Potential for Detection of Ultra-High-Energy Cosmic Ray Sources

University of Crete

Department of Physics

Senior Thesis

Author: Alexandra Pouliasi

Supervisor: Prof. Vasiliki Pavlidou

04/07/2020

Contents

1 Abstract

Finding the sources of Ultra-High-Energy Cosmic Rays (the most energetic particles in our Universe) still remains a great challenge. Their discovery will not only give us insight as to what mechanism (cosmic accelerator) could possibly accelerate these particles to energies up to 10^{20} eV, but will also possibly reveal new particle physics at these highest energies.

One of the main difficulties in locating the sources of Ultra-High-Energy Cosmic Rays (UHECRs) is that they get deflected by the Galactic magnetic field as they travel towards the Earth, coupled with the fact that our knowledge of the 3-dimensional structure of the Galactic magnetic field is limited.

Once the 3-d properties of the Galactic magnetic field are better understood, the trajectories of UHECRs through the galaxies can be reconstructed and followed backwards (backtracking), to "correct" the observed arrival directions of UHECRs, bringing them closer to their sources.

In this work, we study how to treat such corrected arrival directions. In particular, we seek to understand when a doublet (two events with nearby arrival directions) are consistent with coming from the same source, and when two or more events are consistent with coming from some specific, well-motivated candidate source.

2 Introduction

Cosmic rays are a form of high-energy radiation, mainly originating outside the Solar System [1] and even from distant galaxies.[2] Upon impact with the Earth's atmosphere, cosmic rays can produce showers of secondary particles that sometimes reach the surface. Data from the Fermi Space Telescope (2013)[3] have been interpreted as evidence that a significant fraction of primary cosmic rays originate from the supernova remnants of stars.[4] Specifically, there is enhanced gamma ray emission from regions of the sky that are nearby supernova remnants and dense gas clouds. There, newly accelerated cosmic rays interact with gas producing π^0, which in turn decay into gamma rays, which Fermi detects. Active galactic nuclei also appear to produce cosmic rays, based on observations of neutrinos and gamma rays. In this thesis we focus on ultra-high-energy cosmic rays (UHECRs), the highest-energy particles known. More specifically, UHECR are cosmic rays with an energy greater than 1 EeV (10^{18} electronvolts, approximately 0.16 joules), far beyond both the rest mass and energies typical of the bulk of cosmic ray particles.

2.1 Obstacles

Unfortunately, the study of UHECRs involves a lot of obstacles.

Firstly, UHECRs can not be directly detected so they are investigated through their interactions with atmospheric molecules that produce extensive air showers of secondary particles. Their composition is probed by examining these air showers using hadronic models that have only been tested for center-of-mass energies of the order of 10 TeV. Collisions of UHECRs with atmospheric nuclei have center-of-mass energies of 100 TeV, an order of magnitude higher than the energies probed by the Large Hadron Collider (LHC) at CERN. Therefore, determining their composition involves large uncertainties.

Another obstacle is their extremely low flux which, for energies above a few times 10^{18} eV, is at one particle per km^2 per year, while at the highest energies (above a few times 10^{19} eV) falls to one particle per km^2 per century.

Furthermore, since they are composed of charged particles, they are deflected by Galactic and extragalactic magnetic fields, and as a result their apparent arrival directions cannot be directly traced back to their sources. Additionally, knowledge about the Galactic magnetic field (GMF) is quite

limited, as current observables that explore it are integrated along the line of sight. These observables are used to acquire best-fit parameters for global models including a random component [8]. Information about the extragalactic magnetic fields is even more uncertain.

2.2 Promise

However, despite the difficulties, UHECRs are a promising field of research.

One of the reasons is that they challenge theoretical astrophysicists since their energies are the highest that any astrophysical accelerator can even in principle reach. At the moment there is no other way of studying particle physics at the highest center-of-mass energies probed by UHECR collisions in the Earth's atmosphere, so UHECR observations can probe new physics at these highest energies.

Another reason is that they provide a possible tool for studying and mapping Galactic and extragalactic magnetic fields since they travel through them.

In addition, UHECRs experience energy losses as they travel through the infrared, optical, and ultraviolet extragalactic background light. As a result, by studying these energy losses we can study extragalactic background light itself.

Moreover, through the interactions of UHECRs, during their propagation from their sources to the Earth, neutrinos are produced, known as cosmogenic neutrinos, which constitute the only guaranteed high-energy neutrino signal in the Universe. Detecting and studying these cosmogenic neutrinos could enlighten us about the distribution of UHECR sources in the Universe, their cosmological evolution, and the energy spectrum of each single source.

2.3 Backtracking

In order to correct the effect of the deflection of UHECRs by the Galactic magnetic field (GMF) a method called backtracking is used. The method involves the reconstruction of the trajectory of the cosmic ray through the Galaxy and its magnetic field, in order to obtain its original arrival direction. Ideally, if the galactic field is measured with high accuracy, backtracking will allow us to improve the agreement between the "corrected" arrival directions of UHECRs and the location of their sources. However, such measurements do not exist at present. Large effort is being invested in estimating

the promise of backtracking, by evaluating its performance, i.e. the uncertainty that could be achieved in the "corrected" of the arrival directions as a function of accuracy of magnetic-field measurements, cosmic-ray charge, and cosmic ray energy. This is done by using hypothetical measurements of the GMF, based on values received from the most up-to-date GMF models. As shown in [8] the effectiveness of the correction depends on the particle rigidity and arrival direction, and can vary significantly depending on the GMF model used.

3 Characteristics of UHECRs

3.1 Energy Spectrum

The flux of cosmic rays is shown in Fig. 1 as a function of energy [10]. It follows approximately a power law with slope close to –3.

In the past two decades, relatively large amounts of data have been collected, so we get an improved energy spectrum, as seen in Fig 2. Some interesting features of the spectrum are a steepening of the spectrum, where the slope changes from ~ -2.7 to ~ -3, known as the "knee", at the energy of $E_{knee} \sim 3 \times 10^{15}$eV and a flattening of the spectrum, where the slope becomes ~ -2.6, known as the "ankle", at the energy of $E_{ankle} \sim 3 \times 10^{18}$eV.

Studies of the composition of cosmic rays at high energies support that the existence of the "knee" is most likely due to reaching the maximum energy that the Galactic accelerators, which contribute dominantly to energies below E_{ankle} , can achieve. At energies lower than E_{knee} cosmic rays have been shown to be of Galactic origin, while at energies higher than E_{ankle} cosmic rays are expected to be of extragalactic origin, meaning that the transition between Galactic and extragalactic cosmic rays is expected to occur somewhere between the knee and the ankle. The existence of the "ankle" can be interpreted in two ways: 1) if we assume mixed or heavy nuclei composition (iron) around E_{ankle} , then the "ankle" represents the transition between Galactic and extragalactic cosmic rays, while 2) in the case of lighter composition (protons) at E_{ankle} the "ankle" is caused by pair production energy losses during propagation.

Finally, at energies above $\sim 3 \times 10^{19}$ eV a suppression in the spectrum occurs, which is most likely caused by energy losses due to interaction of cosmic rays with the CMB as they propagate the intergalactic medium. This effect is known as Greisen-Zatsepin-Kuzmin (GZK) effect. However, it is still not easy to confirm whether the suppression is truly caused by the GZK effect, by reaching the maximum energy of extragalactic accelerators [7], or by a combination of the two events.

It should be noted that presenting the data as flux multiplied by a power of the energy ($E^{2.5}$ in Fig 2) is quite usual because the features that are not very evident in a normal plot of the flux versus energy become more visible. However, it has the disadvantage of coupling the horizontal and vertical axes, thus enhancing the differences in the energy calibration of the different experiments [10].

In Fig 1 and Fig 2, the equivalent center-of-mass (c.m.) system energy for proton–proton collisions at high-energy accelerating machines is also indicated [10].

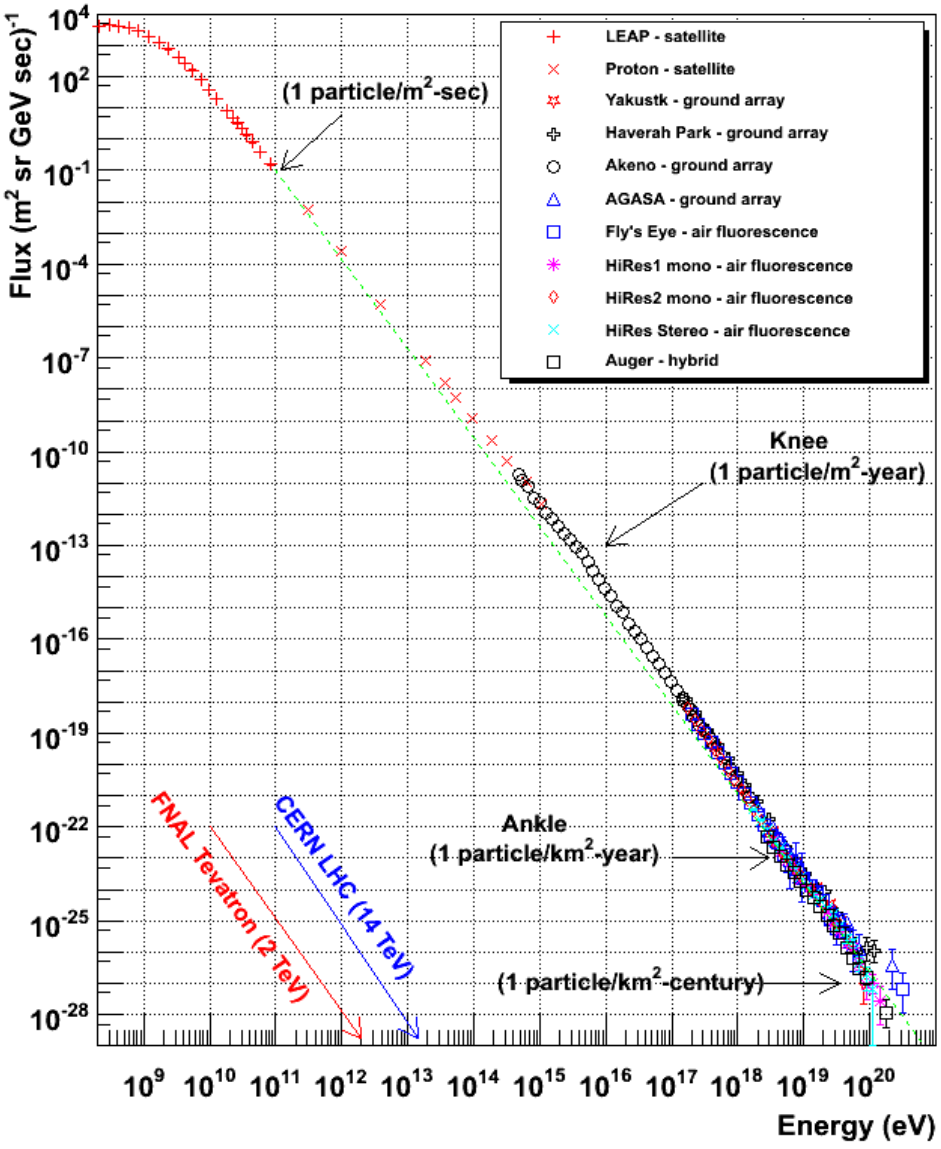

Figure 1: Flux of primary cosmic rays as a function of energy [11].

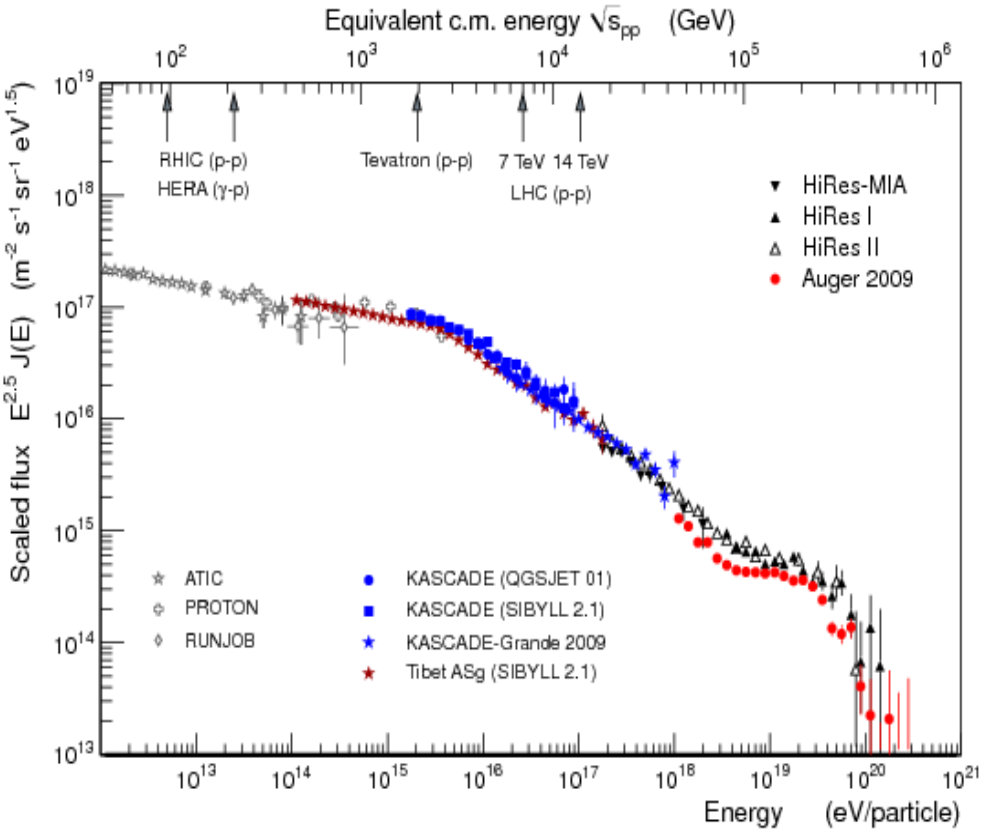

Figure 2: Recent compilation of the cosmic ray energy spectrum [12]. The flux is multiplied by the power law $E^{2.5}$.

3.2 Flux

As shown in Fig. 1, the flux of UHECRs is extremely low. At energies around 10^{18} eV the flux is one particle per km^2 per year which falls even more at energies higher than 10^{19} eV at one particle per km^2 per century [5]. Since the first observation of an UHECR particle, by the University of Utah's Fly's Eye Cosmic Ray Detector [6], at least fifteen similar events have been recorded. These very high energy cosmic ray particles are very rare. For this reason, UHECR studies happen in the very-low-statistics regime, and all correlations require careful statistical evaluation to determine their significance.

3.3 Composition

Since the detection of cosmic rays is indirect, their composition is derived by measuring the penetration depth of the primary in the atmosphere. It is expected that heavier primaries have a bigger cross-section, so they experience their first interaction earlier (higher in the atmosphere). Also their interaction depth fluctuates less because they consist of more nucleons, each of which initiates its own chain of interactions, so fluctuations average out. Therefore, in order to probe composition we measure the mean atmospheric depth when the shower reaches its maximum development (maximum number of particles), $\langle X_{max} \rangle$ (in units of g/cm^2), which depends on the primary's energy, E, and its atomic mass number, A, as follows: $\langle X_{max} \rangle \sim ln(E/A)$ [see Fig 3 (left)].

Furthermore, the spread in X_{max} can provide additional information about the composition of cosmic rays [see Fig 3 (right)], since for light-primary showers X_{max} at a given energy fluctuates about $\langle X_{max} \rangle$ more than that of heavy-primary showers.

It should be noted that the above calculations are based on shower simulations at energies which have not been studied in accelerators and therefore contain significant uncertainties.

As deduced from Fig 3, the cosmic ray composition is lighter at low energies and heavier at high energies.

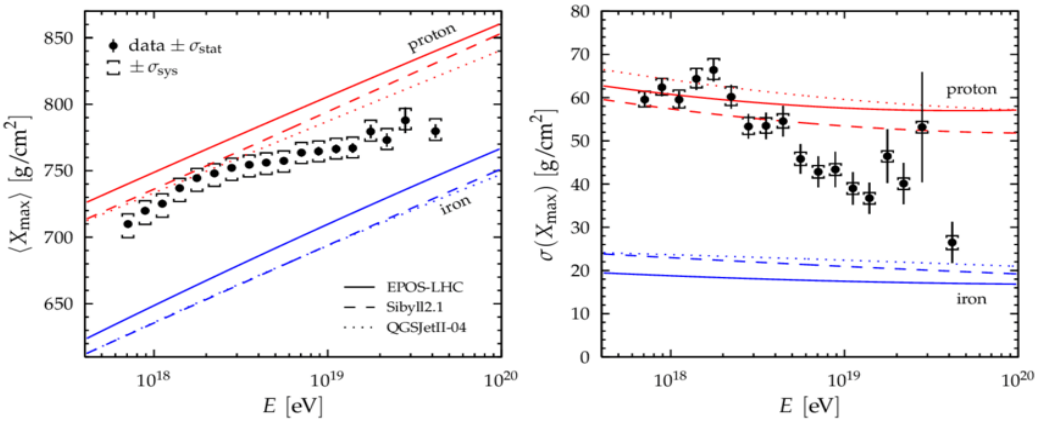

Figure 3: The mean (left) and the standard deviation (right) of the measured X_{max} distributions, by Auger, as a function of energy compared to air-shower simulations for proton and iron primaries using different hadronic models [13].

3.4 Anisotropies

Assuming that the suppression of the spectrum above 4×10^{19} eV is indeed caused by the GZK effect, then UHECRs coming from cosmological distances, which are isotropically distributed, won't be able to reach the Earth. As a result, since the distribution of matter in the local Universe is inhomogeneous, and provided that cosmic ray deflections in the intervening galactic and extragalactic magnetic fields are small enough, the UHECRs that reach us are expected to be anisotropically distributed and spatially correlated with the local large-scale structure [5] [14] (see Fig 4).

The blue gradient of Fig 4 visualizes the exposure of the Auger ground array (darker colors represent higher exposure). An isotropic distribution of arrival directions would result in higher density of events (within Poisson uncertainties) in the direction of the sky with higher exposure. Instead, the density of points is higher in the directions of local extragalactic matter, showing anisotropy [5].

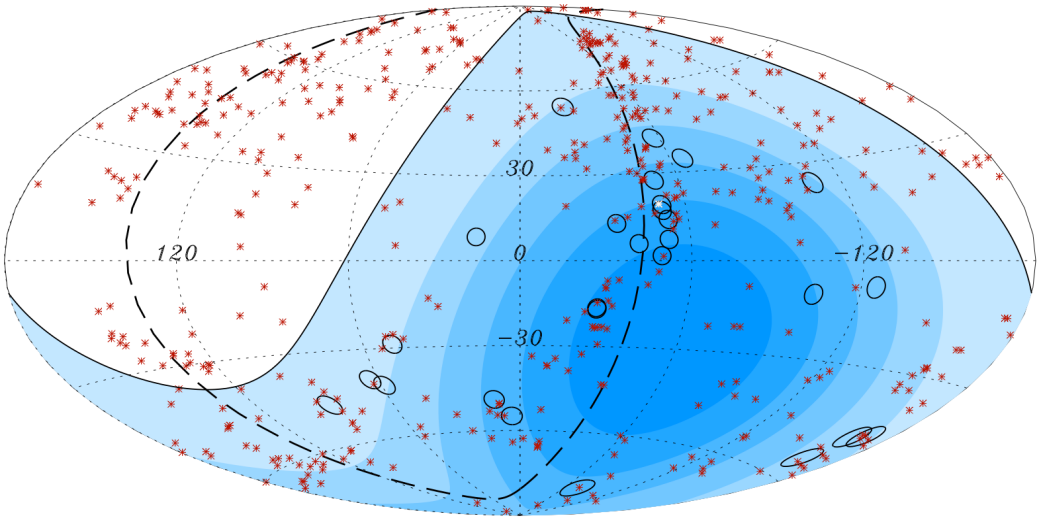

Figure 4: Map in galactic coordinates with circles of 3.2^o radius centered on the arrival directions of the 27 Auger events with energy $E > 57$ EeV (taken from [14]). The positions of the 442 AGN from the 12th edition of the Véron-Cetty and Véron catalog with redshifts $z \leq 0.017$ ($D < 71$Mpc) are shown with red stars (one of the closest objects, Cen A, is shown in white). Shading within the solid line (the Auger field of view for $\theta < 60^o$) indicates regions of equal exposure. The dashed line denotes the supergalactic plane.

3.5 Potential Sources

Due to the high energies of UHECRs their candidate sources are limited. That is because if a particle escapes from the region where it was being accelerated, it will be unable to gain more energy. This situation imposes a limit on the partice's maximum energy that can be expressed as follows: $\epsilon_{max} = qBR$ where q is the electric charge of the accelerated particle, B is the magnetic field, and R is the size of the accelerator. The equation for ϵ_{max} is obtained by demanding that the gyroradius of the particle, $r_g = \epsilon/(qB)$, not exceed the size of the acceleration region. This is a general geometrical criterion known as the Hillas criterion, and is useful in selecting potential acceleration sites.

Fig 5 is an example of a Hillas plot which, for a given maximum energy ϵ_{max} of the accelerated particle, shows the relation between the source's magnetic field strength B and its size R. Sources above the top line are able to accelerate protons up to 10^{21} eV, while sources above the bottom line are able to accelerate iron up to 10^{20} eV. [9]

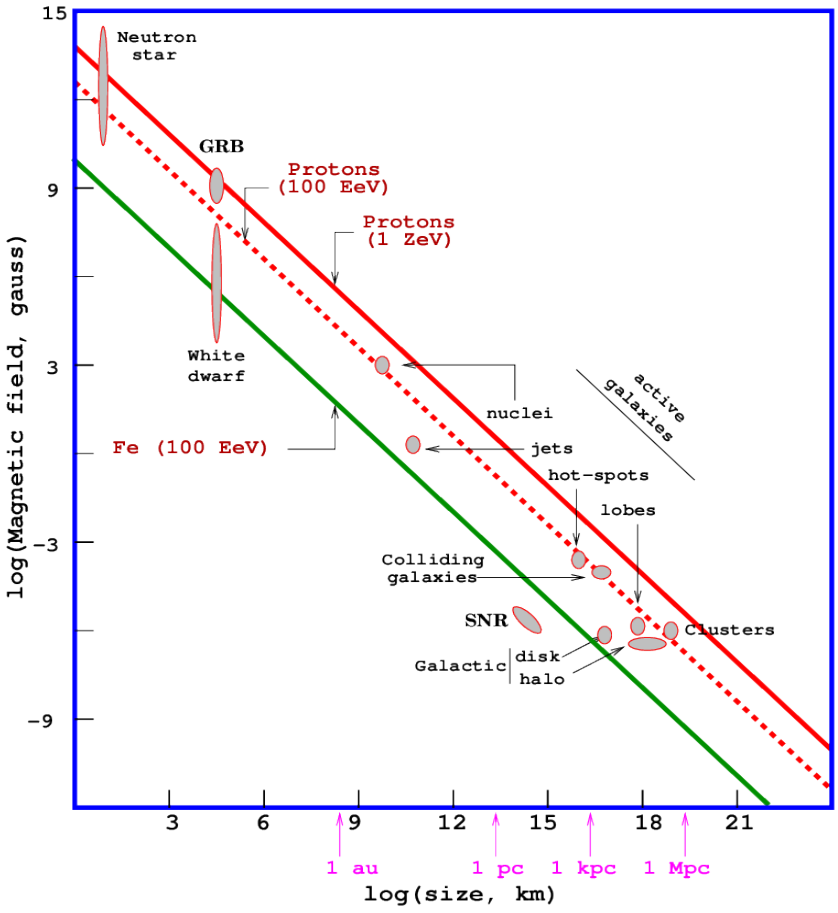

Figure 5: "Hillas plot" showing the magnetic field versus size of potential UHECR sources. Acceleration of protons to 1 $ZeV = 10^{21}$ eV, or protons or Fe nuclei to 100 EeV $= 10^{20}$ eV require conditions above the respective line. [9]

4 This work

In this work, we perform a statistical analysis of simulated corrected arrival directions of UHECRs that could result from a backtracking procedure. Our main concern is to construct and test a procedure that will determine whether a number of detected events of UHECRs are produced by the same source as well as determine what this source may be. If magnetic field backtracking is performed for a single event, it will yield a probability distribution for the position in the sky of the source of that event. We want to evaluate the information that can be obtained by multiple such backtrackings and resulting probability distributions of sky positions. In particular, we are interested in the specific case of doublets. A doublet is defined as two events of high energy ($E > 10^{19}$ eV) with very small angular separation. Usually, the limiting angular separation to consider a pair of events as a "doublet" is the accuracy in the determination of arrival directions of the Observatory that detected the events. If backtracking of the events through the magnetic field of the Galaxy has been performed, the limiting angular separation is usually the positional uncertainty of the backtracking.

Our goal is to estimate a) whether these two events are consistent with being produced by the same source and b) whether the locations of the events are consistent with their source being some specific, well-motivated candidate.

4.1 Methodology

In order to address this problem we use the Monte Carlo method. In particular, we assume that, by using backtracking, we have found the probability distribution for the arrival direction of each event. That is to say its most likely "corrected" arrival direction and the uncertainty in this arrival direction. Then, we choose a coordinate system (θ, ϕ) and we draw numbers from these 2 distributions (with $\sigma_{\theta_1}, \sigma_{\phi_1}$ and $\sigma_{\theta_2}, \sigma_{\phi_2}$ respectively) and get the position of the 2 events, θ_1, ϕ_1 and θ_2, ϕ_2 respectively.

For the purposes of this paper, in order to get a view of the results of our methodology, we use mock data, for which we choose the simple case of bivariate normal distribution (two-dimensional) for the position distribution of each event. In this case, the probability density function of a two-dimensional vector $[\theta\ \phi]$ is

$$f(\theta,\phi) = \frac{1}{2\pi\sigma_\theta\sigma_\phi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(\theta-\mu_\theta)^2}{\sigma_\theta^2} + \frac{(\phi-\mu_\phi)^2}{\sigma_\phi^2} - \frac{2\rho(\theta-\mu_\theta)(\phi-\mu_\phi)}{\sigma_\theta\sigma_\phi}\right]\right) \quad (1)$$

where ρ is the correlation between θ and ϕ and where $\sigma_\theta > 0$, $\sigma_\phi > 0$. In this case, $\boldsymbol{\mu} = \begin{pmatrix}\mu_\theta \\ \mu_\phi\end{pmatrix}$ is the mean vector and $\boldsymbol{\Sigma} = \begin{pmatrix}\sigma_\theta^2 & \rho\sigma_\theta\sigma_\phi \\ \rho\sigma_\theta\sigma_\phi & \sigma_\phi^2\end{pmatrix}$ is the covariance matrix [15].

In our analysis, we choose both mean values μ_θ, μ_ϕ equal to zero, and we also choose spherical covariance, for which the covariance matrix is a multiple of the identity matrix (here we use the identity matrix, so $\sigma_\theta = 1$, $\sigma_\phi = 1$ and $\rho = 0$).

Therefore, we move on to answering the following two questions.

4.2 Question a): Are these two events produced by the same source?

This question cannot be answered with certainty. In general, studies of arrival directions of UHECRs tend to focus on the following variation of the question: "If there are no anisotropies in the UHECR sky, i.e. if all arrival directions are equally probable, modulated by the pattern on the sky of an observatory's exposure, then what is the probability that we would get, by chance, two events with arrival directions so close to each other?". This question has been exhaustively addressed in the literature, e.g. [16].

However, the process of backtracking, which introduces new uncertainties, due to the backtracking itself, in the arrival directions of UHECRs, generates the need for addressing a different kind of question. "Raw" arrival directions of UHECRs are known to deviate from the location of their sources, due to systematic deflection in the magnetic field. Corrected arrival directions will not - on average. Their discrepancies from the location of their sources will be due to backtracking uncertainty alone. We would therefore like to consider apparent clustering of events as possibly meaningful, as long as the distances of these events from each other are not "too large" compared to the uncertainties of backtracking. Therefore, the new question which we address, and which is the subject of this question, is: "Are two events *consistent* with

coming from a single source?". In other words, we will give the two events the "benefit of a doubt" that they do, in fact, come from the same source. Unlike the usual analysis, our null hypothesis will be that a doublet is indeed a pair of events coming from a single source. And we will calculate how far apart the corrected arrival directions of the two events have to be, given the uncertainties of each event's backtracking, for this null hypothesis to be rejected (i.e. for us to decide that the pair is not consistent with coming from the same source), at a certain level of significance.

In order to do this, we follow the following steps.

- We assume we know where the source is located on the sky at $(\theta,\phi) = (\mu_\theta,\mu_\phi)$.
- The detected events that form a doublet will have arrival directions (after a presumed backtracking correction) (θ_1,ϕ_1) and (θ_2,ϕ_2) which will differ from (μ_θ,μ_ϕ) because of uncertainties in backtracking. We assume that these uncertainties are σ_θ, σ_ϕ.
- To simulate the detection of a doublet from the source and the subsequent backtracking, we draw values of (θ_1,ϕ_1) and (θ_2,ϕ_2) from the distribution of Eq. (1).
- We calculate the angular distance between the two events of the doublet, $r_{12} = \sqrt{(\theta_1-\theta_2)^2+(\phi_1-\phi_2)^2}$ and plot the distribution of r_{12}. For the small deflections we consider here, this is an adequate approximation.
- We calculate the probability that r_{12} is larger than a certain value (which will be chosen arbitrarily, since we have no real data yet).
- Finally, we calculate which value of r_{12} contains: 68% (r_{12-68}), 95% (r_{12-95}) and 99.7% ($r_{12-99.7}$) of the probability (1σ, 2σ and 3σ respectively). For example, if two events are found with a corrected (after backtracking) distance from each other larger than $r_{12-99.7}$, this means that the hypothesis that the two events come from the same source is rejected at the 3σ confidence level.

4.3 Question b): Is the source of these two events some specific candidate?

Here, we focus on a different aspect of the problem. Sometimes, a particularly well-motivated source candidate arises. It can, for example, be a source that is very nearby, and that belongs to a class of objects that have been proposed as likely sites of UHECR acceleration. This can be, for example, a nearby AGN, a nearby starburst galaxy, or a galaxy cluster.

To deduce whether the source of two events is some specific candidate, we follow the following steps.

- First of all, we have to assign that specific candidate a position. We assume that the position of the specific candidate is at the origin of our coordinate system, meaning $\theta = \phi = 0$.
- Then, we calculate the distance of each event from the specific candidate, i.e. $r_1 = \sqrt{\theta_1^2 + \phi_1^2}$ and $r_2 = \sqrt{\theta_2^2 + \phi_2^2}$, and plot the distributions of r_1 and r_2. Ideally, these two distributions should be identical. However, we use a finite number of draws from the distributions (in particular 10^7) which cannot be arbitrarily large due to the fact that the code needs to be relatively fast when run and not too computationally demanding. That leads to slightly different distributions for r_1 and r_2.
- We calculate $r_{tot} = \sqrt{r_1^2 + r_2^2}$ and plot its distribution.
- Finally, as we did for r_{12}, we calculate which value of r_1, r_2, r_{tot} contains: 68%, 95% and 99.7% of the probability (1σ, 2σ and 3σ respectively). For example, if an event is found, after backtracking, at a distance larger than $r_{1-99.7}$ from the candidate source, then the hypothesis that it comes from that source can be rejected at the 3σ level. As another example, if two events are found, after backtracking, at distances r_1, r_2 from the source, such that $r_{tot} < r_{tot-99.7}$, then the hypothesis that this is a doublet coming from the particular candidate cannot be rejected since it is consistent with the data at the 3σ level.

5 Results

Our results are shown in Fig 6 - Fig 9 as well as Table 1.

From Figures 6 - 9 we can calculate the values of r_{12}, r_1, r_2 and r_{tot} for which our results are consistent with our hypotheses, within a certain percentage (68% which is equivalent to 1σ, 95% which is equivalent to 2σ and 99.7% which is equivalent 3σ). Our first hypothesis (which corresponds to Question a) is that the two events are coming from a single source and our second hypothesis (which corresponds to Question b) is that the the source of these two events is some specific candidate. So Figure 6 corresponds to the first hypothesis and Figures 7 - 9 correspond to the second.

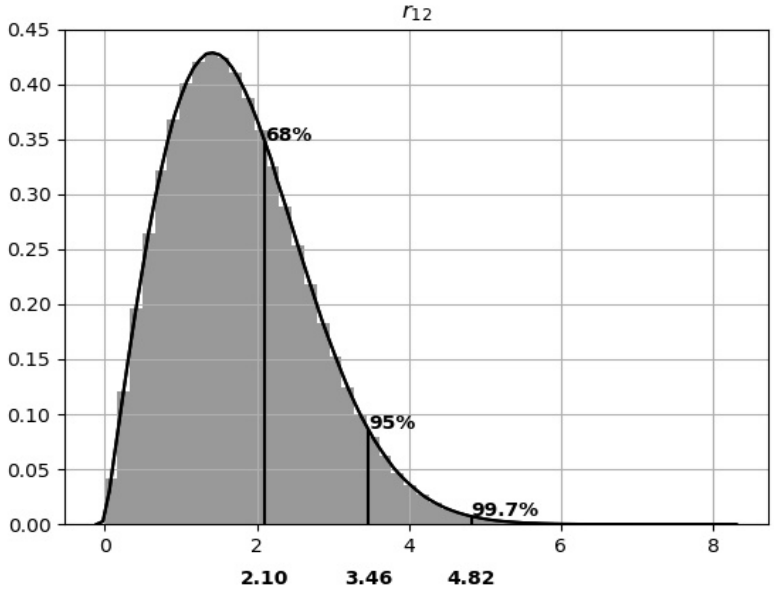

Figure 6: Distribution of r_{12}

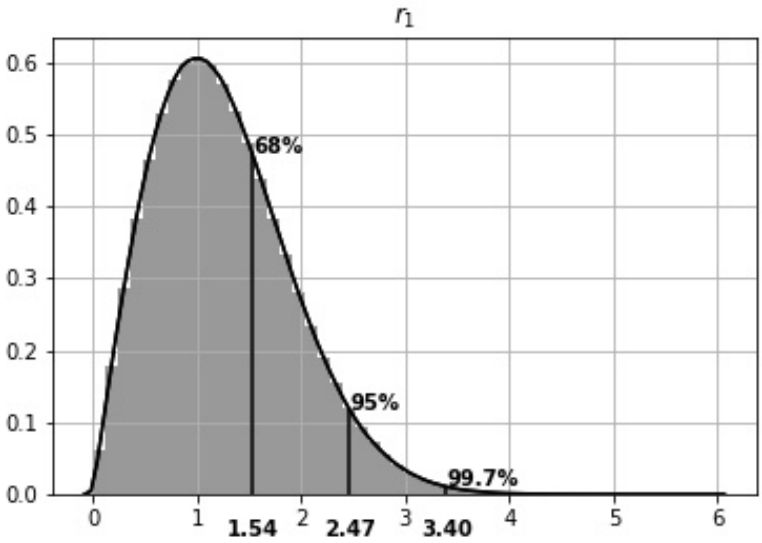

Figure 7: Distribution of r_1

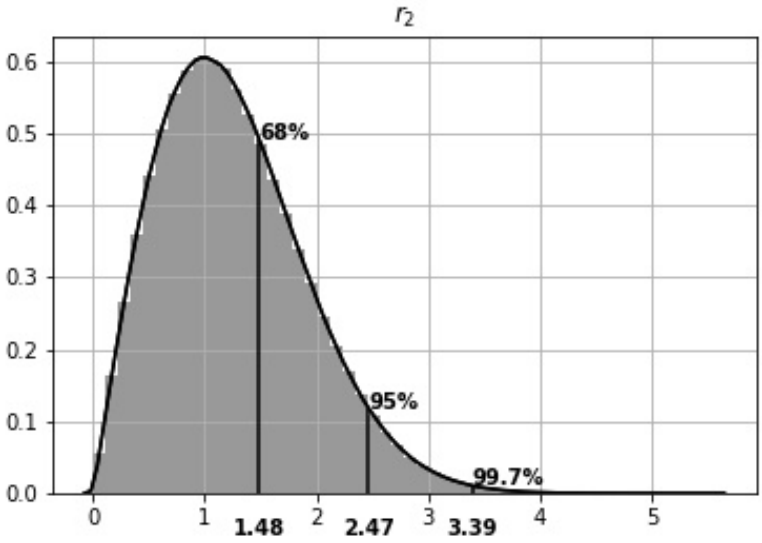

Figure 8: Distribution of r_2

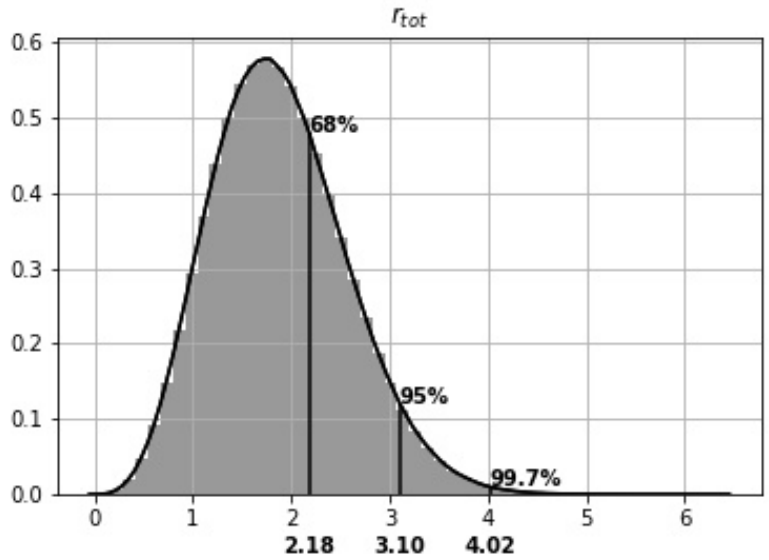

Figure 9: Distribution of r_{tot}

	r_{12} (Fig. 5)	r_1 (Fig. 6)	r_2 (Fig. 7)	r_{tot} (Fig.8)
68%	2.10	1.54	1.48	2.18
95%	3.46	2.47	2.47	3.10
99.7%	4.82	3.40	3.39	4.02

Table 1: Values of r_{12}, r_1, r_2 and r_{tot} which contain 68%, 95% and 99.7% of the probability (in units of $\sigma_\theta = \sigma_\phi$).

Let's assume that θ and ϕ are measured in degrees and so the backtracking accuracy is $\sigma_\theta = \sigma_\phi = 1$ degree. Then our results can be interpreted as following.

Fig 6 indicates that, for example, an observed doublet consisting of two events that after backtracking have arrival directions that are larger than 4.82 degrees apart is inconsistent at the 3σ level (99.7%) with being a "real" doublet (i.e., both events coming from the same source).

Fig 7 (equivalently Fig 8) indicates that, for example, an event that after backtracking has a distance from the origin larger than 3.40 degrees (equiva-

lently 3.39 degrees) is inconsistent at the 3σ level (99.7%) with coming from a specific candidate (which is located at the origin).

Fig 9 indicates that, for example, an observed doublet consisting of two events for which after backtracking the square root of the sum of the squares of their distances from the origin ($r_{tot} = \sqrt{r_1^2 + r_2^2}$) is larger than 4.02 degrees is inconsistent at the 3σ level (99.7%) with coming from a specific candidate (which is located at the origin).

Notes:

- The plots in Fig 6 - Fig 9 are made using Python. The code is provided in Appendix A.
- The values shown in Figures 6 - 9 are not representative of an actual result since we did not use real data.
- In Figures 6 - 9, r_{12}, r_1, r_2, r_{tot} are in units of $\sigma_\theta = \sigma_\phi$.

6 Discussion

Unfortunately, we have not yet collected real data on the magnetic field which would allow us to perform the backtracking process. However, once we do, we will be ready to apply our method of estimating the position of the sources of UHECRs. At the moment, we can qualitatively see that by using the probability distributions shown in Figures 6 - 9 for a specific problem of a doublet we can calculate a) the probability that the two events are produced by the same source (hypothesis a) as well as b) the probability that the source of these two events is some specific candidate (hypothesis b). More specifically, if the probability that r_{12} of a real doublet is greater than the observed value ($r_{12} > r_{12-observed}$) is small, then we can deduce that the two events are not produced by the same source. Similarly, if the probability that r_1 or r_2 or r_{tot} from events coming from the candidate source is greater than the observed value is small, then we can deduce that the source of either or both of the events is not that specific candidate. Therefore, we can answer the questions posed in 4.2 and 4.3.

The same logic of statistical testing can (and ideally, if we collect enough UHECR data, should) be extended to more than two events. The more events we can find being consistent with coming from a specific candidate, the more confident we can be about that candidate being a UHECR source. If we want to test whether N events come from a specific source, we calculate the value of r_{tot} (for N number of events $r_{tot-N} = \sqrt{r_1^2 + r_2^2 + ... + r_N^2}$) which contains 99.7% of the probability (r_{crit}). So, if the measured r_{tot} is smaller than r_{crit} then we can say that the data are consistent with the hypothesis that the source of the events is some specific candidate, or in other words, that the hypothesis cannot be rejected at the 3σ level.

Therefore, we should plot the value of r_{crit}, with respect to the number of events, N. However, for N events the value of σ is $\sigma_N^2 = \sigma_1^2 + \sigma_2^2 + ... + \sigma_N^2$. And since $\sigma_1 = \sigma_2 = ... = \sigma_N = \sigma$ we have $\sigma_N = \sigma\sqrt{N}$. What's more, we define r_{crit} as the value of r_{tot} which contains 99.7% of the probability, so $r_{crit} = 3\sigma$ for a single event. As a result, for N events the value of r_{crit} is $r_{crit-N} = 3\sigma_N = 3\sigma\sqrt{N}$ so it depends linearly on $\sqrt{N}$. That being the case, we plot the value of r_{crit-N}^2 with respect to N, in order to have a linear dependence (Fig 10).

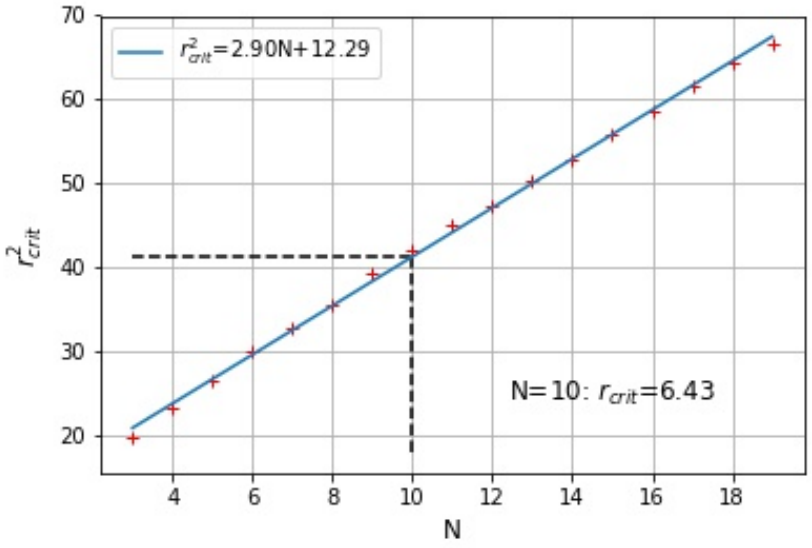

Figure 10: r^2_{crit-N} - N

From Fig 10 we see that r^2_{crit-N} increases linearly with N. For example, if we think there are N=10 candidate events from a source, their arrival directions after backtracking should have an r_{tot} of less than 6.43 degrees in order to be consistent at the 3σ level (99.7%) with coming from a specific candidate.

Moreover, by fitting a line in our plot we can then use it to extrapolate results for an even larger number of events than the ones we used to make the plot. In other words, we can find the r_{crit} for an arbitrarily large N simply by using the fitted line's equation, without having to do the whole analysis. The line equation has a slope of 2.90 and an interception of 12.29 .

Finally, we are anticipating new data that are going to give us the opportunity to test our method and use its results. Determining the sources of UHECRs will open new pathways to investigate particle acceleration and physics of high-energy collisions. Therefore, it will undoubtedly be a major accomplishment not only for Astrophysics, but also for Fundamental Physics.

7 Acknowledgements

I would like to acknowledge Prof. Vasiliki Pavlidou, who assigned me this project and trusted my potential. I wish to thank her for guiding me through the process of solving this problem, for reviewing this document and helping me refine it, and also for keeping me motivated.

Appendices

A Python Code

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy

np.random.seed(40)
plt.rcParams['axes.labelsize'] = 12

size=int(1e7)

mean1 = [0, 0]
cov1 = [[1, 0], [0, 1]]
x1 , y1 = np.random.multivariate_normal(mean1, cov1, size).T

mean2 = [0, 0]
cov2 = [[1, 0], [0, 1]]
x2 , y2 = np.random.multivariate_normal(mean2, cov2, size).T


#Question a): Are these two events produced by the same
                                source?
r12 = np.sqrt((x1-x2)**2+(y1-y2)**2)
sns.distplot(r12, bins=None, hist=True, kde=True, norm_hist=
                                True, color='k')
plt.title('$r_{12}$')

#Which r_12 contains 68%, 95%, 99.7%
p = sns.kdeplot(r12, color='k')
x,y = p.get_lines()[0].get_data()
cdf = scipy.integrate.cumtrapz(y ,x, initial=0)

nearest_068 = np.abs(cdf-0.68).argmin()
x_68 = x[nearest_068]
y_68 = y[nearest_068]
plt.vlines(x_68, 0, y_68, color='k')
plt.text(x_68, y_68, '68%' , color='k' ,weight='bold')
plt.text(x_68, -0.05, '{:04.2f}'.format(x_68), color='k',
                                weight='bold',
                                horizontalalignment='center',
```

```
                                verticalalignment='center')

nearest_095 = np.abs(cdf-0.95).argmin()
x_95 = x[nearest_095]
y_95 = y[nearest_095]
plt.vlines(x_95, 0, y_95, color='k')
plt.text(x_95, y_95, '95%' , color='k', weight='bold')
plt.text(x_95, -0.05, '{:04.2f}'.format(x_95), color='k',
                                weight='bold',
                                horizontalalignment='center',
                                verticalalignment='center')

nearest_0997 = np.abs(cdf-0.997).argmin()
x_997 = x[nearest_0997]
y_997 = y[nearest_0997]
plt.vlines(x_997, 0, y_997, color='k')
plt.text(x_997, y_997, '99.7%' , color='k', weight='bold')
plt.text(x_997, -0.05, '{:04.2f}'.format(x_997), color='k',
                                weight='bold',
                                horizontalalignment='center',
                                verticalalignment='center')
plt.grid()
plt.savefig('r_12.jpg')
plt.show()

print('x_68 =',x_68)
print('x_95 =',x_95)
print('x_997 =',x_997)

#Question b): Is the source of these two events some specific
                                 candidate?

r1 = np.sqrt(x1**2+y1**2)
sns.distplot(r1, bins=None, hist=True, kde=True, norm_hist=
                                True, color='k')
plt.title('$r_1$')

#Which r_1 contains 68%, 95%, 99.7%
p = sns.kdeplot(r1, color='k')
x,y = p.get_lines()[0].get_data()
cdf = scipy.integrate.cumtrapz(y ,x, initial=0)

nearest_068 = np.abs(cdf-0.68).argmin()
x_68 = x[nearest_068]
y_68 = y[nearest_068]
```

```
plt.vlines(x_68, 0, y_68, color='k')
plt.text(x_68, y_68, '68%' , color='k' ,weight='bold')
plt.text(x_68, -0.05, '{:04.2f}'.format(x_68), color='k',
                                weight='bold',
                                horizontalalignment='center',
                                verticalalignment='center')


nearest_095 = np.abs(cdf-0.95).argmin()
x_95 = x[nearest_095]
y_95 = y[nearest_095]
plt.vlines(x_95, 0, y_95, color='k')
plt.text(x_95, y_95, '95%' , color='k', weight='bold')
plt.text(x_95, -0.05, '{:04.2f}'.format(x_95), color='k',
                                weight='bold',
                                horizontalalignment='center',
                                verticalalignment='center')


nearest_0997 = np.abs(cdf-0.997).argmin()
x_997 = x[nearest_0997]
y_997 = y[nearest_0997]
plt.vlines(x_997, 0, y_997, color='k')
plt.text(x_997, y_997, '99.7%' , color='k', weight='bold')
plt.text(x_997, -0.05, '{:04.2f}'.format(x_997), color='k',
                                weight='bold',
                                horizontalalignment='center',
                                verticalalignment='center')
plt.grid()
plt.savefig('r_1.jpg')
plt.show()

print('x_68 =',x_68)
print('x_95 =',x_95)
print('x_997 =',x_997)

r2 = np.sqrt(x2**2+y2**2)
sns.distplot(r2, bins=None, hist=True, kde=True, norm_hist=
                                True, color='k')
plt.title('$r_2$')

#Which r_2 contains 68%, 95%, 99.7%
p = sns.kdeplot(r2, color='k')
x,y = p.get_lines()[0].get_data()
cdf = scipy.integrate.cumtrapz(y ,x, initial=0)
```

```
nearest_068 = np.abs(cdf-0.68).argmin()
x_68 = x[nearest_068]
y_68 = y[nearest_068]
plt.vlines(x_68, 0, y_68, color='k')
plt.text(x_68, y_68, '68%' , color='k' ,weight='bold')
plt.text(x_68, -0.05, '{:04.2f}'.format(x_68), color='k',
                                weight='bold',
                                horizontalalignment='center',
                                verticalalignment='center')

nearest_095 = np.abs(cdf-0.95).argmin()
x_95 = x[nearest_095]
y_95 = y[nearest_095]
plt.vlines(x_95, 0, y_95, color='k')
plt.text(x_95, y_95, '95%' , color='k', weight='bold')
plt.text(x_95, -0.05, '{:04.2f}'.format(x_95), color='k',
                                weight='bold',
                                horizontalalignment='center',
                                verticalalignment='center')

nearest_0997 = np.abs(cdf-0.997).argmin()
x_997 = x[nearest_0997]
y_997 = y[nearest_0997]
plt.vlines(x_997, 0, y_997, color='k')
plt.text(x_997, y_997, '99.7%' , color='k', weight='bold')
plt.text(x_997, -0.05, '{:04.2f}'.format(x_997), color='k',
                                weight='bold',
                                horizontalalignment='center',
                                verticalalignment='center')
plt.grid()
plt.savefig('r_2.jpg')
plt.show()

print('x_68 =',x_68)
print('x_95 =',x_95)
print('x_997 =',x_997)

r_tot=np.sqrt(r1**2+r2**2)
sns.distplot(r_tot, bins=None, hist=True, kde=True, norm_hist
                                =True, color='k')
plt.title('$r_{tot}$')

#Which r_tot contains 68%, 95%, 99.7%
p = sns.kdeplot(r_tot, color='k')
```

```
x,y = p.get_lines()[0].get_data()
cdf = scipy.integrate.cumtrapz(y ,x, initial=0)

nearest_068 = np.abs(cdf-0.68).argmin()
x_68 = x[nearest_068]
y_68 = y[nearest_068]
plt.vlines(x_68, 0, y_68, color='k')
plt.text(x_68, y_68, '68%' , color='k' ,weight='bold')
plt.text(x_68, -0.07, '{:04.2f}'.format(x_68), color='k',
                                weight='bold',
                                horizontalalignment='center',
                                verticalalignment='center')

nearest_095 = np.abs(cdf-0.95).argmin()
x_95 = x[nearest_095]
y_95 = y[nearest_095]
plt.vlines(x_95, 0, y_95, color='k')
plt.text(x_95, y_95, '95%' , color='k', weight='bold')
plt.text(x_95, -0.07, '{:04.2f}'.format(x_95), color='k',
                                weight='bold',
                                horizontalalignment='center',
                                verticalalignment='center')

nearest_0997 = np.abs(cdf-0.997).argmin()
x_997 = x[nearest_0997]
y_997 = y[nearest_0997]
plt.vlines(x_997, 0, y_997, color='k')
plt.text(x_997, y_997, '99.7%' , color='k', weight='bold')
plt.text(x_997, -0.07, '{:04.2f}'.format(x_997), color='k',
                                weight='bold',
                                horizontalalignment='center',
                                verticalalignment='center')
plt.grid()
plt.savefig('r_tot.jpg')
plt.show()

print('x_68 =',x_68)
print('x_95 =',x_95)
print('x_997 =',x_997)

# Plot of r_crit vs N

N=range(3,20)
N=list(N)
r_all=[]
```

```
r_crit=[]
r_tot_all=[]

mean = [0, 0]
cov = [[1, 0], [0, 1]]

for i in range(max(N)):
    x , y = np.random.multivariate_normal(mean, cov, size).T
    r = np.sqrt(x**2+y**2)
    r_all.append(r)

r_all=np.array(r_all)  #list of r1, r2, r3, ... etc
r_all2=r_all**2        #list of r1^2 , r2^2, ... etc

for i in N:
    r_tot = np.sqrt(sum(r_all2[:i]))
    r_tot_all.append(r_tot)

for r_tot in r_tot_all:
    p = sns.kdeplot(r_tot, color='k')
    x,y = p.get_lines()[0].get_data()
    cdf = scipy.integrate.cumtrapz(y ,x, initial=0)
    nearest_0997 = np.abs(cdf-0.997).argmin()
    x_997 = x[nearest_0997]
    y_997 = y[nearest_0997]
    plt.vlines(x_997, 0, y_997, color='k')
    plt.text(x_997, y_997, '99.7%' , color='k', weight='bold'
                                    )
    r_crit.append(x_997)
    plt.show()
    plt.clf()

r_crit=np.array(r_crit)
s, i = np.polyfit(N, r_crit**2, 1)

x=np.arange(min(N),max(N)+1)
line = s*x+i
n=10 # random

plt.plot(N, r_crit**2, "r+")
s1='$r_{crit}^2$='
plt.plot(x, line, label=s1+'{:04.2f}N+{:04.2f}'.format(s,i))
plt.hlines(s*n+i, min(N), n, linestyles ="dashed", color='k')
plt.vlines(n, s*(min(N)-1)+i, s*n+i, linestyles ="dashed",
                                color='k')
```

```
plt.ylabel('$r_{crit}^2$')
plt.xlabel('N')
plt.legend(loc='upper left')
plt.grid()
s2='N=10: $r_{crit}$='
plt.text(n+5, 25, s2+'{:04.2f}'.format(np.sqrt(n*s+i)), color
                                ='k',fontsize=12,
                                horizontalalignment='center',
                                verticalalignment='center')
plt.savefig('r_crit.jpg')
plt.show()

print('N=10: r_crit=',np.sqrt(n*s+i))
print('\nslope =',s ,'\ninterception =',i)
```

References

[1] Sharma (2008). *Atomic And Nuclear Physics.* Pearson Education India. 478.

[2] (21 September 2017). *Detecting cosmic rays from a galaxy far, far away.* Science Daily.

[3] Ackermann, M.; Ajello, M.; Allafort, A.; Baldini, L.; Ballet, J.; Barbiellini, G.; Baring, M. G.; Bastieri, D.; Bechtol, K.; Bellazzini, R.; Blandford, R. D.; Bloom, E.D.; Bonamente, E.; Borgland, A. W.; Bottacini, E.; Brandt, T. J.; Bregeon, J.; Brigida, M.; Bruel, P.; Buehler, R.; Busetto, G.; Buson, S.; Caliandro, G. A.; Cameron, R. A.; Caraveo, P. A.; Casandjian, J. M.; Cecchi, C.; Celik, O.; Charles, E.; et al. (15 February 2013). *Detection of the Characteristic Pion-Decay Signature in Supernova Remnants.* Science. 339 (6424): 807–811. arXiv:1302.3307. Bibcode:2013Sci...339..807A. doi:10.1126/science.1231160. PMID 23413352.

[4] Ginger Pinholster (13 February 2013). *Evidence Shows that Cosmic Rays Come from Exploding Stars.*

[5] Pavlidou,V. (2011). *New Results from Ultra High Energy Cosmic Ray Telescopes.* Hipparchos. 2(8): 15-21.

[6] Jui, Charles C.H. (22 March 2000). *Introduction to Cosmic Rays.* High Resolution Fly's Eye. University of Utah. p. 21. Retrieved 15 November 2012.

[7] Kotera, K. & Olinto, A.V. (2011) *The Astrophysics of Ultrahigh Energy Cosmic Rays.* Annual Review of Astronomy and Astrophysics 49 http://arxiv.org/abs/1101.4256

[8] Magkos,G. & Pavlidou,V. (2018). *Deflections of ultra-high energy cosmic rays by the Milky Way magnetic field: how well can they be corrected?* [arXiv:1802.03409]

[9] M. Kachelriess (2008), *Lecture Notes on High Energy Cosmic Rays* [arXiv:astro-ph/0801.4376].

[10] Giorgio Matthiae (9 July 2010). *The cosmic ray energy spectrum as measured using the Pierre Auger Observatory* New Journal of Physics. 12

[11] Swordy, S. (October 2001). *The Energy Spectra and Anisotropies of Cosmic Rays.* Space Science Reviews. 99(1-4): 85-94. https://doi.org/10.1023/A:1013828611730

[12] Blümer, T. Engel, R. Hörandelb, J. (October 2009) *Cosmic rays from the knee to the highest energies* Elsevier. 63(2): 293-338 https://doi.org/10.1016/j.ppnp.2009.05.002

[13] Bellido,J. (2017). *Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above 1017.2 eV and Composition Implications.* ICRC2017, 40-46. Bexco, Busan, Korea. [arXiv:1708.06592]

[14] Vorobiov, Serguei (2008). textitThe Pierre Auger Observatory - a new stage in the study of the ultra-high energy cosmic rays astro-ph. [arXiv:0811.0752]

[15] UIUC, Lecture 21. *The Multivariate Normal Distribution, 21.5:"Finding the Density".*

[16] K. Kawata1, A. di Matteo, T. Fujii, D. Ivanov, C.C.H. Jui, J.P. Lundquist, J.N. Matthews, S. Ogio, H. Sagawa, G.B. Thomson, P. Tinyakov, I. Tkachev and S. Troitsky (2019). *TA Anisotropy Summary* EPJ Web of Conferences 210, 01004 https://doi.org/10.1051/epjconf/201921001004