Statistics of clustering of ultra-high energy cosmic rays and the number of their sources

S. L. Dubovsky[a], P. G. Tinyakov[a,b] and I. I. Tkachev[c]
[a] *Institute for Nuclear Research, 60th October Anniversary prospect 7a, 117312 Moscow, Russia*
[b] *Institute of Theoretical Physics, University of Lausanne, CH-1015 Lausanne, Switzerland*
[c] *ETH, Hönggerberg, CH-8093, Zurich, Switzerland*

Observation of clustering of ultra-high energy cosmic rays (UHECR) suggests that they are emitted by compact sources. Assuming small ($< 3^\circ$) deflection of UHECR during the propagation, the statistical analysis of clustering allows to estimate the total number of the sources S, including those which have not yet been observed directly. When applied to astrophysical models involving extra-galactic sources, the estimate gives $S \sim 400$ inside the sphere of the radius ~ 50 Mpc. This is too large for models which associate the production of UHECR with exceptional galaxies such as AGN, powerful radio-galaxies, dead quasars, and for models based on gamma ray bursts.

PACS numbers:

I. INTRODUCTION

Recent analysis of arrival directions of ultra-high energy cosmic rays (UHECR) reveals groups of events (clusters) with arrival directions lying within $\sim 3^\circ$, the typical angular resolution of the experiment. The set of 92 observed events with energy $E > 4 \times 10^{19}$ eV contains 7 doublets and 2 triplets [1]. The small probability of chance coincidence, of the order of 10^{-3} [2,1], suggests that clustering is a result of the existence of compact sources. At higher energies, $E > 10^{20}$ eV, one doublet out of 14 events is observed.

Compact sources of UHECR are naturally explained in astrophysical models where they are associated with possible UHECR production sites, such as AGN [3], hot spots of powerful radio-galaxies [4], dead quasars [5] and gamma-ray bursts (GRB) [6]. These models have much in common. They assume that primary particles are protons; the sources of the observed UHECR have, therefore, to lie within the GZK cutoff [7] sphere. For energies $E > 10^{20}$ eV the GZK radius is $R_{\rm GZK} \sim 50$ Mpc [8]. In all these models the distribution of sources in space within the GZK sphere is uniform, while the distribution in intensity does not depend on space and peaks around a certain value.

An important common feature of these models is a small number of sources within the GZK sphere. The number of dead quasars is estimated as $\lesssim 40$ [5]; the number of AGN is $\sim 10\%$ of the total number of galaxies; the latter is ~ 2500 [9]. Most likely, only a small fraction of them is capable of producing UHECR with energies $E > 10^{20}$ eV. In the case of GRB the relevant number of sources is determined by the rate γ of GRB and typical time delay τ of UHECR particles. Taking $\tau \lesssim 10^5$ yr and the rate $\gamma \sim 2 \times 10^{-10} h^3$ Mpc^{-3} yr^{-1} [10] gives the number of sources $\lesssim 10$.

The purpose of this letter is to show that the observed clustering favors larger number of sources, provided the propagation of UHECR with energy $E > 10^{20}$ eV is not strongly affected by extra-galactic magnetic fields. The latter assumption is justified if the existing bound on extra-galactic magnetic field $B \lesssim 10^{-9}$ G [11] is valid.

II. STATISTICS OF CLUSTERING

The observable quantities which characterize clustering are $\bar N_m$, the expected numbers of clusters of different multiplicities m (e.g., $\bar N_1$ and $\bar N_2$ are the expected numbers of single and double events, respectively). They depend on the total exposure of the experiment B and the distribution of sources in the apparent luminosity[1] $n(L)$, which is defined in such a way that the number of sources with luminosity from L to $L + dL$ is $dS = n(L)dL$. Since the events which come from the same source at different times are statistically independent, they have the Poisson distribution. Therefore, the expected number of clusters is

$$\bar N_m = \int_0^\infty \frac{(LB)^m}{m!} {\rm e}^{-LB} n(L) dL. \qquad (2.1)$$

This equation implies that the expected total number of events $N_{\rm tot}$ is

$$\bar N_{\rm tot} = \sum_m m \bar N_m = B \int_0^\infty L n(L) dL = B L_{\rm tot}, \qquad (2.2)$$

as it should be. The probability to observe k clusters of multiplicity m is given by the Poisson distribution,

[1] Here and below we mean the luminosity in cosmic rays with energies above some energy threshold. It measures the number of events per unit time per unit area of the detector.

$$P_m(k) = \frac{(\bar{N}_m)^k}{k!} \mathrm{e}^{-\bar{N}_m}. \tag{2.3}$$

Any model of UHECR can be characterized by the distribution of sources in space and intensity $f(r, j)$ (which we assume to be spherically symmetric for simplicity). In order to express $n(L)$ and $\bar{N}_m$ in terms of the distribution function $f(r, j)$, consider the sources at distances from r to $r + dr$. The number of such sources with intensities from j to $j + dj$ is

$$dS = f(r, j)\, 4\pi r^2 dr\, dj. \tag{2.4}$$

Making use of the relation $L = j/4\pi r^2$ and integrating over r one finds $dS = n(L)dL$, where

$$n(L) = (4\pi)^2 \int_0^\infty dr\, r^4 f(r, 4\pi r^2 L). \tag{2.5}$$

Here we have neglected the curvature effects since they are small at distances of order 50 Mpc.

In the case of the astrophysical models, the distribution function $f(r, j)$ is uniform in space and depends only on the intensity, $f(r, j) = h(j)$. The GZK effect, however, makes distant sources fainter by a factor $\exp(-r/R)$. This is taken into account by setting

$$f(r, j) = \mathrm{e}^{r/R} h(j\mathrm{e}^{r/R}). \tag{2.6}$$

The exact value of R can be determined by numerical simulations of UHECR propagation with full account of the energy dependence. We take a conservative estimate $R \simeq 50$ Mpc.

III. NUMBER OF SOURCES

A key parameter which enters the distribution $f(r, j)$ is the normalization or, equivalently, the total number of sources S (the number of sources within the GZK sphere in the case of astrophysical models). When comparing particular models to the experimental data, this is one of the main parameters to be determined. An important information about S can be obtained from statistical analysis of clustering even if the functional form of the distribution $f(r, j)$ is not known. The idea is to find the distribution $f(r, j)$ which corresponds to the *minimum* number of sources S_* with total number of events and the number of events in clusters, $\bar{N}_{\rm cl} \equiv \bar{N}_{\rm tot} - \bar{N}_1$, being fixed. We will show in a moment that in the case $\bar{N}_{\rm cl} \ll \bar{N}_{\rm tot}$ the number S_* is surprisingly large, much larger than the number of the sources already observed.

It is intuitively clear why in the case $\bar{N}_{\rm cl} \ll \bar{N}_{\rm tot}$ the number of sources is much larger than $\bar{N}_{\rm tot}$. In order to produce $\sim N_{\rm tot}$ single events by $\sim N_{\rm tot}$ sources each of them has to be bright enough. But then a large number of doublets would be produced as well. Since this is not the case, i.e. most of the resolved sources are dim and produce at most one event, one concludes that there is a large number of sources which have not yet revealed themselves. Assuming that all sources have the same luminosity L one finds from Eq. (2.1) $\bar{N}_1 \sim S\bar{n}$ and $\bar{N}_2 \sim S\bar{n}^2$, where $\bar{n} = LB$ is the average number of events produced by one source. Therefore, $S \sim \bar{N}_1^2/\bar{N}_2 \sim \bar{N}_{\rm tot}^2/\bar{N}_{\rm cl}$, i.e. much larger than $\bar{N}_{\rm tot}$. Using methods described in the Appendix it is possible to show that the case of equal luminosities corresponds to the absolute minimum of S. However, this distribution is unphysical. Many realistic situations correspond to a homogeneous distribution of sources in space when more distant sources are fainter; consequently, their number has to be even larger than predicted by the above estimate.

In astrophysical models the distribution $f(r, j)$ is given by Eq. (2.6) containing one unknown function $h(j)$. The minimum density of sources is determined by minimizing over $h(j)$. As is shown in the Appendix, the minimum is reached at the delta-function distribution

$$h(j) = h_* \delta(j - j_*), \tag{3.1}$$

where j_* is the intensity of the sources and h_* is their spatial density. The unknown parameters h_* and j_* can be related to $\bar{N}_{\rm tot}$ and $\bar{N}_{\rm cl}$ by making use of eqs.(2.1) and (2.2). Introducing the notations

$$S_* = (4\pi/3)R^3 h_* \tag{3.2}$$

for the number of sources inside the sphere of the radius R and $\nu_* = Bj_*/(4\pi R^2)$ for the number of events from one source at the distance R, one has the following equations,

$$\bar{N}_{\rm tot} = 3S_*\nu_* \, , \tag{3.3}$$

$$\bar{N}_1 = 3S_*\nu_* \int_0^\infty dx \exp\left(-x - \nu_* x^{-2} \mathrm{e}^{-x}\right). \tag{3.4}$$

These equations can be solved perturbatively at small $\bar{N}_{\rm cl} \ll \bar{N}_{\rm tot}$. One finds

$$\nu_* \simeq \frac{1}{\pi} \frac{\bar{N}_{\rm cl}^2}{\bar{N}_{\rm tot}^2}, \tag{3.5}$$

$$S_* \simeq \frac{\pi}{3} \frac{\bar{N}_{\rm tot}^3}{\bar{N}_{\rm cl}^2}. \tag{3.6}$$

If $\bar{N}_{\rm cl} \ll \bar{N}_{\rm tot}$, the minimum number of sources S_* is indeed much larger than $\bar{N}_{\rm tot}$ and, therefore, is much larger than the number of sources already observed. From Eq. (3.5), each source produces much less than 1 event in average.

IV. DISCUSSION

Let us apply these arguments to the observed events with energies $E > 10^{20}$ eV. In this case, $N_{\rm tot} = 14$ and $N_{\rm cl} = 2$. The solution to eqs.(3.3) and (3.4) is

$$S_* \sim 400. \qquad (4.1)$$

This number is large as compared to the number of sources expected inside the GZK sphere in most of astrophysical models. However, it should be interpreted with care. One may expect large statistical fluctuations because both $N_{\rm tot}$ and $N_{\rm cl}$ are small.

In order to address this issue quantitatively, let us find the model which has the largest probability $p(S)$ to reproduce the observed clustering at fixed number of sources S. To this end, consider the set of models which are described by eqs.(2.6) and (3.1) and are characterized by two parameters h_* and j_*. Fixing S is equivalent to fixing h_*. There remains a freedom of changing j_*. The probability to reproduce the observed data is maximum for some j_*; this probability is $p(S)$. By construction, there are no models with S or smaller number of sources in which the probability to reproduce the observed data is larger than $p(S)$.

probability	14 events 1 doublet		30 events 1 doublet		60 events 1 doublet	
p	S	ν	S	ν	S	ν
0.1	15	0.51	210	0.065	2000	0.012
0.01	2.1	5.7	41	0.38	450	0.058
0.001			14	1.3	170	0.15

TABLE I. Minimum number of sources S inside the sphere of radius $R \sim 50$ Mpc and corresponding source intensity in the units of $\nu = Bj_*/(4\pi R^2)$, which are required to reproduce the observed clustering with given probability p for the real experimental data (1 doublet out of 14 events) and for two hypothetical data sets with larger number of events (one doublet out of 30 events and one doublet out of 60 events).

There is some ambiguity in defining what is "to reproduce the observed data". In the case at hand we request that the number of singlets is 12 or larger, the number of doublets is 1 or smaller, and the number of clusters with the multiplicity 3 and larger is zero. Eq. (2.3) determines the probability $p(h_*, j_*)$ of such clustering as a function of two parameters h_* and j_*. The probability $p(S)$ is found by maximizing $p(h_*, j_*)$ at fixed S. We have performed this calculation numerically. The results are summarized in Table 1 in the form of lower bounds on the number of sources inside the sphere of the radius R. We also present the source intensity in the units of $\nu = Bj_*/(4\pi R^2)$, i.e. the number of events from a single source at the distance R.

The model where observed clustering occurs with the probability 1% has ~ 2 sources inside the 50 Mpc sphere. In this model, most of the observed 14 events are produced by the sources which are further than 50 Mpc and thus have to be bright enough. This is reflected in Table 1 from which we see that these sources would produce in average ~ 6 events each if placed at 50 Mpc in the absence of the GZK cutoff.

It is worth noting that the numbers of Table 1 correspond to the extreme situation when the distribution of sources is given by Eq. (3.1) with a particular value of j_*. In realistic models, the distribution of sources in intensity is usually spread over an order of magnitude at least. There may also be constraints on intensity of the sources. In these cases, the bounds on the number of sources are stronger than in Table 1.

When the new large-area detectors like the Pierre Auger array [12] will start operating, the number of observed events will increase and the statistical errors in determination of the number of sources will go down. At the same time, the number of directly observed sources will grow. One may ask if the bounds on S may become statistically significant before astrophysical models are excluded by direct counting of sources. To show that the answer is positive, we have performed calculations for two hypothetical situations, 1 doublet out of 30 events, and 1 doublet out of 60 events. The results are also listed in Table 1. The bounds obtained from statistics of clustering grow much faster than the number of events.

To summarize, the statistical analysis of clustering may provide tight constraints on astrophysical models of UHECR when the number of clusters is small. In this situation, a key quantity is the density of sources which can be bound from below in a model-independent way. The bound grows very fast with the number of single events above $E = 10^{20}$ eV and is potentially dangerous for astrophysical models which associate production of UHECR with GRB or exceptional galaxies such as AGN, powerful radio-galaxies and dead quasars.

Our method equally applies to models in which UHECR are produced in the Galactic halo, or in which primary particles are immune to the background radiation. In these cases there is no damping exponent in Eq. (2.6). The relation (3.6) remains valid with a different numerical coefficient of order one. Detailed analysis of the first case shows that statistical properties of clustering of UHECR are compatible with clumpiness of super-heavy dark matter in decays of which UHECR are produced. It is interesting to note that in the second case our method counts the number of UHECR sources within the cosmological horizon, which is unaccessible by other means.

ACKNOWLEDGMENTS

The authors are grateful to K.A. Postnov for the references concerning the number of galaxies and to V.A. Rubakov and M.E. Shaposhnikov for valuable comments and discussions.

APPENDIX: MINIMUM NUMBER OF SOURCES

Consider the problem in general terms. First note that by changing the integration variable in Eq. (2.5) one can show that any distribution is equivalent to a factorazable one. So, let us take the distribution of sources in the form

$$f(r,j) = g(r)h(j). \tag{4.2}$$

Let us fix $g(r)$ and minimize the number of sources

$$S = 4\pi \int_0^\infty r^2 g(r) dr \int_0^\infty h(j) dj$$

with respect to the distribution $h(j)$ under the constraints fixing $\bar{N}_{\rm tot}$ and $\bar{N}_1$,

$$B \int_0^\infty g(r) dr \int_0^\infty j h(j) dj = N_{\rm tot}, \tag{4.3}$$

$$B \int_0^\infty dr dj \; j h(j) g(r) \exp\left(-\frac{Bj}{4\pi r^2}\right) = N_1. \tag{4.4}$$

This is equivalent to minimizing the functional

$$W = 4\pi \int_0^\infty dj h(j) \int_0^\infty g(r) dr \left\{ r^2 + \lambda \frac{Bj}{4\pi} \right.$$

$$\left. -\mu \frac{Bj}{4\pi} \exp\left(-\frac{Bj}{4\pi r^2}\right) \right\} - \lambda N_{\rm tot} + \mu N_1 \tag{4.5}$$

with respect to $h(j)$. Here λ and μ are the Lagrange multipliers.

The functional (4.5) is linear in $h(j)$; denote the coefficient by $G(j)$,

$$G(j) = \int_0^\infty g(r) dr \left\{ r^2 + \lambda \frac{Bj}{4\pi} - \mu \frac{Bj}{4\pi} \exp\left(-\frac{Bj}{4\pi r^2}\right) \right\}.$$

At those values of j where $G(j)$ is negative, the minimum of W is at $h(j) \to \infty$. The latter, however, is not compatible with eqs.(4.3) and (4.4). Therefore, at the minimum the values of λ and μ have to be such that $G(j)$ is non-negative.

At those values of j where $G(j)$ is positive, the minimum of W is reached at $h(j) = 0$. If $G(j)$ is positive at all j, then $h(j)$ is identically zero and eqs.(4.3) and (4.4) are again violated. Therefore, λ and μ must be such that $G(j)$ touches zero at some j_*. The function $h(j)$ is non-zero only at this point. Thus, the minimum number of sources corresponds to the situation when all of them have the same intensity j_*, and we arrive at the delta-function distribution, eq.(3.1).

It remains to show that for a given positive function $g(r)$ satisfying $\int g(r) dr < \infty$ the Lagrange multipliers λ and μ can always be chosen in such a way that $G(j)$ is positive everywhere except an isolated point. To this end rewrite $G(j)$ in the following form,

$$G(j) = C + \lambda F(j), \tag{4.6}$$

where $C = \int r^2 g(r) dr$ is a positive constant and the function $F(j)$ depends only on the ratio μ/λ,

$$F(j) = \frac{Bj}{4\pi} \int_0^\infty g(r) dr \left\{ 1 - \frac{\mu}{\lambda} \exp\left(-\frac{Bj}{4\pi r^2}\right) \right\}.$$

The behavior of the function $F(j)$ is the following. At $j \to 0$ it goes to zero. At small j it is negative if $\mu/\lambda > 1$ and positive otherwise. At $j \to \infty$ it grows linearly with j, the coefficient being $B/4\pi \int g(r) dr > 0$. Therefore, at $\mu/\lambda > 1$ the function $F(j)$ must have an absolute minimum at some $j_* > 0$ (which is a function of μ/λ). Then it is clear from eq.(4.6) that by choosing $\lambda = -C/F(j_*) > 0$ one can set $G(j)$ to zero in that particular point. The argument can be easily generalized to the case of infinite number of sources, $\int g(r) r^2 dr = \infty$.

In order to apply this argument to the case of astrophysical models, one should find the factorizable distribution $\tilde{f}(r,j)$ which produces the same $n(L)$ as eq.(2.6). This can be done by substituting eq.(2.6) into eq.(2.5) and changing the integration variable according to

$$r^2 \exp(r/R) = x^2. \tag{4.7}$$

The result reads

$$\tilde{f}(x,j) = g(x) h(j),$$

where

$$g(x) = (1 + r(x)/2R)^{-1} {\rm e}^{-3r(x)/2R}$$

and $r(x)$ is defined by eq.(4.7).

[1] Y. Uchihori et al., *Cluster analysis of extremely high energy cosmic rays in the northern sky*, astro-ph/9908193

[2] M. Takeda *et al.*, Phys. Rev. Lett. **81** (1998) 1163, astro-ph/9807193.

[3] K. Mannheim, Astropart. Phys. **3**, 295 (1995)

[4] J.P. Rachen, P.L. Biermann, Astron.Astrophys. **272** (1993) 161.

[5] E. Boldt and P. Ghosh, MNRAS **307** (1999) 491 (astro-ph/9902342).

[6] E. Waxman, Phys.Rev.Lett. **75** (1995) 386-389

[7] K. Greisen, Phys. Rev. Lett. **16** (1966) 748; G.T. Zatsepin and V.A. Kuzmin, Pisma Zh. Eksp. Teor. Fiz. **4** (1966) 144;

[8] S. Lee, Phys. Rev. **D58** (1998) 043004;

[9] J. Loveday *et.al.*, Astrophys. J. **390** (1992) 338.

[10] S. Mao and H.J. Mo, A&A **339** (1998) L1.

[11] P.P. Kronberg, Rep. Prog. Phys., **57** (1994) 325.

[12] J.W. Cronin, Nucl. Phys. B (Proc. Suppl.) **28B** (1992) 213.