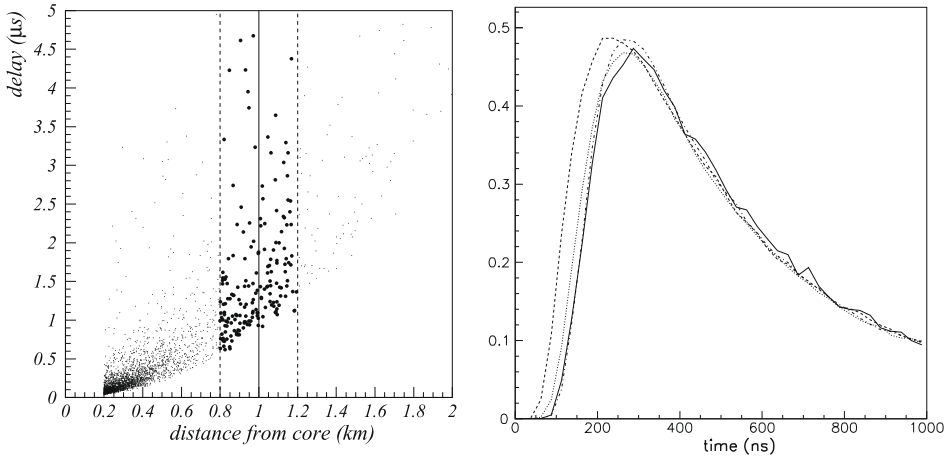

Fig. 8. Sampling biases on time structure. Left: origin of the distortion. The particles in the sampling region (between the dashed lines) are used to represent the distribution in t at 1 km from the core (solid line). *NB: the curvature of the front has been exaggerated for clarity*. Right: effect of a time correction. Solid: "true" distribution of the arrival times of the muons in a vertical shower (obtained with $\delta = 0.05$); the meaningless oscillations are due to the poor statistics; dashed: raw distribution ($\delta = 0.3$); dotted: after a "scaling" correction in r_d/r_p; dash-dotted: with a quadratic curvature correction in addition (see text).

and the second term:

$$\frac{\tilde{\tau}_d}{r_d}\frac{\int_{r_d-\delta r_d}^{r_d+\delta r_d} r^2 f(r)\,\mathrm{d}r}{\int_{r_d-\delta r_d}^{r_d+\delta r_d} r f(r)\,\mathrm{d}r} \simeq \tilde{\tau}_d\left(1-\frac{\eta-1}{3}\delta^2\right) \tag{9}$$

With the usual values of η, both terms have a negative bias of the order of δ^2 in relative value. If the weight correction suggested in formula (7) is applied, $rf(r)$ is replaced by a constant in the previous calculation: the first term has no bias and the second one a relative bias $\delta^2/3$.

Another useful quantity is the start time of the signal, needed for the evaluation of the front curvature. In a simplified approach, it coincides with the time of the first particle. The bias cannot be computed in a simple analytical formula, but it is clearly negative, because the region $r < r_d$ provides particles with times more concentrated than at r_d, after an earlier front time. In the asymptotic case where an infinite number of particles hit the detector, the first one arrives at $\tau_f(r_d - \delta r)$ instead of $\tau_f(r_d)$: the bias is of first order in δ. The weight correction (7) increases the mean value of the start time, but it does not change the asymptotic behaviour mentioned above.

The shape of the signal as a function of time undergoes a complex distortion: the superposition of different slices in r, with an increasing delay, enhances the spread of the times, but the larger density at low r favours a narrower distribution. An explicit expression of the time profile would be needed to obtain analytical expressions for variables related to it, such as the fractional time t_x where x denotes a percentage of the integrated signal. Monte Carlo evaluations will be performed in Section 4.

In the approximation given above (parametrized front and time distribution scaling with r), a correction may be applied to the delay:

$$\tau' = (\tau - \tau_f(r_p))\frac{r_d}{r_p} + \tau_f(r_d) \tag{10}$$

The value of R depends on the zenith angle, the nature and the energy of the primary particle, and has non-negligible fluctuations, so that the correction cannot be fully satisfactory. It was applied for a simulated vertical shower to the arrival time of the muons at ground: the distribution is quite asymmetric (steep rise followed by a quasi-exponential decrease), so the sampling distortion is mainly a negative shift of the rise front. Fig. 8 shows this bias for a large δ (0.3) and the efficiency of the correction (10) which restores quite an acceptable shape if the time resolution is 10 ns or more.

3.4. Corrections on the directions

Taking ground particles in an extended sampling region induces an artificial spread on the incident directions. Here again, the biases on the signal are of second order in δ and α. The exact evaluation depends on the type and the geometry of the detector, and also on the nature and the energy of the particle; in most cases, the biases are small. Here we will only discuss possible corrections.

Assuming the radial symmetry to hold in first approximation, a ground particle at $r_p,\ \zeta_p$ in shower frame should be globally rotated by an angle $\Delta\zeta = \zeta_d - \zeta_p$, and then moved from r_p to r_d, to represent a particle entering the detector. The first operation is straightforward: apply a rotation $\Delta\zeta$ around $\vec{n}_s$ to $\vec{n}_p$. The second one is more problematic and depends on a model for the dependence in r of the divergence γ_p of the particles (angle of $\vec{n}_p$ with $\vec{n}_s$); γ_p depends also on the nature and the energy of the particle, so there is no simple general recipe to do the correction.

Even with such corrections, the opening angle α of the sampling region cannot be fully relaxed, because the radial symmetry is severely broken by the presence of the ground. In most shower simulations (as in the real world in first approximation) the ground acts as a perfect absorber, so $\vec{n}_p$ is always down-going: this condition is clearly not invariant by rotation around $\vec{n}_s$, or by a modification of γ_p. As a consequence, the procedure described above may fail in some cases: if the direction is up-going after correction, the particle has to be rejected; conversely, a virtual down-going particle may have been missed because it would have been up-going before correction.

Moreover, if the direction is modified, the flux on ground (that is, the statistical weight w) has to be modified accordingly. The invariant quantity is the "proper flux" (number of particles through a surface perpendicular to $\vec{n}_p$). The flux on the ground is the proper flux multiplied by $\cos(\theta_p)$ (vertical component of $\vec{n}_p$): if the correction goes from θ_p to θ'_p, the weight should be multiplied by $\cos(\theta'_p)/\cos(\theta_p)$. This factor may be arbitrarily large and induce additional artificial fluctuations (see the discussion in Section 5.4). As a last complication, the effect of the absorption in ground is not simply to suppress up-going particles, at a given horizontal level, in a freely developing radially symmetric shower: in such a picture, some down-going particles (especially the photons and electrons/positrons) may have ancestors below the ground level, and then they should be suppressed if the ground is absorbing.

In practical conditions, the situation is more favourable. Moderately inclined showers have very few up-going particles, and these

have low energies. Very inclined showers contain essentially well collimated muons and their direction is practically unmodified by a small rotation around $\vec{n}_s$. As a conclusion, the best compromise is to chose moderate values of α (typically, less than 20°) and to check whether the rotation around $\vec{n}_s$ and the change on the divergence γ_p give significant modifications to the quantities of physical interest; if yes, and if the modification has undesirable effects (particles becoming nearly horizontal, large modification of weights), it is better to reduce α and not to apply corrections.

4. Monte Carlo evaluation of some biases

4.1. Operating conditions and simplified detector simulation

Vertical proton showers ($E_{prim} = 6 \times 10^{19}$ eV) were simulated using AIRES. The relevant thinning parameters are the thinning ratio ϵ and the thinning weight factor W_f. Here we used $\epsilon = 10^{-7}$ and $W_f = 1$: this ensures a much better statistical quality than the standard Hillas algorithm [1] with the same ratio, against a moderate increase of computing time. The detectors were regularly spaced on each of four circles, at distances $r = 500$, 1000, 1500 and 2000 m. The sampling region was defined, by default, by $\delta = 0.1$, and $\alpha = 15°$; other values were used to evaluate the dependence of the biases on the sampling parameters. No correction was applied.

To study the effects of a poor thinning (large ϵ), we applied to the ground file produced by AIRES a "post-thinning" procedure: the particles were randomly kept with a probability p, and the weights were multiplied by $1/p$ ($p = 0.1$ and 0.01).[4] This is not exactly equivalent to running AIRES with a thinning energy divided by p, but here it was necessary to work on a unique shower to get rid of the shower-to-shower fluctuations; in practice, the statistical quality is the same.

Whenever the amplitude of the signal is relevant, a simplified simulation of the Auger Cherenkov tanks is used: for incident photons, the pair production and the Compton scattering are fully simulated; for charged particles the energy loss is taken constant and the light deposition is supposed to be proportional to the range in water. This procedure may be wrong in the case of subrelativistic particles, with reduced Cherenkov rate; it also neglects energy dependent effects like δ-ray production and bremsstrahlung. However it reproduces very well the structure of the trace and it is adequate to make comparisons between different selections of the resampling parameters. The signal is defined as the number of photo-electrons produced in the PMTs. Most photons and electrons have an energy below 100 MeV at entrance, and give a signal roughly proportional to this energy. At 250 MeV they are equivalent to one vertical through-going muon. At large energies, the energy deposition is not fully contained in the tank and the signal increases slowly with the initial energy.

4.2. Biases due to the size of the sampling region

The azimuthal dependence of the density is smooth at any zenith angle θ, and values of α less than 20° are small enough to reproduce the asymmetries. Here we explore only the biases due to large values of δ, using vertical showers. In that case we can take $\alpha = 180°$. We have tried seven values of δ (0.025, 0.05, 0.1, 0.15, 0.2, 0.25 and 0.3) to estimate, for muons and photons, the number per detector (N_μ, N_γ) and the integrated signal (S_μ, S_γ). We have also examined the start time t_0 and the fractional times.

Fig. 9 shows that the biases on N_μ, N_γ, S_μ, S_γ are practically negligible up to $\delta = 0.1$ and may be significant above 0.2; the dependence is compatible with a second-order effect in δ, as expected (the irregular structure for N_μ, S_μ at 2000 m for $\delta < 0.1$ is due to the poor statistics of input muons in these conditions).

Fig. 10 suggests the same conclusion for the global time structure: no strong effect is expected for the rise time (t_{50}–t_{10}) or the fall time (t_{90}–t_{50}); however, the bias on t_0 increases with the distance: this may result in a bias on the front curvature. From this point of view, the condition $\delta \leqslant 0.1$ is strongly required, unless a curvature correction is applied when shifting the time of arrival.

4.3. Biases related to the level of thinning

In addition to the levels 10^{-7}, 10^{-6}, 10^{-5} obtained directly or with the"post-thinning" procedure as explained above, we can obtain a lower value by exploiting the exact radial symmetry of a vertical shower. Using a full crown as sampling zone, that is taking $\alpha = 180°$ instead of $\alpha = 15°$, we obtain a level equivalent to $(15/180) \times 10^{-7}$. This level is referred as 10^{-8} hereafter.

By construction, the algorithm does not induce any dependence of the number of particles and the integrated signal on the thinning level ϵ: this may be seen in Fig. 11. On the contrary, the time structure may be modified, because the time smearing of clones does not compensate the effects of an undersampling (it is *not* intended to do so). Fig. 12 shows that the start time (in practice the time of the first sampled particle) is larger for a poor thinning, as expected, but the effect is still very small down to 10^{-6}. The fractional times have a negative (but not dramatic) bias, increasing with the thinning ratio.

The main effect of a poor thinning will appear on the *fluctuations* of the measured quantities. The effects seen in Fig. 13 (top) may seem to be large, especially on the photon signal (dominating at small distances); however, if several detectors are used to measure the global size of the shower, an artificial fluctuation of a few percent on a detector very close to the core, although it is well larger than the natural one, has no dramatic consequence. The precision on the global event parameters is limited by the statistical errors on the signals in other detectors. Then we can conclude that $\epsilon \leqslant 10^{-7}$ is fully satisfactory, and with 10^{-6} the effects are well below the systematic errors on energy measurement.

A similar conclusion is valid for the time structure (see Fig. 13, bottom): at 1000 m or more, the values obtained at 10^{-8}, 10^{-7} and 10^{-6} are practically the same. At 500 m, the thinning at 10^{-8} seems to give lower values, but the difference is not really significant: at this distance, t_0 is within a few 10 ns, and the fractional times are within a few 100 ns; then a variation of one percent is generally well below the time resolution of the detector. Moreover, if the exact position of the core is not known, an extreme precision on a distance dependent quantity is useless. In practice, the errors on the significant quantities computed from the time structure (at the event level) are often dominated by the error on the core position and the fluctuations on the farther stations.

4.4. Peak recognition

As explained in Section 2.2, in case of large resampled weights, several clones of an input particle are generated, with a time smearing designed to separate them without affecting the global structure: this condition is automatically fulfilled if the quantity σ is much smaller than 1. However, a peak recognition procedure may be distorted, because the signals of the clones may overlap more than in the real world. The practical consequences depend on the time resolution of the detector: in the following discussion we assume a time unit of 25 ns, as in the Auger Surface Detector.

[4] To reduce the storage size, AIRES applies by default such a procedure to particles very close to the shower core, with $p = (r/r_m)^2$ if $r < r_m$. In all our examples, r_m was less than 200 m, so this has no consequences for our results.

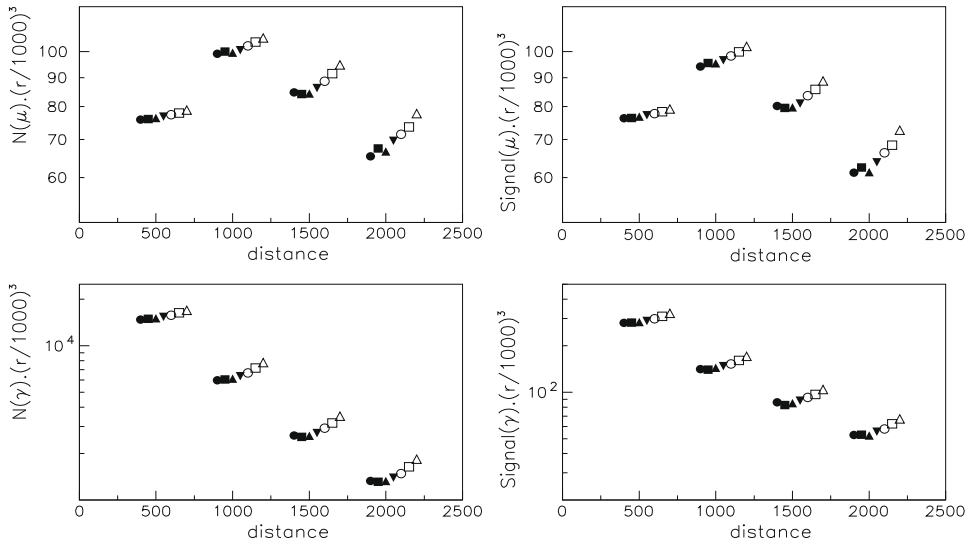

Fig. 9. Estimated number of muons/photons and integrated signals, as a function of the radial extension of the sampling zone (δ = 0.025, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, from left to right), for each of the four values of the distance (500, 1000, 1500 and 2000 m). To reduce the vertical scale, the values are multiplied by $(r/1000)^3$.

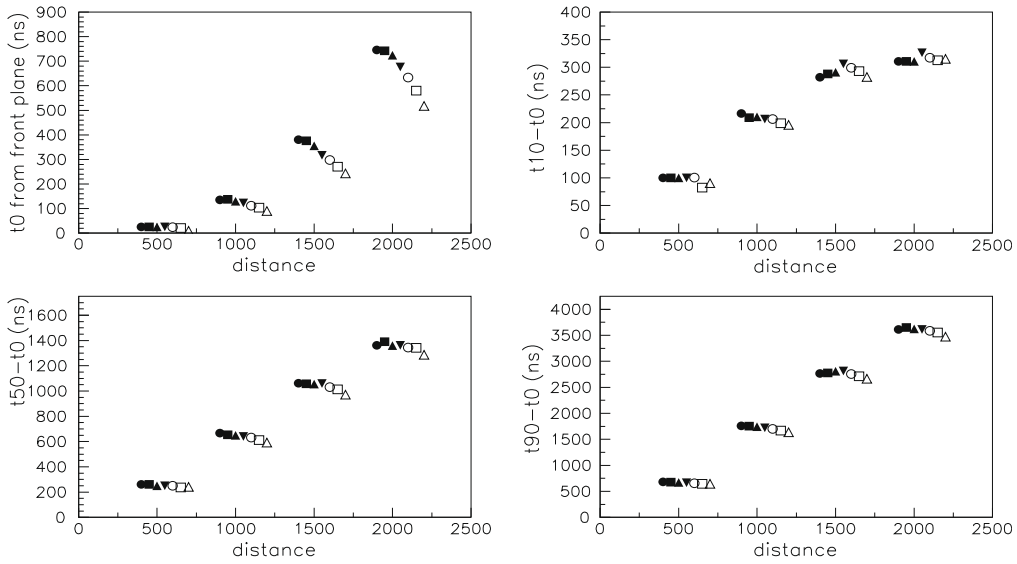

Fig. 10. Estimation of the time structure (starting time and fractional times), as a function of the radial extension of the sampling zone, with the same conventions as in Fig. 9.

Fig. 14 shows the effect of the smearing algorithm defined by Eq. (4). Fig. 15 shows the average time separation between successive muons. At 500 m from the core, it is of the order of a few ns, and there is no hope to separate the peaks; at 1000 m, the typical separation is of the order of 20 ns, comparable to the time smearing of clones. Fig. 16 shows examples of muon repartition, before and after smearing, with a thinning of 10^{-6}: it is clear that the presence of clones does not generate artificial substructures.

4.5. Inclined or low energy showers

For inclined showers the biases on averaged values and fluctuations are expected to be similar. The most serious problem was the estimation of t_0: at large zenith angle the front gets thinner and flatter, then the bias on t_0 should decrease.

For a given thinning ratio, going to a lower primary energy automatically reduces the weights. The artificial fluctuations are so less important. For example, at short distances, the problem of large resampled weights is suppressed.

5. Artificial fluctuations

In this study we want to establish semi-analytical expressions of the fluctuations of the signal in a detector; our aim is:

- to establish practical criteria to define acceptable general conditions of thinning and resampling,
- to insert statistical tools in the simulation chain, in order to check that the artificial fluctuations do not play an important role in the final result, for a given shower at a given position.

5.1. Classification of the fluctuations

We classify the natural fluctuations according to their order of occurrence:

(1) The first steps of the hadronic and electromagnetic cascade produce the *shower-to-shower* fluctuations, which are

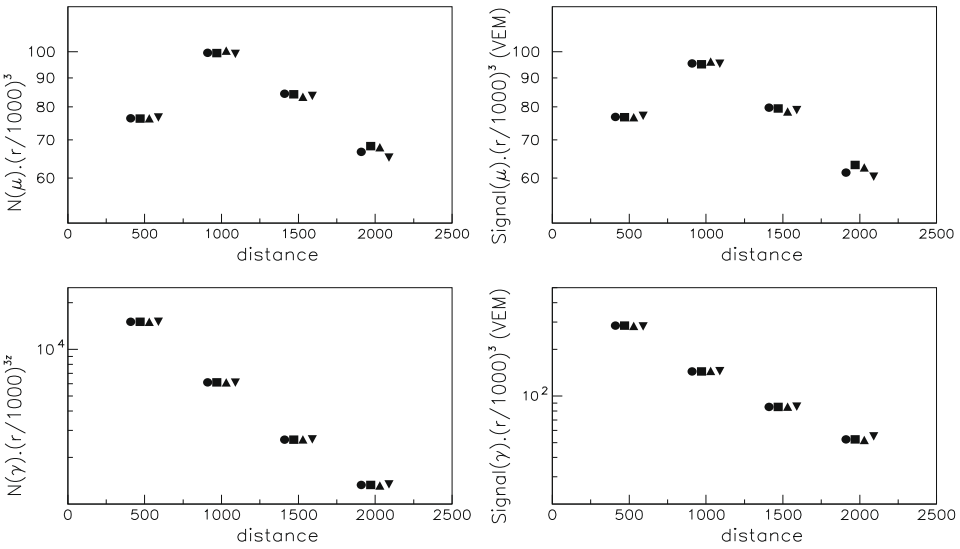

Fig. 11. Estimation of the number of particles and the integrated signal, as a function of the thinning level. For each distance, from left to right: 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}. The unit for the signal (VEM) corresponds to one vertical muon in the middle of the tank.

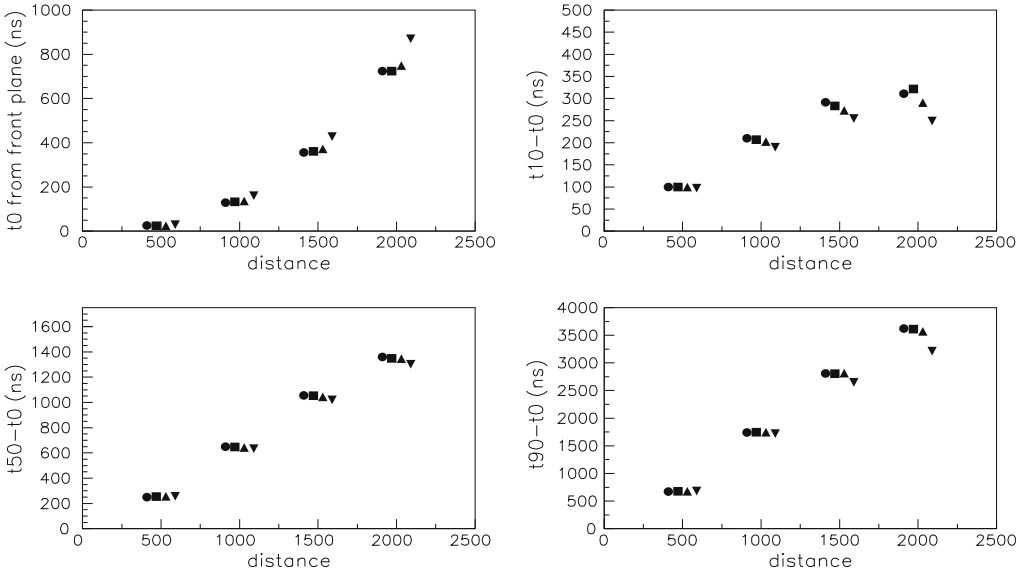

Fig. 12. Estimation of the time structure as a function of the thinning level (same conventions as in Fig. 11).

irreducible. When many ground detectors are hit, they are the main source of uncertainty on individual energy measurement and identification.

(2) There may be fluctuations at *intermediate scales* (between the detector size and the full shower size), for example due to small subshowers generated by a high energy particle. They are difficult to measure with a sparse array of detectors; the thinning procedure washes them out at a scale depending on the thinning level. We suppose here that they are negligible.

(3) At the *detector* level, there is a finite number of particles, following a Poisson statistics. We want to consider here the case where the signal of one particle extends over a large spectrum, so that the distribution of the total signal cannot be reduced to a Poisson or Gauss law.

(4) The fluctuations in the final detecting device (for example on the number of photo-electrons in a PMT) are supposed to be negligible compared to the previous contributions. In the same way, we ignore the contribution of the electronic chain of amplification and acquisition.

Then, we have to define quantitatively the artificial fluctuations occurring in the simulation chain because of the thinning-resampling procedure. They may be evaluated with the same tools as the natural ones, considering sampling regions as virtual detectors. We give practical estimators using the ground particle files themselves.

5.2. Fluctuations of the signal in a detector

5.2.1. Analytical expressions

Let s be the "signal" produced in a given detector by one particle entering it, including the products of the cascade if any. The probability density $f(s)$ is the combination of the density of ground par-

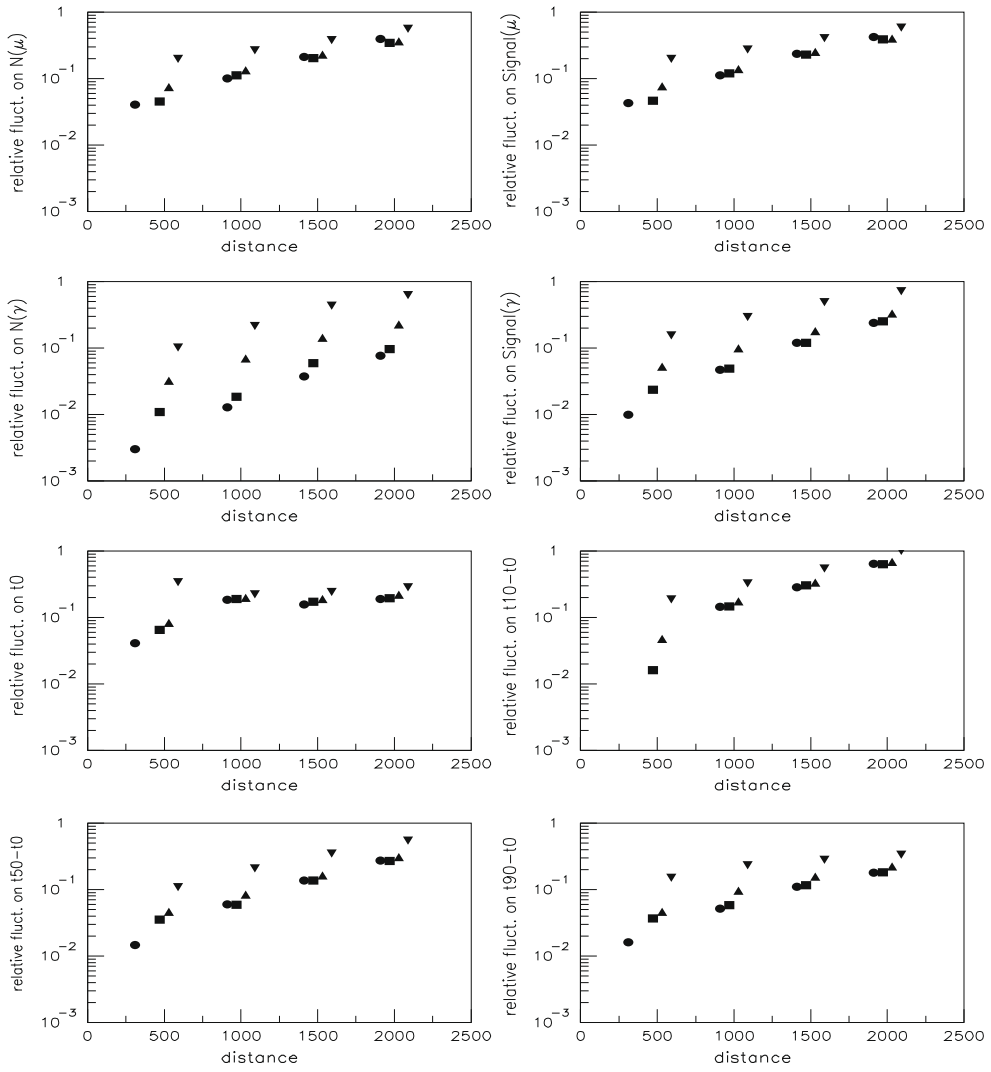

Fig. 13. Fluctuations of the number of particles, the integrated signal and the time structure, as functions of the thinning level (same conventions as in Fig. 11).

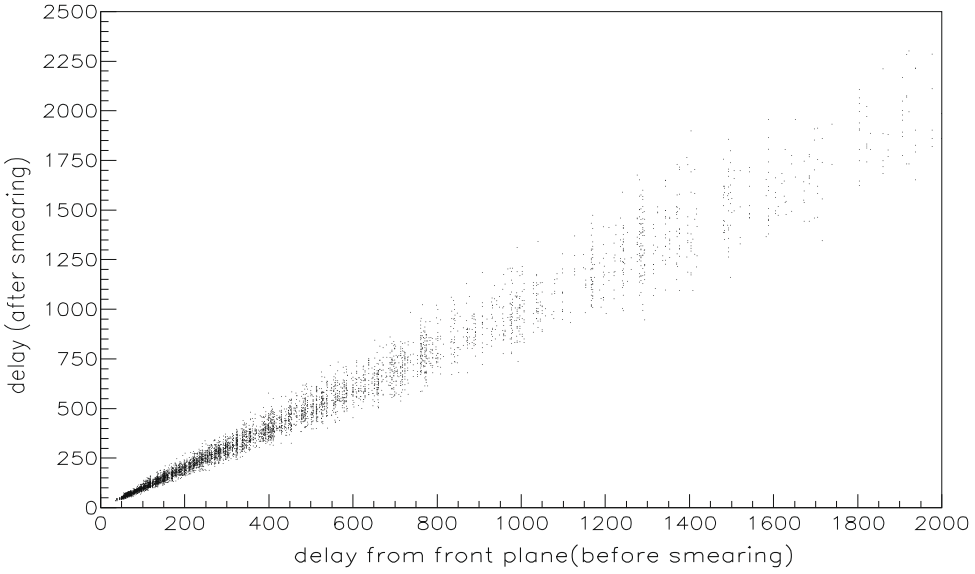

Fig. 14. Smearing applied to the arrival times of clones, with $\sigma = 0.1$.

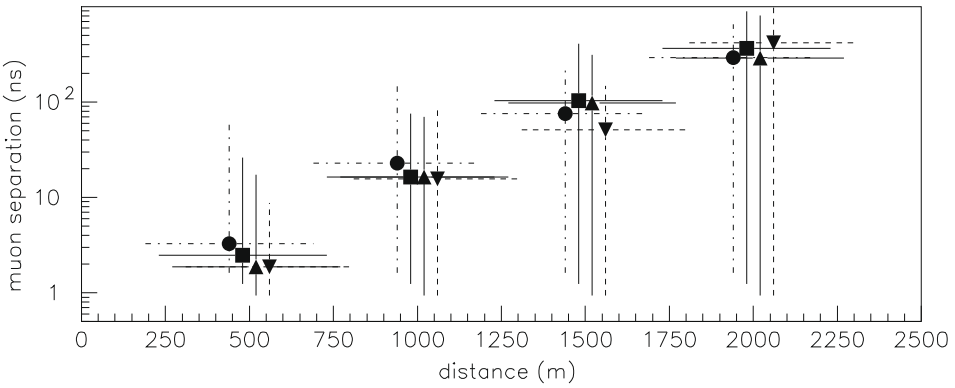

Fig. 15. Time separation of muons at 500, 1000, 1500 and 2000 m from core, in a vertical shower at 10^{19} eV. The different symbols correspond to a thinning at $\epsilon = 10^{-5}$, 10^{-6}, 10^{-7} and 10^{-8} from left to right, and the vertical bars indicate the dispersion.

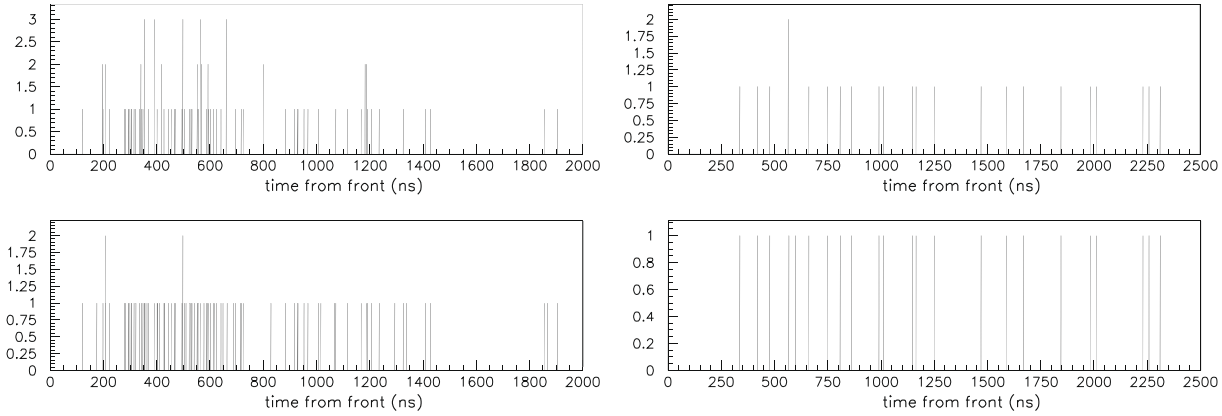

Fig. 16. Repartition of muons before smearing of clones (top) and after (bottom). Left: at 1000 m from core; right: at 1500 m.

ticles in parameter space[5] and the fluctuations of the detector response. The mean signal per particle is:

$$\bar{s} = \int_0^\infty f(s) s \, ds \tag{11}$$

We want to compute the mean value and the variance of the signal S_d in the detector. With k particles entering the detector, we have:

$$\overline{S_{(k)}} = k\bar{s} \tag{12}$$

$$\overline{S_{(k)}^2} = \sum_{i=1}^{k} \overline{s_i^2} + \sum_{i \neq j} \overline{s_i s_j} = k\overline{s^2} + k(k-1)(\bar{s})^2 \tag{13}$$

The number k follows a Poisson law $p_k = \exp(-\overline{N_d}) \cdot (\overline{N_d})^k / k!$ with an expectation value $\overline{N_d}$. Then, summing over all possible values of k, we obtain:

$$\overline{S_d} = \sum_{k=1}^{\infty} p_k k\bar{s} = \overline{N_d}\bar{s} \tag{14}$$

$$\overline{S_d^2} = \sum_{k=1}^{\infty} p_k (k\overline{s^2} + k(k-1)\bar{s}^2) = \overline{N_d}\overline{s^2} + (\overline{N_d}\bar{s})^2 \tag{15}$$

hence the variance:

$$V(S_d) = \overline{S_d^2} - (\overline{S_d})^2 = \overline{N_d}\overline{s^2} \tag{16}$$

For later use, we introduce here a *relative variance* of S_d as:

$$\mathscr{RV}(S_d) = \frac{V(S_d)}{(\overline{S_d})^2} = \frac{\overline{N_d}\overline{s^2}}{(\overline{N_d}\bar{s})^2} = \frac{\int f(s) s^2 \, ds}{\overline{N_d}(\int f(s) s \, ds)^2} \tag{17}$$

We can write the relative fluctuation, with $\sigma(S_d) = \sqrt{V(S_d)}$, as:

$$\frac{\sigma(S_d)}{\overline{S_d}} = \frac{s^*}{\bar{s}\sqrt{\overline{N_d}}} \tag{18}$$

where

$$s^* = \sqrt{\int f(s) s^2 \, ds} \tag{19}$$

is the mean quadratic signal per particle. This formula is a combination of:

- a *poissonian* behaviour: if all particles give the same signal, $s^* = \bar{s}$: we obtain the usual factor $1/\sqrt{\overline{N_d}}$.
- a *spectral factor*: $s^*/\bar{s} = \sqrt{1 + V(s)/(\bar{s})^2}$ depending on the shape of f.

5.2.2. An example of Monte Carlo evaluation

To evaluate the fluctuations in practical conditions, we need to estimate the intrinsic detector response (without resampling). The artificial fluctuations can be suppressed using a low thinning energy and a large sampling area. Here we have taken $\delta = 0.2$, with a weight correction according to formula (7). To increase the area of the sampling region, we take $\alpha = 180°$. The azimuthal asymmetry is ignored and the average values may be slightly biased, but the relative fluctuations are not strongly modified. The results of next section will show *a posteriori* that the artificial fluctuations are low in these conditions. To estimate the quantities $\overline{N}, \bar{s}$ and s^* as functions of the distance from the shower axis, we used an exist-

[5] We can consider separately the particles of different nature (photons, electrons, muons,...) and then add up their contributions; it is equivalent to work directly on the whole set of particles.

ing shower library generated with AIRES, using QGSJET01 [6] as the high energy model, with a thinning level $\epsilon = 10^{-6}$ and a thinning weight factor $W_f = 0.15$ (very strong weight limitation). We have taken five proton showers with $E_{prim} = 10^{19}$ eV, $\theta = 45°$. The same detector simulation was applied as in Section 4. The signal is defined here as the number of photo-electrons produced in one PMT. The results are summarized in Fig. 17, where the contribution of muons is given separately to illustrate the effect of the spectral factor: the difference between s and s^* is mainly due to photons and electrons: their spectral factor is larger and decreases slowly with r; the overall factor (including the muons) has the same behaviour and remains less or equal to 4. Similar results were obtained with vertical showers at the same energy.

5.3. Artificial fluctuations

5.3.1. Integrated amplitude

We want to evaluate the fluctuation on the total signal S_{sr} that would be found in a virtual detector covering exactly the sampling region, by summing the contributions of the weighted particles. The distribution of the ground particles in weight w and signal s is described by a normalized density function $g(s,w)$. If $\overline{N_{sr}}$ is the mean number of *weighted* particles in the sampling region then the mean number of *regenerated* particles in this region is

$$\overline{N_{reg}} = \overline{N_{sr}} \int\int g(s,w) w \, \mathrm{d}s \, \mathrm{d}w \tag{20}$$

and the mean signal deposited by these particles in this region is:

$$\overline{S_{sr}} = \overline{N_{sr}} \int\int g(s,w) ws \, \mathrm{d}s \, \mathrm{d}w \tag{21}$$

To obtain the variance of S_{sr}, the computation done in Section 5.2.1 can be applied again, the signal per *weighted particle* being now ws instead of s. The estimator S_{sr} of the total signal in the sampling region has a variance defined by:

$$V(S_{sr}) = \overline{N_{sr}} \int\int g(s,w)(ws)^2 \, \mathrm{d}s \, \mathrm{d}w \tag{22}$$

and a *relative variance* :

$$\mathscr{RV}(S_{sr}) = \frac{V(S_{sr})}{(\overline{S_{sr}})^2} = \frac{\int\int g(s,w)(ws)^2 \, \mathrm{d}s \mathrm{d}w}{\overline{N_{sr}} \left(\int\int g(s,w) ws \, \mathrm{d}s \, \mathrm{d}w\right)^2} \tag{23}$$

The distribution of the signal per regenerated particle, that is the function $f(s)$ introduced in 5.2.1 at the detector level, may be rewritten, after normalization, as:

$$f(s) = \frac{\int g(s,w) w \, \mathrm{d}w}{\int\int g(s',w) w \, \mathrm{d}s' \, \mathrm{d}w} \tag{24}$$

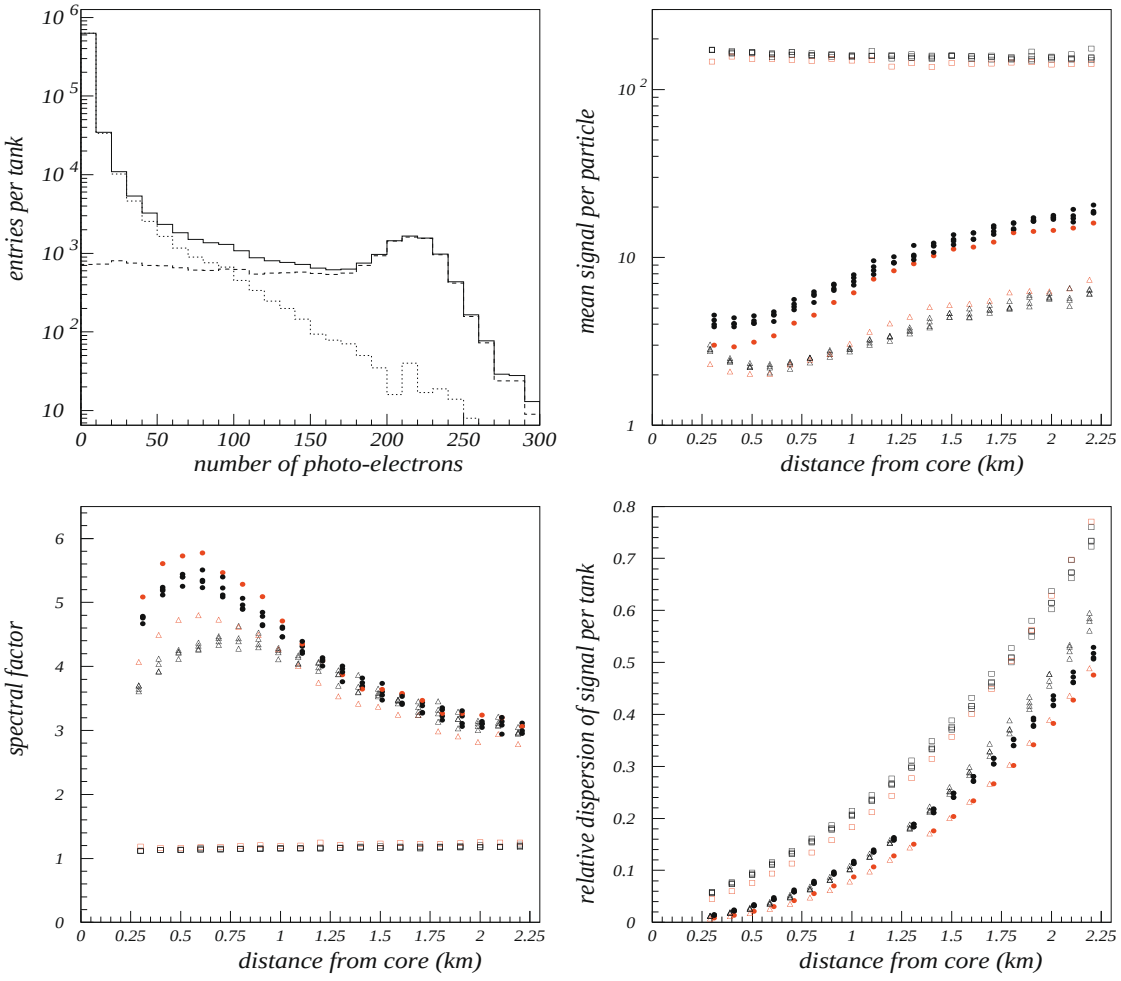

Fig. 17. Fluctuations in a tank, evaluated from 5 proton showers at 10^{19} eV, $\theta = 45°$. Top left: Spectrum of the signal per particle (at 1 km from axis). dashed: muonic; dotted: electromagnetic; Top right: Mean signal per particle. Squares and dashed line: muonic; triangles: electromagnetic; black circles: global; Bottom left: spectral factor (same symbols); Bottom right: Relative fluctuation of the signal in the tank (same symbols). *NB One of the showers had an exceptionally large X_{max}; it is represented by grey symbols.*

Then we can rewrite the relative variance of the signal in the detector Eq. (17) as:

$$\mathscr{RV}(S_d) = \frac{\int s^2\,\mathrm{d}s\left(\int g(s,w)w\,\mathrm{d}w / \int\int g(s',w)w\,\mathrm{d}s'\,\mathrm{d}w\right)}{\overline{N_d}\left(\int s\,\mathrm{d}s \int g(s,w)w\,\mathrm{d}w / \int\int g(s',w)w\,\mathrm{d}s'\,\mathrm{d}w\right)^2} \tag{25}$$

$$= \frac{\int\int g(s,w)ws^2\,\mathrm{d}s\,\mathrm{d}w \int\int g(s,w)w\,\mathrm{d}s\,\mathrm{d}w}{\overline{N_d}\left(\int\int g(s,w)ws\,\mathrm{d}s\,\mathrm{d}w\right)^2} \tag{26}$$

For a sampling ratio $R = \mathscr{A}_d/\mathscr{A}_{sr}$, we have, using Eq. (20): $\overline{N_d} = R\overline{N_{reg}} = R\overline{N_{sr}} \int\int g(s,w)w\,\mathrm{d}s\,\mathrm{d}w$ This gives:

$$\mathscr{RV}(S_d) = \frac{\int\int g(s,w)ws^2\,\mathrm{d}s\,\mathrm{d}w}{R\overline{N_{sr}}\left(\int\int g(s,w)ws\,\mathrm{d}s\,\mathrm{d}w\right)^2} \tag{27}$$

We can now express a condition on the artificial fluctuations: $\mathscr{RV}(S_{sr})$ should be small compared to the relative variance of the signal in the detector; combining Eq. (23)–(25), (and) (27), we obtain:

$$\mathscr{RV}(S_{sr}) \ll \mathscr{RV}(S_d) \Longleftrightarrow R\cdot\mathscr{F} \ll 1 \tag{28}$$

where we have introduced a *weight spectral factor*

$$\mathscr{F} = \frac{\int\int g(s,w)(ws)^2\,\mathrm{d}s\,\mathrm{d}w}{\int\int g(s,w)ws^2\,\mathrm{d}s\,\mathrm{d}w} \tag{29}$$

The factor $\mathscr{F}$ may be seen as the mean value (over the *regenerated* particles) of the weight of the parents, weighted by the square of the signal.[6] If all weights are concentrated around w_0, then $\mathscr{F} \simeq w_0$, and the condition (28) simply means that the number of *weighted particles* in the sampling region is much larger than the number of *regenerated particles* in the detector, which is the same requirement ($w_r \ll 1$) as expressed in Section 2.4. If all particles give nearly the same signal, $\mathscr{F} \simeq (w^*)^2/\bar{w}$, where w^* is the quadratic mean weight, at least equal to the mean weight $\bar{w}$, and possibly much larger if the distribution of w has a large relative dispersion, especially if it has a long tail at high values. In the general case, the interpretation of $\mathscr{F}$ is more complex. As expected, a few abnormally large weights (especially on particles giving a large signal) may waste the statistical quality of the regenerated signal.

It is clear that the weight limitation is a good way to avoid large artificial fluctuations. A limitation applied directly to the product ws would be even better, but it is more difficult to implement in the shower simulation, because the signal produced by a ground particle cannot be evaluated in advance, when setting the probability to keep its ancestors in the cascade. In practice, the thinning algorithms allow to put a more severe weight limitation on particles which may have descendants giving a large signal in the ground detectors (e.g. mesons, which will produce muons, in the case of Cherenkov tanks). In that respect, the shower simulation may be tuned to a given type of detector.

As in Section 4, we can estimate the statistical quality of the sampling on a given simulated ground particle file, by replacing integrals by weighted sums over the particles. Fig. 18 shows the results on the same showers as in Section 4: the weight spectral factor is of the order of 1000. Similar results were obtained using CORSIKA [4] as shower simulator. In these conditions, to suppress the artificial fluctuations well below the natural ones requires a sampling ratio less than 10^{-4} (0.1 km^2 per detector in the case of the Auger Surface Detector). If we adopt $\delta = 0.1$ and $\alpha = 15^\circ$ for a good safety against the biases, this condition is satisfied for $r \geqslant 1000$ m, but not at short distances. As discussed in Section 2.4, this may be acceptable for the global reconstruction of a shower. For fixed values of ϵ and W_f, the weights and $\mathscr{F}$ are approximately proportional to E_{prim}: at larger energies lower values of ϵ (or W_f) are requested, in principle, to obtain the same value of $R\mathscr{F}$ as a function of r. However, relatively large artificial fluctuations may be acceptable in one or a few detectors within a given experiment; the other sources of errors should be studied to evaluate the impact of the artificial fluctuations in the final analyses.

Within a resampling program, it is possible to include a statistical quality check, for each detector, by just computing sums over the *regenerated* particles: a good estimation of $\mathscr{F}$ is given by $\sum w_i s_i^2 / \sum s_i^2$, where w_i is the initial weight of the *parent* particle.

5.3.2. Time structure

There are different ways to describe the time dependence at large scale. To evaluate fluctuations, the simplest point of view is to consider the ratio of partial sums, over different time intervals. The fluctuation on a partial sum follows exactly the same rules as the fluctuation on the full integrated signal, and the previous expressions may be used without change. Moreover, if the intervals do not overlap, the fluctuations are independent. If the relative fluctuations are small, usual approximations hold, for example:

$$\left(\frac{\sigma_{S_1/S_2}}{S_1/S_2}\right)^2 = \left(\frac{\sigma_{S_1}}{S_1}\right)^2 + \left(\frac{\sigma_{S_2}}{S_2}\right)^2 \tag{30}$$

Most of the shape estimators are in some sense functions of partial sums; we suppose here that the fractional times, which are defined in another way, have a similar behaviour with respect to the fluctuations. If the ratio of artificial fluctuations to natural ones is acceptable in all time intervals, it should be also acceptable for a shape estimator. We just need to evaluate the weight spectral factor as a function of the time. It is not constant, because the composition, measured through the muons/electromagnetic ratio, and the energy spectrum are variable; however it is not expected to vary by a large factor from the beginning to the end of the signal; with the Auger Surface Detector, Fig. 19 shows that at medium distances (1–2 km from the core) the spectral factor remains below 4 throughout the signal, and the weight spectral factor does not exceed a few hundreds.

Finally we can conclude that the impact of artificial fluctuations on the time shape is practically the same as on the full integrated signal.

5.4. The problem of nearly horizontal particles

As mentioned in Section 2, the effective area of a detector with non-horizontal parts includes a term proportional to $\tan(\theta_p)$, and weighted integrals over the direction diverge if the density does not vanish at $\theta_p = \pi/2$: as a consequence, the fluctuations are not guaranteed to be kept under control. In practice, in vertical showers, there a few nearly horizontal particles at ground level, but they are essentially low energy photons or electrons, giving very small signals. Very inclined showers contain mainly well collimated muons: in that case the standard sampling region, for given values of δ and α, is scaled by a factor $1/\cos(\theta_s)$ which compensates approximately the factor $\tan(\theta_p)$.

However, the problem remains in theory for divergent particles, for example in a subcascade generated by the bremsstrahlung or the decay of a energetic muon. It is inherent in the shower simulation package, because a quasi horizontal flux is undersampled on the horizontal plane defining the ground, and the statistical degradation is irreversible. The only solution would be, *within the shower simulation package*, to sample particles also in some vertical surfaces when defining the "ground" particle file. The design of such surfaces is delicate, because many pieces are needed to cover the effective extension of the shower at ground (accounting for the detector efficiency).

[6] It is interesting to note that if the weight is anticorrelated to the energy, as is generally the case with weight limitation methods, then $\mathscr{F}$ is *smaller* than a "simple" mean, weighted by s.

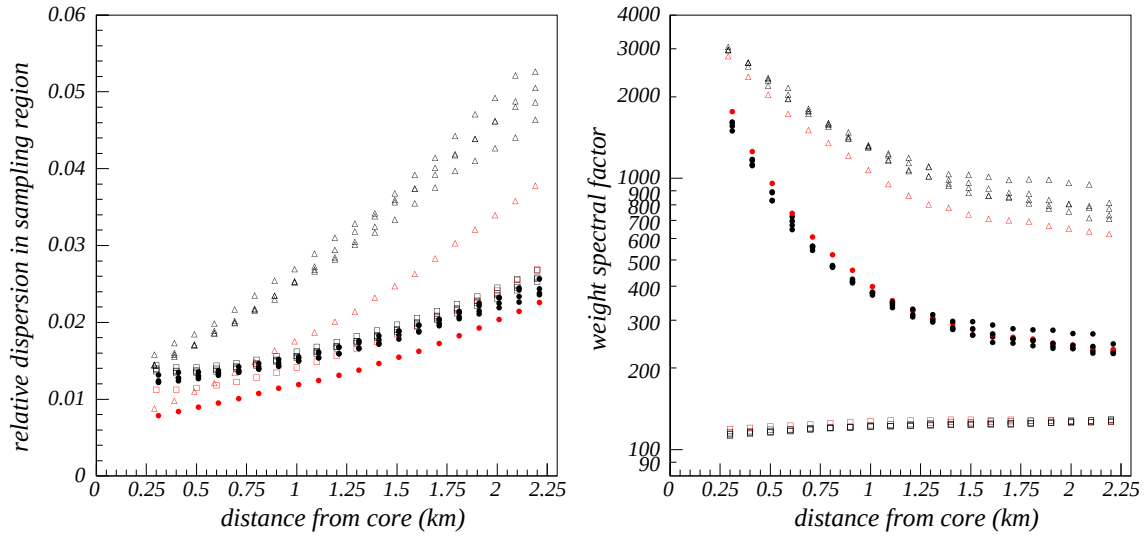

Fig. 18. Artificial fluctuations in a detector (same showers as in Fig. 17): Left: relative fluctuation of the total signal in a sampling region defined by $|r - r_d|/r_d < 0.2$. Squares: muonic; triangles: electromagnetic; black circles: global; Right: weight spectral factor (same conventions).

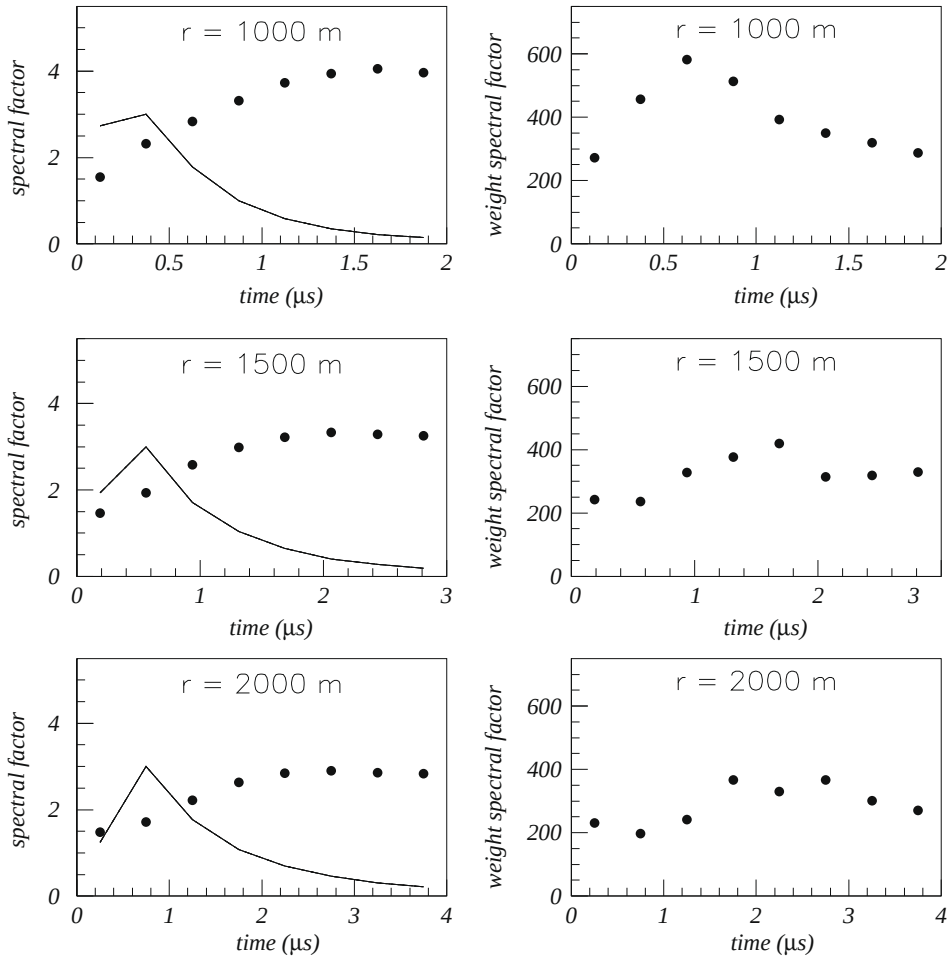

Fig. 19. Artificial fluctuations in a detector as a function of the time (delay after the shower plane), averaged from 5 proton showers (the same as in Fig. 17), at different distances from the core: Left: spectral factor $s^*/\bar{s}$ in the detector; the solid line gives the profile of light emission; Right: weight spectral factor $\mathscr{F}$ (see text).

6. Summary and conclusions

The packages used to simulate giant air showers generated by cosmic rays of ultra high energy include a thinning algorithm and produce sets of weighted particles. A resampling or unthinning procedure has been designed to obtain a fair description of the particles entering a ground detector without using tabulations or parametrizations. The principle is to define a sampling region

centered on this detector, to translate to the detector area all particles within this region, rescaling their weight to reproduce the local flux, with a correct distribution in the parameter space (energy, direction, time). The sampling region needs to be small enough to ensure that the local average does not induce a bias of the flux or a distortion of physically relevant distributions. But it should be large enough to avoid large artificial fluctuations, that is, it should contain enough particles to make the Poisson fluctuation on their number negligible, and to avoid spoiling the statistical quality by large weights.

To reduce the biases, some corrections may be applied, either to the weights of the input particles, or to their parameters (e.g. the time of arrival), within some general assumptions that do not suppress the shower-to-shower fluctuations. However, it is recommended to maintain the sampling region within 10% of the distance of the detector to the shower axis, and within 15° in azimuthal direction around this axis.

The time structure of the signal – its overall shape and significant peaks – may be difficult to restore if some particles with large weights have to be cloned. A time smearing procedure has been proposed to solve the problem in a satisfactory way in a realistic detector with a finite time resolution.

To give a quantitative estimation of the impact of artificial fluctuations on the statistical errors on measured quantities, a mathematical criterion has been defined, based on the *weight spectral factor*. As expected, the statistical quality is better if the distribution of the weights has no tail at high values, especially for particles giving a large contribution to the signal. This criterion was applied to the simulation of the Auger Surface Detector with a standard shower library produced by AIRES. The artificial fluctuations were found to be small compared to the natural ones, except at short distance from the core, and/or at highest primary energies. In such a case, the signal in a detector is very large, so the natural fluctuations are small in relative value, while the limitation of the sampling region does not allow to reduce the artificial ones at a comparable level. However, this does not imply that the resampling method cannot be used, or that very low weights (then huge input files) are needed: a shower is reconstructed with several detectors at various distances, and artificial fluctuations on one of them may have a negligible effect on the error on the global parameters. Moreover, the systematic errors may dominate the reconstruction, especially at highest energies, so that the statistical errors do not need to be reproduced exactly, except for dedicated studies on limited samples of simulated showers.

The problem of the divergence of the resampled weights for quasi-horizontal particles seems to be practically avoided in real inclined showers, which contain essentially well collimated particles. But this potential flaw should be kept in mind; a possible solution at the level of the shower simulation is proposed.

It is clear from this study that it is impossible to define in an absolute way "good thinning parameters" or a "good resampling method". The thinning parameters (especially the weight limitation) need to be adapted to the detector. The definition of the resampling region and the possible bias corrections should be evaluated in realistic conditions against the various sources of errors. Some exotic configurations (for example a deep horizontal shower induced by a neutrino) may need a specific procedure going beyond the standard picture of a radial distribution around an axis.

Acknowledgements

This work was initially motivated by the development of the Auger Observatory Project, and stimulated by the rich scientific context of the cosmic ray community. I thank my colleagues of the Auger Collaboration and the developers of the shower simulation packages AIRES and CORSIKA for many helpful discussions, and also for criticisms of first versions of the method, leading to technical improvements, especially in the bias correction procedures. The detector simulation has inherited some algorithms from the codes written in the 90's by C.L. Pryke [7]. I am grateful to the referees for pointing out imperfections of the first version, technical mistakes and inconsistencies between notations, and to Jim Matthews for carefully reading the revised version.

References

[1] A.M. Hillas, Nucl. Phys. B (Proc. Suppl.) 52B (1997) 29.
[2] M. Kobal, The Pierre Auger Collab., Astropart. Phys. 15 (2001) 259.
[3] S.J. Sciutto, For the thinning method, in: Proc. 27th Int. Conf. on Cosmic Rays, Hamburg 2001, vol. 1, 1999, p. 237. Available from: <astro-ph/9911331>.
[4] D. Heck et al., Report FZKA 6019, 1998. <http://www-ik.fzk.de/corsika/>.
[5] J. Abraham et al., Pierre Auger Collab., Nucl. Instrum. Methods A 523 (2004) 50.
[6] N.N. Kalmykov, S.S. Ostapchenko, A.I. Pavlov, Nucl. Phys. B (Proc. Suppl.) 52B (1997) 17.
[7] C.L. Pryke, PhD Thesis, 25th Int. Conf. Cosmic Rays, Durban (1997), 5, 209.