FERMILAB-PUB-07-467-A

Clustering properties of ultrahigh energy cosmic rays and the search for their astrophysical sources

A. Cuoco,[1] S. Hannestad,[1] T. Haugbølle,[1] M. Kachelrieß,[2] and P. D. Serpico[3]

[1] *Department of Physics and Astronomy, University of Aarhus, Ny Munkegade, DK–8000 Aarhus, Denmark*
[2] *Institutt for fysikk, NTNU, N–7491 Trondheim, Norway*
[3] *Center for Particle Astrophysics, Fermi National Accelerator Laboratory, Batavia, IL 60510-0500, USA*

(Dated: September 18, 2007)

The arrival directions of ultrahigh energy cosmic rays (UHECRs) may show anisotropies on all scales, from just above the experimental angular resolution up to medium scales and dipole anisotropies. We find that a global comparison of the two-point auto-correlation function of the data with the one of catalogues of potential sources is a powerful diagnostic tool. In particular, this method is far less sensitive to unknown deflections in magnetic fields than cross-correlation studies while keeping a strong discrimination power among source candidates. We illustrate these advantages by considering ordinary galaxies, gamma ray bursts and active galactic nuclei as possible sources. Already the sparse publicly available data suggest that the sources of UHECRs may be a strongly clustered sub-sample of galaxies or of active galactic nuclei. We present forecasts for various cases of source distributions which can be checked soon by the Pierre Auger Observatory.

PACS numbers: 98.70.Sa, 98.54.Cm FERMILAB-PUB-07-467-A

I. INTRODUCTION

The identification of the sources of ultrahigh energy cosmic rays (UHECRs) and, more generally, the question if astronomy with charged particles is possible are two important unresolved problems of astroparticle physics. The answer to the latter question depends both on the magnitude of deflections in magnetic fields (which in turn depends also on the chemical composition of UHECR primaries) and on the number density and the luminosity of UHECRs sources. Consensus has not yet emerged on the origin and the amplification mechanisms of primordial magnetic fields, nor on the present magnitude and structure of extragalactic magnetic fields outside of galaxy cluster cores [1, 2]. Uncertainties from modeling strong interactions prevent a clean determination of the fraction of heavy nuclei in UHECRs above $E \gtrsim 10^{18}\,\mathrm{eV}$ [3]. Thus theoretical predictions about the chances of charged particle astronomy differ drastically, and the answer has to come from experiment.

There are various pieces of evidence in the available experimental data. The AGASA data contain several small-scale clusters, i.e. clusters of events within its experimental angular resolution [4]. This result triggered a series of works studying the auto-correlation of UHECR data at small angular scales [5, 6] or correlating UHECRs with potential astrophysical sources [7]. For instance, the best-fit value for the density n_s of UHECR sources found in Ref. [6] is $(1-3)\times10^{-5}/\mathrm{Mpc}^3$, while the 2σ confidence region ranges from $2\times10^{-6}/\mathrm{Mpc}^3$ to $\sim 10^{-2}/\mathrm{Mpc}^3$, i.e. up to the density of galaxies. The large statistical error of this estimate comes mainly from the small number of doublets with less than 3–5 degrees separation, while deflections in magnetic fields of more than a few degrees would result in a systematic overestimation of n_s. Correlation analyses of the small UHECR data set have their own problems: In order to avoid a too large number of potential sources per angular search bin, one has to choose either a very high energy cut or a very specific test sample, e.g. a small subset of all Active Galactic Nuclei (AGN). Although some studies found significant correlations, in particular with BL Lacs, these results have remained controversial.

A second piece of evidence are anisotropies on medium scales. The authors of Ref. [8] analyzed the available data set of CR arrival directions from the HiRes stereo, AGASA, Yakutsk and SUGAR experiments with energies $E \geq 4\times10^{19}$ eV in the HiRes energy scale. They found evidence at $\sim 3\sigma$ C.L. for anisotropies on the scales of 10–35 degrees, with a clear minimum of the chance probability around 20–30 degrees. This result is consistent with the theoretical expectations for anisotropies associated with large-scale structures (LSS) from Ref. [9]. Further studies showed that the correlations are best explained if UHECR sources are either over-biased with respect to normal galaxies and/or if the cosmic ray horizon is smaller than expected for rectilinearly propagating protons [10].

Intriguingly, similar findings seem to emerge from an analysis of the preliminary data from the Pierre Auger Observatory (PAO). For 64 events with $E > 4\times10^{19}$ eV the data presented in Ref. [11] show a surplus of clustering in the broad range from 7 to 30 degrees. The distribution has its minimum at 7 degrees with a second, broad minimum between 19–24 degrees and is quite similar to the distribution with 57 events in Ref. [8]. Remarkably, the PAO data contain also "8 doublets separated by less than 7 degrees in the 19 highest energy events", i.e. with $E > 5.75\times10^{19}$ eV [11].

Finally, the last piece of evidence comes from the UHECRs energy spectrum. In particular, the long controversy about the continuation of the spectrum [12] beyond the Greisen-Zatsepin-Kuzmin (GZK) cutoff [13, 14] seems to be finally solved by the latest data from HiRes [15] and

the PAO [16] that both detect with an high confidence level ($> 5\sigma$) a prominent steepening in the spectrum compatible with the GZK attenuation. Complemented by the new PAO stringent limit on the fraction of UHECRs photon primaries [17] the data are clearly pointing toward a "standard" scenario in which the bulk of UHECRs sources have an astrophysical origin in the nearby universe, with more exotic top-down scenarios playing at most a sub-dominant role. This evidence makes timely a detailed study of possible UHECRs astrophysical sources of the kind addressed in the following.

The main aim of the present work is to compare the auto-correlation function of potential UHECR sources with these early results of the PAO on UHECR arrival directions, and to provide forecasts of the clustering expected for different classes of sources which can be checked shortly by the PAO. We shall also comment on how the clustering features of the public available world data-set compare with expectations. The timeliness of this analysis is due to the fact that, while increasing evidence is accumulating in favor of an astrophysical origin of UHECRs, it is still unclear how accurately one can identify the sources of UHECRs and what are the best tools to do so. Here we advocate the importance of a global comparison, i.e. a comparison on all angular scales, of the observed auto-correlation function of arrival directions with the expectations for different source and primary scenarios. At first glance, cross-correlation tests with source catalogues might appear as the ideal tool to identify the UHECR sources, but if used alone they could be insufficient or misleading. First, the angular resolution of $\mathcal{O}(1^\circ)$ of UHECR observatories is poor for astronomical standards. Additionally, UHECRs are plagued by non-negligible magnetic field deflections, except maybe at the highest energies observed and for proton primaries. Non-spurious signals in a cross-correlation analysis can only be expected if the *overall* magnetic deflection of UHECRs is below the size of the angular bin used. Small-scale auto-correlation studies are less sensitive to magnetic fields since only the relative deflections between pairs of events enter. Yet, since the overall spreading induced by the magnetic fields is unknown, an auto-correlation analysis limited to the first angular bin (whose size is chosen a priori, e.g. motivated by the angular resolution of the observatory) may be unsuccessful and/or have an ambiguous interpretation. By definition, however, if UHECR astronomy is possible at all—at least in a statistical sense—sufficiently large angular scales in the auto-correlation function should reflect the analogous properties of the sources. As a more ambitious goal, a global analysis also offers the possibility to infer the average size of deflections at the chosen energy scale, and thus to perform a kind of magnetic field reconstruction.

For this line of reasoning to be effective, one has to show first that the auto-correlation functions of different potential UHECR sources differ significantly and thus might be used to identify the UHECR sources. This task is addressed in Sec. II, where we discuss several astronomical catalogues. For the study of UHECR anisotropies, catalogues should cover the largest possible fraction of the sky, ideally complete in distance up to redshift $z \sim 0.1$. We shall see that this is rarely the case at present, which forces us to limit most of our quantitative comparisons to the sample of UHECRs with the highest energies: As a rule of thumb, the higher the UHECR energy, the smaller the energy-loss horizon (and the magnetic deflections), and the more reliable the catalogue. Turning this into quantitative statements requires however some assumptions on the nature of primary particle and the absolute energy scale of experiments. In Sec. III, we perform a comparison with the PAO results under the hypothesis of proton primaries and using two different assumptions on the energy scale. There, we also comment on the the world data-set of available data from the HiRes, AGASA, Yakutsk, and SUGAR experiments. Finally, we discuss our results and conclude in Sec. IV.

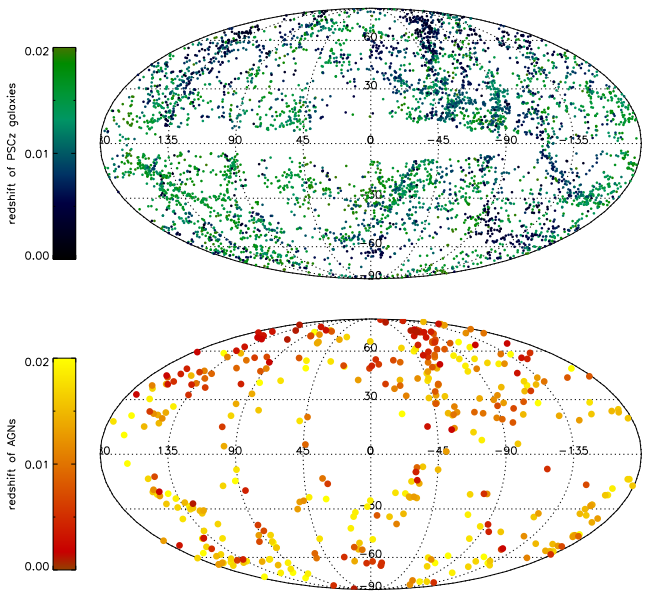

FIG. 1: Top panel: PSCz galaxies within $z = 0.02$ color coded in black to green according to increasing redshift. Bottom panel: AGNs within $z = 0.02$ color coded in red to yellow according to increasing redshift. Both panels are in Galactic Coordinates.

II. GALAXY AND AGN CORRELATION FUNCTIONS

A. Astronomical catalogues

Among the astrophysical objects most often proposed as UHECR sources are AGNs in general, specific subclasses like Blazars, Radio or Seyfert galaxies, Gamma Ray Bursts (GRB) or young neutron stars (for a review see e.g. [18]). All these sources follow the LSS of matter,

although with different and scale dependent biases. In order to understand how large this bias is for different source classes, we examine now the clustering properties of normal galaxies (which may host candidates like neutron stars) and AGNs in the nearby universe. We use the PSCz catalogue [19] as a sample of the galaxy distribution and the 12th edition of the Véron-Cetty & Véron (VCV) catalogue [20] for the AGNs. We also study the clustering properties of several sub-samples, imposing cuts in absolute magnitude for the galaxy catalogue and subdividing the AGN catalogue into Seyfert galaxies of type 1 (S1), type 2 (S2) and LINERs (S3), according to the classification reported in VCV catalogue itself. In Fig. 1 the various kinds of AGNs and the PSCz galaxies within $z = 0.02$ are shown in Galactic Coordinates, color coded according to their redshift. The empty region along the Galactic Plane is the so-called avoidance region due the presence of the Milky Way and does not reflect an intrinsic lack of objects.

The details of the PSCz catalogue, in particular a description of the mask and of the selection function, are summarized in Ref. [19]. The B-band magnitudes reported in the catalogue are, however, biased and show systematic offsets within different regions of the sky where the galaxy magnitudes have been taken from different catalogues with different calibrations. To overcome this problem, we match sources in the PSCz catalogue with sources in the 2MASS extended source catalogue [21], to get accurate magnitudes in the infrared ($2.15\,\mu$m) K-band. This is done by requiring that a PSCz galaxy is inside the 20 mag arcsec^{-2} isophote in the K-band of the matching 2MASS galaxy. We find that $\sim 80\%$ of the galaxies in the PSCz catalogue have a counterpart in the 2MASS XSC and discard the others. We then construct various sub-samples of the catalogue performing cuts in absolute magnitude using the distance modulus relation $M = m - 5\log d_{L,\mathrm{Mpc}} - 25 \sim m - 43.16 + 5\log h - 5\log z$ (where h is the reduced hubble parameter and the redshift dependent K-correction, negligible for $z < 0.03 - 0.04$, has not been considered). For these sub-samples we also empirically build new selection functions using a smooth weight function chosen in order to reproduce the redshift distribution of the subsample. In the top panel of Fig. 2 we show the fraction $f(z)$ of galaxies from the PSCz catalogue for the subsamples obtained with luminosity cuts $M < -24, -24.5$ in redshift bins of width 0.005. In the same panel we also show the complete galaxy sample with no cut imposed. The PSCz catalogue is flux complete and from the figure it can be seen that the brightest sub-samples are essentially also volume complete out to $z \sim 0.02$. In contrast, the full sample shows prominent signatures of volume-incompleteness already at very low redshift.

Differently from the PSCz catalogue, the VCV catalogue is a compilation of observations and is known to suffer increasing incompleteness with increasing redshift. We assume that at least in the very nearby universe (i.e. $z \lesssim 0.02$) the catalogue can be considered fairly complete and we build selection functions analogously to the PSCz sub-samples described above. It will later be seen that our results are quite robust against this assumption. We find that: (i) within $z = 0.02$ the AGN sample follow closely the matter distribution, as expected (see Fig.1); (ii) although the overall catalogue is fairly complete within this distance, there are some subsets of AGNs which suffer from significant incompleteness and therefore selection bias. In the bottom panel of Fig. 2 we show the fraction of S1, S2, and S3 galaxies from the VCV catalogue in redshift bins out to $z = 0.03$, together with the Poisson fluctuation due to the finiteness of the sample (the sample of active galaxies within $z = 0.02$ includes ~ 500 AGNs of which ~ 150 S1, ~ 200 S2, and ~ 80 S3). The S1 and, to a minor extent, the S2 galaxies show a behavior which is reasonably close to the expected z^2 increase of a truly complete sample. Since they constitute by far the largest fraction of AGNs in the VCV catalogue, the catalogue as a whole can be regarded as at least rea-

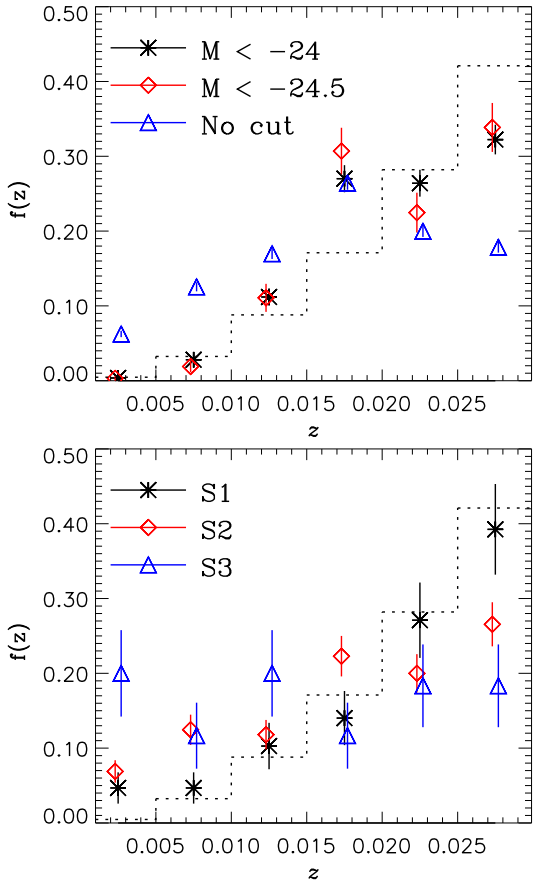

FIG. 2: Top Panel: The fraction $f(z)$ of galaxies from the PSCz catalogue is shown for various luminosity cuts $M < -24, -24.5$ and with no cut in redshift bins of width 0.005. Bottom Panel: The fraction $f(z)$ of S1, S2, and S3 AGNs. The dotted histogram shows the expected behavior of a volume-complete catalogue, i.e. with differential fraction $\propto z^2$. The error bars are the Poisson fluctuations from the number count.

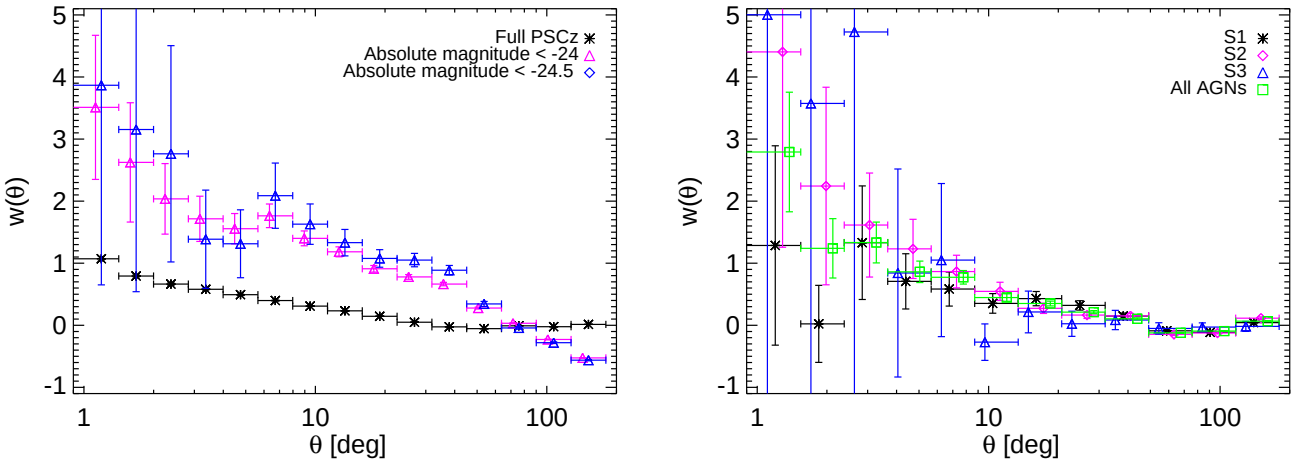

FIG. 3: The auto-correlation function $w(\vartheta)$ for galaxies and for differents cuts in magnitude (left) and AGNs and different subsamples (right) as function of ϑ with $1\,\sigma$ error bars.

sonably complete out to $z \sim 0.02$. We note however that the S3 sub-sample seems the most affected by incompleteness effects, a point which is further discussed below. It is difficult to estimate the number density of AGNs in the VCV catalogue given that, by construction, the astrophysical sources are included without any specific selection rule. As a rough estimate, assuming volume-completeness with 500 AGNs within $z = 0.02$ gives $n_s \sim 5 \times 10^{-4}\mathrm{Mpc}^{-3}h^3$. The same density corresponds also to the galaxies brighter than $M_{\rm cut} = -24$, while 50 uniformly distributed sources, always within the same volume, would give $n_s \sim 5 \times 10^{-5}\mathrm{Mpc}^{-3}h^3$. For ordinary galaxies the number density depends strongly on the assumed $M_{\rm cut}$ and typically ranges from $10^{-3}\mathrm{Mpc}^{-3}h^3$ to values as high as $10^{-1}\mathrm{Mpc}^{-3}h^3$ for Milky Way-like galaxies. In any case, the exact value of the number density of the UHECR sources depends on the exact horizon containing the sources and thus on issues like the nature of the primaries and the absolute energy scale. Thus, in the following we will focus the attention mainly on the absolute number of sources (above the assumed CR energy threshold) and to their bias/overdensity with respect the distribution of matter, as principal observables.

Finally, we briefly discuss the case of GRBs as UHECR sources. The observed rate of GRB is $R_{\rm obs} \approx 0.5 \times 10^{-9}/(\mathrm{Mpc}^3\mathrm{yr})$ according to Ref. [22]. However, deflections in the extragalactic magnetic fields (EGMF) lead to time delays τ that in turn increase the effective density of GRBs as $n_s = \tau R_{\rm obs}$. The clustering properties of GRBs are in general quite different from those of AGNs and massive galaxies. Long duration GRBs which make up about 2/3 of all GRBs are associated with supernova events in extremely massive stars and therefore their distribution essentially follows the star formation rate. Star forming galaxies are mainly spirals and irregulars which are less clustered than average galaxies in the PSCz catalogue. The remaining 1/3 of the GRBs which are most likely the result of binary collisions have a distribution which is close to the one of SN Ia's, but are not considered very likely sites for the UHECR acceleration. Thus GRBs cluster less than average PSCz galaxies and, in the following considerations, we shall use randomly distributed sources as a rough template for their clustering properties.

B. Correlation functions

An important point to prove for the following arguments is that different astrophysical catalogues of candidate UHECR sources have sufficiently different clustering properties. To that purpose, we calculate in this section the auto-correlation function $w(\vartheta)$ of the various samples. In the past, a commonly employed estimator for the auto-correlation has been the intuitive $DD/RR - 1$. This estimator is however sub-optimal especially for the estimation of variance [24], while an optimal estimator is given by [23, 24]:

$$w(\vartheta) = \left\langle \frac{DD - 2DR + RR}{RR} \right\rangle , \qquad (1)$$

where D denotes the data-set and R a randomly generated data-set with the same bias characteristics as the data (same mask, same selection function, same exposure, etc.), while the quantities DD, DR, RR are the normalized pair counts in each angular bin around ϑ. The brackets indicate that the final $w(\vartheta)$ is an ensemble average over many random realizations. Note that for data consistent with a random distribution $w(\vartheta)$ is zero within the errors. The resulting auto-correlation functions $w(\vartheta)$ are shown in the left panel of Fig. 3 for galaxies and in

the right panel for AGNs, without weights for the sources (i.e. without selection effects and attenuation) and using the same masking for all sets in order to have an unbiased comparison. The errors in each bin can be estimated as [24]

$$\mathrm{rms}[w(\vartheta)] = [1 + w(\vartheta)] \left[\frac{1}{n(n-1)/2 \, \langle RR \rangle} \right]^{1/2}, \quad (2)$$

where n is the number of points in the data set D, and hence $n(n-1)/2$ the total number of unique pairs.

Both samples show a strong auto-correlation at small scales, although the clustering of normal galaxies is quite less pronounced ($w_{\rm AGN}(1^\circ)/w_{\rm gal}(1^\circ) \simeq 3$). We will see that in the relation to the small scale clustering seen by the PAO, this difference already tightly constrains the possible contribution of normal galaxies as sources of the highest energy CRs. The situation changes when bright sub-samples of the PSCz galaxies are considered whose clustering properties more nearly resemble those of the AGNs. This is not surprising given that most of the brightest galaxies are in fact AGNs, and the two samples thus are not truly independent. Regarding the AGNs sub-samples it can be seen that the clustering of S1 objects shows no strong differences to the one of all AGNs; by contrast the S2 and S3 subtypes show a stronger auto-correlation on the smallest scales, $\vartheta \lesssim 3^\circ$. Note, however, that the AGN samples, having a smaller number of objects, have in general also larger error bars; in this case, since Poisson statistics makes the errors on $w(\vartheta)$ decrease for increasing ϑ, intermediate scales $\vartheta \sim 10$–30° might be optimal to distinguish between different sources. This is especially true for UHECRs when the statistics is very limited and/or the smearing at the smallest scale by magnetic fields are important.

The above results are in general quite in good agreement with other more detailed studies of the AGNs bias properties. In particular, the clustering properties of AGNs have been studied extensively for example in Kaufmann *et al.* [25] using the SDSS catalogue. Their findings are that AGNs are far more common in massive galaxies and that the AGN correlation function resembles that of massive early-type galaxies, which is similar to what we find for the low-redshift VCV sample.

III. COMPARISON WITH THE DATA AND FORECAST FOR AUGER

We turn now to study the clustering of the various source samples considered at rather small scales, comparing them with the existing observations. In particular, we shall focus on one of the most remarkable findings reported by the PAO, namely that of "8 doublets within 7 degrees out of the 19 highest energy events" [11].

Because of the limited UHECR statistics available, and to have a direct comparison with the Auger findings, in this section we shall use as main observable $\mathcal{C}(\vartheta)$ a slightly modified version of the function $w(\vartheta)$ of the previous section, defined as:

$$\mathcal{C}(\vartheta) = \left\langle \sum_{i=2}^{N} \sum_{j=1}^{i-1} \Theta(\vartheta - \vartheta_{ij}), \right\rangle , \quad (3)$$

i.e. the cumulative number of pairs within the angular distance ϑ, where Θ is the step function (with $\Theta(0) = 1$), N the number of CRs considered, and ϑ_{ij} is the angular distance between events i and j. Although $\mathcal{C}(\vartheta)$ introduces further correlations between different angular bins, the use of cumulative countings instead of differential ones has the great advantage of sensibly reducing the dependence from unknown magnetic field deflections, a crucial point for UHECRs astronomy. The ensemble average is performed over a large number $M = 10^5$ of Monte Carlo sets. The events are extracted randomly from the catalogue under consideration and we take into account the PAO exposure as described in Ref. [26] assuming as characteristic parameters for the PAO $\zeta_{\rm max} = 60^\circ$ as maximal zenith angle for a CR event and $\delta_{\rm PAO} = -35^\circ$ for the PAO latitude location. The selection effects of the catalogue and the attenuation due to CR propagation are included assigning proper weights for the sources, which in turn are used as emission probabilities in the simulation. Here a hypothesis on the nature of the particles and on the overall energy scale enters. To illustrate this point, in Table I we report the distance $D_{1/2}$ from which 50% of the UHECR flux comes, for different assumptions about energy and chemical composition, assuming uniformly distributed sources and rectilinear propagation (see e.g. [27]). Note that the injection spectral index has only a minor effect on $D_{1/2}$ in the energy range considered.

In the following we assume proton primaries and use the propagation window function $W(z, E_{\rm cut})$ as calculated in Ref. [9]. Given the importance of the clustering signal observed by the PAO, we shall focus mostly on the case of $N = 19$ events. In order to study the sensitivity to the assumed energy scale, we shall consider two cases: (i) the preliminary calibration of the energy scale presented by the PAO is correct, and thus the 19 highest

Species	$E_{\rm cut}/10^{19}$ eV	$D_{1/2}$/Mpc	$z_{1/2}$
p	5.0	160	0.037
p	6.0	100	0.023
p	8.0	40	0.009
^{28}Si	6.0	30	0.007
^{56}Fe	6.0	80	0.019
^{56}Fe	8.0	45	0.011

TABLE I: The distance $D_{1/2}$ (or equivalently redshift $z_{1/2}$) within which 50% of the UHECR flux above $E_{\rm cut}$ comes for different assumptions on energy and chemical composition, assuming isotropic and uniform sources and rectilinear propagation. Adapted from plots in [27].

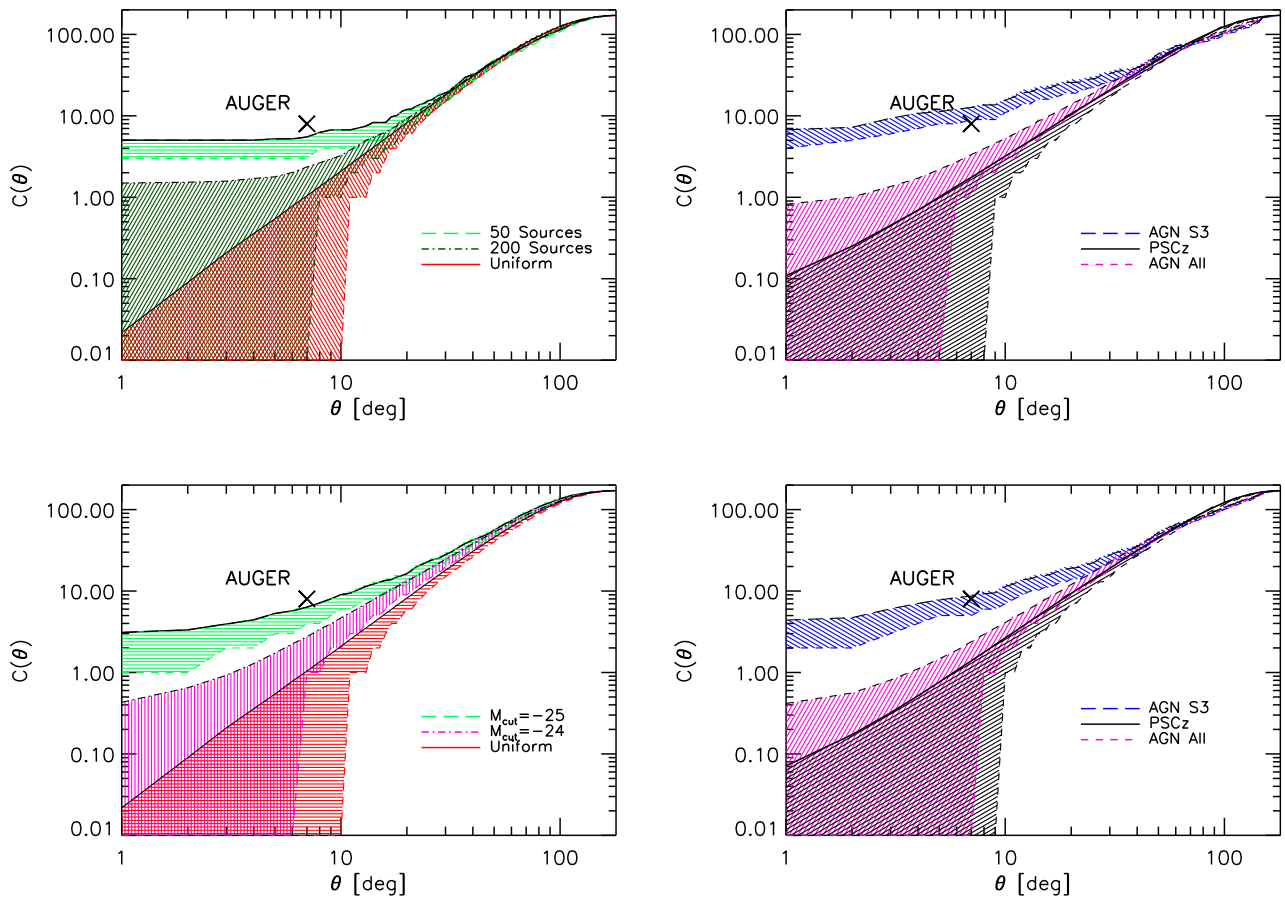

FIG. 4: The average $\mathcal{C}(\vartheta)$ and its 1 σ variation for the case of 19 events and: finite number of sources isotropically distributed compared with the continuous limit (top-left panel); Galaxies with different cuts in magnitude (bottom-left panel); AGN subclasses compared with Galaxy distribution (top-right panel); all the previous cases assume a cut in the window function at $E_{\rm cut} = 8 \times 10^{19}$ eV. The bottom-right panel is the same of the third one, but for $E_{\rm cut} = 5.75 \times 10^{19}$ eV. Notice that the error regions are highly asymmetric for $\vartheta \leq 40^\circ$ (see Figs. 5 -7) and the upper 1σ error regions shown in the plots almost coincide with the mean.

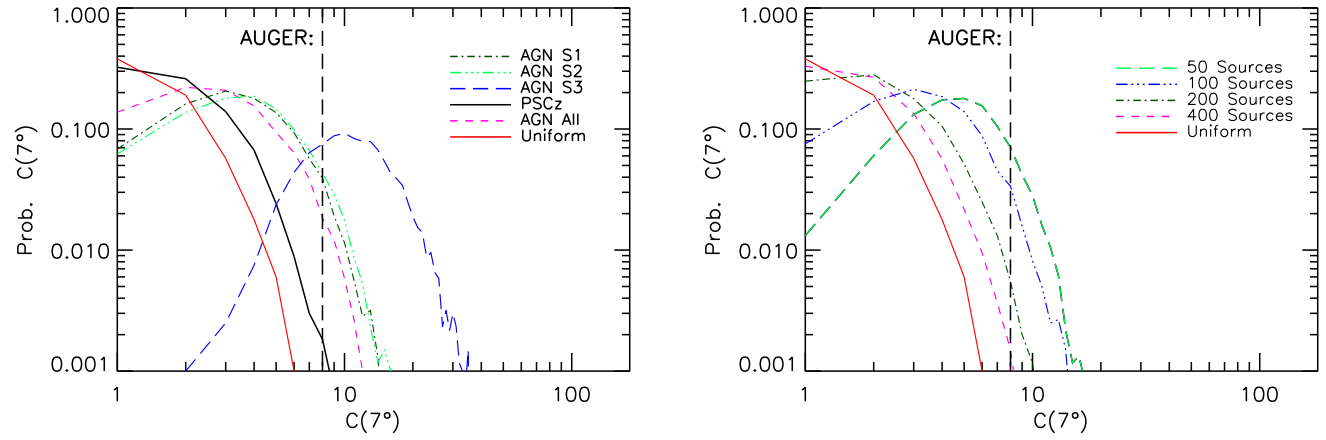

FIG. 5: Probability distribution of $\mathcal{C}(7^\circ)$, for the case of 19 events, an energy cut of $E_{\rm cut} = 8 \times 10^{19}$ eV and the case of different astrophysical models considered (left panel) and a finite number of uniformly distributed sources (right panel).

energy events have energies above $E_{\rm cut} = 5.75 \times 10^{19}$ eV; (ii) UHE air-shower experiments are affected by an over-

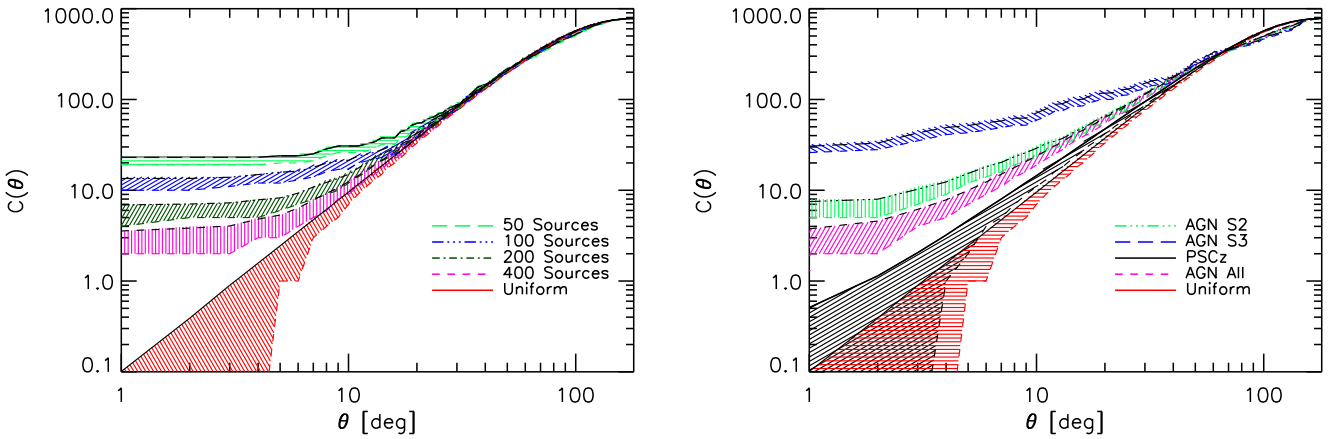

FIG. 6: As in top panels of Fig. 4, but for a statistics of 40 events.

all uncertainty in their energy calibration, whose normalization might be obtained however by requiring that they reproduce correctly spectral features of a model, in our case the dip model [28]. The correction factor found with this method is 1.4 for the PAO energy scale, in which case the highest energy events have energies above $E_{\rm cut} \simeq 8 \times 10^{19}\,{\rm eV}$.

For the galaxy samples we use the PSCz selection function, while for the AGN samples we adopt the approximation $\psi(z) = z^\alpha/z^2$ with $\alpha \simeq 0.4$ as the best-fit slope to the redshift distribution of the samples (see Fig. 2). Although rather crude, this simple approach can be justified by the fact that, especially at relatively small angular scales, the clustering properties are only slightly affected by the exact choice of the selection and propagation weights. As a check, we verified that even the extreme choice of neglecting all selection and propagation effects does not change appreciably the expected mean and distribution of the number of pairs. This approximation breaks down when the UHECR horizon becomes much larger than the distance up to which a catalogue is complete, a point we shall come back to later.

Our results for the average $\mathcal{C}(\vartheta)$ and their 1 σ variations for the case of 19 events are shown in Fig. 4. From top-left to bottom-right, we present the case of a finite number of sources uniformly distributed together with the continuous limit; AGN sub-classes compared with galaxy distributions and an isotropic sky; and galaxies with different cuts in magnitude. While the first three panels assume $E_{\rm cut} = 8 \times 10^{19}\,{\rm eV}$, the bottom right panel is the same as the second one, but for $E_{\rm cut} = 5.75 \times 10^{19}\,{\rm eV}$. We note that the strong clustering observed by the PAO is quite exceptional, and both a uniform random distribution (corresponding to the limit of an infinite number of sources) and the galaxy distribution predict in general a too small number of pairs within 7°. Active galactic nuclei and in particular their sub-samples are much more likely to produce the degree of clustering observed by the PAO. The same happens for the sub-samples of bright galaxies where the brightest galaxies (and thus the set with the smallest number density n_s) provide the best match with the expected clustering. In case of a lower energy scale, the horizon is larger, the sky more isotropic, and especially the LSS and the isotropic sky hypothesis have even more trouble in explaining the observations. Notice however that the S3 sample, which seems the AGN sub-sample most consistent with the high number of pairs, may suffer from a strong selection bias in the VCV catalogue: LINERs are comparatively weak AGNs, which are preferentially detected at low z (see Fig. 2). An additional problem is that most LINERs are outside the field of view of the PAO, and since their total number is much smaller than the one of S1 and S2 AGNs, cosmic variance plays a significant role (as for any other sample made of a small number of objects). Although not manifest from the plots of Fig. 4, another caveat is that, apart from the isotropic case with an infinite number of sources, virtually all models are consistent with the observations at the 3 σ level. The exact confidence levels are illustrated in Fig. 5, where the full distribution within 7° for various model are compared to the Auger result. Also, we had to concentrate on the largest fluctuation in the PAO data-set, since this is the only presently available information. At different angles, we must expect less significant clustering. That said, it is interesting that, as shown in Fig. 6, with a statistics doubled with respect to the one analyzed in [11] the errors should become sufficiently small to rule out most cases.

Clearly the discrimination power between different source models would be greatly improved, if the expected functions $\mathcal{C}(\vartheta)$ were compared to UHECR data not only at a single angular scale but on a range of values. Already a comparison of the correlations at a second angle may be enough to distinguish among different cases. It is likely that a global comparison (based e.g. on a χ^2 method or a Kolmogorov-Smirnov test) of the correla-

tion functions would provide a powerful diagnostic tool. This is one of the main results of our work and deserves a specific example. For the present purposes, it is sufficient to illustrate this point by studying the distribution of the expected number of pairs within 7° (to stick with the most notable finding of the PAO) and 30° (a typical intermediate scale) for several models. In particular, in table II we report the average values $\mathcal{C}(7^\circ)$ and $\mathcal{C}(30^\circ)$ and the 2 σ lower and upper limits—denoted by $\mathcal{C}_-$ and $\mathcal{C}_+$ respectively—for the expected number of pairs in different models and $N = 19$ data. In table III and table IV we report analogous quantites for $N = 40$ and $N = 60$, respectively. We consider three models: (a) 100 uniformly distributed sources, which mimicks GRBs; (b) the SWIFT subclass of AGNs; (c) the galaxies brighter than $M_{\rm cut} = -24.5$. These have been chosen to be basically consistent with the 7° Auger data. Noticeably, even with the 19 events the models (a) and (b) show significant differences at 30°, a difference that become quite large and easily testable with a modest improve in statistics to 40 events. The latter two models are instead almost degenerate from the point of view of clustering properties, which does not come as a surprise since the two samples have a similar number of objects and most of them fall in both subsamples (i.e., they are not independent). Note further that the distributions are generally quite non-Gaussian with a prominent tail toward an higher number of pairs.

Model	$\mathcal{C}_-(7^\circ)$	$\mathcal{C}(7^\circ)$	$\mathcal{C}_+(7^\circ)$	$\mathcal{C}_-(30^\circ)$	$\mathcal{C}(30^\circ)$	$\mathcal{C}_+(30^\circ)$
100 GRBs	0	3	8	10	17	19
SWIFT AGNs	0	3	9	12	21	35
$M_{-24.5}$ Gals	0	3	9	13	25	39

TABLE II: Observables related to the distribution of the expected number of pairs within 7° and 30° and $N = 19$ events (see text for details on the notation). The different models reported are: 100 GRBs, SWIFT AGNs, Galaxies with $M_{\rm cut} = -24.5$.

Model	$\mathcal{C}_-(7^\circ)$	$\mathcal{C}(7^\circ)$	$\mathcal{C}_+(7^\circ)$	$\mathcal{C}_-(30^\circ)$	$\mathcal{C}(30^\circ)$	$\mathcal{C}_+(30^\circ)$
100 GRBs	10	21	32	68	82	96
SWIFT AGNs	9	18	31	85	105	151
$M_{-24.5}$ Gals	6	13	25	81	110	162

TABLE III: As in Table II, but for $N = 40$ events.

Model	$\mathcal{C}_-(7^\circ)$	$\mathcal{C}(7^\circ)$	$\mathcal{C}_+(7^\circ)$	$\mathcal{C}_-(30^\circ)$	$\mathcal{C}(30^\circ)$	$\mathcal{C}_+(30^\circ)$
100 GRBs	27	40	65	161	194	233
SWIFT AGNs	26	44	67	207	261	320
$M_{-24.5}$ Gals	21	32	52	195	268	330

TABLE IV: As in Table II, but for $N = 60$ events.

As a final comment, we notice that due to the limited information available we have restricted the study to the analysis of cumulative number of pairs. However, the above approach can be easily generalized to higher order statistics, like the cumulative counting of the proper number of doublets, of triplets, etc. [29]. A combined use of these tools will likely provide even a more robust and stringent constrain on the nature and number of UHECR sources.

A. Repeaters vs. small scale clustering

An important issue for the future study of UHECR sources is to disentangle the case where an excess of pairs should be attributed to multiple events from a single point source from the case where the excess is produced by the small-scale correlation of two or more sources, as discussed in the previous sections for AGNs and bright galaxies.

For a large fraction of the models considered above, the predicted clustering is mostly due to the intrinsic correlations of the sources rather than being caused by multiple emissions from single sources[1]. A formal but useful way to illustrate this point is to look at the probability distribution of $\mathcal{C}(0^\circ)$, i.e. the number of pairs for $\vartheta = 0^\circ$ (repeaters). In Fig. 7 we report this distribution for the case of 19 events and an energy cut of $E_{\rm cut} = 8 \times 10^{19}\,{\rm eV}$. We show both the case of a finite number of uniformly distributed sources and the case of different astrophysical models considered. In the former case, there should be less than about 50 UHECR sources within the horizon in order to observe a dominant fraction of repeaters as origin for the doublets.

Of course, deflections in magnetic fields and experimental resolution effects prevent an experimental determination of $\mathcal{C}(0^\circ)$. As an example, we ask if the strong clustering signal in the PAO data coming from pairs within 7° is compatible with repeaters. The following arguments show that, without additional information, both the hypotheses of repeaters and of small-scale clustering are consistent with the findings. The correlation functions $w(\vartheta)$ of the astronomical catalogues previously discussed are peaked at small scales, and actually strongly peaked below one degree for the AGN samples. Yet, this peak in $w(\vartheta)$ will be shifted in the excess signal in $\mathcal{C}(\vartheta)$ to larger scales, first of all because the event numbers scales approximately as $N \propto \vartheta^2$. Moreover, we expect the small-scale signal in the data to be washed away partly because of the angular resolution of the detector (that in the PAO data-set mentioned is $\sim 1^\circ$), and mostly because of deflections in the Galactic magnetic

[1] As a consequence, we note that a strong intrinsic small scale clustering within the experimental angular resolution might prevent an unambiguous identification of the source of a given event even in absence of magnetic deflections.

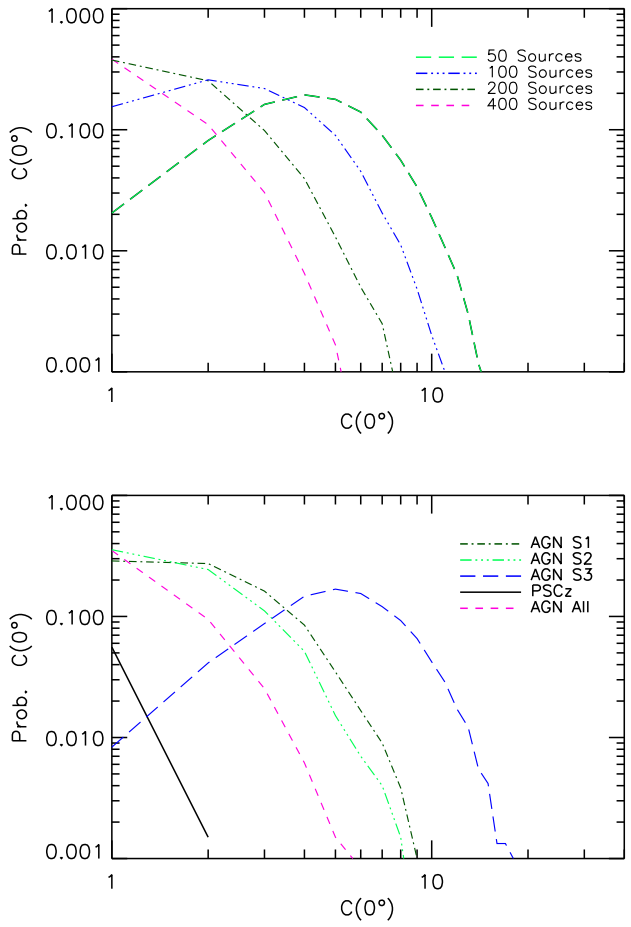

FIG. 7: Probability distribution of $\mathcal{C}(0^\circ)$, for the case of 19 events, an energy cut of $E_{\rm cut} = 8 \times 10^{19}$ eV and a finite number of uniformly distributed sources (top panel) and the case of different astrophysical models considered (bottom panel). Notice that a continuous distribution of uniformly distributed sources predicts $P < 10^{-3}$ for $C(0^\circ) = 1$. The related curve would thus not be visible in the plots.

field (GMF) and possibly extragalactic ones. Even protons of the considered energy are expected to suffer average deflections in the GMF of order $\sim 3^\circ$ [30]. Since the Auger exposure peaks near the Galactic Center region, which in typical GMF models is associated with larger deflections, separations of point-like sources of protons up to 7° by the GMF alone cannot be excluded. Other reasons for the large separation angle are deflections in extragalactic magnetic fields, or an intermediate-heavy chemical composition of UHECRs that would lead to increased deflections in the GMF.

These arguments emphasize once more the importance of a global comparison of the auto-correlation function to perform a robust diagnostics.

B. Comparison with old experiments

Given the importance of a global comparison of the auto-correlation function, one may wonder if the already existing publicly available data offer additional insight. The only sufficiently large data set that is publicly available is the one used for the first time in Ref. [8]. It consists of ~ 100 events with energies $E \geq 4 \times 10^{19}$ eV from the HiRes stereo [31, 32], AGASA [33], Yakutsk [34] and SUGAR [35] experiments. Here, we have rescaled the absolute energy scale of each experiment [8, 36] by requiring that they reproduce correctly the dip spectral feature [28].

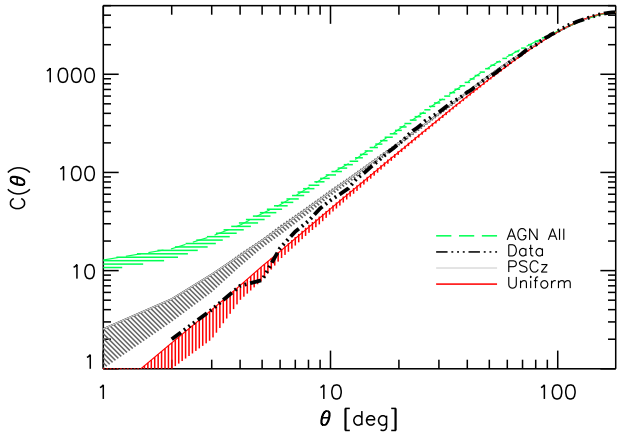

FIG. 8: Cumulative number of pairs and $1\,\sigma$ error regions as in Figs.4, 6 for the set of pre-Auger CR data compared with AGNs and PSCz galaxies within $z = 0.02$, and with a uniform expectation.

Unfortunately, even at energies $E \sim 5 \times 10^{19}$eV, for the case of protons and a rectilinear propagation, sources beyond redshift $z \sim 0.04$ contribute about half of the flux, see Table I. At those distances the catalogues are known to be incomplete, and although we correct for the selection function, we cannot apply the method previously outlined in a reliable, quantitative way. Nonetheless, in Fig. 8 we compare for illustrative purposes the function $\mathcal{C}(\vartheta)$ computed for this data-set with the corresponding expectations from uniformly distributed sources and for AGNs as well as the PSCz Galaxy Catalogue within $z \lesssim 0.02$.

It can be seen that the auto-correlation of the data presents an excess at $10^\circ - 30^\circ$ with respect to a uniform distribution corresponding to the medium scale clustering signal found in [8]. Although a quantitative analysis would likely provide a poor fit, one instead recognizes a qualitative similarity in the pattern of the data function and that of the samples following the LSS, i.e. AGNs and galaxies. (Note, however, that a comparison below $\vartheta \simeq 10^\circ$ is not very meaningful, given the poor angular resolution of SUGAR, for example.) Defi-

nitely, more statistics at higher energy and better quality data are needed from UHECR experiments, while deeper and more complete large-angle surveys would be welcome from the astrophysical community.

IV. DISCUSSION AND CONCLUSIONS

We have examined the clustering properties of ordinary galaxies, GRBs and AGNs with the aim of finding characteristic features which may shed light on them as possible sources of the UHECRs. Our auto-correlation studies have shown that—consistently with what is known from the much larger SDSS galaxy and AGN catalogues, see e.g. [25, 37, 38, 39]—nearby AGNs exhibit much stronger small-scale clustering than average galaxies. The same is true for the brightest galaxies in the PSCz catalogue that are mainly big elliptical (plus some starburst galaxies). Since many of them do overlap with known AGNs, these two samples are not truly separate and the small-scale clustering of bright galaxies it is not surprising. Unfortunately, both the overdensity overlap of physically different classes of sources and the pronounced small-scale clustering of many source candidates does not play in favor of a clear source identification of UHECR sources e.g. by cross-correlation analyses.

We have argued that the auto-correlation function of different source classes differs considerably on all scales and may be used as a tool to identify the sources of UHECRs. Since the PAO has not yet published sufficient information on their observed events, we were restricted to perform a more conventional analysis considering just one bin of the auto-correlation function. At present, the most likely interpretation of the evidence reported by Auger of "8 doublets separated by less than 7 degrees in the 19 highest energy events" is that the sources of UHECRs are either a strongly clustered sub-sample of AGNs, or a sparse population of more or less isotropically distributed sources (as e.g. GRBs), possibly with pairs of events within 7° coming from the same objects. From our results, it is however clear that a comparison on all angular scales would disentangle the two cases. In principle, once the source population giving rise to UHECRs is identified, the magnetic field deflection required to smear-out the original auto-correlation function might be fitted and used for studies of the (extra-) galactic magnetic field.

With a statistics as low as twice the preliminary sample analyzed by the Auger collaboration, we expect that a first conclusive discrimination among source populations should be possible. Measuring the difference between e.g. different subclasses of AGNs as sources of cosmic rays appears to be more difficult and requires more complete catalogues within the near ($z \lesssim 0.1$) universe and much larger statistics.

Probably, opening the era of UHECR astronomy will require a combined advance in many aspects of UHECR physics, from reducing the uncertainty on the absolute energy scale to robust constraints on the chemical composition of the primaries. At the same time, the field would also benefit from advancements in the astrophysics of magnetic fields, like constraints on the Galactic magnetic field and refined simulations of extragalactic ones. No doubt, however, that once born UHECR astronomy will pay-off as an unprecedented diagnostic tool for the study of high energy non-thermal universe, as well as for measuring otherwise inaccessible extragalactic magnetic fields.

Acknowledgments

A.C., S.H. and M.K. thank the Max-Planck-Institut für Physik in Munich for hospitality and support during the initial phase of this work. P.S. acknowledges support by the US Department of Energy and by NASA grant NAG5-10842. This work was supported in part by the European Union under the ILIAS project, contract No. RII3-CT-2004-506222.

[1] G. Sigl, F. Miniati and T. Enßlin, Phys. Rev. D **68**, 043002 (2003) [astro-ph/0302388]; Phys. Rev. D **70**, 043007 (2004) [astro-ph/0401084].

[2] K. Dolag, D. Grasso, V. Springel and I. Tkachev, JETP Lett. **79**, 583 (2004) [Pisma Zh. Eksp. Teor. Fiz. **79**, 719 (2004)] [astro-ph/0310902]; JCAP **0501**, 009 (2005) [astro-ph/0410419].

[3] A. A. Watson, Nucl. Phys. Proc. Suppl. **136**, 290 (2004) [astro-ph/0408110]; T. Abu-Zayyad *et al.* [HiRes-MIA Collaboration], Astrophys. J. **557**, 686 (2001) [astro-ph/0010652]; R. U. Abbasi *et al.* [The High Resolution Fly's Eye Collaboration], Astrophys. J. **622**, 910 (2005) [astro-ph/0407622].

[4] M. Takeda *et al.*, Astophys. J. **522**, 225 (1999) [astro-ph/9902239].

[5] Y. Uchihori *et al.*, Astropart. Phys. **13**, 151 (2000) [astro-ph/9908193]; S. L. Dubovsky, P. G. Tinyakov and I. I. Tkachev, Phys. Rev. Lett. **85** (2000) 1154 [astro-ph/0001317]; Z. Fodor and S. D. Katz, Phys. Rev. D **63** (2001) 023002 [hep-ph/0007158]; P. G. Tinyakov and I. I. Tkachev, JETP Lett. **74**, 1 (2001) [Pisma Zh. Eksp. Teor. Fiz. **74**, 3 (2001)] [astro-ph/0102101]; P. Blasi and D. de Marco, Astropart. Phys. **20**, 559 (2004) [astro-ph/0307067].; D. De Marco, P. Blasi and A. V. Olinto, JCAP **0601** (2006) 002 [arXiv:astro-ph/0507324]; C. B. Finley and S. Westerhoff, Astropart. Phys. **21**, 359 (2004) [astro-ph/0309159]; H. Yoshiguchi, S. Nagataki and K. Sato, Astrophys. J. **614**, 43 (2004) [astro-ph/0404411]. W. S. Burgett and M. R. O'Malley, Phys. Rev. D **67** (2003) 092002

[arXiv:hep-ph/0301001].
[6] M. Kachelrieß and D. Semikoz, Astropart. Phys. **23**, 486 (2005) [astro-ph/0405258].
[7] P. G. Tinyakov and I. I. Tkachev, JETP Lett. **74**, 445 (2001) [Pisma Zh. Eksp. Teor. Fiz. **74**, 499 (2001)] [astro-ph/0102476]; N. W. Evans, F. Ferrer and S. Sarkar, Phys. Rev. D **67**, 103005 (2003) [astro-ph/0212533]; D. S. Gorbunov *et al.*, JETP Lett. **80**, 145 (2004) [Pisma Zh. Eksp. Teor. Fiz. **80**, 167 (2004)] [astro-ph/0406654]; D. S. Gorbunov, P. G. Tinyakov, I. I. Tkachev and S. V. Troitsky, JCAP **0601**, 025 (2006). [astro-ph/0508329]. J. Miralda-Escude and E. Waxman, Astrophys. J. **462** (1996) L59 [arXiv:astro-ph/9601012]; S. Singh, C. P. Ma and J. Arons, Phys. Rev. D **69** (2004) 063003 [arXiv:astro-ph/0308257]; N. W. Evans, F. Ferrer and S. Sarkar, Astropart. Phys. **17** (2002) 319 [arXiv:astro-ph/0103085];
[8] M. Kachelrieß and D. V. Semikoz, Astropart. Phys. **26**, 10 (2006) [astro-ph/0512498].
[9] A. Cuoco *et al.*, JCAP **0601**, 009 (2006) [astro-ph/0510765].
[10] A. Cuoco, G. Miele and P. D. Serpico, Phys. Rev. D **74** (2006) 123008 [astro-ph/0610374]; arXiv:0706.2864 [astro-ph].
[11] S. Mollerach *et al.* [PAO Collaboration], to appear in *Proc. "30th International Cosmic Ray Conference", Mérida, Mexico, 2007*, arXiv:0706.1749 [astro-ph].
[12] M. Takeda *et al.*, Phys. Rev. Lett. **81**, 1163 (1998) [astro-ph/9807193].
[13] K. Greisen, Phys. Rev. Lett. **16** (1966) 748.
[14] G. T. Zatsepin and V. A. Kuzmin, JETP Lett. **4**, 78 (1966) [Pisma Zh. Eksp. Teor. Fiz. **4**, 114 (1966)].
[15] R. Abbasi *et al.* [HiRes Collaboration], arXiv:astro-ph/0703099.
[16] T. Yamamoto [Pierre Auger Collaboration], arXiv:0707.2638 [astro-ph].
[17] D. V. Semikoz and P. A. Collaboration, arXiv:0706.2960 [astro-ph].
[18] D. F. Torres and L. A. Anchordoqui, Rept. Prog. Phys. **67**, 1663 (2004) [astro-ph/0402371].
[19] W. Saunders *et al.*, Mon. Not. Roy. Astron. Soc. **317**, 55 (2000) [astro-ph/0001117].
[20] M.-P. Véron-Cetty and P. Véron, Astron. Astroph. **455**, 773 (2006).
[21] T. H. Jarrett, T. Chester, R. Cutri, S. Schneider, M. Skrutskie and J. P. Huchra, Astron. J. **119** (2000) 2498 [arXiv:astro-ph/0004318].
[22] M. Schmidt, Astrophys. J. **552**, 36S (2001) [astro-ph/0101163].
[23] C. Blake, A. Pope, D. Scott and B. Mobasher, Mon. Not. Roy. Astron. Soc. **368** (2006) 732 [arXiv:astro-ph/0602428].
[24] S. D. Landy and A. S. Szalay, Astrophys. J. **412** (1993) 64.
[25] G. Kauffmann *et al.* [SDSS Collaboration], Mon. Not. Roy. Astron. Soc. **346**, 1055 (2003) [astro-ph/0304239].
[26] P. Sommers, Astropart. Phys. **14** (2001) 271 [astro-ph/0004016].
[27] D. Harari, S. Mollerach and E. Roulet, JCAP **0611**, 012 (2006) [astro-ph/0609294].
[28] V. Berezinsky, A. Z. Gazizov and S. I. Grigorieva, Phys. Rev. D **74**, 043005 (2006) [hep-ph/0204357]; Nucl. Phys. Proc. Suppl. **136**, 147 (2004) [astro-ph/0410650]; astro-ph/0702488, to appear in *Proc. "30th International Cosmic Ray Conference", Mérida, Mexico, 2007*.
[29] D. Harari, S. Mollerach and E. Roulet, JCAP **0405** (2004) 010 [arXiv:astro-ph/0404304].
[30] M. Kachelrieß, P. D. Serpico and M. Teshima, Astropart. Phys. **26**, 378 (2006) [astro-ph/0510444].
[31] R. U. Abbasi *et al.* [High Resolution Fly's Eye Collaboration (HiRes)], Astrophys. J. **610**, L73 (2004) [astro-ph/0404137].
[32] Talk of S. Westerhoff at the CRIS-2004 workshop "GZK and Surrounding", Catania, Italy, `http://www.ct.infn.it/cris2004/talk/westerhoff.pdf`
[33] N. Hayashida *et al.*, astro-ph/0008102.
[34] Talk of M. Pravdin at the 29th Int. Cosmic Ray Conference, Pune 2005, `http://icrc2005.tifr.res.in/htm/PAPERS/HE14/rus-pravdin-MI-`
[35] M. M. Winn *et al.*, J. Phys. G **12**, 653 (1986); *ibid.* 675 (1986), see also the complete catalogue of SUGAR data in "Catalogue of highest energy cosmic rays No. 2", ed. WDC-C2 for Cosmic Rays (1986).
[36] M. Kachelrieß and D. V. Semikoz, Phys. Lett. B **577**, 1 (2003) [astro-ph/0306282].
[37] A. Constantin and M. S. Vogeley, Astrophys. J. **650**, 727 (2006) [astro-ph/0601717].
[38] L. J. Kewley, B. Groves, G. Kauffmann and T. Heckman, Mon. Not. Roy. Astron. Soc. **372**, 961 (2006) [astro-ph/0605681].
[39] P. N. Best *et al.*, Mon. Not. Roy. Astron. Soc. **379**, 894 (2007) [astro-ph/0611197].