On the Origin of Ultra High Energy Cosmic rays (UHECR)

A. Calvez
Department of Physics
California Polytechnic State University
San Luis Obispo, CA 93407, USA

ABSTRACT

We reproduce statistical analyses of HiRes and AGASA data on the clustering and correlation of UHECR, we discuss the techniques used, there significance, and their validity, and we describe the application of these techniques on the future data obtained for the Pierre Auger collaboration, in the search for the origin of UHECR.

I. Introduction and Motivation

The existence of cosmic ray particles with energies above 10^{20}eV is one of the intriguing phenomena of astroparticle physics. Speculations about their origin, and the process by which they can be accelerated have been made [1, 2, 3], but so far no definite conclusion has been reached.

The mystery surrounding these particles is increased by some classical theories that put restrictions on the distance of cosmic ray's sources, and on the arrival directions as a function of energy. The Greisen-Zatsepin-Kuzmin (GZK) cutoff [2, 3] limits the average distance traveled by cosmic rays above 6×10^{19}eV to about 50Mpc before it looses 20% of that energy, hence reducing the number of possible allowed sources. Furthermore, it is believed that the interstellar medium is filled with magnetic fields (on the order of microgauss for galactic magnetic field, and nanogauss for extragalactic magnetic field), which would cause charged particles to be deflected. However, the magnitude of the deflection would be strongly related to the amount of energy that the particle carries.

Figure 1a and b are pictorial representation of the deflection as a function of energy in galactic and extragalactic environments. As it may bee seen from the figures the path of the particles begins to be straight at energies neighboring 10^{20}eV, which is really close to where the GZK cutoff starts to take place, leaving a very small window of opportunity in correlating the cosmic rays with possible sources.

A small amount of data containing energy and incoming direction of the cosmic rays has been accumulated by various experiments, *i.e.* the High Resolution Fly's Eye (HiRes), or the Akeno Giant Air Shower Array (AGASA), and the Pierre Auger Project should add to the already existing information in a near future; extensive analyses have been performed to find clustering in the arrival direction of the cosmic rays (autocorrelation), or correlation between cosmic rays and various catalogues of astronomical objects, yet no statistically significant results have been produced, and the origin of UHECR remains a mystery.

In this paper we will reproduce some of the previous analyses that have been performed and discuss possible methods that could be used in the analysis of the Pierre Auger data.

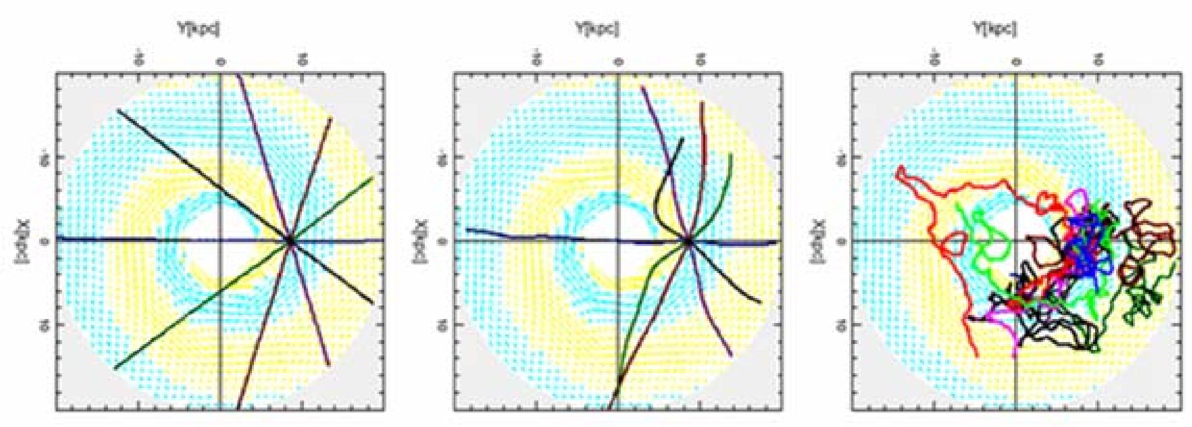

Figure 1a. Deflection of charged particles in galactic magnetic field. From left to right the plots show particles with energies of 10^{20}eV, 10^{19}eV, and 10^{18}eV.

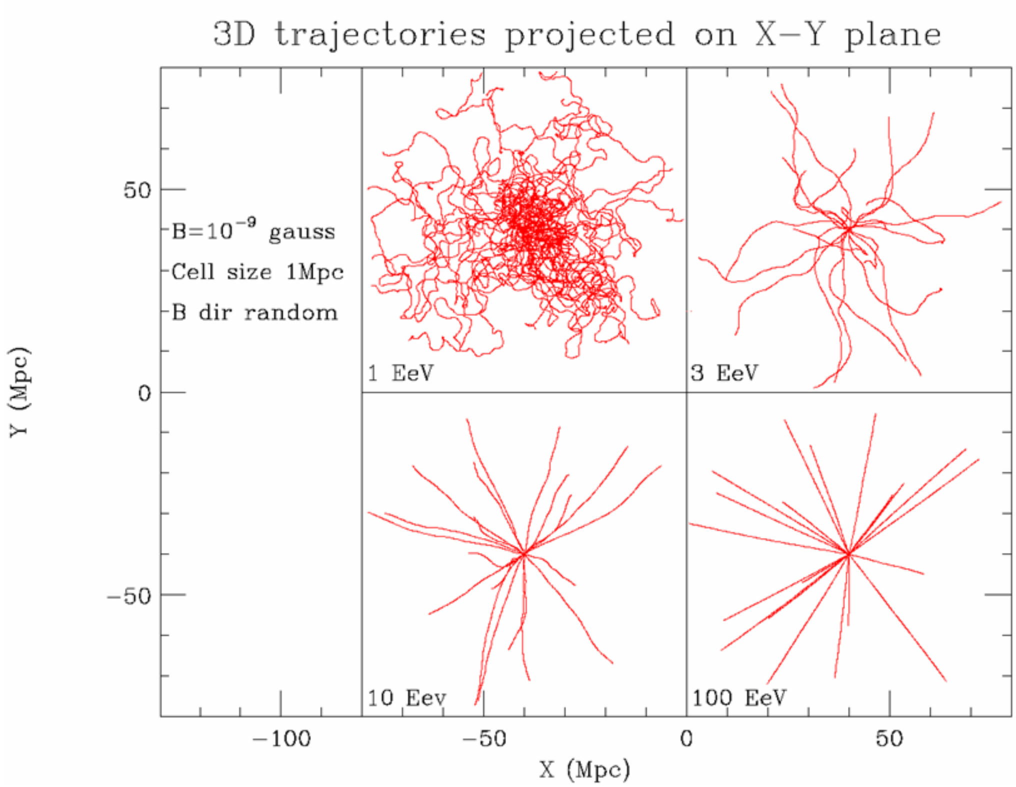

Figure 1b. Deflection of charged particles by extragalactic magnetic fields as a function of energy [10].

II. Methods and Tools

In order to ascertain the significance of the correlation between the origin of UHECR, and the position of objects in the sky (in our case BL Lacertae objects), the chance probability of the correlation must be determined. In other words, it is necessary to calculate the probability of finding correlation between a cosmic ray's origin and a BL Lac object by luck.

The chance probability may be established by means of a Monte Carlo simulation: Given *n* cosmic rays in a data set, and a set of objects in the sky with, for a given angular separation δ, N correlations; a set of *n* cosmic ray's arrival is generated randomly using predefined acceptance on right ascension and declination. For any angle δ, a large number N_{MC} of Monte Carlo simulations are created (i.e. the actual value of N_{MC} must be determined on the go, for example to estimate a probability of 10^{-4} the value of N_{MC} must be on the order of 10^{-5}) and the number of correlations is recorded each time. Given N_{Cor} the number of correlated events during each simulation, the quantity N_{ex} is defined as the number of times the Monte Carlo simulation produces at least as many or more correlations than the original data set (i.e. N_{ex}=Number of Monte Carlo with $N_{Cor}>N$). The Chance probability P for a given angle δ is given by the following equation.

$$P_\delta = \frac{N_{ex}}{N_{MC}} \tag{1}$$

The importance of the correlation may be evaluated by realizing that minimizing the chance probability maximizes the significance of the correlation since it minimizes the probability of correlation between an object and a cosmic ray by luck.

In reality, there are two ways of analyzing the data through the Monte Carlo simulation. The first techniques involves minimizing the chance probability by analyzing the data many time over a large set of angular separation, and by making various cuts on the data (i.e. cosmic ray's energy, object's distance and luminosity, and so on). Unfortunately, this technique involves the use of penalty factors. When analyzing a data set many times over many variations of parameters, the chances of finding a strong correlation's significance increases with the number of trials performed. To compensate for this phenomenon, it is necessary to adjust the chance probability by multiplying by a penalty factor, the value of which depends on the number of trials performed and the number of combinations of parameters. The other reason for using penalty factors has to do with the fact that making cuts *a posteriori* (i.e. deciding on a set of parameters after the data has been scanned), is not statistically valid. The steps described above can be summarized as follow, the data is scanned, and a set of parameters is decided: a hypothesis is obtained out of a set of data. Then the significance is estimated based on this hypothesis: the hypothesis is used to analyze the data that was used to make the hypothesis. This way of analyzing results can easily be assimilated to circular reasoning, hence the use of penalty factors.

On the other hand, if the data is analyzed for a single set of cut, no penalty factor is required. Hence, by making *a priori* cuts, *i.e.* cuts made before the analysis of the data, and based on physical arguments (angular resolution of the detector, energy of the cosmic rays, object's necessary acceleration power); then by analyzing the data set using this particular cut, it is possible to obtain a single chance probability, and a single significance without being subject to penalty factors.

The technique described above can be used in determining the significance of correlation between cosmic rays origin and BL Lacs objects, but it can similarly be used to measure the amount of clustering in a certain data set. This is done by setting N as the number of auto-correlated points in the data set (i.e. points that are separated by less than δ), and by then defining N_{Cor} and N_{ex} as the number of auto-correlated points in the set of randomly generated data points and the number of sets with $N_{Cor}>N$. It is then easy to fine the corresponding the resulting chance probability using Equation 1.

The software developed for this analysis was written in C++, and involved reading the various catalogues of events and the randomly generated set into arrays, making the necessary comparisons and calculation, and writing the results, like the chance probability as a function of angular distance, or the correlations between the BL Lacs and the cosmic rays, to text files; then, ROOT was used to read in these text files and generate the necessary plots. The C++ program takes as arguments the number of Monte Carlo sets to be simulated, the range of angles and the angular step size that the simulation must be ran with, and the desired bin size for the chance probability plots.

III. Reproduction of Previously Obtained Results

The analysis presented in the previous section has previously been applied to data sets from the HiRes and AGASA experiments. These analyses were reproduced as a practice and are presented in details here.

a. HiRes Analysis

The HiRes collaboration has never explicitly published the information regarding their highest energy events. The set of 271 events with energies above 10^{19}eV has been presented in the form of a skymap [4], but individual energies were never divulged. Obtaining the coordinates of the incoming cosmic rays was a challenge in itself as an inverse Hammer-Aitoff projection had to be performed to recover the right ascension and declination from the postscript source code. In other words, the coordinates of the points had to be de-convolved from the map presented in Figure 2.

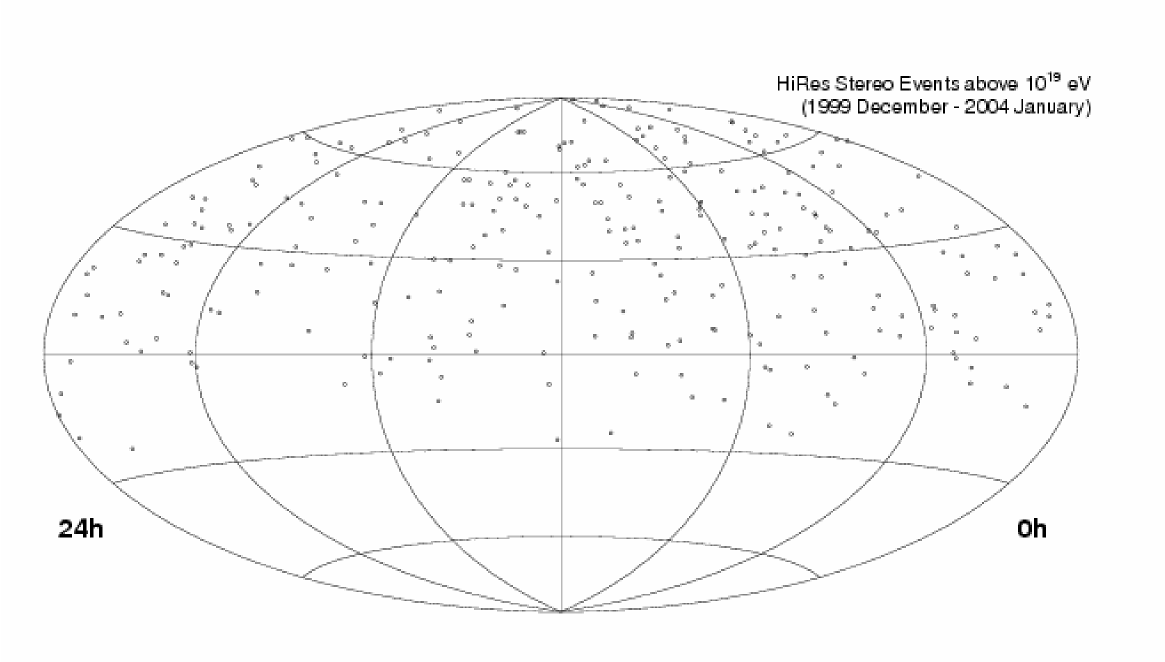

Figure 2. Map of the 271 cosmic rays with energy above 10^{19}eV for the HiRes experiment [4].

Once the coordinates of the cosmic rays were retrieved, it was possible to use a catalogue of BL Lacs [5] to investigate correlation between cosmic rays and BL Lacs objects. The BL Lac Catalogue was modified from its original content as only those 157 confirmed BL Lacs with Visual magnitude smaller than 18 were kept.

Correlation is measured as a function of the angular separation δ, between the cosmic ray's incoming direction and the position of the BL Lac. In other words, given an angular separation δ, we calculate the number of cosmic ray-BL Lacs pairs with angular separation less than δ. This type of analysis of the HiRes data has previously been published [6]; following the technique described in the previous section we will reproduce this analysis. The number of correlated pairs was recorded as a function of δ, where δ ranged from 0^0 to 5^0 in step size of 0.1^0. The BL Lac catalogue was then compared to the Monte Carlo generated set of 271 events to determine the chance probability as a function of angular distance δ.

Note that the HiRes experiment is located in the northern hemisphere at latitude of 40^0 N. Furthermore, HiRes is a fluorescence experiment, hence functioning only during moonless night; consequently the acceptance (or probability distribution of observing a cosmic ray coming from a certain right ascension and declination), is not flat. The acceptance of the HiRes experiment has been published [4], and is presented in Figure 3a.

The data presented in Figure 3 is composed of the actual HiRes data points, along with a Monte Carlo simulation of the acceptance based on the ratio of points in different angle bins. The Monte Carlo simulation of this data was also reproduced, and is presented in Figure 3b. Both graphs have the same general shapes but have slight differences in the distribution of the Monte Carlo. The differences may be the result of a difference in the number of Monte Carlo events produced, or differences in the periodicity of the random number generator. The importance of this information comes from the importance of being able to randomly generate events that have a chance of being observed by the detector.

For every angle δ, 10^5 Monte Carlo set of events were generated (i.e. each Monte Carlo set being a reproduction of the original set of data) and compared to the modified catalogue of BL Lacs. The chance probability (as defined in Equation 1) was calculated and plotted as a function of the angular distance.

The results presented in [6] along with the reproduced results are presented in Figure 4a and b. Note that the graphs look very similar in shape, yet some of the magnitudes seem a little off, furthermore the analysis performed in the paper by Tinyakov and Tkachev only makes use of 156 confirmed BL lacs. These dissimilarities are again attributed to possible differences in the periodicity of the random number generator, but the differences could also be the result of differences in the binning of the acceptance for the distribution of events in right ascension and declination.

Nevertheless, the magnitude of the differences is always rather small.

The theoretical minimum [6] of the chance probability happens at $\delta=0.8^0$, for a chance probability $P_{0.8}=4x10^{-4}$ in the original analysis, and 10^{-3} in the reproduced analysis. At $\delta=0.8^0$ the real data produces 11 correlations with the set of modified BL Lacs for 3 correlations predicted by the Monte Carlo, while an angle of $\delta=1^0$, produces 13 correlations between the real data and the modified BL Lac catalogue, while the Monte Carlo predicts 5 correlations.

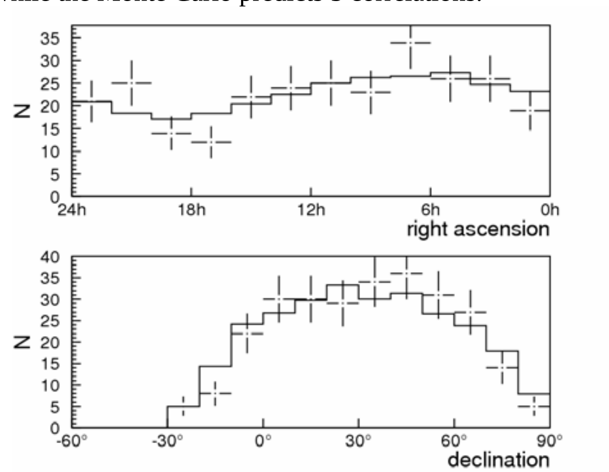

Figure 3a. Acceptance in right ascension and declination for the HiRes experiment, the point represent the actual HiRes data points, and the solid line represents the Monte Carlo simulation obtained from the angular distribution of the data.

Further analysis of the data was performed using the reproduced distribution of cosmic rays. Auto-correlation is defined as the clustering, or the correlation of the data with itself. To measure the chance probability of auto-correlation, one must compare the number of events in the real data that are correlated with another real data point, with the number of similar correlations produced in the randomly generated data set. The result of this analysis is presented in Figure 5.

As it may be seen from Figure 5, there is not much relevance to the auto-correlation of the data as the minimum chance probability is really high, and implies that there is no significant clustering in the HiRes data at energies of 10^{19}eV at any angle.

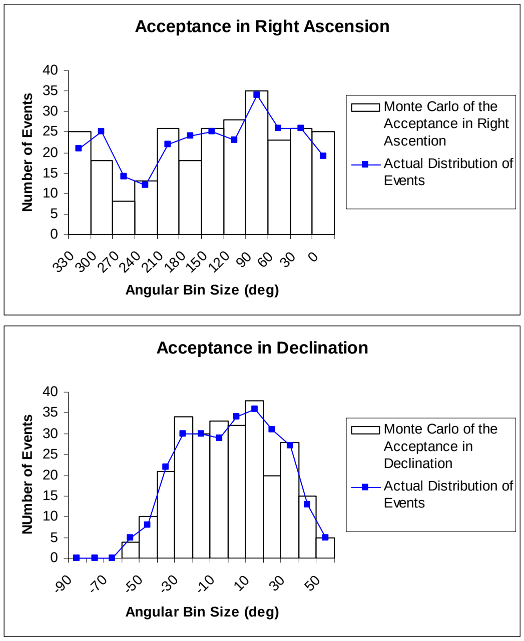

Figure 3b. Monte Carlo simulation of the acceptance in right ascension and declination of the angular distribution of high energy cosmic rays

b. AGASA Analysis

Unlike the HiRes experiment, the AGASA experiment published complete information about their 57 highest energy cosmic rays [7]. Similarly to HiRes, an angular analysis has been performed [8] focusing on an analysis of the auto-correlation of the data. We reproduced such analysis, and also performed an analysis of the correlation between the data and the modified catalogue of BL Lacs used during the analysis of the HiRes data.

Similarly to HiRes, the acceptance of the AGASA experiment had to be determined. Unlike HiRes such information hasn't been released for AGASA; however this information may still be determined. First of all, AGASA is an array of ground detectors functioning continuously, hence having a homogeneous view of the sky in right ascension; therefore, the acceptance in the right ascension should be flat.

On the other hand, the AGASA is located at latitude of 35^0N. Fortunately there exists an equation that yields the theoretical distribution of events as a function of the declination along with the maximal zenith angle of the experiment [9],

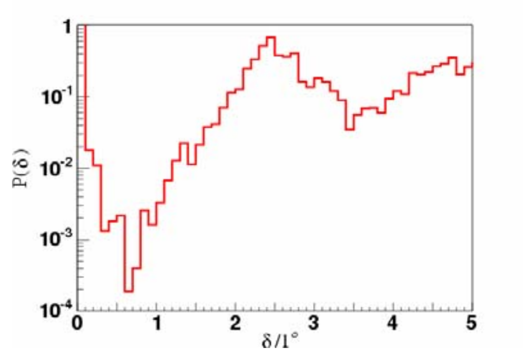

Figure 4a. Chance probability as a function of angular distance for the correlation of the highest HiRes energy events with the catalogue of 156 BL Lacs object.

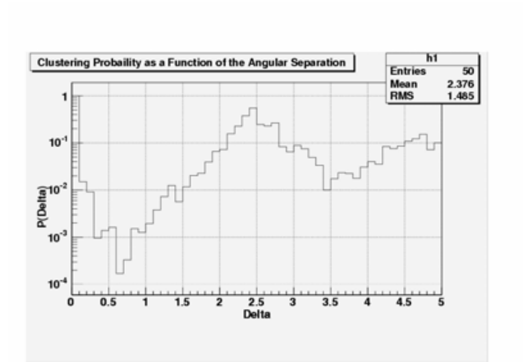

Figure 4b. Reproduced chance probability for the correlation of HiRes highest energy events with the set of 156 BL Lacs.

$$a(\delta) \propto \cos\delta \left(\cos a_0 \cos\delta \sin\alpha_m + \alpha_m \sin a_0 \sin\delta \right) \qquad (2)$$

where α_m is given by

$$\alpha_m = \begin{cases} 0 & if\ \xi > 1 \\ \pi & if\ \xi < -1 \\ \cos^{-1}\xi & otherwise \end{cases} \qquad (3)$$

and

$$\xi = \frac{\cos\theta_m - \sin a_0 \sin\delta}{\cos a_0 \cos\delta} \qquad (4)$$

where a_0 is the declination of the site, and θ_m is the maximum zenith angle of detection of the detector. The resulting distribution is presented in Figure 6.

Figure 5. Auto-correlation of the 271 highest energy events of HiRes.

With this information it is now possible to generate a set of randomly generated cosmic rays matching the distribution of the real data, and proceed with a similar analysis. However, there is a difference between the analysis performed with HiRes data and the AGASA data. The AGASA experiment has published the magnitude of their highest energy events. It is therefore possible to scan over a range of angular distances as well as a range of energy cuts.

The 57 events reported by AGASA were ranked in order of decreasing energies, and cuts were made in as a function of number of events in the cut. A cut of 3 would include the 3 highest energy events.

The auto-correlation analysis was performed for every one of the energy cuts, and plotted as a three dimensional histogram. The results are presented in Figure 7a and b.

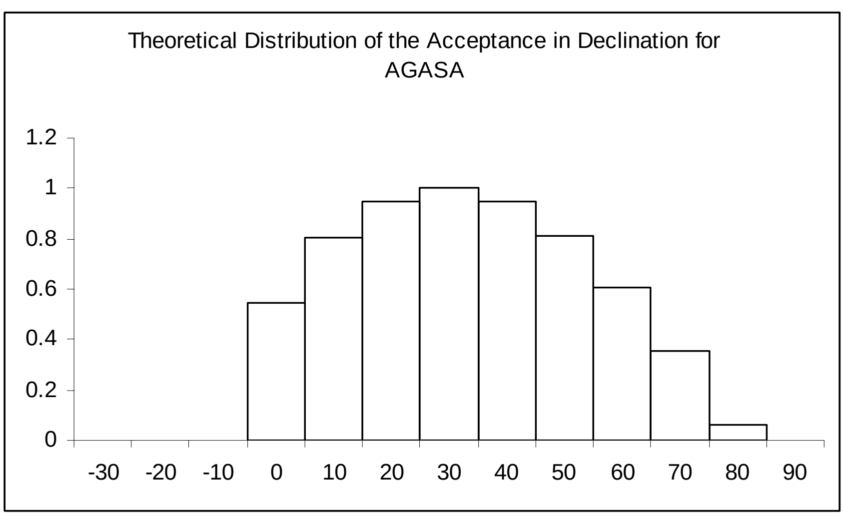

Figure 6. Theoretical distribution of the acceptance for the AGASA data based on Equation 2, knowing that the maximum zenith angle is 45^0, and the declination is 35^0.

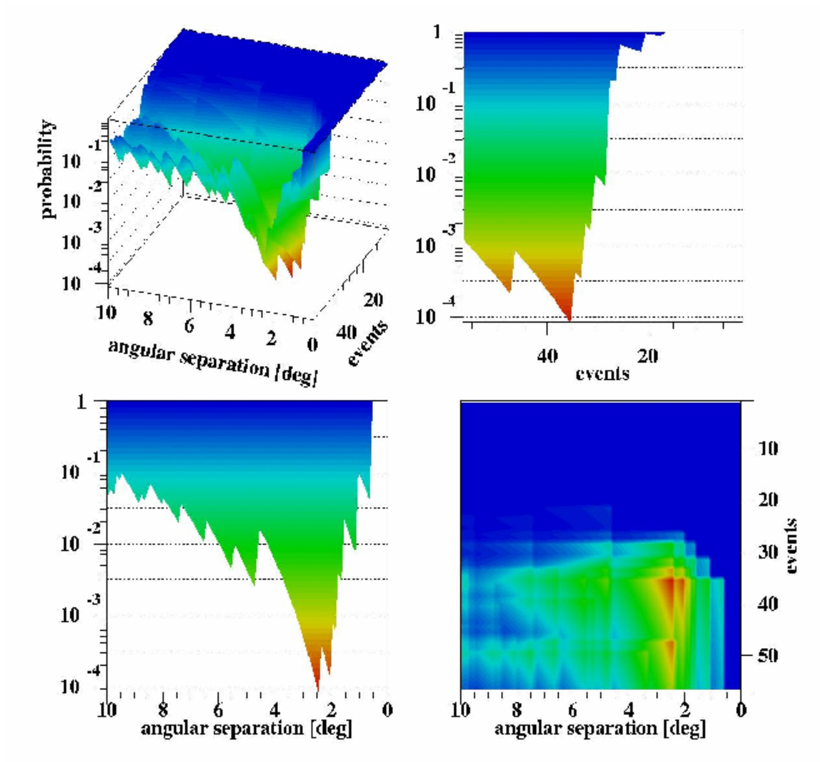

Figure 7a. AGASA Auto-Correlation results obtained by scanning over Angular separation and Energy cuts and previously published [8]. The four plots are just different views of the 3 dimensional graph presented in the top left corner.

The plots produced by Finley and Westerhoff [8], yield a minimum chance probability $P_\delta=8.4x10^{-5}$ at an angular distance of $\delta=2.5^0$ and energy cut of 36 (equivalent to an energy of $E=4.89x10^{19}eV$).

The results obtained in our reproduction are reasonably similar with again differences in the magnitudes $(P_\delta=1.5x10^{-4})$, and exact positions of the minimum $(\delta=2.5^0)$, but once again these differences could be explained by differences in the random number generator, and in the binning of both the acceptance and the final plots.

The correlation of the AGASA data with the modified catalogue of BL Lacs object was also tested, and is presented in Figure 8.

The significance of the results presented in Figure 8 could be considered to be really low. The minimum chance probability occurs at an angular separation of 2.5^0, and has a magnitude around $P_{2.5}=1.6x10^{-2}$; 10 correlations were obtained in the real data, while the Monte Carlo predicted 5.

However, when this result is put in perspective of the previously examined HiRes data, this result is not so bad, and could yield very meaningful results assuming more data is obtained. A similar analysis [12] found, using 22 BL Lacs obtained through cuts [13] on visual magnitude, red shift, and radio flux that $P_{2.5}=3x10^{-4}$. Note that an analysis of the HiRes data using the same cut doesn't yield any significance [6].

IV. Analysis of the Current Auger Data

Both of the previous analyses were made in a debatable manner. Both sets of analysis scanned the data over a range of angular separation (as well as energy for the AGASA experiment), however no penalty factor was assigned to the chance probability of each experiment. Note that all previously mentioned papers [6, 8, 12, 13] mention the existence of penalty factors, and the prescription for using these penalties is well defined. When a scan is performed over a range of values for any parameter, penalty factors must be used. When the value of the parameter can be chosen *a priori* based on a physical argument, no penalty factor is required.

The HiRes analysis for example did not make use of them for they assumed that neutral particles were being observed by a detector. On the other hand, Finley and Westerhoff described a prescription [8] on how to use these penalties to analyze the AGASA data, and the results presented in Figure 6a, and 6b need to be multiplied by penalty factors. The chance probability at $\delta=2.5^0$ is really $P_{2.5}=0.3\%$ as a penalty of 3 must be applied for scanning over energies, and a penalty of 10 for scanning over the angular separation. The correlation study of AGASA with BL lacs objects predicts $P_{2.5}=3x10^{-4}$; this value is obtained by calculating a penalty factor for scanning over a range of angles, however no penalty was assigned for scanning over

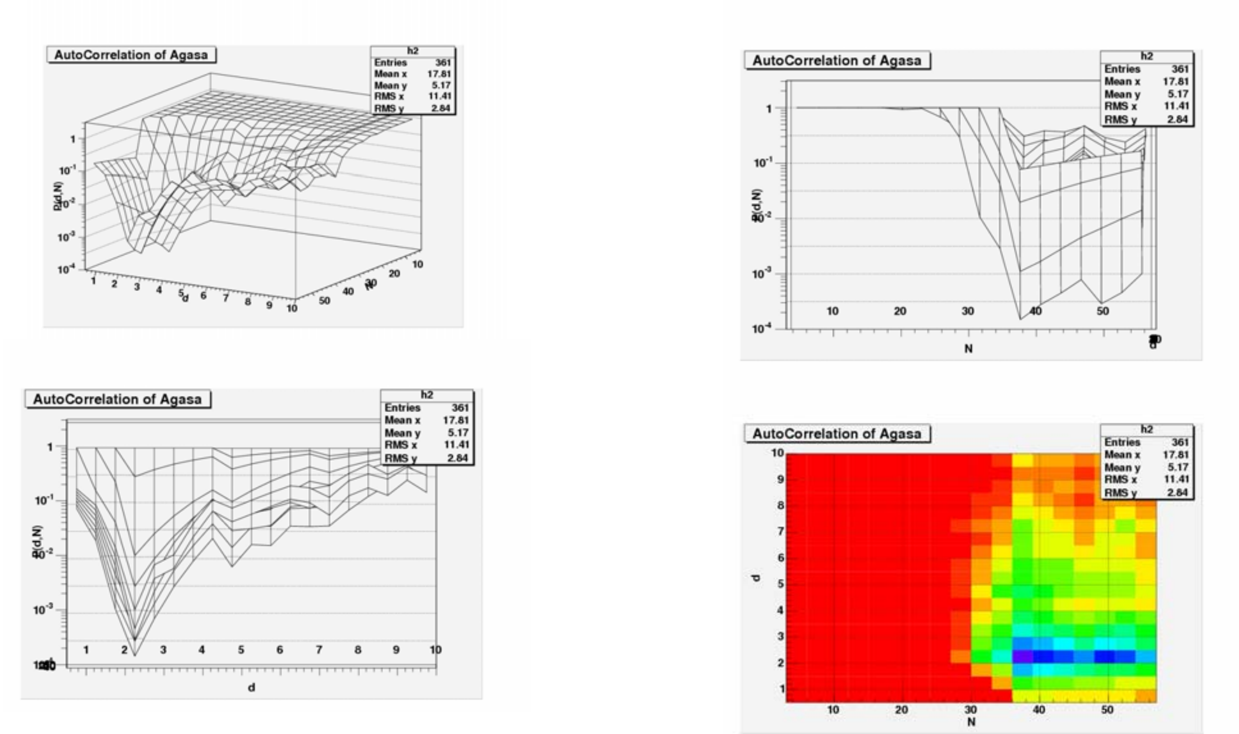

Figure 7b. Reproduction of the results obtained by Finley and Westerhoff regarding the chance probability as a function of angular distance and energy cut.

different cuts on BL Lacs. Note that the result obtained through our analysis doesn't require the use of penalty factor we assumed neutral particles.

The goal with the Auger data is to make *a posteriori* decisions by making use of penalty factors as described by Finley and Westerhoff [8, 11]. Furthermore, it is hoped that the hybrid data (detection of a single event through both fluorescence and ground array detection) of Auger, will be able to resolve the dissimilarities between AGASA, and HiRes. The procedure that will be used for the statistical analysis of Auger involves generating through Monte Carlo simulation, a large number of data sets with the same exposure as the detector (i.e. based on determined acceptance). These data sets can then be used to generate a table of values P_{mc}, where $P_{mc}(N, \theta, n)$ is the fraction of sets in which the first N events contain exactly n points separated by angle less than θ. For every N and θ, the number of pairs n_p in the data can be recorded and the probability P_{data} of finding n_p or more pairs at (N, θ) can be calculated based on the following equation.

$$P_{data}(N,\theta) = \sum_{n=n_p}^{\infty} P_{mc}(N,\theta,n) = 1 - \sum_{n=0}^{n_p-1} P_{mc}(N,\theta,n) \qquad (5)$$

This may be achieved by performing the same scan over n_{mc} Monte Carlo data set, and by identifying $P^i_{\min} = P^i\left(N^i_c, \theta^i_c\right)$ for each trial. The number of trial n_{mc}^* for which $P^i_{\min} \leq P_{\min}$ can be recorded, and the chance probability of observing P_{min} can be calculated as follow.

$$P_{chance} = \frac{n^*_{mc}}{n_{mc}} \qquad (6)$$

This technique includes correction factor for the scan over both variables, because it requires the splitting of the simulated data, hence the hypothesis is not tested against the data that was originally used to generate the hypothesis.

Unfortunately, at this point not enough is known about the Auger experiment to produce any significant analysis. Not enough high energy cosmic rays have been detected to produce any strong significance in the analysis, the angular resolution of the detector is not known at this point, and not a lot of BL Lac object have been detected in the southern hemisphere.

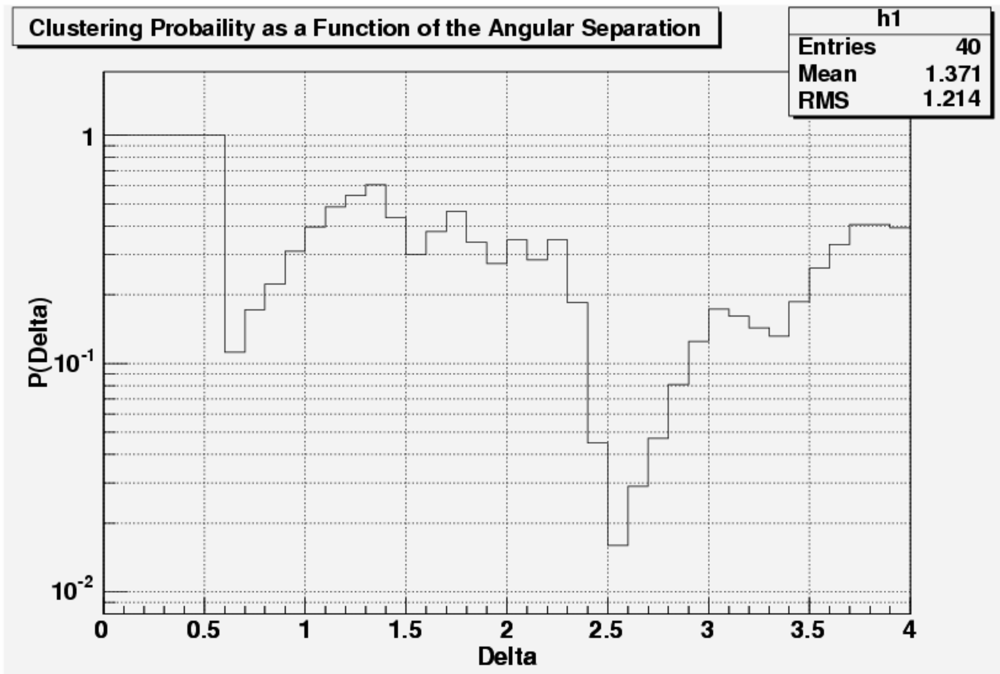

Figure 8. Plot of the chance probability of correlation with BL Lacs objects as a function of angular separation for the AGASA data.

V. Conclusion

The statistical analysis of cosmic ray data is in theory a rather simple task, unfortunately the small amount of information available at high energies make the actual study a challenging process on which most people do not agree. The technique that will be used to study the Pierre Auger data has been carefully chosen so as to be as unbiased as possible, in order to obtain the true significance of the results. It is hoped that these results will answer some of the important questions regarding cosmic rays' origin, and their acceleration mechanism.

VI. Reference

[1] R. J. Protheroe, and R. W. Clay, arXiv: astro-ph/0311466
[2] L. Anchordoqui, T. Paul, S. Reucroft, and J. Swain, arXiv: astro-ph/0206072
[3] P. Bhattacharjee, and G. Sigl, arXiv: astro-ph/9811011
[4] R. U. Abbasi et al. [The High Resolution Fly's Eye Collaboration], arXiv: astro-ph/0404137
[5] M. P. Véron-Cetty and P. Véron, ESO scientific report (2000); M. P. Véron-Cetty and P. Véron, Astron. Astrophys. 374 (2001) 92
[6] D. S.Gorbunov, P.G.Tinyakov, I. I.Tkachev, and S.V.Troitsky, arXiv: astro-ph/0406654
[7] N. Hayashida et al, arXiv: astro-ph/0008102
[8] C. Finley, and S. Westerhoff, arXiv: astro-ph/0309159
[9] P. Sommers, arXiv: astro-ph/0004016
[10] J. Cronin, aXiv: astro-ph/0402487
[11] S. Westerhoff, arXiv: astro-ph/0408343
[12] P. G. Tinyakov, and I. I. Tkachev, JETP Lett, **74**, *1* (2001)
[13] P. G. Tinyakov, and I. I. Tkachev, arXiv: astro-ph/0102476