



# Using fractal dimensionality in the search for source models of ultra-high energy cosmic rays

B.T. Stokes<sup>\*</sup>, C.C.H. Jui, J.N. Matthews

*Department of Physics and High Energy Astrophysics Institute, University of Utah, Salt Lake City, UT, USA*

Received 8 August 2003; accepted 8 December 2003

## Abstract

Although the existence of cosmic rays with energies extending well above  $10^{19}$  eV has been confirmed, their origin remains one of the most important questions in astro-particle physics today. Several source models have been proposed for the observed set of ultra high energy cosmic rays (UHECRs). Yet none of these models have been conclusively identified as corresponding with all of the available data. One possible way of achieving a global test of anisotropy is through the measurement of the information dimension,  $D_1$ , of the arrival directions of a sample of events.  $D_1$  is a measure of the intrinsic heterogeneity of a data sample. We will show how this method can be used to take into account the extreme asymmetric angular resolution and the highly irregular aperture of a monocular air-fluorescence detector. We will then use a simulated, isotropic event sample to show how this method can be used to place upper limits on any number of source models with *no* statistical penalty.

© 2004 Elsevier B.V. All rights reserved.

PACS: 98.70.Sa; 95.55.Vj; 95.85.Ry; 95.35.+d

Keywords: Cosmic rays; Anisotropy; Fractal dimensionality; Centaurus A; Dark matter halo; Air fluorescence

## 1. Introduction

The observation of ultra high energy cosmic rays (UHECRs) has now spanned nearly three decades. Over that period, many different source models have been proposed to explain the origin of these remarkable events. Recently, the akeno giant air shower array (AGASA) reported clustering at small angular scales for the events that were observed above  $4 \times 10^{19}$  eV [1]. However, this result

could not be confirmed by the high resolution Fly's Eye (HiRes) air fluorescence detector [2] despite the fact that HiRes-I's monocular exposure was more than twice that of AGASA within the pertinent energy range [3]. AGASA has also reported a possible excess of events in the direction of the galactic center for events with energies around  $10^{18}$  eV [4]. However, HiRes reported that it did *not* see anisotropies when examining harmonics in right ascension, a priori determined point sources, or enhancement in the supergalactic plane. Furthermore, prior analysis by original Fly's Eye showed no evidence of anisotropies when dependencies in galactic latitude and longitude, harmonics in

<sup>\*</sup> Corresponding author.

E-mail address: [stokes@cosmic.utah.edu](mailto:stokes@cosmic.utah.edu) (B.T. Stokes).

right ascension, excess maps, and specific a priori determined point sources were examined [5,6]. However, Fly’s Eye did find a small apparent excess in the direction of the galactic plane. [7]. In 1995, it was reported by Stanev et al. that the combined data of Haverah Park, Yakutsk, and AGASA showed an excess along the supergalactic plane with several potential point sources for events above  $2 \times 10^{19}$  eV [8].

These conflicting reports call for developing a more global way in which one could determine if a given sample possesses any statistically significant anisotropy. We will show that by considering the information dimension,  $D_1$ , of a given sample, one can simultaneously look for anisotropies at all angular scales greater than the angular resolution of the sample by considering the intrinsic heterogeneity of that particular data sample. This method is quite robust in that it can easily accommodate both asymmetric angular resolutions and irregular apertures. Furthermore, in the event that a sample is shown to be consistent with an isotropic distribution, this method can be used to place upper limits on possible source models. Because only a single measurement is taken of the actual data, any number of source models can be considered simultaneously without incurring statistical penalties.

## 2. Calculating fractal dimensions of a data sample

Fractal dimensionality is a simple measure of the scaling symmetry of a structure. By measuring the fractal dimension of a data sample, one can examine its self-consistency at different levels of magnification. There are several ways of exploiting this idea. From a computational perspective, the simplest is to use box-counting. For the most general case, the capacity dimension,  $D_c$  [9], one partitions the sample space into equi-sized and equi-shaped “boxes” with edge size  $\epsilon$ :

$$D_c = \lim_{\epsilon \rightarrow 0^+} \frac{\log N(\epsilon)}{\log 1/\epsilon}. \quad (1)$$

Here,  $N(\epsilon)$  is the minimum number of “boxes” with edge size  $\epsilon$  necessary to cover one’s sample.

The capacity dimension has a serious limitation: It only looks for the presence of the sample within

the available space and does not consider variations in the density of the sample at a given point in space. In cases where the density may differ within the sample space, the appropriate alternative is to use the information dimension,  $D_1$  [10,11]:

$$D_1 = - \sum_{i=1}^N \lim_{\epsilon \rightarrow 0^+} \frac{P_i(\epsilon) \log P_i(\epsilon)}{\log 1/\epsilon}, \quad (2)$$

where  $P_i(\epsilon)$  is the probability of finding a data point in the  $i$ th box of edge size  $\epsilon$ . This is a particularly suitable measurement when considering a data set consisting of UHECR arrival directions with finite angular resolution.

It should be noted that  $D_c$  and  $D_1$  are both particular cases of the  $q$ -dimension [12,13],

$$D_q = \frac{1}{1-q} \lim_{\epsilon \rightarrow 0^+} \frac{\log I(q, \epsilon)}{\log 1/\epsilon}, \quad (3)$$

where

$$I(q, \epsilon) = \sum_{i=1}^N P_i(\epsilon)^q. \quad (4)$$

We can then see that  $D_c = \lim_{q \rightarrow 0} D_q$  and that  $D_1 = \lim_{q \rightarrow 1} D_q$ .

## 3. Application to arrival direction distributions for UHECRs

In principle, it is simple to calculate the information dimension for a given sample of events. However, two complications arise when considering a set of arrival directions of UHECRs. First, the event directions are not known with complete precision. This makes the determination of  $D_1$  as  $\epsilon \rightarrow 0$  meaningless. Secondly, the determination of  $D_1$  requires that the sample space be divided into equi-sized and equi-shaped bins. For a spherical surface, this is simply not possible. Nevertheless, there are workable solutions for both of these problems.

### 3.1. Angular resolution

We will consider a hypothetical monocular air-fluorescence detector. This detector is largely based upon the characteristics of the Hires-I

detector [14,15]. Our detector observes events with an angular resolution that is described by a highly asymmetric 2-d Gaussian. For a monocular air fluorescence detector, angular resolution consists of two components:  $\sigma_1$ , in the determination of the angle,  $\psi$ , within the plane of reconstruction, and  $\sigma_2$ , in the estimation of the plane of reconstruction itself. Fig. 1 illustrates how this geometry would appear with a particular plane of reconstruction and a particular value for  $\psi$ . Intuitively, we can see that we should be able to determine the plane of reconstruction quite accurately. However, the value of  $\psi$  is more difficult to determine because it is dependent on the precise results of the monocular reconstruction [14,15].

The actual parameterizations of  $\sigma_1$  and  $\sigma_2$  assumed are as follows:

$$\sigma_1 = 20^\circ e^{-1.5 \log_{10} E_{\text{EeV}}} + 4^\circ \quad (5)$$

and

$$\sigma_2 = 100^\circ e^{-0.5 \Delta\chi} + 0.4^\circ. \quad (6)$$

Here,  $E_{\text{EeV}}$  is the primary energy of the shower in EeV. For the purpose of this study, the energy will

be allowed to vary between  $10^{18.5}$  and  $10^{20}$  eV with a differential spectral index of  $-2.7$ . In this scenario, a shower with a primary energy of  $10^{18.5}$  eV will have  $\sigma_1 = 13.4^\circ$ , while a shower with a primary energy of  $10^{20}$  eV will have  $\sigma_1 = 5.0^\circ$ . This difference can be attributed to the fact that larger showers have better defined profiles and a better signal-to-noise ratio.

The factor,  $\Delta\chi$ , in Eq. (6) is the angular track length (in degrees) of the shower as observed by the detector, which is allowed to vary between  $8^\circ$  and  $30^\circ$ . A shower with an observed track length of  $8^\circ$  will have  $\sigma_2 = 2.2^\circ$ , while a shower with an observed track length of  $30^\circ$  will have  $\sigma_2 = 0.4^\circ$ ; a longer track-length leads to a more accurate determination of the plane of reconstruction. The distribution of  $\sigma_2$  values is largely independent of energy because while higher energy showers do lead to more longitudinal development, they are also brighter, which allows one to observe them at greater distances. These competing factors lead to  $\Delta\chi$  distributions that are virtually identical across the observed spectrum.

In general, it should be noted the distributions of  $\sigma_1$  and  $\sigma_2$  values are relatively insensitive to the differential spectral index that is chosen. We ascertained this by considering two sets, one with a differential spectral index of  $-2.5$  and one with a differential spectral index of  $-3.5$ . Even for a variation that was much larger than the accepted range of experimental values for the the UHECR spectrum [14–17], the value of  $\bar{\sigma}_1$  increased by only 11%. The value of  $\bar{\sigma}_2$  remained unchanged. This can be explained by realizing for a steeply falling spectrum, the overwhelming majority of observed events in either case will occur in the first half decade of the measurement. This is a very small effect compared to the expected statistical fluctuations that would occur between two consecutive sets of observations.

For the purpose of calculating  $D_1$ , we can treat the arrival direction of each individual shower as a two dimensional elliptical Gaussian distribution with the parameters  $\sigma_1, \sigma_2$ . The size of a bin's edge,  $\epsilon$ , will be allowed to take on a series of values,  $\Delta\theta$ , which will be in an interval corresponding to a scale length of the sample or in the case of a smooth distribution, the smallest value that is

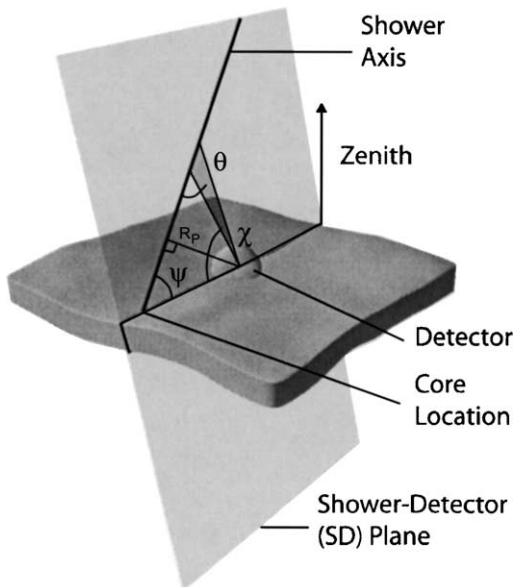


Fig. 1. The geometry of reconstruction for a monocular air fluorescence detector.

computationally feasible. For finite events samples, we will use  $\Delta\theta \simeq 0.5^\circ$ . In the case of smooth distributions we will use a computationally limited value of  $\Delta\theta = 1/6^\circ$ . The number of points in each shower direction distribution,  $N_{\text{Dist}}$ , will be determined by the mean value,  $\langle n_i \rangle$ , necessary to assure that the fractional Gaussian fluctuations of the count,  $n_i$  in each bin, do not on average, exceed a predetermined value (i.e. for 5% fluctuations,  $N_{\text{Dist}} \simeq 500$ ). For each value of  $\epsilon$ , the probability,  $P_i(\epsilon)$ , for the  $i$ th bin will be:

$$P_i(\epsilon) = \frac{n_i}{N_{\text{Dist}} \cdot N_{\text{Shower}}}. \quad (7)$$

We calculate  $D_I$  for each value of  $\epsilon$ :

$$D_I(\epsilon) = - \sum_{i=1}^N \frac{P_i(\epsilon) \log P_i(\epsilon)}{\log 1/\epsilon}. \quad (8)$$

We then determine  $D_I$  to be  $\langle D_I(\epsilon) \rangle$  over the specified interval of  $\epsilon$  values.

### 3.2. Latitudinal binning

For the purpose of calculating  $D_I$ , it is necessary that all bins be equi-sized and equi-shaped as we vary the size of the side of the bins,  $\epsilon$ . While it is impossible to achieve completely this criterion on the surface of a sphere, we will be able to do so approximately by adopting a latitudinal binning scheme.

Latitudinal binning is achieved by first dividing the sky into  $N_\delta$  declinational ( $\delta$ ) bands where each band has a width

$$\Delta\theta = \frac{\pi}{N_\delta}. \quad (9)$$

For each declinational band, the sky is then divided into  $N_{\alpha,\delta}$  bins in right ascension ( $\alpha$ ) where:

$$N_{\alpha,\delta} = \left[ \frac{2\pi \int_{\delta_1}^{\delta_2} \cos \delta d\delta}{(\Delta\theta)^2} \right] = \left[ \frac{2(N_\delta)^2 \int_{\delta_1}^{\delta_2} \cos \delta d\delta}{\pi} \right]. \quad (10)$$

The solid angle,  $\Delta\Omega_\delta$  of each bin (in steradians) is:

$$\Delta\Omega_\delta = \frac{2\pi \int_{\delta_1}^{\delta_2} \cos \delta d\delta}{N_{\alpha,\delta}} \quad (11)$$

with a minimum value of  $(\Delta\theta)^2$  (at the equator) and a maximum value of  $\frac{\pi}{3}(\Delta\theta)^2$  (at the poles) regardless of the value of  $N_\delta$ . This provides us with bins that are all almost the same area and nearly square-shaped (with the exception of three triangular bins at each pole). The total number of bins in the sky can be approximated by:

$$N_{\text{sky}} \simeq 4\pi \left( \frac{N_\delta}{\pi} \right)^2 = \frac{4}{\pi} N_\delta^2. \quad (12)$$

In Fig. 2, we visualize the latitudinal binning technique for a series of different  $N_\delta$  values.

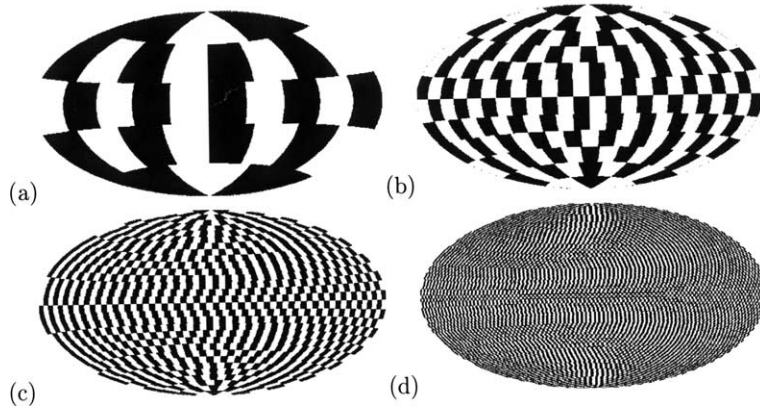


Fig. 2. Hammer-Aitoff projection of latitudinal bins for four different values of  $N_\delta$ —(a)  $N_\delta = 5$ ; (b)  $N_\delta = 12$ ; (c)  $N_\delta = 30$ ; (d)  $N_\delta = 90$ .

### 3.3. Application to the calculation of $D_I$

We can now apply the preceding machinery to the calculation of  $D_I$ : First, we need to normalize the event count in each bin by its respective bin area,  $\Delta\Omega_\delta$

$$P_i(\epsilon) = \frac{n_i(\Delta\theta)^2}{N_{\text{Dist}}N_{\text{Shower}}\Delta\Omega_\delta} = \frac{n_i\pi^2}{N_{\text{Dist}}N_{\text{Shower}}(N_\delta)^2\Delta\Omega_\delta}. \quad (13)$$

If we then realize that  $\epsilon = \frac{1}{N_\delta}$ , we can obtain:

$$D_I(N_\delta) = -\frac{1}{\log N_\delta} \sum_{i=1}^N P_i(N_\delta) \log P_i(N_\delta). \quad (14)$$

This expression is reminiscent of the general formula for entropy from statistical mechanics:

$$S = -k \sum_r p_r \log p_r; \quad (15)$$

where  $p_r$  is the probability of a particle being the  $r$ th state and  $k$  is the Boltzmann constant, which can be thought of as a scaling constant based upon the intrinsic scale of the given particle. The information dimension,  $D_I$ , is an analogous measurement of the heterogeneity of a given data set.

## 4. Calculating $D_I$ for exposures of different source models

### 4.1. Exposure-independent source descriptions

We began by examining four different source models independently of detector exposure. This allows us to calculate the value of the information dimension,  $D_I$ , without consideration to the detector aperture or statistical fluctuations from a finite event sample. The source models are: an isotropic source model, a dipole source model, a model with seven sources superimposed on an isotropic background, and a dark matter halo source model.

#### 4.1.1. Isotropic model

The first model that we will consider is an isotropic source model with distribution:

$$n_{\text{isotropic}} = 1. \quad (16)$$

#### 4.1.2. Dipole model

The second is the Centaurus A dipole source model first proposed by Farrar and Piran [18]. This model has a distribution of arrival directions characterized by a scaling parameter,  $\alpha$ , which can take on any value between  $-1$  and  $+1$  and by  $\theta$ , which is the opening angle between a given event arrival direction and the center of the dipole distribution at Centaurus A. The overall distribution is:

$$n_{\text{dipole}} = 1 + \alpha \cos \theta. \quad (17)$$

#### 4.1.3. Discrete source model

The third model that we will consider is one with seven discrete sources superimposed on an isotropic background. For simplicity's sake, we will assume that all seven sources have an equal intensity indirectly determined by a parameter,  $F_s$ , which will be defined as the fraction of the entire event sample which originates in the seven sources. We will define our source direction to be the centroids of the seven hypothetical point sources proposed by the AGASA collaboration [1]. The equatorial coordinates used for each point source are listed in Table 1.

The arrival directions for the events from each source are assumed to be subjected to magnetic smearing. That is, in the course of traveling through space from the source to the point of observation, the velocity vector of the event is subject to bending in the galactic and extra-galactic magnetic fields. We assumed that this bending produces an apparent source that can be characterized by a Gaussian distribution:

$$P(\Delta\theta) = \frac{\Delta\theta}{\lambda^2} e^{-\frac{(\Delta\theta)^2}{2\lambda^2}}, \quad (18)$$

Table 1  
Coordinates used for the centers of seven discrete sources

Cluster	Right ascension	Declination
C1	01h13m	20.6°
C2	11h17m	56.9°
C3	18h51m	48.2°
C4	04h38m	30.0°
C5	16h02m	23.3°
C6	14h11m	37.4°
C7	03h03m	55.5°

These coordinates correspond to the centers of the seven clusters reported by the AGASA Collaboration [1].

where  $P(\Delta\theta)$  is the probability that an event will be observed with an opening angle,  $\Delta\theta$ , from the nominal direction of the apparent source. We will also assume that the arrival directions of the events from all the sources are subject to the same degree of magnetic smearing, parameterized by  $\lambda_0 = 1.5105 \cdot \lambda$ . The parameter,  $\lambda_0$ , corresponds to the 68% confidence interval in  $\Delta\theta$ . For this paper, we will assume that  $\lambda_0 = 5^\circ$ . It should be noted that the nominal direction of the apparent source is not necessarily the direction of the *actual* source because the possibility exists that the path of the events in question traveled through large regions of homogeneous magnetic fields.

#### 4.1.4. Dark matter halo model

The fourth model that we will consider is a dark matter halo source model. Dark matter halos are characterized by a density profile that is assumed to take the Navarro–Frenk–White (NFW) form [19]:

$$\rho_{\text{NFW}} = \frac{\rho_0}{r(1 + r/r_s)^2}, \quad (19)$$

where  $\rho_0$  is a dark matter density parameter,  $r$  is the distance from the center of the halo, and  $r_s$  is a critical radius. For our source model, we will consider the contribution of only the four closest significant dark matter halos: the Milky Way, M31, LMC, and M33. We will assume that  $\rho_0$  is the same for all four sources and that  $r_s$  scales with the cube root of the luminosity,  $L^{1/3}$ . Thus the

Milky Way will have:  $r_{s,\text{MW}} = 10.0$  kpc, LMC will have  $r_{s,\text{LMC}} = 0.3 \cdot r_{s,\text{MW}} = 3.0$  kpc, M31 will have  $r_{s,\text{M31}} = 1.5 \cdot r_{s,\text{MW}} = 15.0$  kpc, and M33 will have  $r_{s,\text{M33}} = 0.4 \cdot r_{s,\text{MW}} = 4.0$  kpc.

We now calculate the information dimension,  $D_1$ , for each of the our four models. Since these are smooth distributions with no statistical fluctuations, we will only use one, computationally limited value for the number of declinational bands for each model of  $N_\delta = 1080$  (i.e.  $\Delta\theta = (1/6)^\circ$ ), which implies:

$$P_i = n_i \left[ \sum_i n_i \right]^{-1}. \quad (20)$$

Using Eqs. (14) and (20) we can now calculate  $D_1$  for each of the four models. The results are in Table 2 column 1. Note that three of the four values of  $D_1$  exceed the analytical limit of 2 for a 2-D surface.

If we consider the analytic limit for the isotropic case, we have:

$$D_1 = -4 \frac{(N_\delta)^2}{\pi} \frac{P_i \log P_i}{\log N_\delta}. \quad (21)$$

If we then substitute in Eq. (20) we get:

$$D_1 = \frac{2 \log N_\delta + \log 4/\pi}{\log N_\delta}; \quad \lim_{N_\delta \rightarrow \infty} D_1(N_\delta) = 2. \quad (22)$$

The reason we obtain values greater than 2 is because we are working with a finite number of elements on a surface where the total area does not equal  $(N_\delta)^2$ .

Table 2  
Intrinsic  $D_1$  values for the different source models that we examined in Figs. 3 and 5

Source model	1	2
	$D_1$ for source model <i>without</i> detector exposure	$D_1$ for source model <i>with</i> detector exposure
Isotropic	2.035	1.967
Dipole enhancement	2.007	1.945
Seven source	2.033	1.946
Dark matter halo model	1.999	1.978

The estimated values of  $D_1$  with  $N_\delta = 1800$  for the four proposed source models for the entire sky independent of any real detector's aperture (1) and for the four source models superimposed on the estimated aperture of an air-fluorescence detector at  $40^\circ$  N. These values are mathematical descriptors of a data set that are whose number of significant digits are determined by how extensively each bin is sampled (in this case four digits). In the case of a real data set with a finite number of the observations, the number of significant digits is constrained by the fluctuations inherent to the sample size.

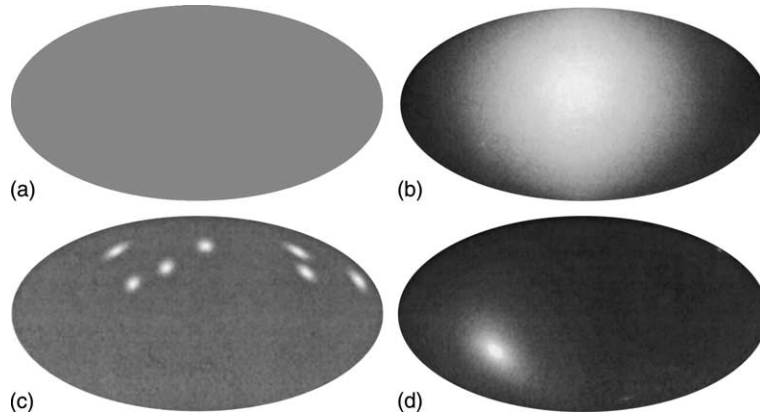


Fig. 3. Density profiles for four different source models—(a) isotropic model; (b) dipole enhancement model ( $\alpha = 1.0$ ); (c) seven source model ( $F_s = 0.28$ ); (d) dark matter halo model ( $r_s = 10$  kpc). All four figures are shown in a Hammer–Aitoff projection of equatorial coordinates (right ascension right to left). The highest density in each panel corresponds to the lightest (red) regions, the lowest density to the darkest (blue) regions. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

#### 4.2. Exposure-dependent source descriptions

For the purpose of this paper, we assume a hypothetical air-fluorescence detector located at  $40^\circ$  N. This analysis assumes an isotropic distribution for the azimuthal component of the arrival directions and a zenith angle distribution that remains constant in time. This is what one would expect for a detector with  $360^\circ$  coverage with identical detector units and stable atmospheric conditions. The acceptance of our detector can thus be defined by two distributions: zenith angle and sidereal time which are shown in Fig. 4. The zenith angle distribution is characterized by 100% acceptance until  $\sim 50^\circ$ , at which point it drops off dramatically due to the lack of a well-defined profile to assist monocular reconstruction. The sidereal time distribution is the combination of the seasonal availability of dark, moonless sky at  $40^\circ$  N and seasonal climatic changes in a desert locale (the rainy season is assumed to extend from February to May which results in a loss of  $\sim 30\%$  of exposure).

By defining acceptance this way, we can calculate the exposure of the detector. By also considering the finite, asymmetric angular resolution,

we can then superimpose the detector exposure upon the various source models that we previously examined (Fig. 3 and Table 2, column 1). We then obtain an effective detector response for the air fluorescence detector for each of our four source models. The results are shown in Fig. 5. It should be emphasized that effective detector effect is due to the *combined* effect of asymmetric sky coverage and angular resolution smearing. A remarkable consequence of this combination is the possibility that point sources can take on an apparently asymmetric shapes due to preferential orientations of the plane of reconstruction for events arriving from a specific location in the physical sky.

We can now determine the value of  $D_1$  using the same method as before. The results are shown in Table 2, column 2. It is interesting to note that the dark matter halo source model now has a larger value for  $D_1$  than the isotropic source model. By looking at Fig. 5, one can verify that the superposition of the detector exposure and source models actually yields a more uniform apparent distribution for the dark matter halo source model than it does for the isotropic source model.

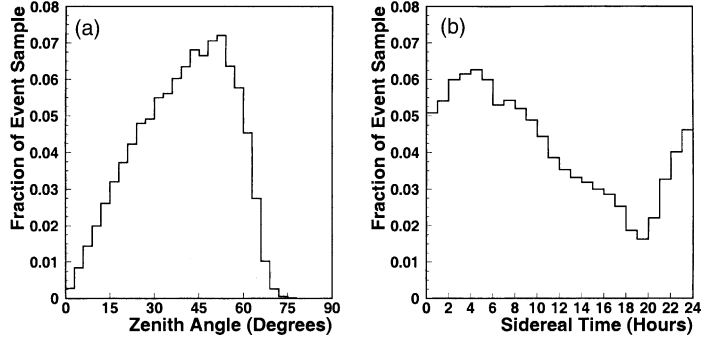


Fig. 4. (a) The distribution of zenith angles for a monocular air-fluorescence detector; (b) the distribution sidereal times for a air-fluorescence detector located at  $40^\circ$  N in a desert locale (right).

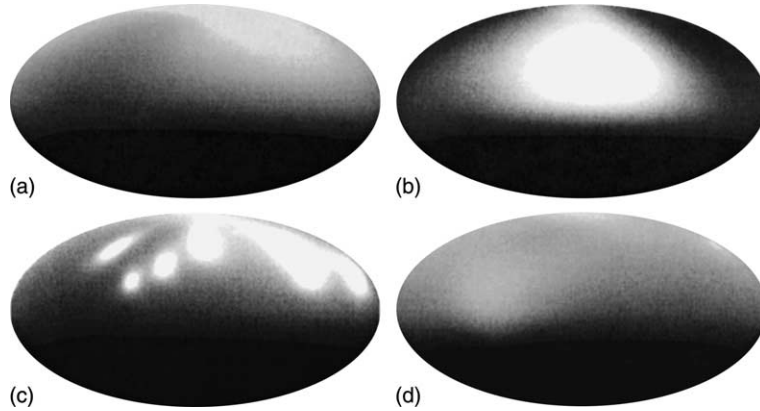


Fig. 5. Effective detector response for the four different source models—(a) isotropic model; (b) dipole enhancement model ( $\alpha = 1.0$ ); (c) seven source model ( $F_s = 0.28$ ); (d) dark matter halo model ( $r_s = 10$  kpc). All four figures are shown in a Hammer–Aitoff projection of equatorial coordinates (right ascension right to left). The highest density in each panel corresponds to the lightest (red) regions, the lowest density to the darkest (blue) regions. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

## 5. Calculating $D_1$ for finite event samples

So far, we have only considered calculating  $D_1$  for smooth distributions. From an experimental standpoint, it is very difficult to collect enough data to obtain a smooth distribution. This is especially true for UHECRs. In order to make a measurement of  $D_1$ , we must first determine what value(s) we should assign to  $N_\delta$ . A reasonable approach is to assign a scale length to our sample. Choosing  $\Delta\theta = 0.5^\circ$ , which approximately reflects the lowest value that can be obtained from  $\sigma_2$  in Eq. (6), yields  $N_\delta = 360$ .

However, it can be beneficial to actually calculate  $D_1$  for a range of values around  $N_\delta$ . In Fig. 6(a), we display the values  $D_1(N_\delta)$  over the range  $N_\delta = [354, 360]$  for two separate simulated sets where  $N_{\text{Shower}} = 500$ . These sets yield very similar values for  $D_1$  over the full range of values for  $N_\delta$ . However, if we examine Fig. 6(b) we can see that for the same two finite samples, the fractional difference between values of  $D_1$  can fluctuate substantially even over small intervals of  $N_\delta$ . While these fluctuations are typically much smaller than the difference in  $D_1$  values between any two sets, we will take  $D_1$  to be  $\langle D_1(N_\delta) \rangle$  for the interval:



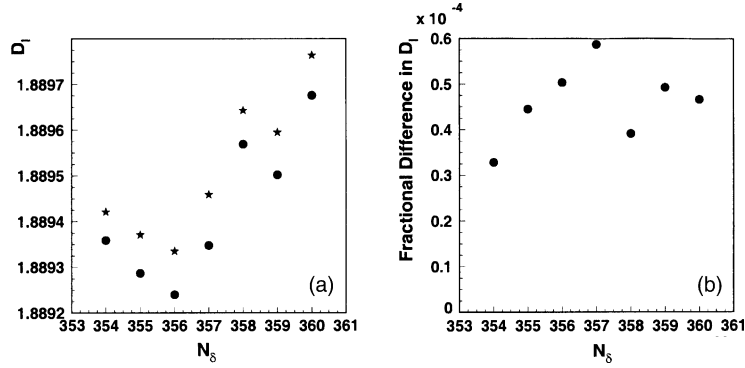


Fig. 6. (a)  $D_1$  over a range of values of  $N_\delta$  for two similar finite event sets that have similar values for  $D_1$  (the dots indicate the values  $D_1$  for the first set while the stars indicate the values  $D_1$  for the second set); (b) the fractional difference in  $D_1$  for the same two sets over the range of  $N_\delta$  values.

$N_\delta = [354, 360]$  in order to minimize the statistical fluctuations in  $D_1$  between individual sets. This range was chosen to optimize our computational ability.

To account for the fact that we no longer have a smooth distribution for the calculation of the values of  $P_i$ , we refer back to Eq. (13) to see how to calculate  $D_1$  for a finite set of elements. This only requires us to determine a value for  $N_{\text{Dist}}$ . We can set this value based upon what value we wish for  $n_i$  in combination with Eq. (12) (i.e.  $\langle n_i \rangle = \frac{1}{(\Delta n)^2}$  where

$\Delta n$  is the fractional Gaussian fluctuation of a bin with  $n_i = \langle n_i \rangle$ ). Then,

$$N_{\text{Dist}} = \frac{\frac{4}{\pi} (N_\delta)^2 \langle n_i \rangle}{N_{\text{Shower}}} \quad (23)$$

If we then combine Eqs. (9), (13) and (23); we obtain:

$$P_i(N_\delta) = \frac{n_i}{\langle n_i \rangle} \frac{\pi^3}{4(N_\delta)^4 \Delta \Omega_\delta} \quad (24)$$

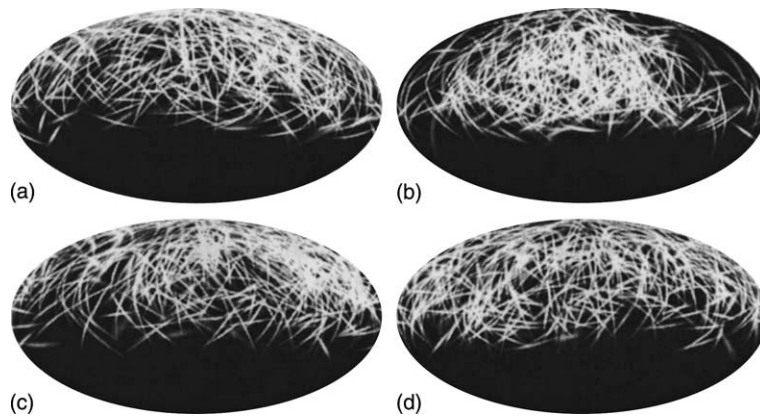


Fig. 7. Simulated 500 event distributions for four different source models—(a) isotropic model; (b) dipole enhancement model ( $\alpha = 1.0$ ); (c) seven source model ( $F_s = 0.28$ ); (d) dark matter halo model ( $r_s = 10$  kpc). All four figures are shown in a Hammer–Aitoff projection of equatorial coordinates (right ascension right to left). The highest density in each panel corresponds to the lightest (red) regions, the lowest density to the darkest (blue) regions. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

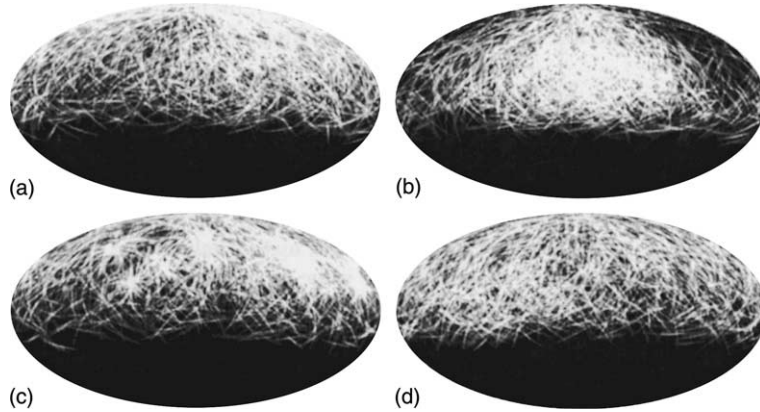


Fig. 8. Simulated 2000 event distributions for four different source models—(a) isotropic model; (b) dipole enhancement model ( $\alpha = 1.0$ ); (c) seven source model ( $F_s = 0.28$ ); (d) dark matter halo model ( $r_s = 10$  kpc). All four figures are shown in a Hammer–Aitoff projection of equatorial coordinates (right ascension right to left). The highest density in each panel corresponds to the lightest (red) regions, the lowest density to the darkest (blue) regions. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

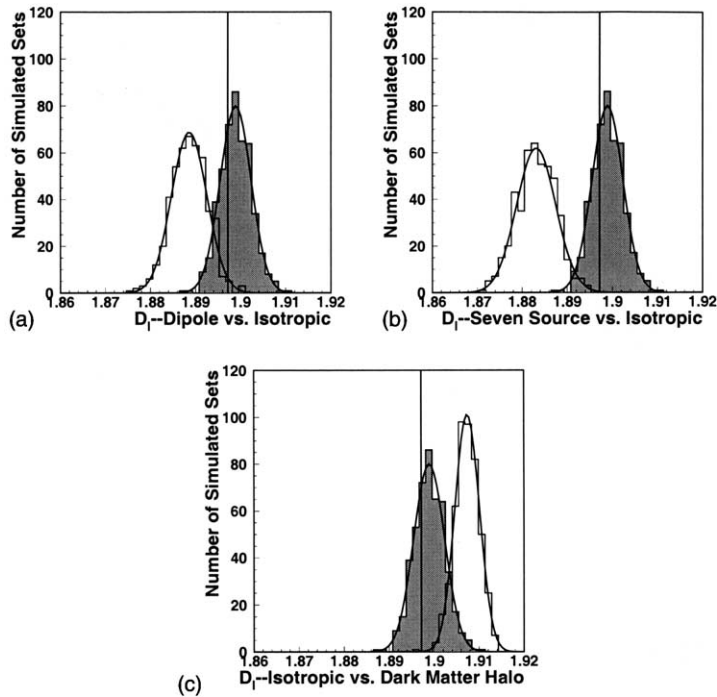


Fig. 9. Comparison of distributions of  $D_1$  values for  $\sim 500$  sets of 500 events between the isotropic source model (shaded) and the other three source models—(a) dipole enhancement model vs. isotropic model ( $\alpha = 1.0$ ); (b) seven source model vs. isotropic model ( $F_s = 0.28$ ); (c) isotropic model vs. dark matter halo source model ( $r_s = 10$  kpc). The vertical line corresponds to the value of  $D_1$  for the “real” 500 event sample. In all cases the distributions of  $D_1$  values fit well to a Gaussian curve.

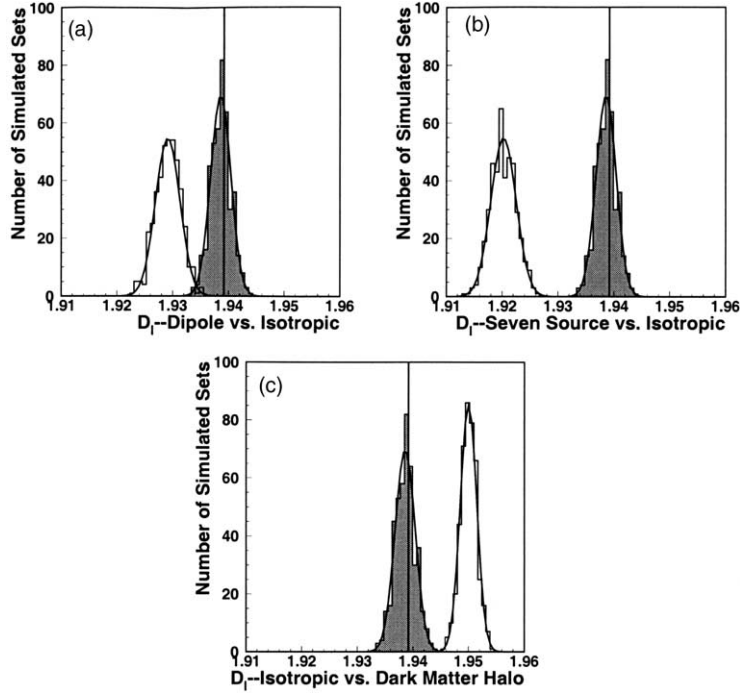


Fig. 10. Comparison of distributions of  $D_I$  values for  $\sim 500$  sets of 2000 events between the isotropic source model (shaded) and the other three source models—(a) dipole enhancement model vs. isotropic model ( $\alpha = 1.0$ ); (b) seven source model vs. isotropic model ( $F_s = 0.28$ ); (c) isotropic model vs. dark matter halo source model ( $r_s = 10$  kpc). The vertical line corresponds to the value of  $D_I$  for the “real” 2000 event sample. In all cases the distributions of  $D_I$  values fit well to a Gaussian curve.

We can then calculate  $D_I$  from Eq. (14):

$$D_I = \left\langle -\frac{1}{\log N_\delta} \sum_{i=1}^N P_i(N_\delta) \log P_i(N_\delta) \right\rangle, \quad (25)$$

$$N_\delta = [354, 360].$$

Thus, If we want  $\langle n_i \rangle = 500$  and  $N_{\text{Shower}} = 500$ , we find that  $N_{\text{Dist}} \simeq 1.65 \times 10^5$ . If  $N_{\text{Shower}} = 2000$ , we have  $N_{\text{Dist}} \simeq 4.1 \times 10^4$ .

We now consider two cases: finite event sets with 500 events and finite event sets with 2000 events. These sets will have the angular resolution characteristics described in Eqs. (5) and (6). The exposure will be modeled via the zenith angle and sidereal time distributions shown in Fig. 4. Fig. 7 contains examples of event sets with all four source models and  $N_{\text{Shower}} = 500$  and Fig. 8 contains examples of events sets with all four previously described source models and  $N_{\text{Shower}} = 2000$ .

In Figs. 7 and 8 one can see that these distributions of arrival directions have a far greater

degree of statistical fluctuation than the smooth distributions shown in Fig. 5. Because of the fluctuations in our simulated event samples, the value of  $D_I$  varies significantly (see Figs. 9 and 10) from one simulated set to the next. In Figs. 9 and 10, we examine the distribution of  $D_I$  values for  $\sim 500$  sets (of 500 and 2000 events respectively). We see that the distribution of 500 event samples have both a lower mean value and larger-width than the 2000 event samples.

## 6. Application to anisotropy analysis

We now need to develop a scheme by which we can apply fractal dimensionality analysis to a real data set. In the case of a real data set, we will be dealing with only *one* value of  $D_I$ . By itself  $D_I$  is insufficient to characterize the data set;  $D_I$  fluctuates a great deal due to variation in  $N_{\text{Shower}}$ . However, a comparison between the value of  $D_I$

for a real event sample and a distribution of  $D_I$  values (with the *same*  $N_{\text{Shower}}$  and  $N_{\text{Dist}}$  values as the real data) for a series of simulated data sets of a given source model does provide a viable measurement of anisotropy.

We can demonstrate this by considering a *single* simulated event sample generated with an isotropic source model. We will suppose this sample to be our “real” data. We consider the isotropic simulated event samples shown in Figs. 7(a) and 8(a). We will once again stipulate that  $\langle n_i \rangle = 500$  which means that in the case of  $N_{\text{Shower}} = 500$ , we have:  $N_{\text{Dist}} \simeq 1.6 \times 10^5$  which leads to  $D_I = 1.89715$  and in the case of  $N_{\text{Shower}} = 2000$ , we have:  $N_{\text{Dist}} \simeq 4 \times 10^4$  which leads to  $D_I = 1.93920$ .

In Figs. 9 and 10, we demonstrated that for a fixed scaling parameter the distribution of  $D_I$  values for a large number of simulated sets fits well to a Gaussian curve. By establishing the relationship

between the distribution of  $D_I$  values and the scaling parameter in each model, we can establish a 90% confidence interval on the scaling parameter for that model.

### 6.1. Dipole enhancement source model

In the case of the dipole enhancement source model in Eq. (17), the scaling parameter is  $\alpha$ . By varying  $\alpha$  between  $-1$  and  $1$ , we develop a curve which will show the relationship between  $D_I$  and  $\alpha$ . By considering the actual value of  $D_I$  for the “real” data set, we then establish a nominal value for  $\alpha$  and a 90% confidence interval. The results for both  $N_{\text{Shower}} = 500$  and  $N_{\text{Shower}} = 2000$  are shown in Fig. 11. In the case of our simulated isotropic set with  $N_{\text{Shower}} = 500$ ,  $\alpha = 0.02 \pm 0.21$  with a 90% confidence interval of  $[-0.29, 0.36]$ . In the case of our simulated isotropic set with  $N_{\text{Shower}} = 2000$ ,  $\alpha =$

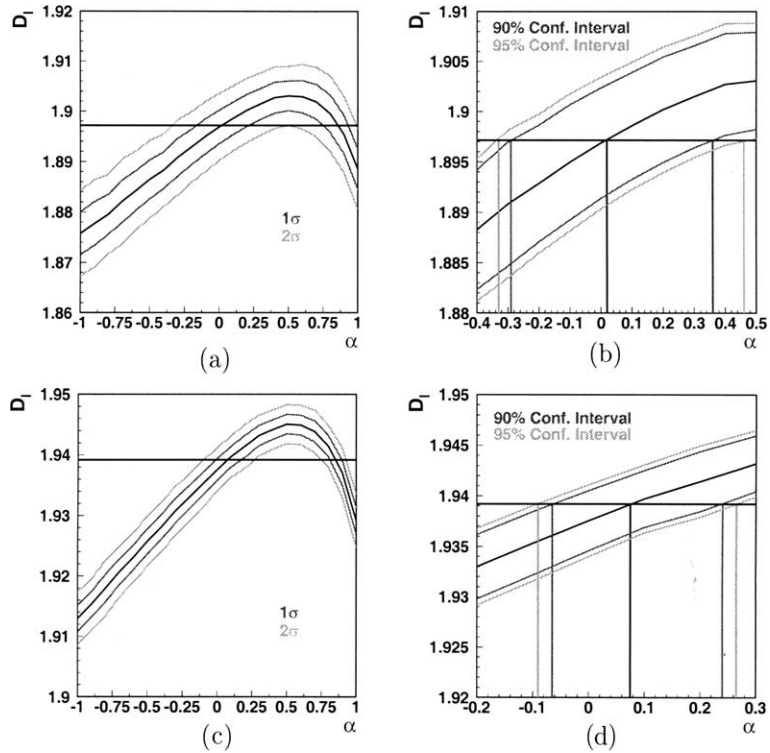


Fig. 11. Study of the dependence of  $D_I$  upon  $\alpha$  for a dipole enhancement source model—(a)  $N_{\text{Shower}} = 500$ ;  $\alpha = 0.02 \pm 0.21$ ; (b)  $N_{\text{Shower}} = 500$  (zoomed); 90% confidence interval:  $[-0.29, 0.36]$ ; (c)  $N_{\text{Shower}} = 2000$ ;  $\alpha = 0.075 \pm 0.085$ ; (d)  $N_{\text{Shower}} = 2000$  (zoomed); 90% confidence interval:  $[-0.065, 0.24]$ . In each case, the solid horizontal line indicates the value of  $D_I$  for the simulated isotropic data set. The vertical lines in (b) and (d) indicate the projection of the nominal value and 90% and 95% confidence intervals of  $\alpha$  on the  $x$ -axis.

$0.075 \pm 0.085$  with a 90% confidence interval of  $[-0.065, 0.24]$ . These results tell us that our “real” data can, at the  $1\sigma$  level of up to 0.2. We notice that in Fig. 11, there are potentially two suitable intervals of  $\alpha$  that possess similar values of  $D_I$ . This emphasizes the necessity of making density plots to assure by visual inspection that the appropriate interval is chosen.

### 6.2. Seven source model

In the case of the seven source model, the scaling parameter is  $F_s$ , the fraction of the total

event sample that is produced by the discrete sources. By varying  $F_s$  between 0 and 0.40, we develop a curve which shows the relationship between  $D_I$  and  $F_s$ . By considering the actual value of  $D_I$  for the “real” data set, we then establish a nominal value for  $F_s$  and a 90% confidence upper limit for  $F_s$ . The results for both  $N_{\text{Shower}} = 500$  and  $N_{\text{Shower}} = 2000$  are shown in Fig. 12. In the case of our simulated isotropic set with  $N_{\text{Shower}} = 500$ , we have a 90% confidence upper limit of  $F_s = 0.16$ . In the case of our simulated isotropic set with  $N_{\text{Shower}} = 2000$ , we have a 90% confidence upper limit of  $F_s = 0.04$ . This tells us that our “real” data

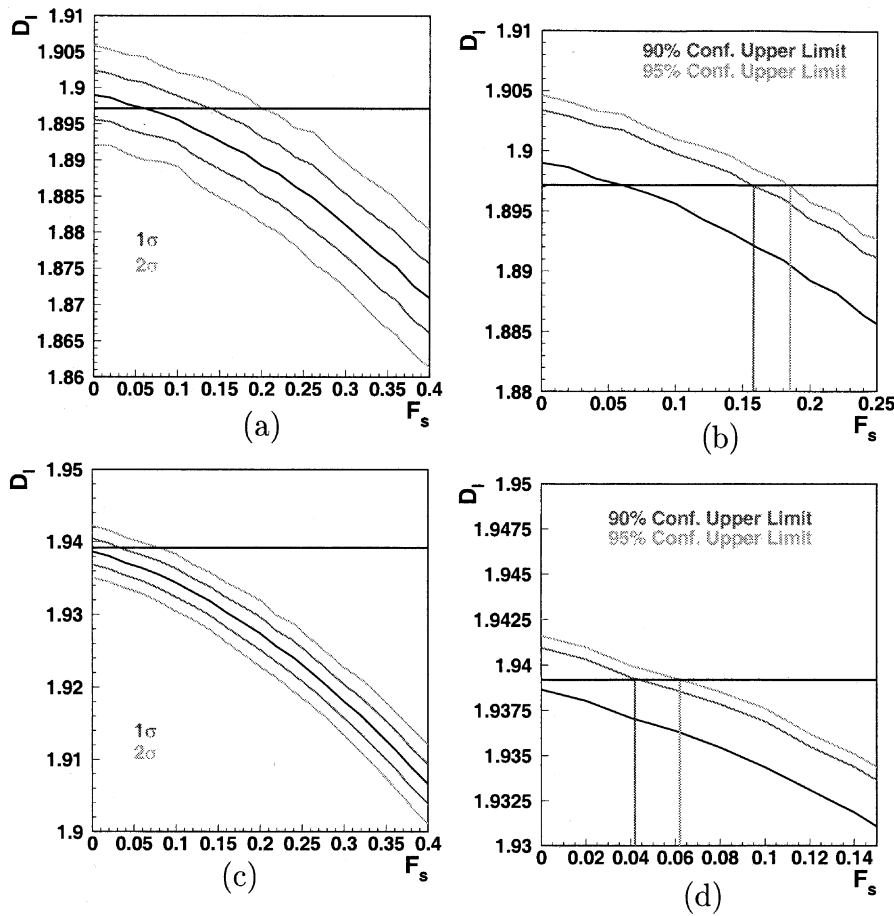


Fig. 12. Study of the dependence of  $D_I$  upon  $F_s$  for the seven source model—(a)  $N_{\text{Shower}} = 500$ ; (b)  $N_{\text{Shower}} = 500$  (zoomed); 90% confidence upper limit = 0.16; (c)  $N_{\text{Shower}} = 2000$ ; (d)  $N_{\text{Shower}} = 2000$  (zoomed); 90% confidence upper limit = 0.04. In each case, the solid horizontal line indicates the value of  $D_I$  for the simulated isotropic data set. The vertical lines in (b) and (d) indicate the projection of 90% and 95% confidence upper limits of  $r_s$  on the x-axis.

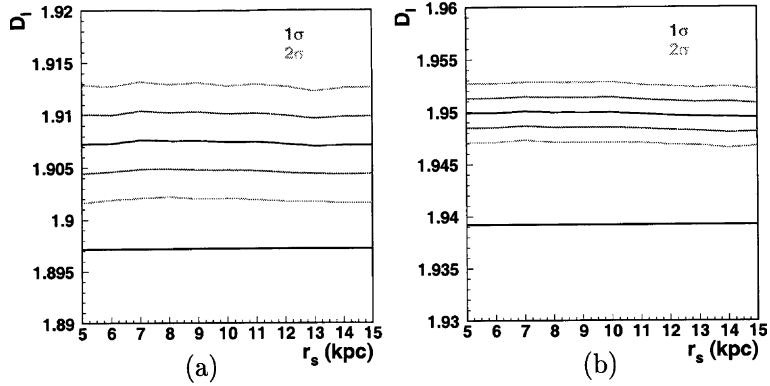


Fig. 13. Comparison of the distribution of  $D_1$  values for a dark matter halo source model with a full range of hypothesized values for  $r_s$ —(a)  $N_{\text{Shower}} = 500$ , the dark matter halo model can be rejected at a level  $\geq 3.6\sigma$ ; (b)  $N_{\text{Shower}} = 2000$ , the dark matter halo model can be rejected at a level  $\geq 7.0\sigma$ . In each case, the solid horizontal line indicates the value of  $D_1$  for the simulated isotropic data set.

can have, at most, 16% (for 500 events) or 4% (for 2000 events) of these events coming from the seven sources.

### 6.3. Dark matter halo source model

In the case of the dark matter halo source model in Eq. (19), the variable parameter is  $r_s$ , the critical radius in the NFW profile [19]. By varying  $r_s$  between 5.0 and 15.0 kpc we develop the curve which will demonstrate the dependence of  $D_1$  upon  $r_s$ . We can then show that the dark matter halo source model can be rejected with at level  $\geq 3.6\sigma$  for  $N_{\text{Shower}} = 500$  and at a level of  $\geq 7.0\sigma$  for  $N_{\text{Shower}} = 2000$  for the full range of hypothesized values for  $r_s$ . The results are shown in Fig. 13.

## 7. Discussion

Fractal dimensionality has several advantages over conventional anisotropy techniques. First of all, it naturally accommodates angular resolution. This is extremely important when considering event sets with asymmetric errors, when analysing event sets with variable values for the angular resolution (e.g. dependent on energy or geometry), or when combining multiple data sets from different detectors for a single analysis.

Another advantage that fractal dimensionality possesses is the ability to accommodate any aper-

ture. Because this method makes a relative comparison between a sample and simulations using the same aperture, the physical aperture is simply folded into the analysis. This once again allows the combination of data from multiple detectors with *very* different apertures. It also allows the analysis of extremely complicated apertures without the need to include normalizing factors that needlessly complicate the predictability of the Poisson fluctuations in the data sample.

Perhaps the most striking feature of the fractal dimensionality method is that it only requires a *single* measurement of one's data. While fractal dimensionality will not always provide better statistical significance than a direct measurement for a specific anisotropy, the fact that one considers only a single measurement of the data, for any number of potential anisotropic models, provides one the means to simultaneously reject or accept all of those models without the ensuing statistical penalty.

A possible way of increasing the sensitivity of the fractal dimensionality method is by considering the general case of  $D_q$ . By varying the value of  $q$  to something other than 1, it might be possible to increase the sensitivity of this method to various anisotropies.

The fractal dimensionality method does have some weaknesses. First and foremost is the potential for multiple solutions as was demonstrated in the dipole source model above. The method cannot be applied blindly. It requires a careful

inspection of both the data sample and the simulated samples in order to resolve possible ambiguities. Another drawback is the amount of computation required to calculate  $D_I$  for a large number of simulated data sets. Producing just the plots in Fig. 11(c) consumed the equivalent of  $\sim 1000$  CPU hours on 1 GHz Athlon machine. Another limitation is the potential for different source models to effectively cancel each other out and yield a potentially deceiving value of  $D_I$  that resembles that of an isotropic source. One solution for this is to separately consider the value of  $D_I$  for different celestial regions in the data. Of course, this will incur a statistical penalty.

Two particularly useful roles for fractal dimensionality analysis are (a) as a first test to ascertain if a sample possesses the same heterogeneity as the expected isotropic background and (b) as a wholly independent observable that can be used to confirm the results of a direct measurement.

Today the HiRes detector continues to accumulate data. Soon the Auger detector will be acquiring data at seven times the rate of HiRes. As the number of observed UHECRs continues to rise, fractal dimensionality will grow ever more effective in its ability to discern the between potential source models in the continuing effort to solve the mystery of UHECRs.

### Acknowledgements

This work is supported by US NSF grants PHY-9321949 PHY-9974537, and PHY-0140688. We would like to thank Professors Glennys Farrar and Benjamin Bromley for invaluable insights that went into the analysis used to produce this paper. We gratefully acknowledge the contributions from

the University of Utah Utah Center for High Performance Computing.

### References

- [1] M. Takeda et al., *Astrophys. J.* 522 (1999) 225, arXiv:astro-ph/9902239.
- [2] J. Bellido et al., in: *Proc. of 27th ICRC, Hamburg*, vol. 1, 2001, p. 364.
- [3] C.C.H. Jui et al., in: *Proc. of 27th ICRC, Hamburg*, vol. 1, 2001, p. 354.
- [4] N. Hayashida et al., in: *Proc. of 26th ICRC, Salt Lake City*, vol. 3, 1999, p. 256 [arXiv:astro-ph/9906056].
- [5] D.J. Bird et al., in: *Proc. of 23rd ICRC, Calgary*, vol. 2, 1993, p. 51.
- [6] D.J. Bird et al., in: *Proc. of 23rd ICRC, Calgary*, vol. 2, 1993, p. 30.
- [7] D.J. Bird et al., *Astrophys. J.* 511 (1999) 738, arXiv:astro-ph/9806096.
- [8] T. Stanev et al., *Phys. Rev. Lett.* 75 (1995) 3056–3059, arXiv:astro-ph/9505093.
- [9] A.N. Kolmogorov, *Dokl. Akad. Nauk. SSSR* 124 (1959) 754; *Math. Rev.* 21 (1959) 2035.
- [10] J. Balatoni, A. Reyni, *Publ. Math. Inst. Hung. Acad. Sci.* 1 (1959) 9.
- [11] A.H. Nayfeh, B. Balachandran, *Applied Nonlinear Dynamics: Analytical, Computational, and Experimental Methods*, Wiley, New York, 1995, pp. 545–547.
- [12] P. Grassberger, *Phys. Lett. A* 97 (1983) 227.
- [13] H.G.E. Hentschel, I. Procaccia, *Physica D* 8 (1983) 435.
- [14] T. Abu-Zayyad et al., *High Resolution Fly's Eye Collaboration*, arXiv:astro-ph/0208243.
- [15] T. Abu-Zayyad et al., *High Resolution Fly's Eye Collaboration*, arXiv:astro-ph/0208301.
- [16] HIREs Collaboration D.J. Bird et al., *Astrophys. J.* 424 (1994) 491.
- [17] M. Takeda et al., *Phys. Rev. Lett.* 81 (1998) 1163, arXiv:astro-ph/9807193.
- [18] G.R. Farrar, T. Piran, arXiv:astro-ph/0010370.
- [19] J.F. Navarro, C.S. Frenk, S.D.M. White, *Astrophys. J.* 462 (1996) 563, arXiv:astro-ph/9508025.