On the Global Anisotropy of Cosmic Ray Data above $4 \times 10^{19} eV$

Soebur Razzaque and John P. Ralston

Department of Physics & Astronomy
University of Kansas, Lawrence, KS 66045, USA

Abstract

The distribution of arrival directions of ultra-high energy cosmic rays may yield clues to their mysterious origin. We introduce a method of *invariant statistics* to analyze cosmic ray data which eliminates coordinate-dependent artifacts. When combined with maximum likelihood analysis, the method is capable of quantifying deviations of the distribution from isotropy with high reliability. We test our method against published AGASA events with energies above $4 \times 10^{19} eV$. Angular cuts from observational limitations are taken into account. A model based on the Fisher distribution reveals the rotation of the Earth with the axis $\hat{n}$ along the direction $(5^h\ 53.36^m,\ 85.78°)$ in (RA, DEC) coordinates, which is within 5° of the equatorial north pole. Global anisotropy of the data, if any, hinges on finer understanding of detector acceptance than what is available from the published literature.

Introduction: A puzzle has existed for more than 30 years regarding cosmic ray events with energies exceeding $4 \times 10^{19} eV$, a value in the range of the GZK bound [1, 2]. The nature of the primary particles causing these events is controversial. If the primaries are protons from cosmological distances, then their sources should be isotropically distributed. The scrambling effects of intervening magnetic fields are difficult to assess, but on very basic physics the magnitudes of the fields (or associated correlation lengths) would have to be exceedingly small to avoid isotropization. Cosmic rays from our own Galaxy core are not considered a viable explanation for the events approaching the GZK limit. If our Galaxy substantially modifies propagation or contributes to production of the primaries, then any consequent anisotropy should be correlated with known features of the Galaxy. On several bases, and in particular on the basis of *symmetry*, one asks whether the highest energy events are *isotropic*, up to the cuts imposed by observational limitations.

Several studies [3, 4, 5] at much lower energies have demonstrated a significant isotropy in arrival directions. The AGASA group has presented an analysis of a sizable (4%) anisotropy of several hundred thousand events with energies above $10^{17} eV$. The study uses right ascension, declination (RA, DEC) coordinates [4] and finds a statistically significant first Fourier moment in $\sin(RA)$. The definition of right ascension however, involves historical human conventions and a choice of axis based on the Earth's orbit. The coordinate system axis defining RA requires two parameters for specification, which are implicitly used in the analysis, potentially affecting the claimed statistical significance of effects. In addition, $\sin\phi$ is not a particular sensitive or informative statistic, potentially diminishing the impact of 216,000 events.

Here we examine all available data from the AGASA experiment with energy above $4 \times 10^{19} eV$. We take care to define certain quantities of the analysis which previously have been used in a coordinate dependent way. Since the coordinate system is a human artifact, this is to be avoided, unless there is a preferred coordinate system based on other information. Our methods are taken from Refs. [6] and a large literature on invariant angular statistics. Recently Sommers [7] has advocated an approach using the same basic concepts.

Statistical comparisons require formulation of a well defined *null hypothesis.* We take particular care on this point. The term "null" often indicates conditions of "no signal", which would be appropriate if one considers anisotropy a signal. More deeply, the scientific method proceeds by ruling out possibilities, rather than attempting to prove particular notions. The symmetries of data are the simplest and purest characterizations capable of being tested, or ruled out. Thus a solid null hypothesis is the foundation of any further claims.

The elimination of coordinate dependent artifacts, the testing of a symmetry-based null, and the objectivity of maximum likelihood analysis create very powerful tools. If this method is valid, then it should be able to determine the systematic bias in real data objectively. The main purpose of the paper is to demonstrate the power of the tools for future use. However we also test our method to find significant evidence for anisotropy in the distribution of the highest energy events, which has been a question of intrinsic interest for many years.

Methodology: The method of maximum likelihood (see Refs. [8, 9] for example) allows one to test objectively between different statistical models. It is good practice to test models which are continuously related to the null when parameters are varied. Continuity in the likelihood test tends to assure that the same features of the data are tested by the null and competing models.

Let $\vec{x}$ denote an element of a data set, for example the coordinates of an observed track. Given a normalized distribution $f(\vec{x})$, the likelihood of the data in the distribution is defined to be the product of the distribution evaluated at each data point over the data set. Maximum likelihood occurs at maximum log likelihood $\mathcal{L}$, defined as

$$\mathcal{L} = \sum_{j}^{N} log[f(\vec{x}_j)], \tag{1}$$

where N is the number of data points.

Hypothesis testing is done by comparing the difference of maximized log-likelihoods (T) for a given data set evaluated with two different distributions $f(\vec{x})$, $f_{\rm null}(\vec{x})$, where $f_{\rm null}(\vec{x})$ is the null distribution. For large $N \gg 1$, the statistic $2T = 2(\mathcal{L}^{\rm max}_{\rm model} - \mathcal{L}^{\rm max}_{\rm null})$ has a chi-squared distribution for p parameters (χ^2_p). We will verify this and also determine the distribution independently by Monte Carlo. Note that it is not possible and not our objective to find the "exact" distribution behind the observed data. Instead, emphasis lies on testing the null distribution. It is sufficient to show that a trial distribution fits the data sufficiently well to rule out the null; one does not imply, or conclude from this that the trial distribution is the last word.

Covariant and Invariant Statistics: For covariant quantities we map the astronomical coordinates $(DEC, RA) \to (\pi/2 - \theta, \phi) \to \hat{x}$, where $\hat{x}$ is a unit vector on the dome of the sky,

$$\hat{x} = (\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta).$$

Naturally $\hat{x}$ transforms like a vector when the coordinate system is changed. The expansion in spherical harmonics are the same thing: vectors are the $l = 1, m$ representations of spherical harmonics. By contracting vector indices it is straightforward to make trial distributions and statistically valid quantities that are scalars under rotations, and therefore independent of the coordinate system.

The procedure is very natural, and so simple that we should expand on alternatives which do not have the same features. It is very common in the subject of "circular statistics" to see quantities such as $<\theta> = \sum_i^N \theta_i/N$, $\Delta\theta = \sqrt{<\theta^2> - <\theta>^2}$. Such quantities can be computed but they are so faulty as to be nearly meaningless. The fault is immediately seen by calculating $<\theta>$ for an isotropic distribution on a circle. The distribution has no preferred orientation, yet the average angle $<\theta> \to \pi$ yields a preferred point, which is unacceptable.

Analysis: We apply our method to the set of 58 published tracks from the AGASA group with energies not less than $4 \times 10^{19} eV$ (Table 1, Refs.[10]). The average angular resolution of the track directions is stated to be 1.8 degrees [3] containing one sigma (68%) of the events. The tracks come from the region of polar angle $100° < \theta < 10°$. This data set could be, in principle, anisotropic in any direction. In this analysis, we do *not* include the bias of the instrument, discussed shortly.

The world's published data consists of some 200 points. However not all are available, while resolution, cuts and data quality varies, and the dangers of combining data from different groups suggest that restricting the study to a homogeneous set is the logical first step. We also want to make it clear that objective and highly significant conclusions can be extracted from carefully constructed tests without needing particularly large data sets. Our methods, of course, extend readily to much larger data sets anticipated from new arrays including HIRES and AUGER.

The most basic model [11, 12] to analyze unimodal spherical data $\{\hat{x}(\theta, \phi)\}$ is the *Fisher distribution*:

$$f_{\text{Fisher}}(\theta, \phi) = \frac{\kappa}{4\pi\sinh\kappa}\sqrt{1 - (\hat{n}\cdot\hat{x})^2}\, e^{\kappa\,\hat{n}\cdot\hat{x}} \qquad (2)$$

where $\hat{n}(\alpha, \beta)$ is the direction of symmetry-breaking axis. This distribution has a long history as the spherical generalization of the venerated von-Mises distribution on the circle. Both distributions have the analytic properties of the Gaussian distribution, as seen by writing

$$e^{-\kappa(\hat{n}-\hat{x})^2/2} \sim e^{\kappa\,\hat{n}\cdot\hat{x}},\ \ \hat{n}^2 = \hat{x}^2 = 1,$$

suppressing normalization factors. Parameter κ is called the *concentration parameter*, and determines the extent that data is concentrated along the *anisotropy axis* $\hat{n}$.

The Fisher distribution has the minimal number of parameters possible on the sphere, and exhibits cylindrical symmetry about the axis $\hat{n}$. The pre-factor $\sqrt{1-(\hat{n}\cdot\hat{x})^2}$ is explained as follows. Choose axis direction $\hat{n}=(0, 0, 1)$ or along the *north pole* . This hides 2 parameters, and reduces the distribution to the simpler form:

$$f_{\text{Fisher}}(\theta,\phi)=\frac{\kappa}{4\pi\sinh\kappa}\sin\theta e^{\kappa\cos\theta}. \tag{3}$$

The pre-factor $\sin\theta$ takes into account the Jacobian of solid angle on the sphere. Cylindrical symmetry about $\hat{n}$ is also obvious now, since ϕ does not appear anywhere in (3). For the limit of small anisotropy $\kappa \ll 1$, the exponential can be expanded

$$e^{\kappa\cos\theta}\sim 1-\kappa\cos\theta,$$

which is one of the distributions discussed by Sommers [7], and which needs the additional Jacobian factor to test anisotropy.

The data set is subject to a cut in declination, $(10^\circ \leq \theta \leq 100^\circ)$. We modify the *Fisher distribution* to take this cut into account as:

$$f^{\text{cut}}_{\text{Fisher}}(\theta,\phi)=\frac{f_{\text{Fisher}}(\theta,\phi)g(\theta)}{\int f_{\text{Fisher}}(\theta,\phi)g(\theta)d\theta d\phi}; \tag{4}$$

$$g(\theta)=[\eta(\theta-10^\circ)-\eta(\theta-100^\circ)], \tag{5}$$

where η is the Heaviside step function. The step functions have an invariant representation we do not bother to write here. Now (4) is the normalized *Fisher distribution* in the cut region and serves as a trial model of anisotropy.

We also need a *null distribution* on the cut region to test our model using the method of maximum likelihood. The *null isotropic distribution* on the sphere with the cut in angle θ is given by,

$$f^{\text{cut}}_{\text{null}}(\theta,\phi)=\frac{\sin\theta\; g(\theta)}{\int\sin\theta\; g(\theta)d\theta d\phi}. \tag{6}$$

The null distribution coincides with $f^{\text{cut}}_{\text{Fisher}}(\theta,\phi)$ in the limit $\kappa\to 0$. Thus when likelihood is evaluated, the variation of parameter κ allows the analysis to be continuously connected with the null.

Results: Evaluation of likelihood for the AGASA data set is straightforward. One finds a maximum log likelihood in the Fisher distribution to be

$$\mathcal{L}_{\text{Fisher}}=-128.10$$

units at

$$\kappa=1.35\pm 0.41;\; \hat{n}=(4.22^\circ, 88.34^\circ).$$

In equatorial coordinates $(RA,\ DEC)$, the axis of concentration lies along the direction $(5^h\ 53.36^m,\ 85.78^\circ)$.

The null hypothesis is evaluated by running the same analysis with $\kappa=0$. The log likelihood is found to be

$$\mathcal{L}_{\text{null}}=-134.39$$

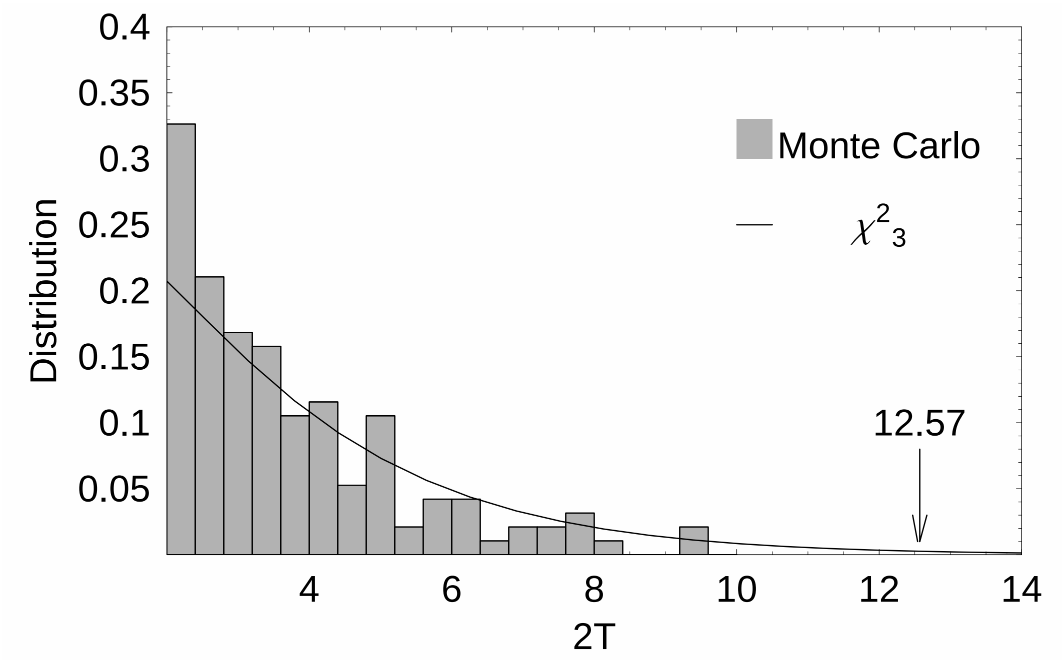

Figure 1: We plot here the $2T$ distribution from Monte Carlo (shaded histogram) and chi-square distribution of 3 parameters (solid curve). The Monte Carlo $2T$ distribution was found from a thousand sets of 58 random tracks and taking twice the difference between maximum log likelihood of the Fisher and the null distributions. The $2T$ value of 12.57 found from actual AGASA data is pointed by the arrow.

units. The quantity $2T^{\max} = 2(\mathcal{L}_{\text{Fisher}} - \mathcal{L}_{\text{null}}) = 12.57$ units.

For large $N \gg 1$ it is known that $2T$ is distributed like $\chi_p^2(2T)$ of p parameters. In practice 30-40 points is sufficient for large N to apply, but the question also depends on the dimensionality of the data. We have 3 parameters in our case, so one should expect a χ_3^2 distribution. This is a prediction of pure statistical theory, without detail about the problem at hand, and something that may be questioned.

Real data sets often contain correlations or irregularities of myriad possible origin. Such features may upset analytic estimates. We therefore made an independent Monte Carlo study of the statistic $2T$ based on the features of the problem. We took one thousand random samples of 58 data points inside the cut region. Randomness was implemented by selecting points from a distribution flat in ϕ, and flat in $\cos\theta$. For each sample of 58 we varied κ and $\hat{n}$ to determine the maximum likelihood, and we calculated the likelihood for $\kappa = 0$; the result was a trial value of $2T$. The distribution of $2T$ values for 1000 runs is shown in Fig 1 along with the χ_3^2 distribution. Agreement is excellent.

Statistical significance is determined by calculating P-values, also known as confidence levels, which are defined as the integrated probability of $2T$ to fluctuate in the null distribution above the value determined for the data. A comparison of P-values is also a method to determine whether the distributions have any unexpected long tails. The P-value for the χ_3^2 distribution to give $2T \geq 12.57$ is 3.9×10^{-4}, or 0.039%. Consistently, zero

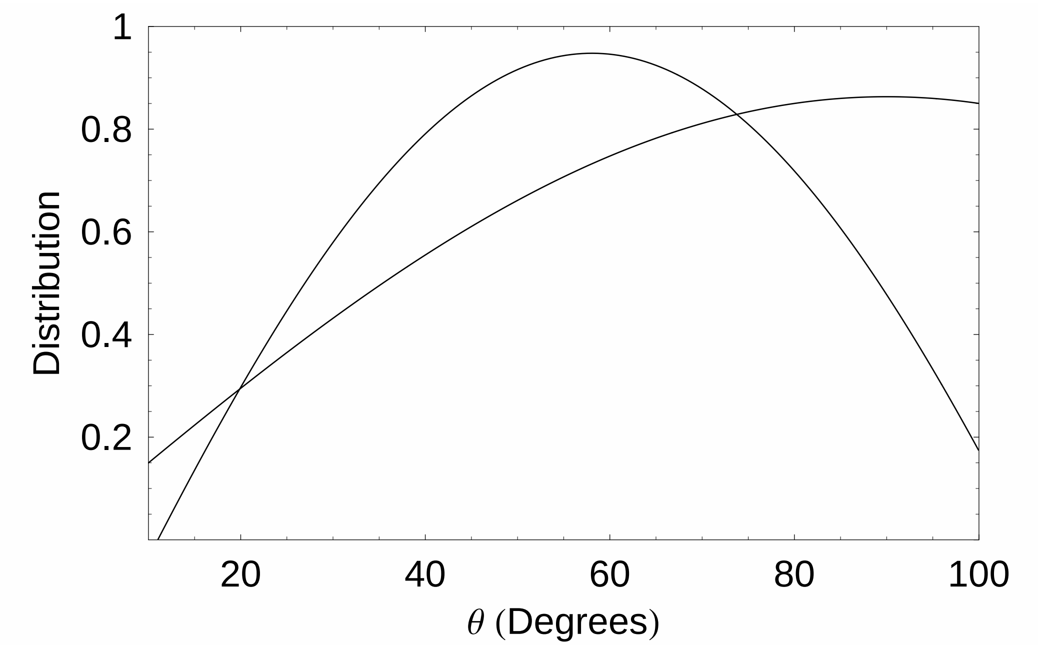

Figure 2: An isotropic distribution in θ and the detector response observed by AGASA. While the isotropic distribution peaks at $\theta = 90$ degrees, the detector response shows a peak at around $\theta = 57$ degrees.

examples of $2T \geq 12.57$ were obtained in the Monte Carlo evaluation. In terms of 2-sided Gaussian statistics, which are often used for comparisons, $P = 3.9 \times 10^{-4}$ corresponds to a "3.55 σ" effect.

These results might look like a real signal at first sight. A closer look reveals that the anisotropy axis $\hat{n}$ is along the direction $(5^h\ 53.36^m,\ 85.78°)$ in (RA, DEC) coordinates and is very close to the equatorial north pole. We determined the error in the axis by exploring a contour of parameters $\Delta\hat{n}$ yielding 1/2 unit of ΔT: the error circle is $7.1°$ and includes the north pole. This indicates a possible bias of the instrument.

AGASA [5] gives a distribution in declination of all events above $> 10^{19} eV$. We fit the distribution by a function of the form $(a \sin b\theta + c)$ which is related to the isotropic distribution in θ, namely $d(\cos\theta)$. While the isotropic distribution peaks at $\theta = 90$ degrees, the detector response drops as one move away from $\theta = 57$ degrees. The distributions are plotted in Fig 2.

A revised likelihood analysis with the declination distribution corresponding to the detector response gives log likelihood to be $\mathcal{L}'_{\rm null} = -128.81$ units. This yields $2T = 1.42$, indicating that the previous log likelihood difference is attributable to detector bias. This procedure of determining detector bias is circular in the sense that one can use the distribution in declination from observed data and fit the track distribution without discovering any signal. However, if the detector response is not explicitly known or is not well defined, one can interpret the result as a real signal using the null distribution which is isotropic.

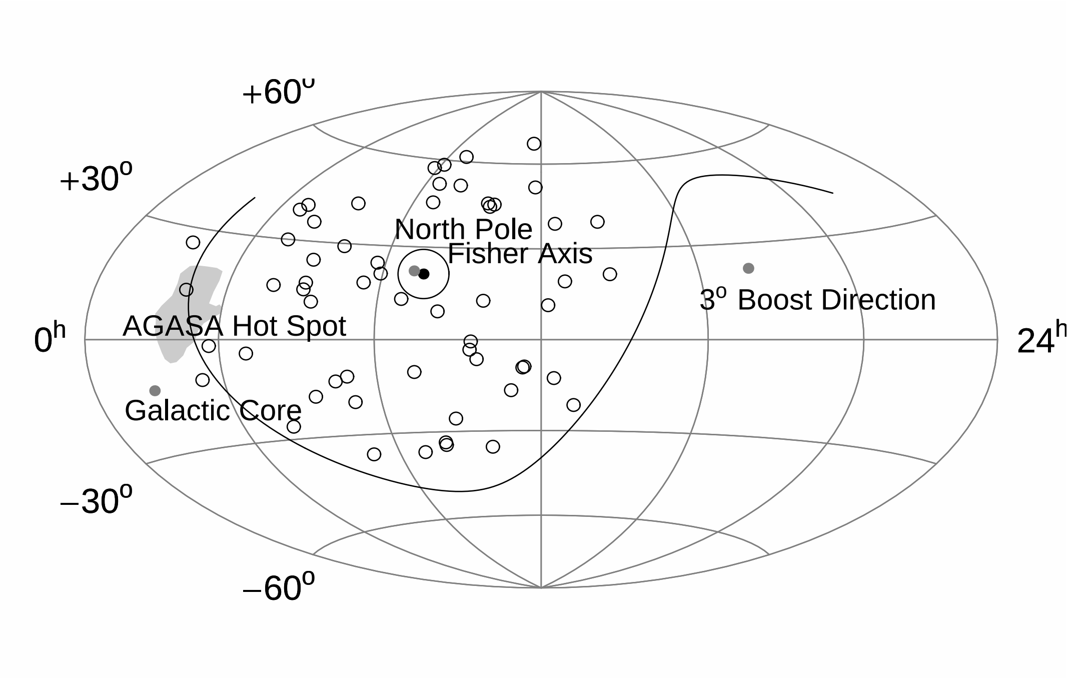

Figure 3: Aitoff-Hammer equal area projection of the sky in Galactic Coordinates. The AGASA tracks are denoted by open circles with a radius equal to their angular resolution of 1.8 degrees. The anisotropic hot spot AGASA group found at lower energies [10] has been shown as the shaded region at the left corner. The Fisher axis of global anisotropic direction we have found is about 90 degrees from the AGASA hot spot and close to the equatorial north pole. The error in the determination of the Fisher axis is shown by the circle We also show the direction of motion relative to 3 degrees cosmic microwave background and the equatorial plane.

Observations: We plotted the data in several coordinate systems to visually examine signals of anisotropy, and to check the fit to the Fisher distribution. Results in Galactic Coordinates and Aitoff-Hammer equal-area projection are shown in Fig 3. The advantage of an equal-area projection is that an isotropic distribution will appear uniform, and anisotropy is not unduly distorted by the coordinates. Galactic coordinates are used so that any correlation with the galaxy plane might be easily seen. The equatorial plane is shown by the solid curve.

The figure shows the Fisher "hot-spot" extracted from the fit, namely the best-fit anisotropy axis $\hat{n}$. As can be seen, the Fisher axis is close to the equatorial north pole and its error cone actually encloses the north pole. Also shown are the locations of some other features of obvious interest: the hot-spot of the lower energy AGASA analysis [5], and the approximate location of the galaxy core. We added the direction of our system's motion relative to the cosmological background radiation, as deduced from the dipole component of the the $3°$ cosmic microwave background, for another comparison.

Conclusion: Unresolved questions regarding the origin of cosmic rays with energies above $4 \times 10^{19} eV$ suggests testing the degree of isotropy of the data. The method we have developed is capable of ruling out isotropy and eliminating any coordinate dependent artifacts. Given the detector bias reported by AGASA, the group's cosmic ray tracks above

$4 \times 10^{19} eV$ are consistent with an isotropic distribution. The question of anisotropy probably hinges on a finer understanding of angular bias than currently available in published sources: for example, error bars on the angular acceptance would be useful.

Acknowledgements: This study springs out of collaboration with Pankaj Jain, who we thank for productive comments and criticsm. Work was supported in part under the Department of Energy grant number DE-85ER40212 by the University of Kansas General Research Fund, and the *Kansas Institute for Theoretical and Computational Science/K*STAR* program.

References

[1] K. Greisen. *Phys. Rev. Lett.*, 16:748, 1966.

[2] G. T. Zatsepin and V. A. Kuzmin. *Sov. Phys. JETP Lett.*, 4:78, 1966.

[3] Y. Uchihori et al. *Astropart. Phys.*, 13:151, 2000.

[4] N. Hayashida et al. astro-ph/9906056, 1999.

[5] M. Takeda et al. astro-ph/9902239, 1999.

[6] A. Virmani, S. Bhattacharya, P. Jain, S. Razzaque, J. P. Ralston, and D. W. McKay. astro-ph/001010235, 2000. *Astropart. Phys.* (in press).

[7] Paul Sommers. *Astropart. Phys.*, 14:271, 2001.

[8] Alan Stuart and J. Keith Ord. *Kendall's Advance Theory of Statistics*, volume 1. Halsted Press, 1994.

[9] Particle Data Group. Review of particle physics. *The European Physical Journal C*, 15, 2000.

[10] N. Hayashida et al. *Astrophys. J.*, 522:225, 1999.

[11] Edward Batschelet. *Circular Statistics in Biology.* Academic Press, 1981.

[12] N. I. Fisher, T. Lewis, and B. J. J. Embleton. *Statistical analysis of spherical data.* Cambridge University Press, 1987.

Table 1: AGASA Events Above 4×10^{19} eV (J2000 coordinates)

Date	Energy ($\times 10^{19} eV$)	RA	DEC	Date	Energy ($\times 10^{19} eV$)	RA	DEC
84/12/12	6.81	$22^h\ 21^m$	38.4°	84/12/17	9.79	$18^h\ 29^m$	35.3°
86/01/05	5.47	$4^h\ 38^m$	30.1°	86/10/23	6.22	$14^h\ 02^m$	49.9°
87/11/26	4.82	$21^h\ 57^m$	27.6°	89/03/14	5.27	$13^h\ 48^m$	34.7°
89/08/16	4.07	$5^h\ 51^m$	58.5°	90/11/25	4.51	$16^h\ 17^m$	$-7.2°$
91/04/03	5.09	$15^h\ 47^m$	41.0°	91/04/20	4.35	$18^h\ 59^m$	47.8°
91/05/31	5.53	$3^h\ 37^m$	69.5°	91/11/29	9.10	$19^h\ 06^m$	77.2°
91/12/10	4.24	$0^h\ 12^m$	78.6°	92/01/07	4.51	$9^h\ 36^m$	38.6°
92/01/24	4.88	$17^h\ 52^m$	47.9°	92/02/01	5.53	$0^h\ 34^m$	17.7°
92/03/30	4.47	$17^h\ 03^m$	31.4°	92/08/01	5.50	$11^h\ 29^m$	57.1°
92/09/13	9.25	$6^h\ 44^m$	34.9°	93/01/12	10.1	$8^h\ 17^m$	16.8°
93/01/21	4.46	$13^h\ 55^m$	59.8°	93/04/22	4.42	$1^h\ 56^m$	29.0°
93/06/12	6.49	$1^h\ 16^m$	50.0°	93/12/03	21.3	$1^h\ 15^m$	21.1°
94/07/06	13.4	$18^h\ 45^m$	48.3°	94/07/28	4.08	$4^h\ 56^m$	18.0°
95/01/26	7.76	$11^h\ 14^m$	57.6°	95/03/29	4.27	$17^h\ 37^m$	$-1.6°$
95/04/04	5.79	$12^h\ 52^m$	30.6°	95/10/29	5.07	$1^h\ 14^m$	20.0°
95/11/15	4.89	$4^h\ 41^m$	29.9°	96/01/11	14.4	$16^h\ 06^m$	23.0°
96/01/19	4.80	$3^h\ 52^m$	27.1°	96/05/13	4.78	$17^h\ 56^m$	74.1°
96/10/06	5.68	$13^h\ 18^m$	52.9°	96/10/22	10.5	$19^h\ 54^m$	18.7°
96/11/12	7.46	$21^h\ 37^m$	8.1°	96/12/08	4.30	$16^h\ 31^m$	34.6°
96/12/24	4.97	$14^h\ 17^m$	37.7°	97/03/03	4.39	$19^h\ 37^m$	71.1°
97/03/30	15.0	$19^h\ 38^m$	$-5.8°$	97/04/10	3.89	$15^h\ 58^m$	23.7°
97/04/28	4.20	$2^h\ 18^m$	13.8°	97/11/20	7.21	$11^h\ 09^m$	41.8°
98/02/06	4.11	$9^h\ 47^m$	23.7°	98/03/30	6.93	$17^h\ 16^m$	56.3°
98/04/04	5.35	$11^h\ 13^m$	56.0°	98/06/12	12.0	$23^h 16^m$	12.3°
98/09/03	4.69	$19^h 36^m$	50.7°	98/10/27	6.11	$3^h 45^m$	44.9°
99/01/22	7.53	$19^h 11^m$	5.3°	99/07/22	4.09	$7^h 39^m$	32.2°
99/07/28	7.16	$3^h 46^m$	49.5°	99/09/22	10.4	$23^h 03^m$	33.9°
99/09/25	4.95	$22^h 40^m$	42.6°	99/10/20	6.19	$4^h 37^m$	5.1°
99/10/20	4.29	$4^h 02^m$	51.7°	00/05/26	4.98	$14^h 08^m$	37.1°