

A method to search for correlations of UHECR masses with the large scale structures in the local galaxy density field

Anatoly Ivanov^a

Shafer Institute for Cosmophysical Research & Aeronomy, Lenin Avenue, 31, Yakutsk 677980, Russia

Abstract. One of the main goals of investigations using present and future giant extensive air shower (EAS) arrays is the mass composition of ultra-high energy cosmic rays (UHECRs). A new approach to the problem is presented, combining analysis of arrival directions with the statistical test of the pairs of EAS samples. An idea of the method is to search for possible correlations of UHECR masses with their presumably separate sources. The method is based on a non-parametric statistical test, specifically Wilcoxon rank sum routine, which does not depend on the populations fitting any parameterized distributions.

1. INTRODUCTION

The origin of ultra-high energy cosmic rays (UHECRs) is a long-standing challenge to astrophysics. The energy spectrum of particles constituting CRs is measured up to and slightly above 10^{20} eV (= 100 EeV) [1–3], but no evidence is revealed till now neither of the sources nor the sort(s) of the highest energy particles, mainly due to the arrival direction distribution which is nearly isotropic [4]. However, some hints were found recently of the possible correlation of UHECR arrival directions with nearby active galactic nuclei (AGN) at energies $E > 56$ EeV [5].

Disputable estimates of the average mass composition of the highest energy CRs were given basing on the depth of EAS maximum, x_m , measured at the Pierre Auger Observatory (PAO, [6]) and the High Resolution Fly's Eye (HiRes, [7]) arrays. One of the possible explanations of the diverging results may be the different average masses of UHECRs observed in different fields of view of instruments.

In this paper, another approach is used to search for possible correlation of UHECR masses with their sources basing on extensive air shower (EAS) observables, namely x_m and the shower age, τ , varying with the mass of primary particles. Here, 'age' means the stage of the cascade development at the detector level, x_0 : $\tau = x_0 \sec \theta / x_m$, where θ is the inclination angle of the shower axis [8]. This method is convenient to reveal different primary particles initiating EAS rather than to search for excess flux from candidate sources of UHECRs. An advantage of the method is that it belongs to some few one able to indicate different primary particles.

2. SEARCHING FOR A DIFFERENCE IN MASS COMPOSITION OF UHECRS ARRIVING FROM COMPLEMENTARY CELESTIAL REGIONS

The method is aimed at the possible difference in the mass of UHECRs arriving from different regions on the celestial sphere. For example, in the model of CR acceleration at the shocks created by expanding

^ae-mail: ivanov@ikfia.ysn.ru

cocoons around AGN [9], the resulting UHECR flux¹ consists of heavy nuclei at energies ~ 10 EeV. Selecting EASs with arrival directions in the large scale celestial structure where the majority of AGN are enclosed, and in the complementary area which contains less or no AGN, and where, presumably, protons dominate, we can verify the applicability of the model.

This method can be used as an extension of the matter tracer² model proposed in [10] to the case where different particles are supposed to be generated in UHECR sources.

Our first task is to test the null hypothesis that there is no difference in mass composition between two regions. We have to compare two distributions of some EAS observable sensitive to composition of the primaries in the given energy range, in order to decide – is there a significant difference or not.

2.1 Non-parametric statistical test of data samples

The distributions of measured EAS parameters usually are not described by the normal distribution, and moreover, have a specific form in each particular case, so that the general approach to the statistical test of the data samples is preferably non-parametric. The meaning of the term refers here to distribution-free methods, which do not rely on assumptions that the data are drawn from a known probability distribution.

In this paper, a Wilcoxon rank sum test (WRST) for a pair of independent samples is used in data analysis. Also known as Mann-Whitney U test in statistics, this is one of the far-famed non-parametric significance tests. It is useful for deciding whether the two samples of observations belong to the same original distribution: the null hypothesis, H_0 , is that the two samples are drawn from a single population.

If we have the pair of independent samples, S and T containing, for simplicity, equal number of measurements, N , then the test ought to detect deviations from H_0 . Combining both samples into a single ascending series of $2N$ ranked observations, where the rank is an item number in a series, i_S and j_T , we get the WRST statistic as a sum of the ranks for observations from one of the samples, S for instance, $w_S = \sum i_S$. The sum of the ranks in a sample T is $w_T = N(2N + 1) - w_S$, and therefore, under H_0 : $\overline{w_S} = \overline{w_T} = 0.5N(2N + 1)$, $\delta w = N\sqrt{(2N + 1)/12}$.

The probability $P(|w_S - \overline{w_S}|)$ is a measure of the significance of deviation from the expected value under H_0 . In the following, $P_{crit} = 0.01$ is assumed as the critical value for rejecting the null hypothesis.

2.2 A method based on the measurement of the depth of EAS maximum

In UHECR domain, the shower maximum is observable directly with air fluorescence detectors. Collaborations working at HiRes [7], PAO [6] and TA [11] have measured $\overline{x_m}$ as a function of energy to estimate the mass composition of CRs. In future detectors like JEM-EUSO, Auger Next and others the technique is planned to be used. In all these experiments the proposed method can be applied to search for correlations of UHECR masses with their sources.

A relation between x_m and CR mass, A , is obvious in a superposition approximation, where an air shower initiated by the nucleus of mass A is treated as a superposition of A nucleons of energy E/A . So, the depth of a shower maximum is [12]: $x_m = x_{18} + D_{ER} \log(E/A)$, where D_{ER} is elongation rate, E is in EeV. Averaging it over the distribution of CR energy and mass, $J(E, A)$, we have

$$\overline{x_m} = x_{18} + D_{ER} \overline{\log E} - D_{ER} \overline{\log A}. \quad (1)$$

Assuming A as the slowly changing function of energy, we can use the relation

$$\Delta \overline{x_m} = D_{ER} \Delta \overline{\log A} \quad (2)$$

in a narrow energy interval.

¹ In the ankle scenario.

² Within ~ 100 Mpc.

Table 1. Lower limits to UHECR mass resolution, $\overline{\Delta A/A}$, of the HiRes detectors.

E_{thr} , EeV	N	QGSjetII	Sibyll2.1	EPOSv1.99
1.6	815	0.041	0.035	0.033
10	125	0.132	0.094	0.086
25	18	0.348	0.247	0.226

Table 2. Minimal difference in average mass, $\overline{\Delta \log A}$, of UHECRs observed by PAO fluorescence detectors, resolvable by WRST.

E_{thr} , EeV	N	QGSjetII	Sibyll2.1	EPOSv1.99
1	1407	0.10	0.08	0.08
10	339	0.26	0.18	0.17
25	47	0.69	0.49	0.45

The accuracy of the estimation (2) can be evaluated in comparison with the model simulation results. It is believed that Monte Carlo codes like CORSIKA [13] with implemented hadronic interaction models give more realistic predictions of EAS than the superposition approximation. In Fig. 2 a comparison is presented of the results of CORSIKA simulations (QGSjetII, Sibyll2.1 and EPOSv1.99 are implemented [14]) with estimations along equation (2). The elongation rates as the functions of energy are taken from corresponding models. It seems that in Sibyll and EPOS models the interactions of nuclei are approximated similar to superposition at $E > 0.1$ EeV. The difference of QGSjet results from that of equation (2) is 8 to 17 g/cm² in the energy interval (0.01,10) EeV.

The resolution of x_m by HiRes-I and HiRes-II telescopes in a particular shower is better than 25 g/cm² above energy 1.6 EeV [7], and is at the level of 20 g/cm² over the energy range considered in PAO measurements [6]. Then, the lower limit to the resolution of the difference in average mass of UHECR, $\overline{\Delta A/A}$, by the air fluorescence detection technique can be estimated using $D_{ER} = dx_m/d \log E_P$. The subscript P has to indicate the energy independent mass composition in this case. The experimental values of $dx_m/d \log E$ are influenced by the possible changes of mass composition with energy. So one has to use the model simulation of D_{ER} in order to estimate the mass resolution.

In Table 1 the results are given for three models and energy thresholds. EAS events observed by HiRes during ~ 6 years that passed the quality selection criteria [7] are used here. The data observed by the PAO fluorescent detectors (FD) are more abundant, so the resolution of $\overline{\Delta \log A}$ is higher. But we have to estimate additionally a WRST resolution applied to the observational data.

In order to use WRST to the analysis of EASs arriving from different celestial regions, we have to compare the rank sum of x_m in the series of events. Due to the limited number of events available at the highest energies of interest, we cannot confine the showers in a narrow bin. Instead, the paired samples can be used in which every EAS from one sample has its counterpart in another sample with the same or close energy. In this case the two samples have the same $\overline{x_m}$ if the original mass compositions are identical, in spite of energy dependent $\overline{A(E)}$. On the other hand, if there is a difference in average mass of original populations, then Δx_m should exist and is approximated by (2).

With known resolving power of WRST (Fig. 1) and D_{ER} we have $\overline{\Delta \log A}$ for the observed number of showers. In Table 2, the results for the PAO FD data are shown. The number of events are sufficient only at the threshold energy 1 EeV. To increase the number of events under analysis, the data of the surface detectors (SD) can be used. This possibility will be discussed in the next section. Obviously, future arrays with rather large aperture will be more convenient for the WRST.

2.3 A method based on the age variance of air showers initiated by different primary particles

There are some shower parameters measured by the SDs of the EAS arrays sensitive to the nuclear composition of the primary particle: muon content, shower front curvature, etc. It seems that one of the

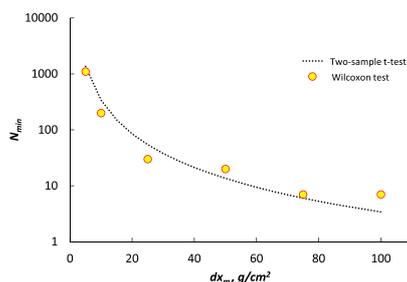


Figure 1. Comparison of the resolving power of Wilcoxon and Student tests. Minimal sample size, N_{min} , required to resolve a difference, dx_m , in the two EAS samples.

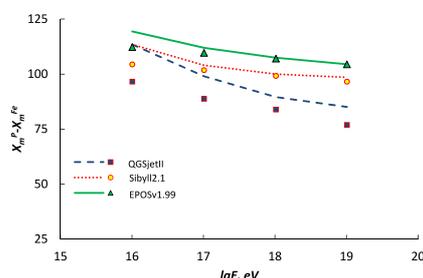


Figure 2. Difference in x_m of showers initiated by hydrogen and iron nuclei. CORSIKA simulation results [14] are shown by curves. Points are for calculation results with equation (2).

most appropriate parameters is the shower age. It is explicitly related to x_m^3 , hence, to the primary mass, and can be estimated in each shower using the universal relation with lateral distribution parameters measured by the SDs [8]. Only electromagnetic component detectors are needed in the surface stations in this case, so the method can be applied to arrays with no muonic and other component detectors.

As in the previous section, having two paired samples of EAS ages one can apply WRST in order to decide: is there an appreciable difference in the mean age of the showers in samples S and T or these samples are drawn from the same distribution (the null hypothesis)? To do so, one has to compare the ranks in samples of ages. Consulting with statistical tables about the deviation of statistic from the expected value, one can accept or reject the null hypothesis at the significance level specified.

Using the age variance instead of Δx_m in EASs initiated by different primary particles one has an additional variable to fix – zenith angle. So, the paired samples should be selected with close or equal energies and zenith angles. This can be done as follows. Initially, one has two subsets of EASs arrived from different sources, and presumably, with different masses. For each shower from the minor subset a counterpart shower should be found in another subset closest in θ and E , then ages of these EASs should be collected in corresponding samples S and T . Original showers should be removed from subsets. Repeating operations until the end of the minor subset, one assembles a pair of samples of size N congruous, in average, at $N \gg 1$, to the circumstances of zenith angle and energy.

A minimal mass difference resolvable by WRST applied to age samples is limited by uncertainty in age estimation and zenith angle. Neglecting the uncertainty of θ measurement in PAO data ($d \cos \theta / \cos \theta < 0.01$), and taking the number of EAS detected in the period between 1.01.2004 and 31.12.2010 [2], we have the estimation of the minimal difference in average mass resolvable by WRST applied to PAO SD data (Table 3).

³ For the fixed zenith angle.

Table 3. Estimation of the minimal difference in average mass, $\overline{\Delta \log A}$, of UHECRs observed by PAO surface detectors, resolvable by WRST.

E_{thr} , EeV	N	QGSjetII	Sibyll2.1	EPOSv1.99
3	64000	0.01	0.01	0.01
10	5000	0.07	0.05	0.04
25	600	0.19	0.14	0.13

Comparison of Tables 2 and 3 shows that the number of events detected with PAO SDs is sufficient to apply WRST to the age variance of EAS primaries in the energy range above 10 EeV, and may be applicable at energies above 25 EeV if there is an explicit difference in masses of UHECRs. Concerning future arrays, large apertures, and additionally, presence of the surface detectors would be sufficient conditions of applicability of the method.

3. APPLICATION OF THE METHOD TO ANALYSIS OF THE YAKUTSK ARRAY DATA

The Yakutsk array is detecting EAS of cosmic rays in the energy interval from 1 PeV to 100 EeV. The array is located at 61.7° N, 129.4° E, 105 m above sea level (1020 gcm⁻²). It consists of 71 surface and 6 underground scintillation detectors of charged particles (electrons and muons), 49 detectors of the air Cherenkov light. Detailed description of the array and results obtained can be found in [3, 15, 16].

An air shower cascading higher in the atmosphere (“old” shower, $\tau > 1$) has a broad and flat lateral distribution of secondary particles at the observational level while “young” one ($\tau < 1$) has steep LDF. Previously, it was shown that the LDF parameters (e.g. slope of the Cherenkov light and charged particles lateral distribution, RMS radius) can be used as the indicators of the shower age [8].

In this work, the LDF slope of charged particles, η , detected with scintillators of the Yakutsk array is used to estimate the shower age in the given energy and zenith angle intervals. The data set used to analyze the slope parameter consists of EAS events collected during the period 1974 to 2004, with a total of 19407 showers selected with energies from 1 to 100 EeV, zenith angles $\theta < 50^\circ$ and axes within the array area. Inclined events beyond 50° are rejected because of substantial fraction of muons in the distribution of charged particles measured. In order to estimate LDF slope of each shower in a set, additional selection criteria were applied: i) at least 4 stations in the core distance interval $r \in (200, 1000)$ m should have particle density above a threshold; ii) the slope calculated using the least square method should be in the interval $\eta \in (-8, 0)$.

The same procedure as in section 2.3 is used to apply Wilcoxon test to paired samples, except for the shower age replaced by the LDF slope here. Sample S consists of showers with arrival directions in supergalactic (SG) “pancake”, or “dumbbell” structure around the SG plane [17]. Sample T consists of all other showers. A hypothesis tested is that UHECR sources in SG pancake emit nuclei, while from other (distant) sources protons arrive. Our task is to ascertain – is there an appreciable difference in average mass of UHECRs in two samples?

In order to reveal the reliability bounds of WRST in this particular case, we have applied the procedure to a pair of samples of EAS events selected in the narrow zenith angle and primary energy intervals ($\Delta\theta = 1^\circ$; $\Delta E = 1$ EeV). Then, a rank sum of LDF slopes in $2N$ shower set is used to distinguish samples at the significance level P_{crit} . Varying a distance in zenith angle or the primary energy between samples we can set a lower limit to the average slope difference resolvable by the WRST. In Fig. 3 this limit is shown as a function of the sample size. Experimental errors in charged particle densities measured by scintillators result in uncertainties of slope differences resolved (shown by vertical bars).

While in Optical Redshift Survey and IRAS 1.2-Jy redshift survey data the density contrast in the local galaxy density field is aligned along SG axes with radius $40 h^{-1}$ Mpc and thickness of

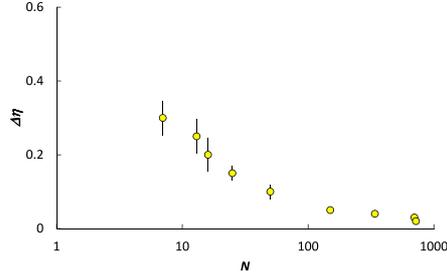


Figure 3. Resolving power of WRST applied to paired LDF slope samples. The minimal difference in slope, $\Delta\eta$, resolvable ($P_{crit} = 0.01$) with given sample size, N .

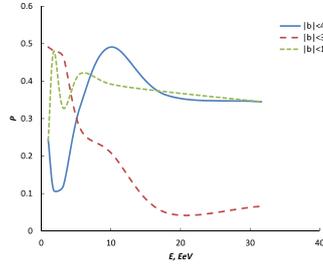


Figure 4. The probability of two LDF slope samples to be drawn from the same distribution.

Table 4. Estimation of the minimal difference in parameters of EASs detected with the Yakutsk array, resolvable by WRST.

E_{thr}, EeV	N	$\Delta\eta$	$\Delta\tau$	$\text{rms}(\tau)$	$\Delta \log A$
1.0	9180	0.007	0.008	0.001	0.07
3.2	1245	0.019	0.021	0.002	0.19
10.0	114	0.062	0.069	0.006	0.63
17.8	43	0.100	0.111	0.022	1.02

$20 h^{-1} \text{Mpc}$ [17], we have used three variants of the pancake angular boundaries: $b_{SG} < 15^\circ, 30^\circ, 45^\circ$ in order not to miss the possible correlation of UHECRs with the SG plane proposed by [18].

The results are shown in Fig. 4. In all cases we cannot reject the null hypothesis – there is no appreciable difference in LDF slopes of EAS samples. Only at $E \sim 20 \text{EeV}$, $b_{SG} < 30^\circ$ there is the minimum of a probability, but it is greater than P_{crit} .

In Table 4, the minimal differences resolvable by WRST are given: in the LDF slope, $\Delta\eta$, in the shower age, $\Delta\tau$, with estimated errors, $\text{rms}(\tau)$, and in average mass, $\Delta \log A$, for the sample sizes depending on the threshold energy. Simulation results with the Sibyll2.1 model are applied in this case.

With the values given in the last column of Table 4, basing on the Yakutsk array measurements, we can set an upper limits to the average mass of UHECRs⁴ arriving from the SG pancake: $\bar{A} < 2$ at $E_{thr} = 1 \text{EeV}$ and $E_{thr} = 3.2 \text{EeV}$, $\bar{A} < 5$ at $E_{thr} = 10 \text{EeV}$, and $\bar{A} < 11$ at $E_{thr} = 17.8 \text{EeV}$.

⁴ Within the framework of the hypothesis accepted.

4. CONCLUSIONS

A new method is developed combining analysis of arrival directions with the statistical test of the pairs of EAS samples. It can be applied in the case if the sky regions are separated where CR mass compositions are different.

The method is based on a non-parametric statistical test – Wilcoxon rank sum test, which does not depend on the populations fitting any parameterized distributions. Two algorithms are proposed: first, using measurements of the depth of EAS maximum position in the atmosphere, and second, relying on the age variance of air showers initiated by different primary particles.

The resolving power of Wilcoxon test is estimated with the distributions of shower maximum and the slope of lateral distribution in EAS. It is shown that the resolving power of Wilcoxon and Student tests are equal in the case of normal distribution of the variable.

A trial run of the test is performed with the Yakutsk array data in order to reveal a possible difference in the age of air showers arriving from the supergalactic pancake and a complementary region. No significant difference is found in the LDF slope parameter of the two subsets of EAS events detected with different threshold energies. The lower limits are set to the LDF slope difference resolvable by the method with the data sample available.

Using the linear correlation between the LDF slope and the shower age, and model simulation results of EAS initiated by different primary particles, upper limits are derived to the possible difference in the mass of UHECRs arriving from the vicinity of the supergalactic plane and the background.

The method developed is quite general and can be applied at other energies and to the various EAS parameters measured at different observatories.

The author is grateful to the Yakutsk array staff for data acquisition and valuable discussions. The Yakutsk array experiment is funded by the Russian Ministry of Education and Science and by RAS; this work is partially supported by RFBR (grants 11-02-00158, 11-02-12193, 12-02-10005).

References

- [1] R.U. Abbasi et al., 2007, Phys. Rev. Lett. (submitted)
- [2] F. Salamida et al., Proc. 32th ICRC, Beijing 1107.4809 (2011)
- [3] V.P. Egorova et al., Nucl. Phys. B (Proc. Suppl.) **136** (2004) 3
- [4] J.W. Cronin, Nucl. Phys. B (Proc. Suppl.) **138** (2005) 465
- [5] The Pierre Auger Collaboration, Science **318** (2007) 938
- [6] P.F. San Luis et al., Proc. 32th ICRC, Beijing (2011)
- [7] R.U. Abbasi et al., Phys. Rev. Lett. **104** (2010) 161101
- [8] A.A. Ivanov, M.I. Pravdin and A.V. Sabourov, Int. J. Mod. Phys. D **20** (2011) 1539
- [9] E.G. Berezhko, Astrophys. J. **698** (2009) L138
- [10] H.B.J. Koers and P. Tinyakov, JCAP **0904** (2009) 003
- [11] Y. Tsunesada et al., Proc. 32th ICRC, Beijing (2011)
- [12] J. Linsley, Proc. 15th ICRC, Plovdiv **12** (1977) 56
- [13] D. Heck et al., *Report FZKA* (Forschungszentrum, Karlsruhe, 1998) 6019
- [14] T. Bergman et al., Astropart. Phys. **26** (2007) 420
- [15] A.A. Ivanov, S.P. Knurenko and I.Ye. Slepsov, New Journ. Phys. **11** (2009) 065008
- [16] A.A. Ivanov, Astrophys. Journ. **712** (2010) 746
- [17] O. Lahav et al., Mon. Not. R. Astron. Soc. **312** (2000) 166
- [18] T. Stanev, astro-ph/0805.1746 (2008)