A GLOBAL AUTOCORRELATION STUDY AFTER THE FIRST AUGER DATA: IMPACT ON THE NUMBER DENSITY OF UHECR SOURCES

A. CUOCO1, S. HANNESTAD1, T. HAUGBOLLE1, M. KACHELRIEB2, AND P. D. SERPICO3,4
1 Department of Physics and Astronomy, University of Aarhus, Ny Munkegade, Bygn. 1520, DK-8000 Aarhus, Denmark
2 Institutte for fysikk, NTNU, N-7491 Trondheim, Norway
3 Center for Particle Astrophysics, Fermi National Accelerator Laboratory, Batavia, IL 60510-0500, USA
4 Physics Division, Theory Group, CERN, CH-1211 Geneva 23, Switzerland
Received 2009 February 21; accepted 2009 June 25; published 2009 August 17

ABSTRACT

We perform an autocorrelation study of the Auger data with the aim to constrain the number density \( n_s \) of ultrahigh energy cosmic ray (UHECR) sources, estimating at the same time the effect on \( n_s \) of the systematic energy scale uncertainty and of the distribution of UHECR. The use of global analysis has the advantage that no biases are introduced, either in \( n_s \) or in the related error bar, by the a priori choice of a single angular scale. The case of continuous, uniformly distributed sources is nominally disfavored at 99% CL and the fit improves if the sources follow the large-scale structure of matter in the universe. The best-fit values obtained for the number density of proton sources are within a factor \( \sim 2 \) around \( n_s \approx 1 \times 10^{-4} \text{Mpc}^{-3} \) and depend mainly on the Auger energy calibration scale, with lower densities being preferred if the current scale is correct. The data show no significant small-scale clustering on scales smaller than a few degrees. This might be interpreted as a signature of magnetic smearing of comparable size, comparable with the indication of a \( \approx 3^\circ \) magnetic deflection coming from cross-correlation results. The effects of some approximations done on the above results are also discussed.

Key words: cosmic rays – large-scale structure of universe – methods: statistical

Online-only material: color figures

1. INTRODUCTION

Evidence is now emerging that ultrahigh energy cosmic rays (UHECRs) have an astrophysical origin, as opposed to being generated in exotic top-down models: the detection of a spectral suppression consistent with the Greisen–Zatsepin–Kuzmin (GZK) effect (Greisen 1966; Zatsepin & Kuzmin 1966) by both HiRes (Abbasi et al. 2008) and Auger (Abraham et al. 2008d) Collaborations, together with the Auger bounds on the UHE neutrino flux (Abraham et al. 2008b), on the photon fraction (Abraham et al. 2008a), and on the anisotropy toward the Galactic center (Aalseth et al. 2007; Aloisio & Tortorici 2008) are all consistent with this scenario. The next step is clearly the identification of the sources of UHECRs, an arena where anisotropy studies play a crucial role. Yet, the limited angular resolution of extensive air shower detectors and especially the deflections that charged particles suffer in astrophysical magnetic fields make the task highly nontrivial. This is especially troublesome given that the UHECRs chemical composition is unknown, that we lack a detailed knowledge of the Galactic magnetic field structure, and, above all, of the very magnitude and structure of extragalactic magnetic fields (EGMF) outside of cluster cores. These limitations—together with the small statistics available at present—suggest that, at least in an initial phase, charged particle astronomy may be studied only with a deep understanding of the small-scale structure of the Galactic magnetic field. This understanding, however, is just beginning to emerge from measurements of Galactic polarized synchrotron radiation (Curran et al. 2008), the recent detection of Galactic cold dark matter, and the potential constraints on the magnetic field structure (McMillan et al. 2009). The detection of Galactic cosmic rays (CRs) with energies up to \( \sim 10^{15} \text{eV} \) (Aguilar et al. 2008) is also consistent with this picture. This is especially troubling given that the UHECRs chemical composition is unknown, that we lack a detailed knowledge of the Galactic magnetic field structure, and, above all, of the very magnitude and structure of extragalactic magnetic fields (EGMF) outside of cluster cores. These limitations—together with the small statistics available at present—suggest that, at least in an initial phase, charged particle astronomy may be limited to the inference on the statistical properties of UHECR sources rather than a detailed study of single accelerators.

In Cuoco et al. (2008b), we found that a global comparison of the two-point autocorrelation function of the data with one of the catalogs of potential sources is a powerful diagnostic tool: this observable is less sensitive to unknown deflections in magnetic fields than the cross-correlation function, while keeping a strong discriminating power among source candidates. In particular, the autocorrelation function of (sub-) classes of galaxies has different biases with respect to the large-scale structure (LSS) of matter. As a result, the best-fit value for the density \( n_s \) of different source classes may differ, especially if only one or a small range of angular scales is considered. Although the bias of different source classes differs maximally at small angular scales, we showed that the statistically most significant differences are at intermediate angular scales, where both the larger number of cosmic ray (CR) pairs and of galaxy pairs lead to relatively smaller error bars. Moreover, the autocorrelation function on larger angular scales becomes less dependent on possible deflections in the Galactic and extragalactic fields.

In this paper, we derive the number density of UHECR sources using the recently published arrival directions and energies of the 27 Auger events (Abraham et al. 2008c) with estimated energy \( E \geq 57 \text{EeV} \) thereby complementing the study (Cuoco et al. 2008b) with a concrete example for a comparison of the global cumulative autocorrelation function of sources and UHECRs. Note that we showed in Cuoco et al. (2008b) that, even in an ideal case where systematics play no major role, roughly three times the number of “useful” events that can be extracted from Abraham et al. (2008c) are required to start distinguishing between different subclasses of sources. Thus, a study of the kind envisaged in Cuoco et al. (2008b) is unrealistic at present. Here, we restrict ourselves to more modest goals: (1) to compare the data against predictions of two toy model cases of uniformly distributed sources and of “normal galaxies” (i.e., sources that have the same distribution as the PSCz catalog (Saunders et al. 2000)) which we shall refer to with the two values for the label \( \kappa = \{\text{uni, LSS}\} \), respectively. (2) To study the effect on the allowed range of \( n_s \) of a systematic error on the energy scale of the UHECR experiment. Note that preliminary results of the clustering of the Auger events has been presented in Mollerach (2007), but astrophysical implications have not been discussed there.
We review the statistical method that we use in Section 2, and apply it in Section 3 to the Auger data, providing some interpretation of the results. In Section 4, we discuss some limitations and caveats of the analysis. Finally, we summarize interpretation of the results. In Section 4, we discuss some

2. STATISTICAL ANALYSIS OF THE DATA

The use of correlation functions is well suited to the study of over- and underdensities of nonuniformly distributed sources and of the resulting anisotropies in the radiation received from them. Since the number of CR events published by Auger is still small, we use in our analysis, following Kachelrieß & Semikoz (2006), the cumulative two-point autocorrelation function $C(\theta)$ defined as

$$C(\theta) = \sum_{i=2}^{N} \sum_{j=1}^{i-1} \Theta(\theta - \theta_{ij}),$$

i.e., the number of pairs within the angular distance $\theta$. Here, $N$ is the number of CRs considered, $\theta_{ij}$ is the angular distance between events $i$ and $j$, and $\Theta$ is the step function (with $\Theta(0) = 1$).

This function is straightforward to compute for the data, and denoted then by $C_1(\theta)$. For a specific model hypothesis $X$, a set of functions $C_k(\theta|X)$ is obtained in the following way: sources with equal luminosities are distributed within a sphere of $180 h^{-1}$ Mpc either uniformly or following the three-dimensional LSS as given by a smoothed version (Cuoco et al. 2006, 2008a) of the PSCz catalog (Saunders et al. 2000). For the LSS case (but not for the uniform case) sources and CR events within the PSCz mask are excluded, leaving 22 CR events. Note that the mask mostly overlaps with the Galactic plane and the bulge region, where larger deflections due to the Galactic magnetic field are expected; the mask is thus not only a catalog limitation, but also implements to some extent a physically motivated angular cut. Finally, each source $k$, at redshift $z_k$, is weighted by the factor

$$\frac{1}{z_k} \int_{E_{cut}(E_{in}, z_k)}^{\infty} E^{-3} dE$$

(2)

accounting for its redshift-dependent flux suppression and the CR energy losses. These are parameterized as a continuous process in term of the function $E_{i}(E_f, z)$ which gives the initial injection energy $E_i$ for a particle leaving the source at $z_k$ and arriving at the Earth with energy $E_f < E_i$. Further details are given in Cuoco et al. (2006). The injection spectral index $s$ is assumed to be the same for all the sources and equal to 2.0. The dependence on $s$ is however weak as shown in more detail in Cuoco et al. (2006). This procedure defines the model, while a single random realization is obtained by choosing the observed number of events from the sources according to the source weights and the declination-dependent Auger experimental exposure.

The model thus directly depends only on $n_s$, and the choice between sources distributed uniformly or according to the PSCz catalog (of course, implicitly it also depends on the assumptions of sufficiently small magnetic field deflections). Additionally, the model depends via the weights of Equation (2) on the type of the primary particle, the energy spectrum of the sources, and the energy scale and cut $E_{cut}$. The latter dependence arises, because the energy scale of CR experiments has a relatively large systematic error that is difficult to determine. In particular, it has been argued (Teshima 2007) that the energy scale of Auger should be shifted up by 30%–40% to obtain agreement with the spectral shape predicted by $e^+e^-$ pair production (Berezinsky et al. 2006) and the CR flux measured by other experiments. Therefore, we use two different values for the energy cut, $E_{cut} = 60$ EeV, assuming that the energy calibration of Auger is correct and $E_{cut} = 80$ EeV inspired by the dip interpretation. We do not include in this work the finite energy resolution of the Auger experiment that is of order 20% in $\Delta E / E$. A finite energy resolution would result in an effective decrease of the nominal energy cut due to the steeply falling CR spectrum and to a larger GZK horizon (Kachelrieß et al. 2007). Similarly, we do not account for the stochasticity of the energy loss in the photopion production process. Both of the effects are subdominant at the moment with the present low statistics while a more careful analysis will be needed in the future when more data are available. For a rough estimate of the influence of both effects on $n_s$ one may compare how the 30% upward shift of $E_{cut}$ from 60 to 80 EeV changes $n_s$. Finally, throughout this work we consider proton primaries, but note that the combination of nuclei with large deflections and few sources has been advocated too (Armengaud et al. 2005; Fargion 2008; Gorbunov et al. 2008). A further brief discussion of this point is postponed to Section 4.2.

The cumulative autocorrelation of the data $C_n(\theta)$, which is a single, one-dimensional function, has now to be compared with the hypothesis $X$, for which we have various Monte Carlo realizations $C_k(\theta|X), k = 1, \ldots, M$ (we use typically $M \sim 10^5$). A standard method to compare data and model is to use angular bins $\theta_i$ so that to substitute the continuous function $C(\theta)$ with the discrete set of values $C(\theta_i)$. The Monte Carlo realizations can then be used to calculate the marginalized probability distribution of each single $C(\theta_i|X)$ or, if required, the joint probability distribution of two $C$ variables $C(\theta_1|X), C(\theta_2|X)$, three $C$ variables or more. In practice, to derive the best-fit value $n_s$ as well as the goodness-of-fit for the chosen hypothesis $X$ a possible way is to calculate the mean $(C(\theta_i|X))$ and the variance $\sigma_i$ per bin as well as the correlation matrix $\Sigma_{ij}$ and then to perform a $\chi^2$ test. However, the difficulty to deal with such a high-dimensional probability space and the generally strong non-Gaussianity of the probability distribution make the $\chi^2$ method clearly not optimal for the problem at hand.

The usual way to circumvent the problem is to use the Monte Carlo to calculate the chance probability to observe stronger clustering than in the data. Given the problem at hand we slightly generalize the method defining two functions: the chance probability to observe stronger $(P_+(\theta|X))$ or weaker $(P_-(\theta|X))$ clustering at the angular scale $\theta$ than in the data given by

$$P_\pm(\theta|X) = \frac{1}{M} \sum_{i=1}^{M} \Theta[\pm(C_i(\theta|X) - C_n(\theta))].$$

(3)

The minimum of the chance probability can then be used as a global estimator for the agreement of the hypothesis with the data. In our case with a scan of $P_+(\theta|X)$ over $\theta$, we obtain the two minima $P_-(\theta|X)$ for the minimum of $P_-(\theta|X)$ and $P_+(\theta|X)$ for the minimum of $P_+(\theta|X)$, respectively. The simplest way to combine these two estimators is then to use the product $P(X) = P_-(\theta)P_+(\theta)$.

However, a drawback of this method is that neither the quantity $P(X)$ nor the single $P_+(\theta)$ is truly probabilities. The scan over the angular scales $\theta$, in fact, introduces a bias which
need to be corrected with the use of a penalty factor (Tinyakov 
& Tkachev 2001, 2004; Finley & Westerhoff 2004). More
precisely, to obtain correct probabilities the identical procedure
as described above needs to be performed with many simulated
data sets. Counting how often smaller values of \( P_s(X) \) and
\( P(X) \) are obtained by chance with respect to the case of the data
provides true penalized chance probabilities \( p_s(X) \), \( p_s(X), \) and
\( p(X) \).

The use of the chance probability tool has often generated
some confusion in the past. An emblematic case is the signifi-
cance of the small-scale clustering in the AGASA data for which
very different estimates ranging from \( \sim 10^{-6} \) to \( \sim 10^{-2} \) have
been reported in various studies (see for example Takeda et al.
1999) depending on the use or not of the penalty and on different
a priori choices of the angular and energy scale of reference (see
Finley & Westerhoff 2004 for a detailed account). The effect of
the scan can then be quite relevant and a major point in the
following is that the effect of the penalty is correctly taken into
account when quoting the constraints on \( n_s \) and the constraints
are further compared with the case of an a priori choice of the
relevant angular scale.

Note, anyway, that the penalty calculation can be avoided if a
particular angular scale is chosen a priori and the values of the
chance probability at this scale are employed. However, the scan
over all angles avoids possible bias, in contrast to the choice
of a single angular scale, which introduces some theoretical
prejudice even if this choice may be physically motivated. In the
case at hand, it is unclear if this choice should be dominated by the \( \sim 1^\circ \)
angular resolution of the detector, as in Takami & Sato (2009)
which implicitly assumes negligible magnetic field deflections;

by a \( \sim 3^\circ \) scale, as suggested by the cross-correlation with
active galactic nuclei (AGNs) revealed by Auger (Abraham et al.
2007); or yet some other scale, as the \( 6^\circ \) separation considered
in Abraham et al. (2008c). To emphasize this dependence, we
summarize the results of this kind of “single-bin” analyses in
Table 1 and compare them with the ones obtained with the global
method, as reported in the last column (see the next section).
Note, in particular, how the \( 6^\circ \) bin chosen in Abraham et al.
(2008c) systematically leads to an overly stringent bound.

3. INTERPRETATION

In Figure 1, we report the results for the quantities \( p_s(X) \)
defined above, where \( X = [n_s, E_{\text{cut}}, \kappa] \), in our case the latter

\[ p_s(X) = \frac{1}{N_T} \sum_{i=1}^{N_T} \delta(X - X_i) \]

where \( N_T \) is the total number of simulated events, \( X_i \) are the
randomly chosen values of the angular and energy scale of reference,
and \( \delta(X - X_i) \) is the Dirac delta function.

Table 1

| \( n_s \times 10^{-4} \text{ Mpc}^{-3} \) | \( \theta_1 \) | \( 6^\circ \) | \( 3^\circ \) | \( 1^\circ \) | Global |
|-----------------|-----------|---------|---------|---------|--------|
| LSS (80)        | 1.3^{+5.7}_{-1.9} | 2.0^{+0.6}_{-0.4} | 2.0^{+\infty}_{-\infty} | 5.0^{+0.5}_{-0.2} | 1.3^{+100}_{-3.8} |
| Uniform (80)    | 0.8^{+0.8}_{-0.8} | 1.2^{+0.8}_{-0.8} | 2.5^{+0.5}_{-0.5} | 5.0^{+0.5}_{-0.2} | 1.4^{+1.4}_{-0.7} |
| Uniform (60)    | 0.3^{+1.3}_{-0.3} | 0.3^{+1.3}_{-0.3} | 0.7^{+0.3}_{-0.3} | 1.5^{+100}_{-100} | 0.8^{+0.6}_{-0.6} |
| LSS (60)        | 0.2^{+1.0}_{-0.12} | 0.3^{+1.7}_{-0.2} | 0.8^{+0.3}_{-0.6} | 1.0^{+0.8}_{-0.8} | 0.5^{+0.5}_{-0.2} |

\[ \text{Note that, for uncorrelated deflections, the window size used for the scan is } \theta_1 \text{, and the energy scale of reference is } E_{\text{cut}}. \]

\[ \text{Note that, for a single angular scale, which introduces some theoretical prejudice even if this choice may be physically motivated. In the case at hand, it is unclear if this choice should be dominated by the } \sim 1^\circ \text{ angular resolution of the detector, as in Takami & Sato (2009) which implicitly assumes negligible magnetic field deflections; actually, since deflections from the source are correlated and the energies of events similar, the relative deflections for a single chemical species would be likely } \lesssim 3^\circ. \]
two being two-valued discrete variables. Because the number of CRs usable for this analysis is still small, all four hypotheses are compatible with the data at the 2σ level for some range of \( n_s \) values. Yet, several interesting conclusions can already be drawn. The best fit is achieved for sources following the PSCz distribution and a source density \( n_s \approx 1 \times 10^{-4} \text{ Mpc}^{-3} \). Also, independently of \( E_{\text{cut}} \), both the penalized probability and the range of \( n_s \) with compatibility at 95% CL are larger for the LSS model than for the uniform case. Reversing the argument, as can be read more quantitatively in Table 1, we can see that the constraints are generally stronger for the uniform cases with respect to the LSS ones, but this is achieved only at the expense of a worse general best fit. The fact that the LSS models fit better than the data is not surprising: most of the Auger events are aligned along the local overdensity known as the supergalactic plane which is suitably reproduced with the use of the PSCz catalog within our LSS scenario.

The case of a uniform distribution of “infinitely many” sources (\( n_s \to \infty \) each with an infinitesimal luminosity) is excluded for both energy cuts at the 95% CL; the upper bound is \( n_s \lesssim (1-3) \times 10^{-4} \text{ Mpc}^{-3} \). This is another way to say that the Auger data are inconsistent with a structureless UHECR sky, independently of the use of a catalog and of a pre-determined angular scale for the search. This is, in our opinion, an important milestone in the development of UHECR astronomy. While the best-fit point for \( n_s \) is approximately a factor 10 higher than that found in earlier studies using AGASA data above \( E_{\text{cut}} > 40 \text{ EeV} \) (in the AGASA energy scale; Yoshiguchi et al. 2003; Blasi & De Marco 2004; Kachelrieß & Semikoz 2005), the shape of the chance probability \( p(n_s) \) agrees: for low values of \( n_s \), \( p(n_s) \) is a steeply decreasing function of \( n_s \), since the probability to observe multiplets from the same source increases fast. In particular, the radius within which 70% of all observed UHECRs with energy above \( E_{\text{cut}} = 80 \text{ EeV} \) are produced is \( R \approx 60 \text{ Mpc} \) (Cuoco et al. 2006). As a result, the number of sources within this radius becomes less than the number of observed CR events for densities smaller than \( n_s \approx 10^{-5} \). Such a scenario would require large deflections (and probably nuclei primary) and thus contradicts our assumptions. On the other hand, \( p(n_s) \) decreases relatively slowly for high densities and only weak constraints can be obtained with the current data set for the maximally allowed value of \( n_s \). Since both an increase of \( E_{\text{cut}} \) and of the bias of the sources lead to a decrease of the effective number of sources inside the GZK volume, large values of \( n_s \) have the strongest constraint in the case of uniformly distributed sources and \( E_{\text{cut}} = 60 \text{ EeV} \) (left, top panel) and weakest for sources following the LSS and \( E_{\text{cut}} = 80 \text{ EeV} \) (right, bottom panel).

In Figure 2, we show the model autocorrelation function with 1σ and 3σ shaded regions for the best-fit uniform and LSS model for \( E_{\text{cut}} = 80 \text{ EeV} \), both corresponding to \( n_s \approx 1.4 \times 10^{-4} \text{ Mpc}^{-3} \). The data at small angular scales, \( \theta \leq 3^\circ \), the data show a deficit of clusters compared to the expectation for the best-fit density from the global analysis. This "tension" is qualitatively present in most models fitting the data. Within our assumptions, this deficit is explained in a natural way by (relative) deflections of this size in magnetic fields. This value is comparable with the 3.2° found in Abraham et al. (2008c) that optimizes the correlations of the same data set with AGNs. The absence of small-scale clusters is also responsible for the shift in the best-fit value for \( n_s \) compared to old analyses using the AGASA data. The result thus shows that, intriguingly, also the autocorrelation function can be employed as a sensitive tool for magnetic field studies. Clearly, however, a more detailed study needs to be complemented with a model of the intervening magnetic fields while, likely, more statistics is required to derive significant constraints.

Our result is in contrast with that of Takami & Sato (2009) which instead reports a small-scale clustering in the Auger data. Notice however that, with respect to our work, the authors of Takami & Sato (2009), although including an explicit treatment of galactic and extragalactic magnetic fields, make use of the noncumulative autocorrelation functions and do not take into account penalties for the scan over the angular scale. Their claim of small-scale clustering within 1° is thus likely to disappear if the angular scan is taken into account and the comparison is made properly with respect to a model effectively fitting the data. Thus, as already noted in Harari (2004), the interpretation of small-scale clustering could be misleading if it is not defined with respect to a proper model of the distribution of sources. Apart from this point, however, the constraints we obtain on \( n_s \)
are roughly in agreement with Takami & Sato (2009). This is mainly due to the very low statistics available at present which is still not sensitive to the exact analysis method employed. We stress however, as noted in Cuoco et al. (2008b), that already with a statistic of ∼3 times the present one a formally correct analysis will become crucial to avoid biased results.

Figure 2 also clarifies why the LSS case gives a better fit to the data: as can be seen, the LSS best-fit model fits nicely to the data, basically within 1σ over all the angular scales, while the best fit case for uniform sources shows a ∼2σ deficit in the broad range 4°–30°. A lower n_s (i.e., a higher clustering) would not help because it would give much more pairs than the data in the 1°–4° range. Thus, at the end, even with the best possible compromise the agreement with the data is only at the 20% level for the best fit (or, equivalently, the uniform model is excluded at the 80% CL).

We summarize in Table 1 the list of best-fit n_s with 95% error bars for the four cases considered and for different choices of the angular scale or for the global autocorrelation analysis. The crucial point to note here is that the derived n_s intervals are sensibly biased with respect to each other for different choices of the a priori angular scale. Thus, the choice of the angular scale crucially affects the result and should be avoided unless the given scale has some strong physical motivation. This problem is of course avoided employing a global comparison. The choice of a single angular scale also affects the error bars which can be both larger or smaller with respect to the global case. This can be easily understood from Figure 3 where we can see that the choice 4°–15° is optimal because excess of clustering is observed for low n_s while a deficit is present for high n_s giving thus the tightest constraints. Again, however, a choice in this range is not a priori motivated so that from the global analysis we get somewhat larger error bars properly taking into account all the angular scales.

Although it is difficult to draw strong conclusions on the nature of the UHECR sources at this moment, it is worth noting that X-ray-selected AGNs with X-ray luminosity L > 10^{45} erg s^{-1} naturally fall in the range n_{AGN} = (1–5) × 10^{-5} Mpc^{-3} (Steffen et al. 2003). At the same time, for AGNs with densities in the range (10^{-3}–10^{-4}) Mpc^{-3}, the required luminosity is of the order L/n_{AGN} ∼ (10^{40}–10^{41}) erg s^{-1} in UHECRs above ∼60 EeV. So, simple energetics arguments are consistent with the inferred values for the density; on the other hand, acceleration efficiency (see, e.g., Ptitsyna & Troitsky 2008) tends to favor some sub-classes of objects and thus a lower inferred density of sources, yet typically within the 95% CL range inferred for n_s. Another effect which might reconcile the small tension is that the sources can be bursting/transient or beamed. In this case, the luminosity requirement is softened (see, e.g., the model proposed by Farrar & Gruzinov 2009). However, the effectively visible number density of sources decreases too, unless some isotropization takes place after acceleration but close to the source (see Sigl 2008, for more details). Of course, these are very preliminary considerations based on a limited number of data. For example, these arguments might be significantly modified when accounting for a more realistic luminosity function. The effects of extragalactic magnetic fields and the consequent UHECRs time delays can also be quite relevant, although at present our knowledge of EGMFs is affected by large uncertainties and needs to be better understood (see Sigl et al. 2003, 2004; Dolag et al. 2004; Kotera & Lemoine 2008; Ryu et al. 2008). Also, in the presence of a heavy-nuclei component and/or of extragalactic magnetic fields, a significant fraction of events might be associated with the nearest AGN Cen A (see, e.g., Gorbunov et al. 2008). Interestingly, signatures in the gamma and neutrino bands are expected to help disentangling among many scenarios (see Sigl 2008; Cuoco & Hannestad 2008; Kachelriß et al. 2009; Becker & Biermann 2009; Halzen & O’Murchadha 2008; Hardcastle et al. 2008), enlarging the realm of multi-messenger astronomy.

4. Discussion on Some Simplifying Assumptions

4.1. The Role of the Energy Cut

The energy cut used by the PAO to produce the sample used in the present analysis was chosen in order to maximize the cross-correlation signal, see Abraham et al. (2008b, 2008c) for details. One may wonder how this selection affects the conclusions of our present work. In principle, the optimal energy cut for the autocorrelation signal and the cross-correlation signal with a given catalog differ from one another, although in the case of a common physical origin one does expect some correlation between them. In particular, it seems reasonable that the optimal cut for a global autocorrelation might reside at a lower energy, since the small-scale displacement from putative sources is not as relevant to the signal as for cross-correlations, and the larger statistics helps. Lacking a direct access to the data, it is hard to estimate quantitatively how large a bias is introduced by focusing on the sample of public available events. From Figure 2 presented in Mollerach (2007), however, one can draw two qualitative conclusions: (1) that a correlation between the two optimal cuts in energy is indeed present; (2) that a slightly different cut, in the range 40 EeV ≲ E_{cut}/EeV ≲ 60, should still lead to an appreciable clustering of the events.

We checked that this is indeed the case by performing the same analysis as before, but now adding a further scan over the range of accessible energy cuts, i.e., any value E > 57 EeV. We plot the results in Figure 4. One can note that in all cases the best fit improves and the constraints on n_s worsen, as expected given the further penalty due to the energy scan. Yet, most qualitative features described in the previous section stay the same: for example, the LSS model is still preferred over a uniform one. It should be also said that if the true minimum of the chance

Figure 3. Contours of equal chance probability p(θ|n_s) for the LSS case and E_{cut} = 80 EeV.

(A color version of this figure is available in the online journal.)
Figure 4. Penalized chance probabilities \(p_-(n_x), p_+(n_x),\) and \(p(n_x)\), for \(E_{\text{cut}} = 60\) EeV (top panels) and \(E_{\text{cut}} = 80\) EeV (bottom panels) taking into account the effect of the energy scan. The left column reports the case for uniformly distributed sources, the right panel for sources following LSS with the bias of the PSCz Galaxy catalog. Also shown is the 95% and 99% confidence levels.

When including (the poorly known) magnetic fields, two additional effects are relevant. (1) For a given energy, the larger the charge the larger the deflection and hence the loss of information at small angular separations. Quantifying how large is this scale is a difficult task. We note that protons of these energies in the sole Galactic field likely suffer a few degrees of absolute deflections, see Kachelrieß et al. (2006), implying a degree-scale smoothing in the relative deflections important for the 2pcf. This is comparable with the angular resolution of the PAO: in this optimal case the whole information in the 2pcf starting at \(\theta_{\text{min}} \sim 1^\circ\) could be used. However, a different Galactic field model (especially toward the Galactic center), the presence of heavy nuclei, and/or significant extragalactic magnetic fields can easily lift \(\theta_{\text{min}}\) by 1 order of magnitude or more. (2) The other effect is that the real path length of the nucleus would be longer than the distance to the source: thus, the distance of accessible sources would be even shorter than estimated from energy-loss considerations. This is particularly relevant if the nucleus spends a lot of time in a magnetized region surrounding the accelerator (e.g., a magnetized cluster in which it is immersed) before escaping to the intergalactic medium.

Finally, if the maximal energy of acceleration of different species of nuclei fall by unfortunate coincidence in the same region expected for the GZK feature, slightly different energy cuts in the data (as well as statistical and systematic errors on the energy scale) might significantly change the expected pattern of anisotropies. The same happens if the proportions of different nuclei accelerated at the source change as a function of the energy. This is in principle a possibility, especially if different...
classes of objects contribute to the events at slightly different energies.

The general pessimistic conclusion is that if several of the above effects are relevant (or perhaps a single one is dominant), the capability of performing UHECR astronomy would be greatly reduced. While one still expects indications for anisotropies, inverting the problem and inferring the source/propagation medium properties would require a much larger statistics (especially in the trans-GZK region): disentangling the different effects is in fact a formidable task.

To provide a glimpse of how some of the above effects alter the reconstruction of \( n_s \) (our main topic here), in Table 2 we report how the constraint on \( n_s \) degrades as a function of a “smoothing angle” \( \vartheta_{\text{min}} \), below which we assume that the 2pcf information is completely lost. The main trend is that, if the smallest angular scales are neglected, it is easier to find parameter configuration fitting the data and, correspondingly, the allowed range for \( n_s \) widens. This had to be expected, given the shape of the correlation functions shown in Figures 2 and 3. In particular, we find that with \( \vartheta_{\text{min}} = 3^\circ \) we obtain almost the same results as in the global case. The case \( \vartheta_{\text{min}} = 10^\circ \) still places useful constraints especially for the \( E_{\text{cut}} = 80 \text{ GeV} \) case, while finally using only the information above \( \vartheta_{\text{min}} = 30^\circ \) basically no constraints on \( n_s \) are obtained. Note however that relative deflection angles of that sort would imply even larger overall deflections, seriously questioning the perspectives of present instruments to perform some form of UHECR astronomy.

## 5. SUMMARY

We used the first UHE data released by Auger to perform a global autocorrelation study of UHECR arrival directions, assuming proton primaries. The major advantage of our tool is that no biases are introduced by the a priori choice of a single angular scale. The main observable we have focused on is the number density \( n_s \) of ultrahigh energy cosmic ray (UHECR) sources. While the global analysis does not bias \( n_s \) by what is the theoretical prior of the “relevant” angular scale, still it is important to establish how the extraction of the allowed range of \( n_s \) from the data depends on a number of other effects, an issue often overlooked in the literature. In particular, here we discussed the systematic energy scale uncertainty and the bias of UHECR sources with respect to LSSs. As a first attempt to extract some information from the data, we compared four hypotheses: a structured universe (following the PSCz catalog) and an isotropic case, each for two possible values for the absolute energy scale of the PAO experiment. The density \( n_s \) is the free continuous parameter in terms of which constraints have been analyzed.

Not surprisingly, we find that the number of CRs usable for this analysis is still small and all four hypotheses are compatible with the data at the 2\( \sigma \) level for some range of \( n_s \). Yet, several interesting observations can be tentatively drawn. The best fit is achieved for sources following the matter tracer distribution, and a source density \( n_s \leq 1 \times 10^{-4} \text{ Mpc}^{-3} \), it is interesting to note that the data show some preference (although not significant, yet) for a structured universe: other recent statistical studies, like Kashti & Waxman (2008) and Koers & Tinyakov (2009), qualitatively agree in that respect. Also, there is an indication that the case of a uniform distribution of “infinitely many” sources \( n_s \rightarrow \infty \) each with an infinitesimal luminosity is excluded for both energy cuts at the 95% CL: the upper bound is \( n_s \leq (1–3) \times 10^{-2} \text{ Mpc}^{-3} \). This is another way to say that early Auger data suggest that data are poorly consistent with a structureless UHECR sky, independently of the use of a catalog and of a pre-determined angular scale for the search.

Compared to a benchmark number density of proton sources \( n_s \leq 1 \times 10^{-4} \text{ Mpc}^{-3} \), a factor ~ 2 lower densities are preferred if the current Auger energy scale is correct, a factor ~ 2 higher value if it is underestimated as required by the dip model. Including the finite energy resolution of the Auger experiment into the analysis will further reduce the best-fit value for \( n_s \). The width of the allowed region is dominated at present by the statistical error due to the small number of events. Nominally, approximately a three times larger sample is needed to reduce the Poisson error below the typical differences between source candidates. Once that level of statistics is reached, other effects will provide the main source of error, a major one being the systematics on the energy scale as we illustrated here. In the future, other effects such as the systematic and statistical errors in the energy determination of UHECR events and the stochastic nature of the photopion process need to be included to correctly determine the best-fit value of \( n_s \). Even considering the above limitations, preliminary conclusions are the following. First, the fit generally improves for sources following the LSS compared to uniformly distributed sources; qualitatively, for AGNs which are known to be even more overdensity biased than the LSS (see, e.g., discussion in Cuoco et al. 2008b), the agreement is expected to improve even further. In particular, the case of an uniform distribution of “infinitely many” sources, i.e., a structureless UHECR sky, is excluded at 99% CL. Second, the absence of clustering on scales smaller than a few degrees is most easily understood as the effect of magnetic smearing of comparable size. This, intriguingly, suggests that autocorrelation studies can be employed as a complementary tool to study galactic (and extragalactic) magnetic fields. Conclusions about the source scenario are complicated by the limited knowledge of the EGMFs and the presence of beaming and/or bursting effects which are difficult to disentangle with the use of the autocorrelation alone. Hence, the true source density could be somewhat lower than the best-fit value found: the energy scale error alone shifts the best-fit value by a factor 2 or 3; luminosity function effects are likely relevant too. The scarce statistics at the moment is a serious limiting factor to constrain more realistic models, but with a greater exposure of currently existing instruments and the multi-messenger combination of gamma-ray and neutrino data, some of these issues will probably be addressed and solved in the near future.

For a significant contamination of nuclei in the sample, a much more complicated analysis is needed, since many more variables enter the game. Qualitatively, one can expect that although future data might allow to disentangle a proton-dominated sample from a more complicated picture, inferring source properties and disentangling them from magnetic effects.
(in a few words, performing UHECR astronomy) should wait for a major jump in exposure, perhaps beyond the capabilities of currently planned instruments. Similar considerations apply unfortunately to the constraints to the cross sections of UHECRs as well.

We are grateful to Günter Sigl for useful discussions. M.K. thanks the Max-Planck-Institut für Physik in Munich for hospitality and support. P.S. is supported by the US Department of Energy and by NASA grant NAG5-10842. Fermilab is operated by Fermi Research Alliance, LLC under contract no. DE-AC02-07CH11359 with the United States Department of Energy.

REFERENCES

Abbasi, R., et al. (HiRes Collaboration) 2008, Phys. Rev. Lett., 100, 101101
Abraham, J., et al. (Pierre Auger Collaboration) 2007, Science, 318, 938
Abraham, J., et al. (Pierre Auger Collaboration) 2008a, Astropart. Phys., 29, 243
Abraham, J., et al. (Pierre Auger Collaboration) 2008b, Phys. Rev. Lett., 100, 211101
Abraham, J., et al. (Pierre Auger Collaboration) 2008c, Astropart. Phys., 29, 188
Abraham, J., et al. (Pierre Auger Collaboration) 2008d, Phys. Rev. Lett., 101, 061101
Aglietta, M., et al. (Pierre Auger Collaboration) 2007, Astropart. Phys., 27, 244
Aloisio, R., & Tortorici, F. 2008, Astropart. Phys., 29, 307
Armengaud, E., Sigl, G., & Miniati, F. 2005, Phys. Rev. D, 72, 043009
Becker, J. K., & Biermann, P. L. 2009, Astropart. Phys., 31, 138
Berezinsky, V., Gazzlov, A. Z., & Grigorieva, S. I. 2006, Phys. Rev. D, 74, 043005
Blasi, P., & De Marco, D. 2004, Astropart. Phys., 20, 559
Cuoco, A., & Hannestad, S. 2008, Phys. Rev. D, 78, 023007
Cuoco, A., Miele, G., & Serpico, P. D. 2008a, Phys. Lett. B, 660, 307
Cuoco, A., et al. 2006, J. Cosmol. Astropart. Phys., JCAP01(2006)009
Cuoco, A., et al. 2008b, ApJ, 676, 807
Dolag, K., Grassa, D., Springel, V., & Tkachev, I. 2004, JETP Lett., 79, 583 [Pisma Zh. Eksp. Teor. Fiz., 79, 719 (2004); J. Cosmol. Astropart. Phys. JCAP01(2005)009]
Fargion, D. 2008, Phys. Scripta, 78, 045901
Farrar, G. R., & Gruzinov, A. 2009, ApJ, 693, 329
Finley, C. B., & Westerhoff, S. 2004, Astropart. Phys., 21, 359
Gorbunov, D. S., Tinyakov, P. G., Tkachev, I. I., & Troitsky, S. V. 2008, arXiv:0804.1088
Greisen, K. 1966, Phys. Rev. Lett., 16, 748
Halzen, F., & O’Murchadha, A. 2008, arXiv:0802.0887
Harari, D., Mollerach, S., & Roulet, E. 2004, J. Cosmol. Astropart. Phys., JCAP05(2004)010
Hardcastle, M. J., Cheung, C. C., Feain, I. J., & Stawarz, L. 2008, MNRAS, 393, 1041
Kachelriess, M., Serpico, P. D., & Teshima, M. 2006, Astropart. Phys., 26, 378
Kachelriess, M., Ostapchenko, S., & Tomass, R. 2009, New J. Phys., 11, 065017
Kachelriess, M., Pariot, E., & Semikoz, D. V. 2007, arXiv:0711.3635
Kachelriess, M., & Semikoz, D. 2005, Astropart. Phys., 23, 486
Kachelriess, M., & Semikoz, D. 2006, Astropart. Phys., 26, 10
Kashi, T., & Waxman, E. 2008, J. Cosmol. Astropart. Phys., JCAP05(2008)006
Koers, H. B. J., & Tinyakov, P. 2009, J. Cosmol. Astropart. Phys., JCAP04(2009)003
Kotera, K., & Lemoine, M. 2008, Phys. Rev. D, 77, 123003
Mollerach, S. (Pierre Auger Collaboration) 2007, Proc. 30th International Cosmic Ray Conference, Mérida, Mexico, ed. R. Caballero, et al. (Mexico City, Mexico: Universidad Nacional Autónoma de México)
Ptitsyna, K., & Troitsky, S. 2008, arXiv:0808.0367
Ryu, D., Kang, H., Cho, J., & Das, S. 2008, Science, 320, 909
Saunders, W., et al. 2000, MNRAS, 317, 55
Sigl, G. 2008, arXiv:0803.3800
Sigl, G., Miniati, F., & Enßlin, T. 2003, Phys. Rev. D, 68, 043002
Sigl, G., Miniati, F., & Enßlin, T. 2004, Phys. Rev. D, 70, 043007
Steffen, A. T., et al. 2003, ApJ, 596, L23
Takami, H., & Sato, K. 2009, Astropart. Phys., 30, 306
Takeda, M., et al. 1999, ApJ, 522, 225
Teshima, M. 2007, in Proc. 30th International Cosmic Ray Conf., ed. R. Caballero, et al. (Mexico City, Mexico: Universidad Nacional Autónoma de Mexico)
Tinyakov, P. G., & Tkachev, I. I. 2001, JETP Lett., 74, 445 [Pisma Zh. Eksp. Teor. Fiz., 74, 499 (2001) (arXiv:astro-ph/0102476)]
Tinyakov, P., & Tkachev, I. 2004, Phys. Rev. D, 69, 128301
Unger, M. (Pierre Auger Collaboration) 2007, in Proc. 30th International Cosmic Ray Conf., ed. R. Caballero, et al. (Mexico City: Universidad Nacional Autónoma de Mexico)
Yoshiguchi, H., Nagataki, S., Tsubaki, S., & Sato, K. 2003, ApJ, 586, 1211 [Erratum: 601, 592 (2004)]
Zatsepin, G. T., & Kuzmin, V. A. 1966, JETP Lett., 4, 78 [Pisma Zh. Eksp. Teor. Fiz., 4, 114 (1966)]