CLUSTERING PROPERTIES OF ULTRA–HIGH-ENERGY COSMIC RAYS AND THE SEARCH FOR THEIR ASTROPHYSICAL SOURCES

A. CUOCO, S. HANNESTAD, AND T. HAGBOLLE
Department of Physics and Astronomy, University of Aarhus, Ny Munkegade, DK-8000 Aarhus, Denmark

AND

M. KACHELRIESS
Institutt for fysikk, NTNU, N-7491 Trondheim, Norway

P. D. SERPICO
Center for Particle Astrophysics, Fermi National Accelerator Laboratory, Batavia, IL 60510-0500

Received 2007 September 25; accepted 2007 November 14

ABSTRACT

The arrival directions of ultra–high-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 autocorrelation function of the data with that of catalogs of potential sources is a powerful diagnostic tool. In particular, this method is far less sensitive than cross-correlation studies to unknown deflections in magnetic fields 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 subsample of galaxies or of active galactic nuclei. We present forecasts for various cases of source distributions which can be checked soon with the Pierre Auger Observatory.

Subject headings: BL Lacertae objects: general — cosmic rays — galaxies: active — galaxies: Seyfert — large-scale structure of universe — methods: statistical

Online material: color figures

1. INTRODUCTION

The identification of the sources of ultra–high-energy cosmic rays (UHECRs) and, more generally, the question of whether 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 UHECR 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 (Sigl et al. 2003, 2004; Dolag et al. 2004, 2005). Uncertainties from modeling strong interactions prevent a clean determination of the fraction of heavy nuclei in UHECRs above \( E \geq 10^{19} \) eV (Watson 2004; Abu-Zayyad et al. 2001; Abbasi et al. 2005). 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 (Akeno Giant Air Shower Array) data contain several small-scale clusters, i.e., clusters of events within its experimental angular resolution (Takeda et al. 1999). This result triggered a series of works studying the autocorrelation of UHECR data at small angular scales (Uchihori et al. 2000; Dubovsky et al. 2000; Fodor & Katz 2001; Tinyakov & Tkachev 2001a; Blasi & de Marco 2004; DeMarco et al. 2006; Finley & Westerhoff 2004; Yoshiguchi et al. 2004; Burgett & O’Malley 2003; Kachelriess & Semikoz 2005) or correlating UHECRs with potential astrophysical sources (Tinyakov & Tkachev 2001b; Evans et al. 2002, 2003; Gorbunov et al. 2004, 2006; Miralda-Escude & Waxman 1996; Singh et al. 2004). For instance, the best-fit value for the density \( n_s \) of UHECR sources found in Kachelriess & Semikoz (2005) is \( (1-3) \times 10^{-3} \) Mpc\(^{-3} \), while the 2 \( \sigma \) confidence region ranges from \( 2 \times 10^{-6} \) to \( \sim 10^{-2} \) 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\( ^\circ \)–5\( ^\circ \) 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 (AGNs). Although some studies found significant correlations, in particular with BL Lac objects, these results have remained controversial.

A second piece of evidence is anisotropies on medium scales. Kachelriess & Semikoz (2006) analyzed the available data set of cosmic-ray (CR) arrival directions from the HiRes (High Resolution Fly’s Eye) stereo, AGASA, Yakuts, and SUGAR (Sydney University Giant Air Shower Recorder) experiments with energies \( E \geq 4 \times 10^{19} \) eV in the HiRes energy scale. They found evidence at \( \sim 3 \) \( \sigma \) c.l. for anisotropies on the scales of 10\( ^\circ \)–35\( ^\circ \), with a clear minimum of the chance probability around 20\( ^\circ \)–30\( ^\circ \). This result is consistent with the theoretical expectations for anisotropies associated with large-scale structures (LSSs) from Cuoco et al. (2006a). Further studies showed that the correlations are best explained if either the UHECR sources are overbiased with respect to normal galaxies or the cosmic-ray horizon is smaller than expected for rectilinearly propagating protons (Cuoco et al. 2006b, 2007).

Intriguingly, similar findings seem to emerge from an analysis of the preliminary data from the Pierre Auger Observatory (PAO). For 64 events with \( E > 4 \times 10^{19} \) eV the data presented in
Mollerach et al. (2008) show a surplus of clustering in the broad range from 7° to 30°. The distribution has its minimum at 7° with a second, broad minimum between 19° and 24° and is quite similar to the distribution with 57 events in Kachelriess & Semikoz (2006). Remarkably, the PAO data also contain eight doublets separated by less than 7° in the 19 highest energy events, i.e., with $E > 5.75 \times 10^{19}$ eV (Mollerach et al. 2008).

Finally, the last piece of evidence comes from the UHECRs’ energy spectrum. In particular, the long controversy about the continuation of the spectrum (Takeda et al. 1998) beyond the Greisen-Zatsepin-Kuzmin (GZK) cutoff (Greisen 1966; Zatsepin & Kuzmin 1966) seems to have finally been solved by the latest data from HiRes (Abbasi et al. 2007) and the PAO (Yamamoto et al. 2008), which both detect with a 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 UHECR photon primaries (Semikoz et al. 2008), the data are clearly pointing toward a “standard” scenario in which the bulk of UHECR sources have an astrophysical origin in the nearby universe, with more exotic top-down scenarios playing at most a subdominant role. This evidence makes timely a detailed study of possible UHECR astrophysical sources of the kind addressed in the following.

The main aim of the present work is to compare the autocorrelation 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 also comment on how the clustering features of the publicly available world data set compare with expectations. This analysis is timely because 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 the best tools are to do so. Here we advocate the importance of a global comparison, i.e., a comparison on all angular scales, of the observed autocorrelation function of arrival directions with the expectations for different source and primary scenarios. At first glance, cross-correlation tests with source catalogs might appear as the ideal tool to identify the sources of UHECRs and what the best tools are to do so. However, turning this into quantitative statements requires some assumptions on the nature of the primary particle and the absolute energy scale of the experiments. In § 3, 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 world data set of available data from the HiRes, AGASA, Yakutsk, and SUGAR experiments. Finally, we discuss our results and conclude in § 4.

2. GALAXY AND AGN CORRELATION FUNCTIONS

2.1. Astronomical Catalogs

Among the astrophysical objects most often proposed as UHECR sources are AGNs in general, specific subclasses such as blazars, radio or Seyfert galaxies, gamma-ray bursts (GRBs), or young neutron stars (for a review, see, e.g., Torres & Anchordoqui 2004). 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 such as neutron stars) and AGNs in the nearby universe. We use the PSCz catalog (Saunders et al. 2000) as a sample of the galaxy distribution and the 12th edition of the Véron-Cetty & Véron (VCV) catalog (Véron-Cetty & Véron 2006) for the AGNs. We also study the clustering properties of several subsamples, imposing cuts in absolute magnitude for the galaxy catalog and subdividing the AGN catalog into Seyfert galaxies of type 1 (S1), type 2 (S2), and LINERs (S3), according to the classification reported in the VCV catalog itself. In Figure 1 the various kinds of AGNs and the PSCz galaxies within $z = 0.02$ are shown in Galactic coordinates, color-coded according to their redshifts. The empty region along the Galactic plane is the so-called avoidance region caused by the presence of the Milky Way and does not reflect an intrinsic lack of objects.

The details of the PSCz catalog, in particular a description of the mask and of the selection function, are summarized in Saunders et al. (2000). However, the $B$-band magnitudes reported in the catalog are biased and show systematic offsets within different regions of the sky where the galaxy magnitudes have been taken from different catalogs with different calibrations. To overcome this problem, we match sources in the PSCz catalog with sources in the Two Micron All Sky Survey (2MASS) extended source catalog (Jarrett et al. 2000) 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 ~80% of the galaxies in the PSCz catalog have a counterpart in the 2MASS Extended Source Catalog and discard the others. We then construct various subsamples of the catalog performing cuts in absolute magnitude using the distance modulus relation $M = m - 5 \log dL_{\text{Mpc}} - 25 - m - 43.16 + 5 \log h - 5 \log z$ (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 subsamples 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 Figure 2 we show the fraction $f(z)$ of galaxies from the PSCz catalog for the subsamples...
obtained with luminosity cuts $M < -24$ and $-24.5$ in redshift bins of width 0.005. In the same panel we also show the complete galaxy sample with no cut imposed. The PSCz catalog 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.

The VCV catalog is different from the PSCz catalog in that it 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., $z \leq 0.02$) the catalog can be considered fairly complete, and we build selection functions analogously to the PSCz subsamples described above. Below we show that our results are quite robust against this assumption. We find that (1) within $z = 0.02$ the AGN sample closely follows the matter distribution, as expected (see Fig. 1); and (2) although the overall catalog 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 Figure 2 we show the fraction of S1, S2, and S3 galaxies from the VCV catalog 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 are 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 catalog, the catalog as a whole can be regarded as at least reasonably complete out to $z \sim 0.02$. We note, however, that the S3 subsample 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 catalog 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 $n_s \sim 5 \times 10^{-4}$ Mpc$^{-3}$ h$^3$. The same density also corresponds to the galaxies brighter than $M_{\text{cut}} = -24$, while 50 uniformly distributed sources, always within the same volume, would give $n_s \sim 5 \times 10^{-5}$ Mpc$^{-3}$ h$^3$. For ordinary galaxies the number density depends strongly on the assumed $M_{\text{cut}}$ and typically ranges from $10^{-3}$ Mpc$^{-3}$ h$^3$ to values as high as $10^{-1}$ Mpc$^{-3}$ h$^3$ for Milky Way–like galaxies. In any case, the value of the number density of the UHECR sources depends on the horizon containing the sources and thus on issues such as the nature of the primaries and the absolute energy scale. Thus, in the following we focus the attention mainly on the absolute number of sources (above the assumed CR energy threshold) and on their bias/overdensity with respect the distribution of matter, as principal observables.

BL Lac AGNs, also popular UHECR source candidates, are very rare in our GZK neighborhoods. The nearest confirmed BL Lac object in the VCV catalog is RXS J05055+0416 at $z = 0.027$. If we include possible BL Lac objects only six objects are found in the VCV within $z \sim 0.03$. If indeed such a small number of sources were responsible for the UHECRs, then even more peculiar clustering signatures should be expected as a large number of triplets.
or even quadruplets, as long as deflections by extragalactic magnetic fields are not too large. At the same time, a cross-correlation between these objects and the UHECR multipoles should become evident. This possibility could then be easily confirmed or disproved by looking at this kind of signature, and we do not discuss it further in the following.

Finally, we briefly discuss the case of GRBs as UHECR sources. The observed rate of GRBs is \( R_{\text{obs}} \approx 0.5 \times 10^{-9} \) (Mpc yr\(^{-1}\)) according to Schmidt (2001). However, deflections in the extragalactic magnetic fields lead to time delays \( \tau \) that in turn increase the effective density of GRBs as \( n_e = \tau R_{\text{obs}} \). The clustering properties of GRBs are in general quite different from those of AGNs and massive galaxies. Long-duration GRBs, which make up about two-thirds 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 catalog. The remaining one-third of the GRBs, which are most likely the result of binary collisions, have a distribution which is close to that of Type Ia supernovae, but they are not considered very likely sites for the UHECR acceleration. Thus, GRBs cluster less than average PSCz galaxies and, in the following considerations, we use randomly distributed sources as a rough template for their clustering properties.

### 2.2. Correlation Functions

An important point to prove for the following arguments is that different astrophysical catalogs of candidate UHECR sources have sufficiently different clustering properties. To that purpose, we calculate in this section the autocorrelation function \( w(\theta) \) of the various samples. In the past, a commonly employed estimator for the autocorrelation has been the intuitive \( DD/RR - 1 \). This estimator, however, is suboptimal especially for the estimation of variance (Landy & Szalay 1993), while an optimal estimator is given by (Blake et al. 2006; Landy & Szalay 1993)

\[
w(\theta) = \left( \frac{DD - 2DR + RR}{RR} \right),
\]

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 \( DD \), \( DR \), and \( RR \) are the normalized pair counts in each angular bin around \( \theta \). The brackets indicate that the final \( w(\theta) \) is an ensemble average over many random realizations. Note that for data consistent with a random distribution \( w(\theta) \) is zero within the errors. The resulting autocorrelation functions \( w(\theta) \) are shown in the left panel of Figure 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 (Landy & Szalay 1993)

\[
rms[w(\theta)] = [1 + w(\theta)]^{1/2} \left[ \frac{1}{n(n - 1)/2(\text{RR})} \right]^{1/2},
\]

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 autocorrelation at small scales, although the clustering of normal galaxies is quite less pronounced \( w_{\text{AGN}}(1')/w_{\text{gal}}(1') \approx 3 \). We show below 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 subsamples 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 AGN subsamples it can be seen that the clustering of S1 objects shows no strong differences from that of all AGNs; by contrast the S2 and S3 subtypes show a stronger autocorrelation on the smallest scales, \( \theta \leq 3' \). 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 \( w(\theta) \) decrease for increasing \( \theta \), intermediate scales \( \theta \sim 10''-30'' \) 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 is important.

The above results are in general in quite good agreement with other more detailed studies of the AGNs’ bias properties. In particular, the clustering properties of AGNs have been studied extensively (e.g., in Kauffmann et al. 2003) using the Sloan Digital Sky Survey (SDSS) catalog. 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.

3. 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 focus on one of the most remarkable findings reported by the PAO, namely, that of eight doublets within 7° out of the 19 highest energy events (Mollerach et al. 2008).

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

\[
C(\vartheta) = \left\{ \sum_{i=2}^{N} \sum_{j=1}^{i-1} \Theta(\vartheta - \vartheta_{ij}) \right. ,
\]

i.e., the cumulative number of pairs within the angular distance \( \vartheta \), where \( \Theta \) is the step function [with \( \Theta(0) = 1 \)], \( N \) is the number of CRs considered, and \( \vartheta_{ij} \) is the angular distance between events \( i \) and \( j \). Although \( C(\vartheta) \) introduces further correlations between different angular bins, the use of cumulative countings instead of differential ones has the great advantage of significantly reducing the dependence on unknown magnetic field deflections, a crucial point for UHECR astronomy. The ensemble average is performed over a large number \( M = 10^5 \) of Monte Carlo sets. The events are extracted randomly from the catalog under consideration, and we take into account the PAO exposure as described in Sommers (2001) assuming as characteristic parameters for the PAO \( \gamma_{\text{max}} = 60° \) for the maximal zenith angle for a CR event and \( \vartheta_{\text{PAO}} = -35° \) for the PAO latitude location. The selection effects of the catalog and the attenuation due to CR propagation are included by 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 1 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., Harari et al. 2006). Note that the injection spectral index has only a minor effect on \( D_{1/2} \) in the energy range considered.

| Species | \( E_{\text{cut}}/10^{19} \text{ eV} \) | \( D_{1/2}/\text{Mpc} \) | \( z_{1/2} \) |
|---------|---------------------------------|-----------------|----------|
| \( p \) | 5.0 | 160 | 0.037 |
| \( ^{28}\text{Si} \) | 6.0 | 100 | 0.023 |
| \( ^{56}\text{Fe} \) | 8.0 | 40 | 0.009 |
| \( ^{56}\text{Fe} \) | 8.0 | 45 | 0.011 |

Note.—Given is the distance \( D_{1/2} \) (or equivalently redshift \( z_{1/2} \)) within which 50% of the UHECR flux above \( E_{\text{cut}} \) comes for different assumptions on energy and chemical composition, assuming isotropic and uniform sources and rectilinear propagation. Values are adapted from plots in Harari et al. (2006).

In the following we assume proton primaries and use the propagation window function \( W(z, E_{\text{cut}}) \) as calculated in Cuoco et al. (2006a). Given the importance of the clustering signal observed by the PAO, we focus mostly on the case of \( N = 19 \) events. In order to study the sensitivity to the assumed energy scale, we consider two cases: (1) the preliminary calibration of the energy scale presented by the PAO is correct, and thus the 19 highest energy events have energies above \( E_{\text{cut}} = 5.75 \times 10^{19} \text{ eV} \); (2) ultra–high-energy air shower experiments are affected by an overall uncertainty in their energy calibration, whose normalization might be obtained by requiring that they correctly reproduce spectral features of a model, in our case the dip model (Berezinsky et al. 2004, 2006, 2008). 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 \( E_{\text{cut}} \approx 8 \times 10^{19} \text{ eV} \).

For the galaxy samples we use the PSCz selection function, while for the AGN samples we adopt the approximation \( v(z) = z^{\alpha} \), with \( \alpha \approx 0.4 \) 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 appreciably change 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 catalog is complete, a point we come back to below.

Our results for the average \( C(\vartheta) \) and their 1σ variations for the case of 19 events are shown in Figure 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 subclasses compared with galaxy distributions and an isotropic sky, and galaxies with different cuts in magnitude. While the first three panels assume \( E_{\text{cut}} = 8 \times 10^{19} \text{ eV} \), the bottom right panel is the same as the second one, but for \( E_{\text{cut}} = 5.75 \times 10^{19} \text{ eV} \). 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 subsamples are much more likely to produce the degree of clustering observed by the PAO. The same happens for the subsamples of bright galaxies where the brightest galaxies (and thus the set with the smallest number density \( n_s \)) provide the best match to the expected clustering. In the case of a lower energy scale, the horizon is larger and the sky more isotropic, and especially the LSS and the isotropic sky hypotheses have even more trouble explaining the observations. Note, however, that the S3 sample, which seems to be the AGN subsample most consistent with the high number of pairs, may suffer from a strong selection bias in the VCV catalog: 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 that 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 Figure 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 3σ level. The exact confidence levels are illustrated in Figure 5, where the full distribution within 7° for various models are compared to the Auger results. In addition, we had to concentrate on the largest fluctuation in the PAO data set, since at present this is the only available information. At different angles, we must expect less significant clustering. That said, it is interesting that, as shown in Figure 6, with a statistics doubled with respect to that analyzed in Mollerach et al. (2008) 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 $C(\theta)$ 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 $\chi^2$ method or a Kolmogorov-Smirnov test) of the correlation 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 $7^\circ$ (to stick to the most notable finding of the PAO) and 30$^\circ$ (a typical intermediate scale) for several models. In particular, in Table 2 we report the average values $C(7^\circ)$ and $C(30^\circ)$ and the 2$\sigma$ error regions shown in the plots almost coincide with the mean. [See the electronic edition of the Journal for a color version of this figure.]

![Fig. 4.—Average $C(\theta)$ and its 1 $\sigma$ variation for the case of 19 events and finite number of sources isotropically distributed compared with the continuous limit (top left), galaxies with different cuts in magnitude (bottom left), and AGN subclasses compared with galaxy distribution (top right); all the above cases assume a cut in the window function at $E_{\text{cut}} = 8 \times 10^{19}$ eV. The bottom right panel is the same of the top right one, but for $E_{\text{cut}} = 5.75 \times 10^{19}$ eV. Note that the error regions are highly asymmetric for $\theta \leq 40^\circ$ (see Figs. 5–7), and the upper 1 $\sigma$ error regions shown in the plots almost coincide with the mean. [See the electronic edition of the Journal for a color version of this figure.]](image)

![Fig. 5.—Probability distribution of $C(7^\circ)$, for the case of 19 events and an energy cut of $E_{\text{cut}} = 8 \times 10^{19}$ eV, and the case of different astrophysical models considered (left) and a finite number of uniformly distributed sources (right). [See the electronic edition of the Journal for a color version of this figure.]](image)
lower and upper limits—denoted by \( C_- \) and \( C_+ \), respectively—for the expected number of pairs in different models and \( N = 19 \) data. In Tables 3 and 4 we report analogous quantities for \( N = 40 \) and 60, respectively. We consider three models: (1) 100 uniformly distributed sources, which mimics GRBs, (2) the S2 subclass of AGNs, and (3) the galaxies brighter than \( M_{\text{cut}} = -24.5 \). These have been chosen to be basically consistent with the 7º Auger data. Note that even with the 19 events, models 1 and 2 show significant differences at 30º, a difference that becomes quite large and easily testable with a modest improvement 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 a higher number of pairs.

![Graph](image)

**Fig. 6.—Same as the top panels of Fig. 4, but for a statistics of 40 events. [See the electronic edition of the Journal for a color version of this figure.]**

### TABLE 3

| Model          | \( C_-(7^\circ) \) | \( C_(7^\circ) \) | \( C_+ (7^\circ) \) | \( C_-(30^\circ) \) | \( C_(30^\circ) \) | \( C_+ (30^\circ) \) |
|----------------|-------------------|-----------------|-------------------|-------------------|-------------------|-------------------|
| 100 GRBs ....... | 10                | 21              | 32                | 68                | 82                | 96                |
| S2 AGNs......... | 9                 | 18              | 31                | 85                | 105               | 151               |
| \( M_{-24.5} \) Gal....... | 6                | 13              | 25                | 81                | 110               | 162               |

**Note.—** See the text for details on the notation. The different models reported are 100 GRBs, S2 AGNs, and galaxies with \( M_{\text{cut}} = -24.5 \).

As a final comment, we note that due to the limited information available we have restricted the study to the analysis of a cumulative number of pairs. However, the above approach can be easily generalized to higher order statistics, such as the cumulative counting of the proper number of doublets, of triplets, etc. (Harari et al. 2004). A combined use of these tools will likely provide an even more robust and stringent constraint on the nature and number of UHECR sources.

### 3.1. Repeaters versus 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. A formal but useful way to illustrate this point is to look at the probability distribution of \( C(0^\circ) \), i.e., the number of pairs for \( \theta = 0^\circ \) (repeaters). In Figure 7 we report this distribution for the case of 19 events and an energy cut of \( E_{\text{cut}} = 8 \times 10^{19} \) 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 the origin for the doublets.

Of course, deflections in magnetic fields and experimental resolution effects prevent an experimental determination of \( C(0^\circ) \). As an example, we ask if the strong clustering signal in the PAO data coming from pairs within 7º is compatible with repeaters. The following arguments show that, without additional information, the hypotheses both of repeaters and of small-scale clustering are consistent with the findings. The correlation functions \( w(\theta) \) of the astronomical catalogs previously discussed are peaked at

---

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.
small scales, and actually strongly peaked below $1^\circ$ for the AGN samples. Yet this peak in $w(\varnothing)$ will be shifted in the excess signal in $C(\varnothing)$ to larger scales, first of all because the event numbers scales approximately as $N \propto \varnothing^{-2}$. 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 field (GMF) and possibly extragalactic ones. Even protons of the considered energy are expected to suffer average deflections in the GMF of the order of $\sim 3'\!\!\!$ (Kachelriess et al. 2007). Since the Auger exposure peaks near the Galactic center region, which in typical GMF models is associated with larger deflections, separations of pointlike 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 autocorrelation function to perform a robust diagnostic.

3.2. Comparison with Old Experiments

Given the importance of a global comparison of the autocorrelation function, one may wonder if the already existing public available data offer additional insight. The only sufficiently large data set that is publicly available is that used for the first time in Kachelriess & Semikoz (2006). It consists of $\sim 100$ events with energies $E \geq 4 \times 10^{19}$ eV from the HiRes stereo (Abbasi et al. 2004; Westerhoff 2004), AGASA (Hayashida et al. 2000), Yakutsk (Pravdin et al. 2005), and SUGAR (Winn et al. 1986a, 1986b) experiments. Here we have rescaled the absolute energy scale of each experiment (Kachelriess & Semikoz 2006, 2003) by requiring that they correctly reproduce the dip spectral feature (Berezinsky et al. 2004, 2006, 2008).

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 1). At those distances the catalogs 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 Figure 8 we compare for illustrative purposes the function $C(\varnothing)$ computed for this data set with the corresponding expectations from uniformly distributed sources and for AGNs as well as the PSCz galaxy catalog within $z \leq 0.02$.

It can be seen that the autocorrelation of the data presents an excess at $10'^-30'$ with respect to a uniform distribution corresponding to the medium-scale clustering signal found in Kachelriess & Semikoz (2006). 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 $\varnothing \sim 10'$ is not very meaningful, given the poor angular resolution of SUGAR, for example.) More statistics at higher energy and better quality data are definitely needed from UHECR experiments, while deeper and more complete large-angle surveys would be welcome from the astrophysical community.

4. 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 these objects as possible UHECR sources. Our autocorrelation studies have shown that—consistent with what is known from the much larger SDSS galaxy and AGN catalogs (see, e.g., Kauffmann et al. 2003; Constantin & Vogeley 2006; Kewley et al. 2006; Best et al. 2007)—nearby AGNs exhibit much stronger small-scale clustering than average galaxies. The same is true for the brightest galaxies in the PSCz catalog.
which are mainly big ellipticals (plus some starburst galaxies). Since many of them do overlap with known AGNs, these two samples are not truly separate and the similarities in the small-scale clustering of bright galaxies and AGNs are not surprising. Unfortunately, neither the overdensity overlap of physically different classes of sources nor the pronounced small-scale clustering of many source candidates play in favor of a clear source identification of UHECR sources, for example, by cross-correlation analyses.

We have argued that the autocorrelation 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 performing a more conventional analysis considering just one bin of the autocorrelation function. At present, the most likely interpretation of the evidence reported by Auger of eight doublets separated by less than $7^\circ$ in the 19 highest energy events is that the sources of UHECRs are either a strongly clustered subsample of AGNs or a sparse population of more or less isotropically distributed sources (e.g., GRBs), possibly with pairs of events within $7^\circ$ coming from the same objects. From our results, however, it is 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 autocorrelation function might be fitted and used for studies of the Galactic (or extragalactic) magnetic field.

With 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, for example, different subclasses of AGNs as sources of cosmic rays appears to be more difficult and requires more complete catalogs within the near ($z \leq 0.1$) universe and much larger statistics.

Opening the era of UHECR astronomy will probably 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, such as 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 the high-energy nonthermal universe, as well as for measuring otherwise inaccessible extragalactic magnetic fields.

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 RI3-CT-2004-506222.

**REFERENCES**

Abbasi, R. U., et al. 2004, ApJ, 610, L73
———. 2005, ApJ, 622, 910
———. 2007, Phys. Rev. Lett., submitted (astro-ph/0703099)

Abu-Zayyad, T., et al. 2001, ApJ, 557, 686

Berezinsky, V., Gazizov, A. Z., & Grigorieva, S. I. 2004, Nucl. Phys. B, 136, 147
———. 2006, Phys. Rev. D, 74, 043005
———. 2008, in Proc. 30th Int. Cosmic Ray Conf. (Mérida), in press (astro-ph/0702488)

Best, P. N., et al. 2007, MNRAS, 379, 894

Blake, C., Pope, A., Scott, D., & Mobasher, B. 2006, MNRAS, 368, 732

Blasi, P., & de Marco, D. 2004, Astropart. Phys., 20, 559

Burget, W. S., & O’Malley, M. R. 2003, Phys. Rev. D, 67, 092002

Constantin, A., & Vogeley, M. S. 2006, ApJ, 650, 727

Cuoco, A., D’Abrusco, R., Longo, G., Miele, G., & Serpico, P. D. 2006a, J. Cosmol. Astropart. Phys., 0601, 009

Cuoco, A., Miele, G., & Serpico, P. D. 2006b, Phys. Rev. D, 74, 123008
———. 2007, preprint (arXiv:0706.2864)

DeMarco, D., Blasi, P., & Olinto, A. V. 2006, J. Cosmol. Astropart. Phys., 0601, 002

Dolag, K., Grasso, D., Springel, V., & Tkachev, I. 2004, J. Exp. Theor. Phys. Lett., 79, 583
———. 2005, J. Cosmol. Astropart. Phys., 0501, 009

Dubovsky, S. L., Tinyakov, P. G., & Tkachev, I. I. 2000, Phys. Rev. Lett., 85, 1154

Evans, N. W., Ferrer, F., & Sarkar, S. 2002, Astropart. Phys., 17, 319
———. 2003, Phys. Rev. D, 67, 103005

Finley, C. B., & Westerhoff, S. 2004, Astropart. Phys., 21, 359

Fodor, Z., & Katz, S. D. 2001, Phys. Rev. D, 63, 023002

Gorbunov, D. S., Tinyakov, P. G., Tkachev, I. I., & Troitsky, S. V. 2004, J. Exp. Theor. Phys. Lett., 80, 145
———. 2006, J. Cosmol. Astropart. Phys., 0601, 025

Greisen, K. 1966, Phys. Rev. Lett., 16, 748

Harari, D., Mollerach, S., & Roulet, E. 2004, J. Cosmol. Astropart. Phys., 0405, 010
———. 2006, J. Cosmol. Astropart. Phys., 0611, 012

Hayashida, N., et al. 2000, preprint (astro-ph/0008102)

Jarrett, T. H., Chester, T., Cutri, R., Schneider, S., Skrutskie, M., & Huchra, J. P. 2000, AJ, 119, 2498

Kachelriess, M., & Semikoz, D. V. 2003, Phys. Lett. B, 577, 1
———. 2005, Astropart. Phys., 23, 486
———. 2006, Astropart. Phys., 26, 10

Kachelriess, M., Serpico, P. D., & Teshima, M. 2007, Astropart. Phys., 26, 378

Kaufmann, G., et al. 2003, MNRAS, 346, 1055

Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, MNRAS, 372, 961

Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64

Miralda-Escude, J., & Waxman, E. 1996, ApJ, 462, L59

Mollerach, S., et al. 2008, in Proc. 30th Int. Cosmic Ray Conf. (Mérida), in press (arXiv:0706.1749)

Pruvost, M. I., et al. 2005, 29th Int. Cosmic Ray Conf. (Pune), 243

Saunders, W., et al. 2000, MNRAS, 317, 55

Schmidt, M. 2001, ApJ, 552, 36

Semikoz, D. V., et al. 2008, in Proc. 30th Int. Cosmic Ray Conf. (Mérida), in press (arXiv:0706.2960)

Sigl, G., Miniati, F., & Ensslin, T. 2003, Phys. Rev. D, 68, 043002
———. 2004, Phys. Rev. D, 70, 043007

Singh, S., Ma, C.-P., & Arons, J. 2004, Phys. Rev. D, 69, 063003

Sommers, P. 2001, Astropart. Phys., 14, 271

Takeda, M., et al. 1998, Phys. Rev. Lett., 81, 1163
———. 1999, ApJ, 522, 225

Tinyakov, P. G., & Tkachev, I. I. 2001a, J. Exp. Theor. Phys. Lett., 74, 1
———. 2001b, J. Exp. Theor. Phys. Lett., 74, 445

Torres, D. F., & Anchordoqui, L. A. 2004, Rep. Prog. Phys., 67, 1663

Uchiyori, Y., et al. 2000, Astropart. Phys., 13, 151

Véron-Cetty, M.-P., & Vérón, P. 2006, A&A, 455, 773

Watson, A. A. 2004, Nucl. Phys. B, 136, 290

Westerhoff, S. 2004, in Cosmic Ray Int. Semin., GZK and Surroundings, ed. C. Aramo, A. Insolia, & C. Tuve (Catania: INFN), http://www.ct.infn.it/cris2004/talk/westerhoff.pdf

Winn, M. M., et al. 1986a, J. Phys. G, 12, 653

———. 1986b, J. Phys. G, 12, 675

Yamamoto, T., et al. 2008, in Proc. 30th Int. Cosmic Ray Conf. (Mérida), in press (arXiv:0707.2638)

Yoshiguchi, H., Nagataki, S., & Sato, K. 2004, ApJ, 614, 43

Zatsarin, G. T., & Kuzmin, V. A. 1966, J. Exp. Theor. Phys. Lett., 4, 78