Order statistics of the arrival directions of the highest energy cosmic rays

Dalibor Nosek* **and Jana Nosková**†

**Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic*
†Czech Technical University, Faculty of Civil Engineering, Prague, Czech Republic

Abstract. **We use a measure of clustering derived from order statistics to distinguish between regular and clustered structures of events distributed on the celestial sphere. Applying this method to the highest energy events recorded by the AGASA experiment, their small scale anisotropy is discussed.**

Keywords**: Anisotropy, Order statistics**

I. Introduction

The demonstration that arrival directions of the highest energy cosmic rays (CR) are not distributed uniformly on the sky is of high importance on the way to identify their origin and explain their nature. Traditionally, correlation functions and related measures are favored. Nonetheless, there are many other statistics available characterizing distributions of events on the celestial sphere. Corresponding tests include methods based on the binomial distribution, counts–in–cells tests relying upon Poissonian statistics, spherical harmonics and many others.

The aim of this study is to present a selfconsistent method that can be used to quantify the departure of distributions of CR arrival directions from isotropy. We use a method based on order statistics. It allows us to study angular distances between the arrival directions observed on the sky with the emphasis on searching for their cluster structures. Thus, it provides an easy way how a significance of the anisotropy signal, if it is present in the data, can be established without any information on astrophysical objects and sky maps.

II. Method

To study the distribution of the arrival directions of CR events on the celestial sphere we consider their separation angles. The departure from isotropy is investigated using a method testing an ordered set of these angles.

We proceed in the following steps. First, we use a sky–exposure function of a surface detector to transform measured arrival directions with the property that if the original arrival directions of CR events are spread uniformly on the sky, their new coordinates are distributed uniformly as well. In the second step, we construct empirical distributions of angular distances between transformed arrival directions. Finally, a statistical analysis seeking cluster structures of CR events in distributions of their separation angles is applied.

A. Exposure transformation

Assume that a surface detector is located at latitude δ_0 and that it is fully efficient for particles arriving with zenith angles $\theta \le \theta_{\rm m}$. For a flat detector at a single side with full–time operation there is no exposure variation in right ascension. Then there is a relative sky–exposure function $w = w(\delta)$, $w \in \langle 0, 1\rangle$, that depends only on declination δ, where $\delta \in \langle -\frac{\pi}{2}, \frac{\pi}{2}\rangle$.[1] It is given by the relation [1]

$$w(\delta) \propto \cos\delta_0 \cos\delta \sin\alpha_{\rm m} + \alpha_{\rm m} \sin\delta_0 \sin\delta, \quad (1)$$

where $\alpha_{\rm m} = 0$ if $\xi > 1$, $\alpha_{\rm m} = \pi$ if $\xi < -1$ and $\alpha_{\rm m} = \arccos\xi$ otherwise, and the parameter ξ is defined by $\xi \cos\delta_0 \cos\delta = \cos\theta_{\rm m} - \sin\delta_0 \sin\delta$. Note that $\xi > 1$ and $w(\delta) = 0$ if $\delta > \delta_0 + \theta_{\rm m}$ or $\delta < \delta_0 - \theta_{\rm m}$.

Consider that the arrival directions of CR events are distributed uniformly on the sky. Then, the measured declinations, Δ, are distributed according to the sky–exposure function $w(\delta)$ of the detector,

$$f_\Delta = C^{-1} w(\delta)\cos\delta, \quad C = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} w(\delta')\cos\delta'\,{\rm d}\delta', \quad (2)$$

where C is a normalization constant. If these measured declinations δ are transformed to new declinations ψ, where $\psi \in \langle -\frac{\pi}{2}, \frac{\pi}{2}\rangle$, using the relation

$$\frac{1}{C}\int_{-\frac{\pi}{2}}^{\delta} w(\delta')\cos\delta'\,{\rm d}\delta' = \frac{1}{2}\int_{-\frac{\pi}{2}}^{\psi} \cos\psi'\,{\rm d}\psi', \quad (3)$$

the new arrival directions are distributed uniformly on a sphere. The probability distribution of the new declinations, Ψ, is $f_\Psi = \frac{1}{2}\cos\psi$. This shift is nothing but a transformation of the random variable Δ (measured declination) to the new random variable Ψ (transformed declination) given implicitly by the equality of their cumulative distributions, i.e. $F_\Delta(\delta) = F_\Psi(\psi)$.

The exposure transformation of the 58 arrival directions of the highest energy ($E \ge E_{\rm min} = 38.9$ EeV) CR events recorded by the AGASA experiment [2], [3] is shown in Fig. 1. In this figure, the original and transformed arrival directions in the equatorial coordinates are depicted. Corresponding histograms of measured and transformed declinations are shown in Fig. 2.

[1] For the detector located at latitude δ_0 that observes particles arriving with zenith angles $\theta \le \theta_{\rm m}$, declinations of their arrival directions are such that $\delta \in \langle \delta_{\rm n}, \delta_{\rm m}\rangle$ where $\delta_{\rm n} = \max(\delta_0 - \theta_{\rm m}, -\frac{\pi}{2})$ and $\delta_{\rm m} = \min(\delta_0 + \theta_{\rm m}, \frac{\pi}{2})$.

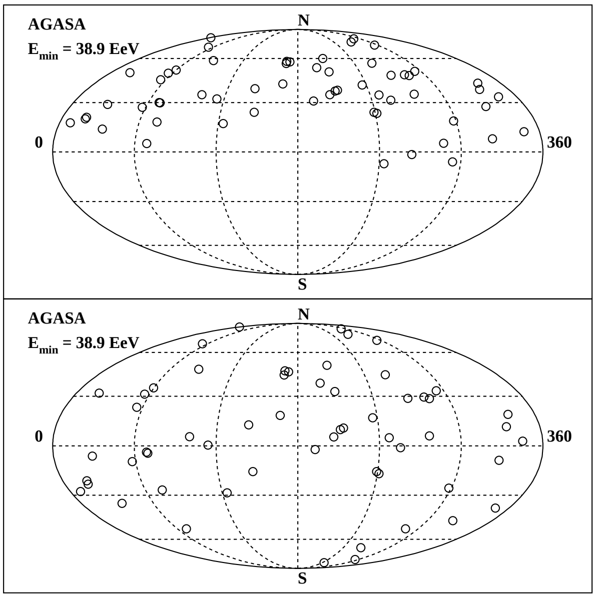

Fig. 1. Measured (top) and transformed (bottom) arrival directions of the 58 AGASA events with energies equal or higher than 38.9 EeV [2], [3].

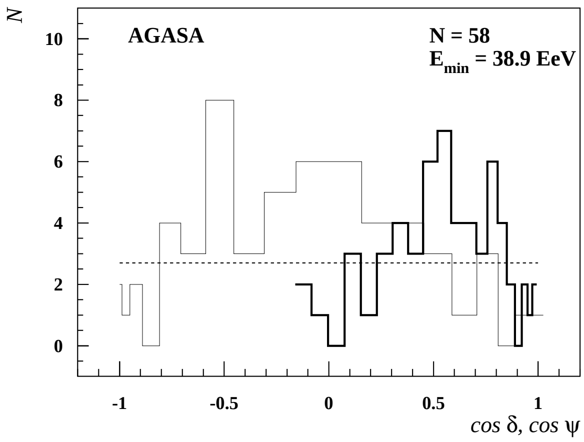

Fig. 2. Histograms of measured (thick) and transformed (thin) declinations of the 58 most energetic AGASA events [2], [3]. A dashed line is for the same number of uniformly distributed arrival directions.

B. Order statistics distribution

In the following, we consider the transformed arrival directions of CR events. They are used to construct empirical distributions of separation angles between a chosen target event and its l–th nearest event.

Let us assume N events uniformly distributed on the surface of a sphere. We arbitrarily choose one event and determine $N-1$ separation angles measured from it to the remaining events. These separation angles, treated here as a random variable Φ, are distributed with the probability function $f_\Phi = \frac{1}{2}\sin\phi$ for $\phi \in \langle 0, \pi\rangle$. It is easy to show that a new random variable, the angular distance $Y = \sin^2\frac{\Phi}{2}$, that also measures the separation of events on a sphere is distributed uniformly on a unit interval, i.e. $Y \propto \mathrm{U}(0,1)$. It follows that the probability distribution of the angular distance between the chosen event and its l–th nearest event, $Y_{(l)}, l = 1, 2, \ldots, N-1$, i.e. its l–th ordered statistic, is given by the beta distribution, $Y_{(l)} \propto \mathrm{B}(l, N-l)$, for more details see Appendix. In this way, an isotropy hypothesis of arrival directions of CR events can be easily verified using the transformed angular distances of an arbitrary order.

In practice, given a sample of N events on a sphere, the distribution of the l–th order statistics of their angular distances can be directly obtained by measuring the angular distances between a chosen target event and all other $N-1$ events, retaining their l–th smallest value, and repeating this procedure for every of N events in the sample. One ends up with the distribution consisting of N angular distances of the l–th order.

The first and second order probability distributions of separation angles between transformed arrival directions of the 58 highest energy AGASA events [2], [3] are depicted in Fig. 3; also distributions expected under the isotropy hypothesis are shown.

C. Order statistics test

The distributions of angular distances of a given order constructed from the transformed arrival directions are further examined. We test whether such distributions are consistent with corresponding distributions originated in arrival directions spread uniformly on a sphere. The Kolmogorov–Smirnov (KS) test using a maximal distance between an empirical and hypothetical distribution or the multinomial (χ^2) test dealing with a sum of squared residuals are usually adopted. However, both these method work asymptotically, i.e. dealing with tens of events, they may lead to a conclusion the significance of which is not under control. For this reason, we use a more advanced method that controls hypothetical and empirical distributions for any number of events. Being based on order statistics it verifies whether a given separation angle can be found in a specific position in an ordered set of these angles.

Assume angular distances of the l–th order, $Y_{(l)}$, originating in N arrival directions spread uniformly on the sky. The random variable $Z = G_l(Y_{(l)})$ that is given by the cumulative distribution of these distances, with $G_{(l)}(x)$ given in Eq. (6) inside Appendix, is distributed uniformly on a unit interval, $Z \propto \mathrm{U}(0,1)$. It follows that the distribution of its k–th ordered statistics, $Z_{(k)}, k = 1, 2, \ldots, N$, is given by the beta distribution, $Z_{(k)} \propto \mathrm{B}(k, N-k+1)$, see Appendix.

Therefore, having an ordered sample of N distances $z_1 \leq z_2 \leq \ldots \leq z_N$, where $z_k = G_l(y_{(l),k})$ and where $y_{(l),k} = \sin^2(\frac{1}{2}\phi_{(l),k})$ are ordered angular distances of the l–th order between tested events, one can easily find out probabilities that the k–th order statistics $Z_{(k)}$ is smaller or larger than the k–th distance z_k. These probabilities (p–values) that measure the significance with which an isotropy hypothesis could be rejected are, respectively, $\Pr(Z_{(k)} < z_k) = G_k(z_k)$ (excess) and $\Pr(Z_{(k)} > z_k) = 1 - G_k(z_k)$ (deficit), where $G_k(x)$ is given in Eq. (6). The former (later) p–values provide

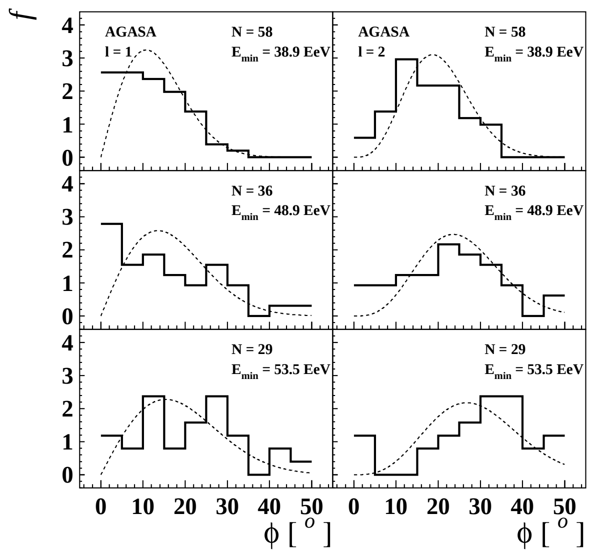

Fig. 3. Probability distributions obtained for the first (left) and second (right) order statistics of separation angles between transformed arrival directions of AGASA events for three threshold energies. Dashed lines are for the same number of uniformly distributed arrival directions.

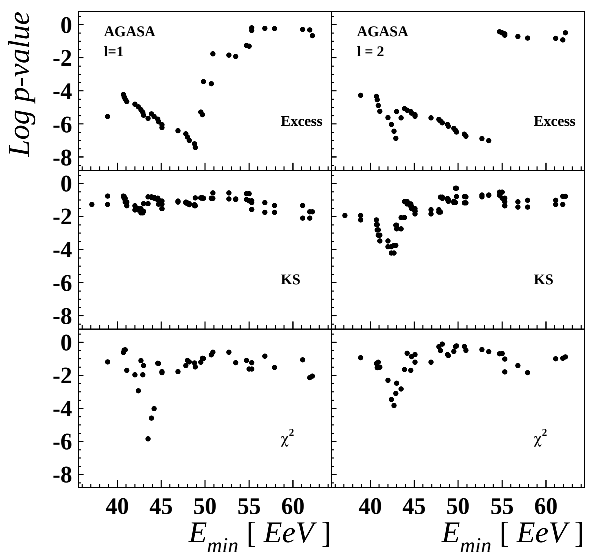

Fig. 4. Test p–values for the first (left) and second (right) order distributions of angular distances are displayed as functions of threshold energy. The smallest p–values for an excess taken over ordered angles (top), results for KS (middle) and χ^2 tests (bottom) are shown.

significances for an excess (deficit) of distances of the l–th order in the region of the k smallest distances. In this way, N tests for the excess and the same number of tests for the deficit can be carried out for each set of angular distances of any order.

In principle, having a sample of N events on a sphere, one can obtain $N-1$ different distributions of angular distances with $l = 1, 2, \ldots, N-1$, each comprising N angles, and compare them with expectations. However, there are altogether $\frac{1}{2}N(N-1)$ angular distances among N events but only $2N-3$ of them are independent.[2] It suggests that testing uniformity of events it suffices to test two sets of angular distances since other tests applied to the same data should depend on them. Once the isotropy of events is rejected by either of these tests, the resultant significance level provides a lower bound of the rejection, because some further tests could refuse the null isotropy hypothesis with a higher significance. However, $N-1$ samples of N separation angles between an event and its l–th nearest event can deviate from an expectation on different angular scales making all of corresponding tests meaningful.

The order statistics analysis treats the angular separations that are not independent. Moreover, a number of dependent test statistics for every ordered set of these distances is examined. In such a case, a true significance with which an isotropy hypothesis could be rejected is usually estimated in Monte Carlo (MC) simulations. The same number of studied events uniformly distributed on a sphere is generated many times and all steps of the analysis are repeated with every such MC sample. The probability that the resultant anisotropy is due to a chance coincidence is then estimated by a fraction of MC samples that yield p–values not larger than the experimental data. This procedure accounts for the dependence within the data and multiple testing as well.

[2] N events on a surface of a sphere are defined by $2N$ coordinates. One event can be placed at the pole and the second one at the zero meridian. There are $N-1$ independent angular distances to the pole and $N-2$ independent angular distances to the second event.

III. Application to the AGASA data set

To illustrate the application of the order statistics method we consider the arrival directions of 58 events with energies higher than $E_{\min} = 38.9$ EeV and zenith $\leq 45°$ measured by the AGASA surface detector [2], [3]. In our analysis, we have tested whether this data set is a likely realization of an isotropic distribution.

In this data set, one triplet and four doublets of events within a separation angle of $2.5°$ were reported by the AGASA experiment [2], [3], and the number of doublets increases to six for $3.5°$. Therefore, we expect a significant departure from isotropy for the first and second order distributions of separation angles.

We started with a set of the 20 most energetic events, corresponding to the threshold energy of 62.2 EeV, and added events with energies in decreasing order one by one up to 58 events. For every energy threshold, distributions of separation angles of different orders have been constructed, examples of the first and second order are shown in Fig. 3. Then, using the cumulative beta distribution, a series of tests seeking an excess or deficit in ordered sets of separation angles have been carried out. This procedure has been repeated 39 times for all energy thresholds. Altogether 1142 and 1293 single tests seeking for the departure from isotropy in the first and second order distributions of separation angles have been performed, respectively. Also higher order statistics have been examined. In order to get an estimation of the

TABLE I
ORDER STATISTICS RESULTS

(l)	$\phi_{(l),k}$	$E_{\min}$	N	$k/k_{\rm exp}$	p–value
1	$2.8°$	48.9 EeV	36	$9/0.72$	$3.8\ 10^{-8}$
2	$3.4°$	53.5 EeV	29	$3/0.0083$	$9.6\ 10^{-8}$
3	$37.5°$	57.9 EeV	23	$3/8.8$	$5.0\ 10^{-3}$
4	$39.0°$	55.3 EeV	26	$2/7.6$	$6.2\ 10^{-3}$
5	$50.7°$	57.9 EeV	23	$2/8.3$	$2.0\ 10^{-3}$

significance of the tests we have generated 10^6 MC sets of 58 arrival directions spread uniformly on a sphere and repeated the whole procedure with every set searching the number of MC trials with not larger p–values.

The results of our analysis are summarized in Table I. The most significant departure from isotropy is found in a set of 36 AGASA events with the threshold energy of $E_{\min} = 48.9$ EeV if the separation angles between two nearest events are analyzed. The smallest p–value taken over all threshold energies and all separation angles is about $3.8\ 10^{-8}$. An excess of small separation angles is found; 9 angles are smaller than $\phi_{(1),9} = 2.8°$ against about 0.72 expected. The analysis of the second order distribution yields 3 angles smaller than $\phi_{(2),3} = 3.4°$ while only less than 0.01 ones are expected, giving the smallest p–value of $9.6\ 10^{-8}$ for $E \geq 53.5$ EeV. The higher order statistics applied to the separation angles of events the examples of which are written in the three bottom lines in Table I give much larger p-values for a deficit found within much larger angular regions.

Using 10^6 MC sets, we estimate that the departure from isotropy in the AGASA data is given by chance with the probability of about 0.46% and 2.20%, respectively, when the first and second order statistics are considered. In both cases, a small scale clustering is observed. Our estimates fit well with the results of the AGASA experiment [2], [3] reporting one triplet and four doublets within a separation angle of $2.5°$ above 38.9 EeV. Although criticized [4], the original analysis reported the probability of observing such clusters by a chance under the isotropy hypothesis to be about 1%.

The p–values for the first and second order angular distances are depicted as functions of threshold energy in Fig. 4. Also p–values of the KS and χ^2 tests, both seeking differences between hypothetical and observed distributions, are shown for comparison in this figure. Since both latter tests work asymptotically, their significances, especially for a moderate number of data treated here, have to be taken with caution.

IV. CONCLUSIONS

We have presented a novel method for testing distributions of arrival directions of CR events on the sky using order statistics without reference to an association with any astrophysical objects or sky maps.

Unlike many other techniques the order statistics method naturally incorporates information stemming from high order correlations. Angular range selection is completely eliminated in our analysis. In some sense, since every order statistics peaks in a typical angular region, the order chosen in the analysis serves as a filter for the angular scale one wants to be focused on. Among further advantages, this method holds exactly for any number of tested events and naturally avoids problems related to binning. However, similarly as most of other relevant techniques, the numerical analysis based on order statistics have a general problem since it treats the angular separations that are not independent.

The order statistics method was shown to be a useful tool for detecting anisotropies and properly accounting for clustering and directional features of CR events distributed on the sky. Consistently with the previous findings [2], [3], we have quantified that the deviation of the highest energy AGASA events from isotropy is significant at the 1% level.

APPENDIX

Assume a real–valued random variable X for the experiment uniformly distributed on a unit interval, $X \propto \mathrm{U}(0,1)$. We perform n independent replications of the experiment to generate a sample of size n of random variables, $(X_1, X_2, \ldots, X_n)$, each with the distribution of the random variable X. Let us denote $X_{(k)}$, $k = 1, 2, \ldots, n$, the k–th smallest random variable of $(X_1, X_2, \ldots, X_n)$. This random variable is usually assigned as the k–th order statistics. Since it is a function of the sample variables it is a statistics, see e.g. Ref. [5].

The probability of having $k-1$ random variables among n ones at the distance $\leq x$ and one random variable among remaining $n-k+1$ ones at the distance $(x, x + \mathrm{d}x)$ is simply

$$\mathrm{d}P_k = \binom{n}{k-1}\binom{n-k+1}{1} x^{k-1}\,\mathrm{d}x(1-x)^{n-k}. \quad (4)$$

Then, the random variable $X_{(k)}$ has the beta distribution, $X_{(k)} \propto \mathrm{B}(k, n-k+1)$, i.e. its probability and cumulative distributions are, respectively,

$$g_k(x) = k\binom{n}{k} x^{k-1}(1-x)^{n-k}, \quad (5)$$

and

$$G_k(x) = \sum_{i=k}^{n} \binom{n}{i} x^i (1-x)^{n-i}. \quad (6)$$

ACKNOWLEDGMENT

This work was supported by the grants MSMT LC527, MSMT LA08016 and MSM0021620859 of the Ministry of Education, Youth and Sports of the Czech Republic. The authors would like to thank Petr Trávníček and Radomír Šmída for their help and support.

REFERENCES

[1] P.Sommers, Astropart.Phys. **14**, 271 (2001).
[2] M.Takeda et al., ApJ **522**, 255 (1999).
[3] N.Hayashida et al., [astro-ph0008102].
[4] C.B.Finley, S.Westerhoff, Astropart.Phys. **21**, 359 (2004).
[5] H.A.David, H.N.Nagaraja, Order statistics, John Willey & Sons, Third Edition, 2003.