Master Thesis

Constant Flux Test for Correlation Signals from the Pierre Auger Observatory

He Xi
University of Wuppertal

November 2009

Abstract

The flux of ultra-high energy cosmic rays correlating with positions of nearby active galactic nuclei (AGN) is investigated by making use of the accumulated exposure at the time of observation of each event, and by applying a Kolmogorov-Smirnov (KS) test to the cumulative distribution of number of events vs. exposure. We find that the observation is well compatible with a constant flux of correlating events. An additional study of the non-correlating events and of all highest-energy events gives KS probabilities in the few-percent region. Some implications are discussed.

Contents

Chapter 1

Introduction

The study of the origin of the ultra-high energy cosmic rays (UHECRs) is one of the fundamental researches of astroparticle physics. The correlation of the ultra-high energy cosmic rays with the positions of nearby active galactic nuclei (AGNs) was found by Pierre Auger Observatory in some earlier time [1], which means they might be the main sources of the UHECRs. There were 18 out of 27 total events correlating with active galactic nuclei, however, only 8 out of 31 were found in the later on period. This could happen with low probability if the theory is not wrong. The ring is alarmed to check if the Pierre Auger Observatory is still in stable condition. One clear way doing this task is to study whether the correlating events arrive with constant flux. This is expected if the "true flux" of these events (arriving at Earth) is constant, and if the observation conditions are stable. Good result indicates the correctness of the two hypotheses. Bad result indicates the "true flux" is not constant, or crucially, points to problems in the observation conditions, or both.

The second chapter of this article describes some basic concepts of astroparticle physics. The cosmic rays, air shower generation (which is important for UHECRs detection), GZK effect and active galactic nuclei are briefly introduced. In chapter three, the detectors of Pierre Auger Observatory are described in details. The main idea of the air shower measurement mechanism, and the uncertainties are given. The problem of the research work that this observatory facing and the way to check on the observation conditions are raised. Chapter four focuses on the statistical method named Kolmogorov-Smirnov test, which does the checking task for our study (whether the correlation signals arrive with constant flux). In the last chapter, we apply this method to the signals captured by Pierre Auger Observatory, analyze the testing results and give the conclusion that the deviation of the arrival rates of the two periods is acceptable under the hypothesis of constant flux. Some other ways of investigations of the time evolution of the correlation signal are given at the end.

Chapter 2

Astroparticle Physics and Cosmic Rays

2.1 A glimpse into astroparticle physics

Astroparticle physics is a branch of particle physics that studies elementary particles of astronomical origin, and their relations with astrophysics and cosmology. It is a relatively new field of research emerging at the intersection of particle physics, astronomy and cosmology.

Its rapid development has led to the design of new infrastructure types. In underground laboratories or with specially designed telescopes, antennas and satellite experiments, astroparticle physicists employ new detection methods to observe a wide range of cosmic particles including neutrinos, gamma rays and cosmic rays at the highest energies.

Astroparticle physics aims to answer fundamental questions, such a question asked frequently is:

- What is the origin of cosmic rays?

Even today the answer to the question is not fully understood. However, some research work illustrates there are correlation between ultra high-energy cosmic rays and the active galactic nuclei[1].For further discussion, a brief introduction of cosmic ray and active galactic nuclei is necessary.

2.2 Cosmic rays

The term *Cosmic Rays (CRs)* was introduced by Millikan, which is a misnomer, as cosmic particles arrive individually, not in the form of a ray or beam of particles. The cosmic rays are energetic particles originating from outer space that impinge on Earth's atmosphere. Almost 90% of all the incoming cosmic ray particles are protons, almost 10% are helium nuclei (alpha particles), and slightly under 1% are heavier elements and electrons (beta minus particles) [2].

The CRs have been subdivided into two types: The *primary CRs* are particles which are produced in distant sources and reach the Earth directly from outer space. *Secondary CRs* are induced by interactions between primary (or even secondary) CRs and molecules in the atmosphere. The discovery of the positron and muon [3] and several mesons (e.g the pions and kaons) and hadrons was made in CR experiments before the field of particle physics was dominated by accelerator physics. After a long period CR particle physics research work stagnating because of the successful of the

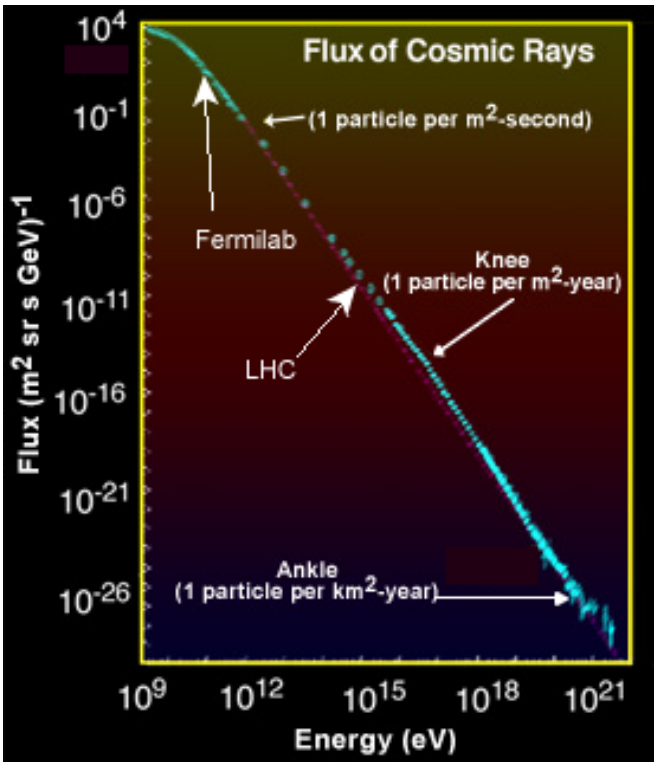


Figure 2.1: Cosmic ray spectrum

accelerators, the discovery of ultra high-energy cosmic rays (UHECRs) bring it back to active, which could own an energy up to 10^9 TeV, while the man-made accelerators can only reach several TeV. The scarcity of the UHECRs should be kept in mind and for sure the spectrum of cosmic rays has been examined in detail.

2.2.1 The Cosmic Ray spectrum

The flux of cosmic ray nuclei has been relatively well-studied at energies below 10^{18} eV. To first approximation, the spectrum is a rapidly falling power law in energy, $d\mathrm{N}/d\mathrm{E} \propto E^{-\alpha}$, with an overall index α of about 2.8. The spectrum does, however, show significant structure (See to Fig 1.1) . At the so-called "knee" a steepening of the spectrum at about 10^{15} eV, the spectral index α changes from about 2.7 to 3.0. A second steepening at about 5×10^{17} eV ($\alpha = 3.3$) is followed by a harder spectral index ($\alpha = 2.7$) above 5×10^{18} eV.

Breaks in the cosmic ray spectrum are typically correlated with changes in the composition. In our current understanding, the decrease of flux in the knee region can be interpreted as a confinement problem: due to their increasing gyroradius, cosmic rays above a critical energy can more easily escape from the Galaxy.

As the gyroradius is proportional to the charge of the cosmic ray particle, this critical energy is smaller for lighter particles. With increasing energy, lighter nuclei will escape earlier, and consequently, the break in the spectrum at the "knee" is accompanied by a change from a mixed to a heavy composition. Measurements of the composition around the "ankle" at 10^{18} eV also indicate a change of the cosmic ray composition towards a lower mean atomic number between 10^{17} and 10^{18} eV, with protons completely dominating the composition at the highest energies. The harder spectrum at 5×10^{18} eV is interpreted as the crossover from Galactic to extragalactic origin of the cosmic rays.

2.2.2 The GZK effect

We expect the cosmic ray spectrum to fall off rapidly around 6×10^{19} eV. This cut-off, first predicted by Greisen (1966) and Zatsepin and Kuz'min (1966) and named the GZK-cutoff, is expected due to the interaction of cosmic ray particles with the 2.7° K cosmic microwave background radiation. The collision of 10^{20} eV protons with 10^{-3} eV photons produces center of mass energies above 100 MeV, which is above the threshold for photo pion production. Subsequently, any proton or nucleus with a travel distance from its origin to the Earth of more than around 50 Mpc suffers severe energy losses, and independent of the original energy will end up with an energy below the GZK cutoff energy. persistent The AGASA[1] cosmic ray experiment has found that the spectrum seems to continue beyond this energy without evidence for a cutoff. This leaves us with a two-fold problem: while it is already difficult to explain how "traditional" astrophysical sources can accelerate protons to energies above 10^{20} eV, the expected energy losses due to interaction with the microwave background require the sources to be relatively nearby, at a distance of 50 Mpc at most. The situation is complicated by the fact that the deflection of protons in Galactic and intergalactic magnetic fields is less than a few degrees at these distances, so cosmic rays should point back to their origin. The distribution, however, seems uniform and shows no strong correlation with the matter distribution in the nearby universe.

2.2.3 Air shower caused by CR

When cosmic ray particles enter the Earth's atmosphere they collide with molecules, mainly oxygen and nitrogen, to produce a cascade of lighter particles, a so-called air shower. The general idea is shown in the Fig. 2.2.3 which shows a cosmic ray shower produced by a high energy proton of cosmic ray origin striking an atmospheric molecule.

This image is a simplified picture of an air shower: in reality, the number of particles created in an air shower event can reach in the billions, depending on the energy and

[1] The Akeno Giant Air Shower Array (AGASA) is a very large surface array designed to study the origin of ultra-high energy cosmic rays. The array is operated by the Institute for Cosmic Ray Research, University of Tokyo at the Akeno Observatory.

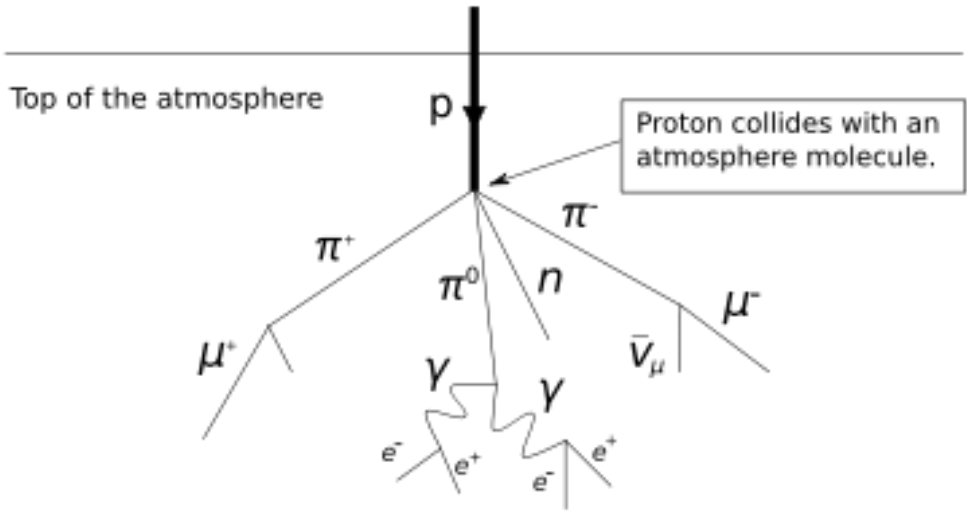


Figure 2.2: An simplified example of air shower. The primary particle is a proton here.

chemical environment (i.e. atmospheric) of the primary particle. All of the produced particles stay within about one degree of the primary particle's path. Typical particles produced in such collisions are charged mesons (e.g. positive and negative pions and kaons).

A simplified model to understand the most important features of the air shower has been introduced by Heitler [4]. In this model, each particle in one interaction process create two particles carrying half of the energy after an interaction length λ. So the number of particles always get doubled and energy of each particle get half of the previous. This sequence continues until the particle energy reaches a critical energy E_c. The particles with energy below E_c only lose energy, get absorbed or decay. If the energy of the primary cosmic ray is E_0, then the maximum number of the particles is given by

$$N_{max} = E_0/E_c \tag{2.1}$$

while the depth of maximum[2] is given by

$$X_{max} = \frac{\lambda}{\ln 2} \ln(\frac{E_0}{E_c}) \tag{2.2}$$

This two equations provide clues to study some important features of the primary cosmic rays by reconstructing the air shower. Observatories such as Pierre Auger are build to capture the air shower. And super computers reconstruct the data collected by detectors to determine the direction and the energy of the original cosmic ray.

[2]The maximum here means when the air shower gets *maximum* particles number.

2.3 Active Galactic Nuclei (AGN)

Although the mechanism of the origin of cosmic rays are not fully understood, some scientific groups provide some research clue. One of the report published by the Pierre Auger Collaboration[3] illustrate the the correlation between the UHECRs and Active Galactic Nuclei [1].

The active galactic nucleus (AGN) here is a compact region at the centre of a galaxy which has a much higher than normal luminosity over some or all of the electromagnetic spectrum (in the radio, infra-red, optical, ultra-violet, X-ray and/or gamma ray wavebands). A galaxy hosting an AGN is called an active galaxy. The radiation from AGN is believed to be a result of accretion of mass by the super massive black hole at the centre of the host galaxy. AGN are the most luminous persistent sources of electromagnetic radiation in the universe, and as such can be used as a means of discovering distant objects; their evolution as a function of cosmic time also provides constraints on cosmological models.

2.3.1 Models of the active nucleus

AGN are found to be compact and persistently extremely luminous. This fact implies the existence of accretion which can potentially give very efficient conversion of potential and kinetic energy to radiation, and a massive black hole which has a high eddington luminosity that provide the observed high persistent luminosity. In the standard model of AGN, cold material close to the central black hole forms an accretion disc. Dissipative processes in the accretion disc transport matter inwards and angular momentum outwards, while causing the accretion disc to heat up. And the accretion disc is expected to peak in the optical-ultraviolet waveband in its spectrum. Although a large fraction of the AGN's primary radiation may be obscured by interstellar gas and dust close to the accretion disc, some other waveband, most likely the infra-red, would be re-radiated. Furthermore, twin highly collimated and fast jets in opposite directions which perpendicular to the disc are produced by accretion discs. So far, the physics of the two jets is not completely understood, but it is a possible source of UHE particles because of the relativistic jet which can produce enormous particle energies assuming first order diffuse Fermi acceleration.

2.3.2 Unification

Unified models of AGN unite two or more classes of objects, based on the traditional observational classifications, by proposing that they are really a single type of physical object observed under different conditions. The currently favoured unified models are 'orientation-based unified models' meaning that they propose that the apparent

[3]The Pierre Auger Observatory is an international cosmic ray observatory designed to detect ultra high-energy cosmic rays. It is located in western Argentina's Mendoza Province.The details will be given in chapter 3.

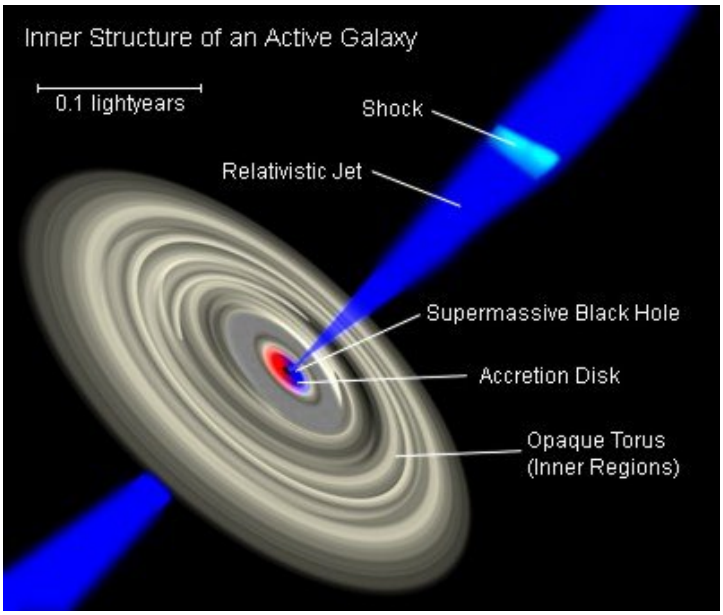


Figure 2.3: The inner structure of an active galaxy

differences between different types of objects arise simply because of their different orientations to the observer.

The Fig. 2.3.2 shows the unification by viewing angle. For an overview of these see [5] and [6].

Since the possible source of UHE particles is the jet from AGN, it is for sure that it has a low possibility to observe UHECRs from the directions of some AGNs, for not all of the jets pointing to the Earth. Surely that a certain deviation is allowed, but once it is close to 90 degree, the chance to capture UHECRs from that AGN would be nearly 0.

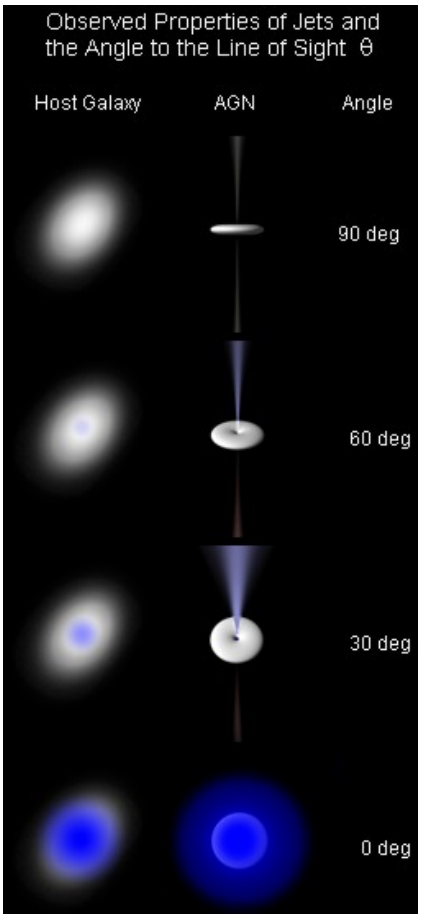


Figure 2.4: Observed properties of jets and the angle to the line of sight θ

Chapter 3

Pierre Auger Observatory

The Pierre Auger Observatory signals and the observation condition are the main topics of this article. This observatory is an international cosmic ray observatory designed to detect ultra-high energy cosmic rays. It is named after the French physicist Pierre Auger, who first introduced the notation of *extensive cosmic ray shower* in late 1930s. This extensive cosmic ray shower, or *extensive air shower (EAS)*, and the companion effect are the observative clues for measuring the energy spectrum, arrival directions and the chemical composition of the primary cosmic rays.

In reality, it is not possible to track the whole composition of the extensive air shower caused by the primary cosmic ray, but detecting the lateral distribution function of particles that reach the ground would be one possible choice. Another technique used is to detect the nitrogen fluorescence in the atmosphere caused by the extensive air shower. The two complementary techniques can form a hybrid technique that enhance the resolution and be valuable in determining systematic errors inherent in both techniques as well as providing more information to determine the particle kind and check hadronic interaction models [8]. The observatory will consist of northern and southern sites located at mid-latitudes in order to achieve a full sky coverage. The Auger collaboration has finished constructing the southern site in Malargüe, located at an elevation of 1400 m in the province of Mendoza, Argentina as shown in Fig. 3.1. It covers an area of 3000 km^2 in order to collect a couple of events above 10^{20} eV per year. The northern site is located in southeast Colorado, United States and the construction time is scheduled to be 2009–2012 [9]. Once finished, the northern site will cover an area of 10000 km^2 with 4000 water-Čerenkov tanks and additional fluorescence telescopes.

3.1 Surface detector

As shown in Fig. 3.1, the surface detector (SD) of the southern array is consist of 1600 water-Čerenkov stations set on a regular triangular grid, with 1.5 km separation each other [10], covering an area as big as 3000 km^2 in order to detect more ultra high energy cosmic rays that with energies above 10^{18} eV. Each of the stations is a cylindrical tanks, filled with 12000 liter of purified water. Once a charged particles that achieve a certain high speed enter the tanks, electromagnetic radiation would be emitted [11]. This electromagnetic radiation is called *Čerenkov light*. The water in the tanks has a high diffuse reflectivity in the wavelength of combined maximum Čerenkov light production, water transmissivity and photocathode sensitivity. Three

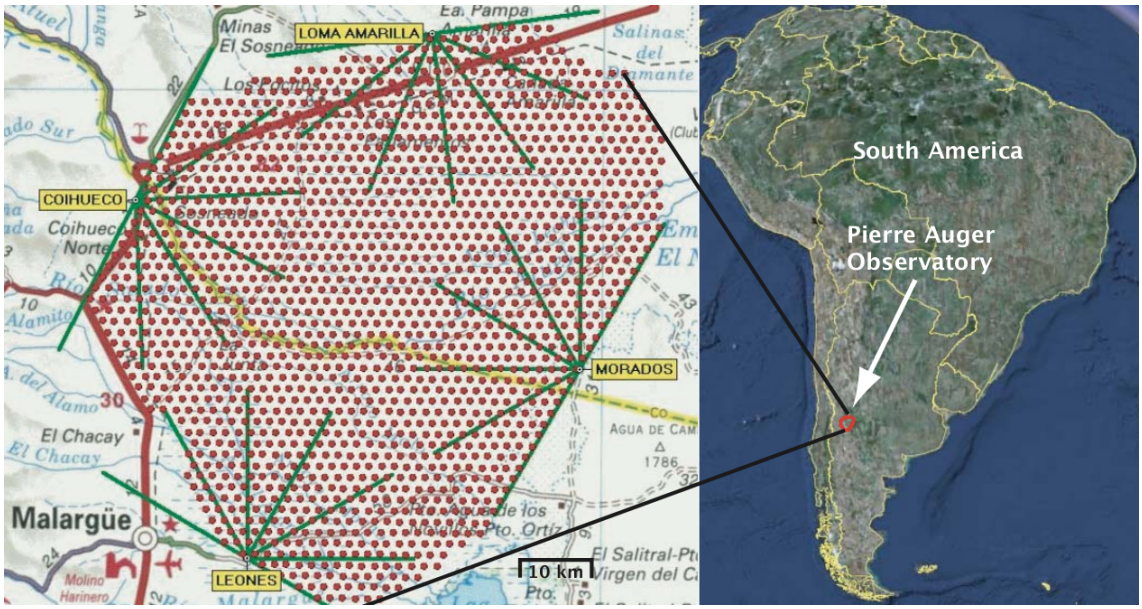


Figure 3.1: A map of the Pierre Auger Observatory with 1600 water tanks (red dots) and four fluorescence telescopes labeled in yellow located next to Malargüe, Argentina.

9″ Photomultiplier tubes (PMTs) are placed on the top of the tank detecting Čerenkov light when particles propagate through the detector. The signals are firstly filtered and then read out by a flash analog digital converter (FADC) that samples at a rate of 40 MHz. Some ring buffer memories are designed to store the data temporarily and processed by a programmable logic device (FPGA) to implement various trigger conditions [12] [13]. The timing information for each station is received from a GPS system located on each tank with timing resolution < 20 ns [14]. The power of each station is supplied by two solar panels, combined with buffer batteries. One picture of one surface detector is shown in Fig. 3.2.

Large amounts of data is generated in each short time, the stored signals are transferred to the Central Data Acquisition System (CDAS) only if a shower trigger has been detected in three adjacent tanks simultaneously. The water tanks of the surface detector are continuously monitored and calibrated by single cosmic muons. By adjusting the trigger rates, the PMT gains are matched to within 5%. For convenience, the number of particles in each tank is defined in units of Vertical Equivalent Muons (VEM) defined as the average charge signal produced by a penetrating down going muon in the vertical direction. The stability and the trigger rates are remarkably uniform over all detector stations [15].

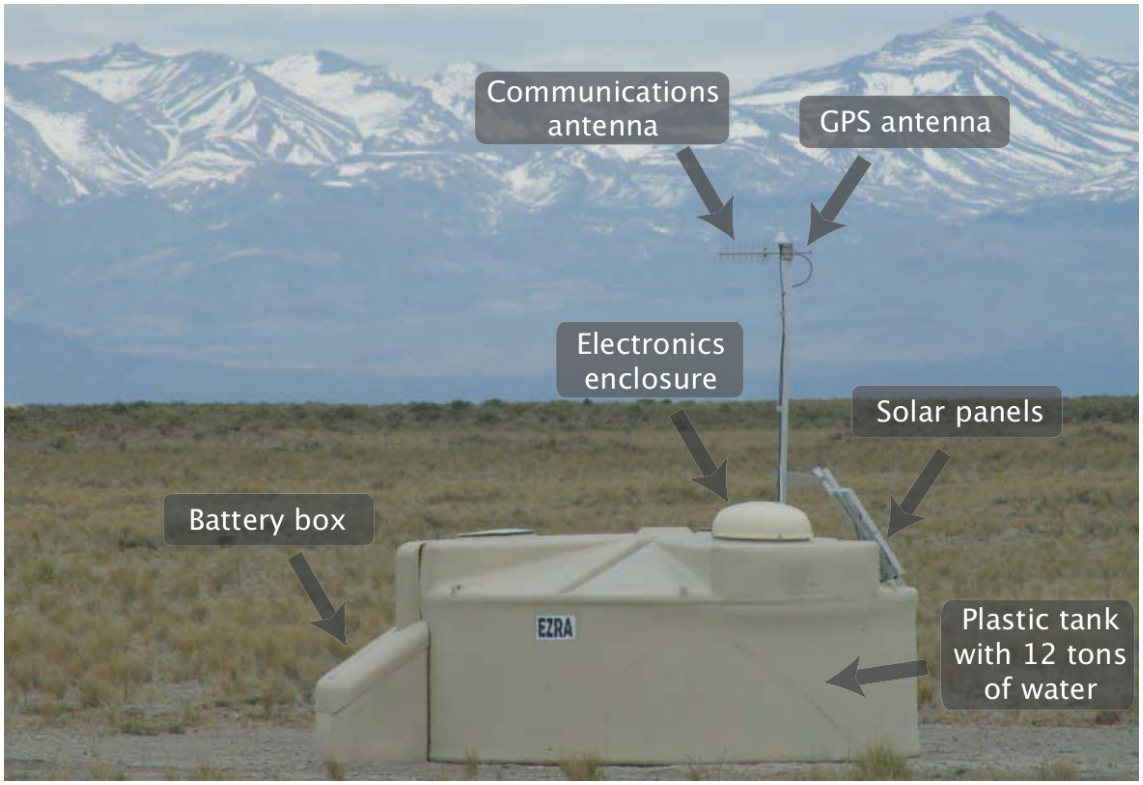


Figure 3.2: One example of surface detector station.

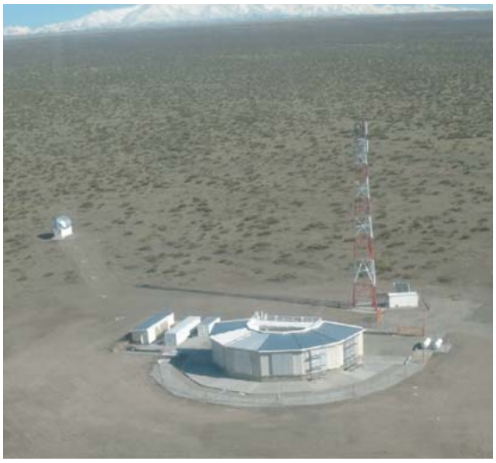
3.2 Fluorescence detector

While charged particles propagating through the atmosphere fluorescence light is emitted by de-excitation processes of nitrogen molecules. The fluorescence detector (FD) of the southern array is conceived to the light. The fluorescence yield is very low, which is approximately 4 photons per meter of electron track [16], but large imaging telescopes are able to detect this light during clear new to half moon nights, resulting in a duty cycle of $\approx 10 - 15\%$.

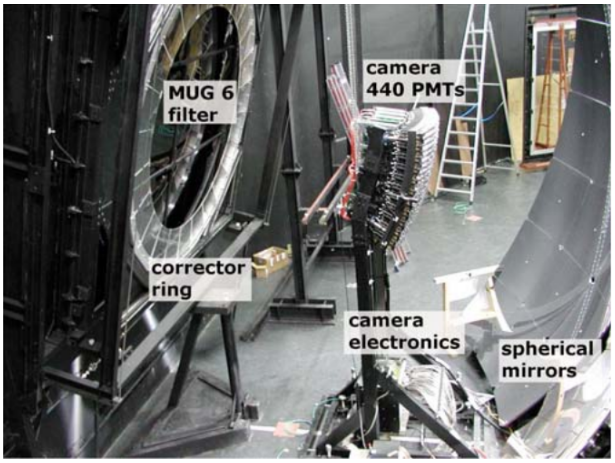
Four different eyes are build up at the corners of the observatory area (named Los Leones, Los Morados, Loma Amarilla and Coihueco) as shown in Fig. 3.1. They are arranged in a way that cover the perimeter of the SD, which enables detection of EAS simultaneously by SD and FD ("hybrid detection"). Each eye consists of 6 independent Schmidt telescopes (bays). Each bays are made of a 440 pixel camera, and each pixel views approximate $1.5^\circ \times 1.5^\circ$ of the sky. The 440 pixels are arranged in a 22×20 matrix to give a field of view of 30° in azimuth and 28.6° in elevation, adding to a 180° view inwards the array of one eye (cf. Fig. 3.1). A 12 m^2 mirror with a radius of 3.4 m is used to reflect the fluorescence light to the camera, which located at the focal surface of the mirror. The telescopes use a Schmidt optics design to avoid coma aberration, with a diaphragm, at the center of curvature of the mirror. The radius of the diaphragm is 1.1 m including a corrector lens with an inner radius of 0.85 m and outer radius of 1.10 m. The effect of the lens is to increase the light collection area by a factor of two while maintaining an optical spot size of 0.5° [17]. To avoid interfering background light each diaphragm has a UV transparent filter that restricts the incoming light to the wavelength range between 300 and 420 nm, which is where the main fluorescence emission lines can be found. To reduce signal losses when fluorescence light crosses PMT boundaries, small light reflectors ("mercedes stars") are placed between the PMTs [18].

The PMT signals are continuously digitized at 10 MHz sampling rate with a dynamic range of 15 bit in total. A FPGA based multi-level trigger system is used to filter traces out of a random background.

To measure air shower energies correctly the fluorescence detectors have to be calibrated and monitored. The *absolute calibration* provides the conversion between the digitized signal (in ADC units) and the photon flux incident on the telescope aperture. This calibration of each telescope is performed three or four times a year. During the calibration a large homogeneous diffuse light source was constructed for use at the front of the telescope diaphragm. This drum shaped source has a diameter of 2.5 m and the emitted light is known from laboratory measurements [18]. The ratio of the drum intensity to the observed signal for each PMT gives the required calibration. The main goal of the relative calibration is to monitor short term and long term changes between successive absolute calibration measurements and to check the overall stability of the FD. The atmospheric conditions must be monitored closely since attenuation of the light from the EAS to the telescope due to molecular (Rayleigh) and aerosol (Mie) scattering has to be corrected. Several methods are currently used to



(a) Photo of fluorescence telescope Loma Amarilla.



(b) Design of the fluorescence telescope [15].

Figure 3.3: Fluorescence detector of the Pierre Auger Observatory

determine the effects in the air at any given time during data taking. The relevant parameters are determined by a Horizontal Attenuation Monitor (HAM), Aerosol Phase Function monitors (APF) and a Laser Illuminated Detection And Ranging system (LIDAR) located at each eye (cf. [19] [20]). There are also cloud and star monitors to detect clouds and track stars and any changes in their intensity caused by changing atmospheric conditions.

3.3 Central laser facility

Another complementary measurement of the aerosol vertical optical depth Vs. height and the uniformity of the atmosphere across the aperture of the array is provided by the central laser facility (CLF) [21]. It is a steerable automatic system which produces regular pulses of linearly polarized UV light at 355 nm. It is located in the middle of the array, 26 km away from Los Leones (cf. Fig. 3.4). In addition, the CLF provides a laser generated "test beam" for the observatory. This system creates an artificial hybrid cosmic ray event by feeding a signal into a nearby tank (Celeste) through a fiber optics cable. The scattered laser light is intense enough to be registered by all eyes thereby providing a real-time confirmation that the FD eyes are functioning and are able to "see" the array center. The time recorded at each detector is used to measure and monitor the relative timing between SD tanks and FD eyes. The stability of that time offset has been measured by previous measurements to be ~ 100 ns [22].

The possibility to determine the shower axis in mono-mode and single-tank hybrid mode offers the ability to test the accuracy of hybrid reconstruction: Since vertical

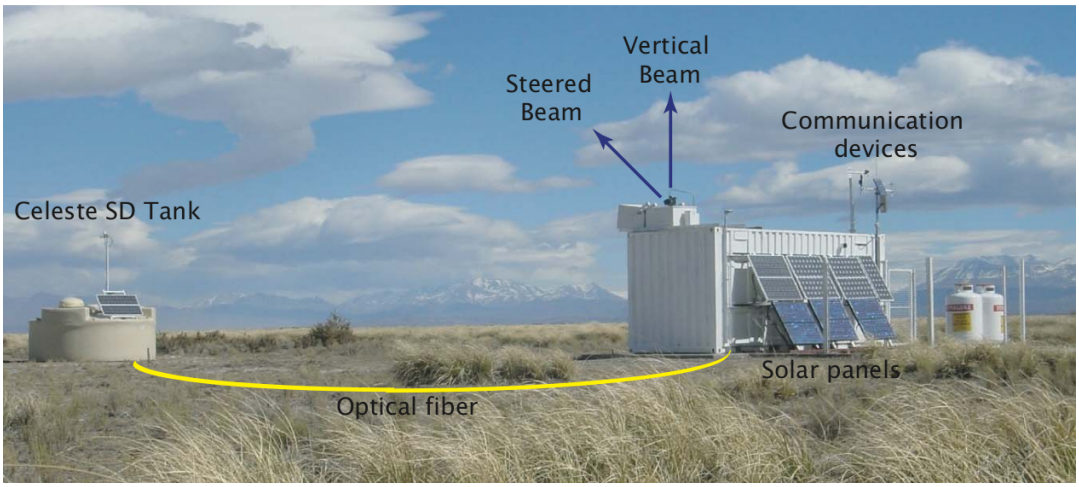


Figure 3.4: The central laser facility [23].

laser shots the location of the CLF could be determined with a resolution of 550 m in mono-mode and after including the timing information of the single water tank, the resolution improved to 20 m without a systematic shift [15].

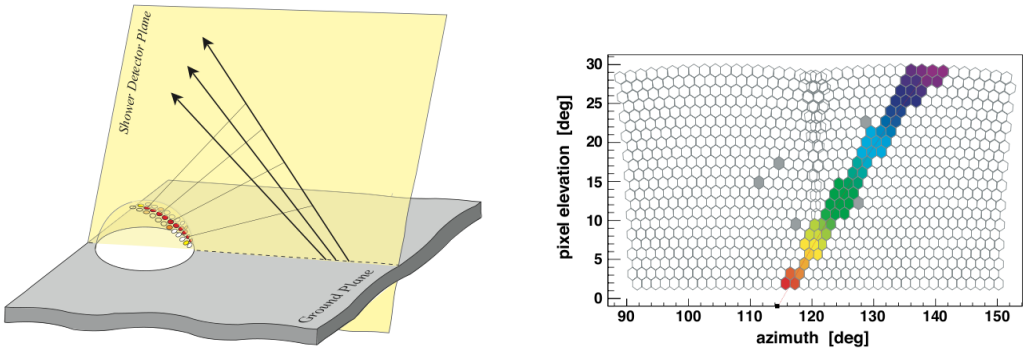
3.4 Reconstruction

With Pierre Auger Observatory, we detect the lateral distributed particles of the air shower that reach the ground by surface detector, and we detect the nitrogen fluorescence generated by the air shower with the fluorescence detector. Both of the information is separation of signals detected by a group of individual photon cameras. Reconstruction of these data is needed to measure some ruling parameters of the air shower. And then we can determine the energy spectrum, arrival directions and the chemical composition of the primary cosmic ray.

The reconstruction of the shower axis is relatively easy and would be the start point. With the data collected by the fluorescence detector, we can get a line-like geometry with the pixels in one eye. This information actually implies the axis of the shower. The arrival time of the signals on each pixel should be also considered in, to determine the direction of it. An typical example of this strategy is show in Fig. 3.5 [8].

In this figure, we can see that the geometry only tells us in which plane is the axis. This plane is called *shower detector plane (SDP)*. To know the exact direction, we need to analysis the arrival time of the fluorescence light at the telescope. It is normally done by finding the closest distance of the axis to the telescope, together with the shower inclination with respect to the SDP. For a detailed discussion of the strategy used, look in [8].

The energy information of the primary particle is involved in the quantity of the



(a) All axis in the same plane induce the same geometry in the camera pixels. This plane is the shower detector plane.

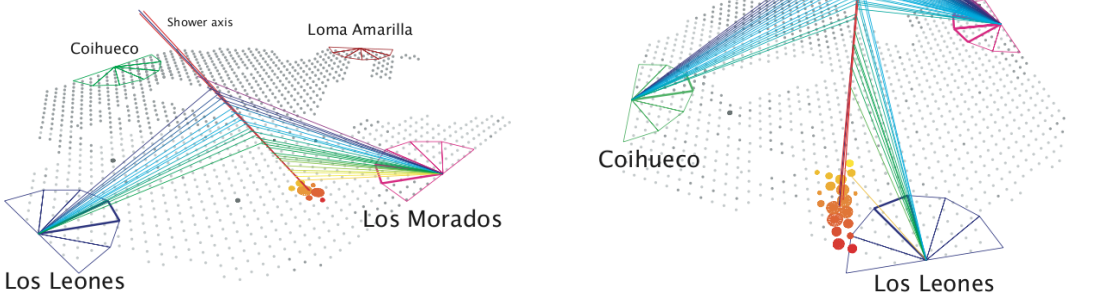
(b) Light track of event 3308259 as seen by two adjacent fluorescence cameras (Los Morados). The arrival time is indicated by color from purple (earlier) to red (later).

Figure 3.5: A typical example of fluorescence detector signal geometry.

photons captured by each pixel of the camera. This is analyzable after we get the direction of the shower axis. Hence in principle, the fluorescence detector is able to achieve the shower measurement without extra information provided by surface detector signals (of course, with lower precision). And four eyes functioning simultaneously cover the full horizon. Air showers faraway are also detectable by them.

Showers that hit the ground in the area where the surface detector located can be detected by SD. One shower is usually capture by several adjacent surface detector tanks. Some reconstruction strategies use then. For the first step, the shower axis is determined by a fit of the arrival times of the shower front at the SD. The angular resolution is defined as the angular aperture around the arrival directions of cosmic rays within which 68% of the showers are reconstructed, which has been verified experimentally [24] [25]. Almost all events with energies above 10 EeV trigger at least six SD stations haveing an angular resolution better than 1° [26] [24]. For the second step, the shower size S is calculated from the signal detected in each surface station and then converted to energy using a linear calibration curve based on the fluorescence telescope measurements [27]. The uncertainty in energy scale for the set of UHE events used in the present analysis is about several tens of percentage due to the systematic errors and relatively low statistics available for calibration in this energy range.

The shower arrives at SD would be also captured by one or several FD eyes. Different eyes besides SD will give different reconstruction results of the air shower, a hybrid geometry reconstruction is applied under this situation and it improves the measurement quite a lot. Some examples of hybrid events are shown in Fig. 3.6and one hybrid reconstruction result is shown in Fig. 3.7.



(a) Stereo event 3031623 seen by Los Leones and Los Morados.

(b) Triple event 3336808 seen by Los Leones, Los-Morados and Coihueco.

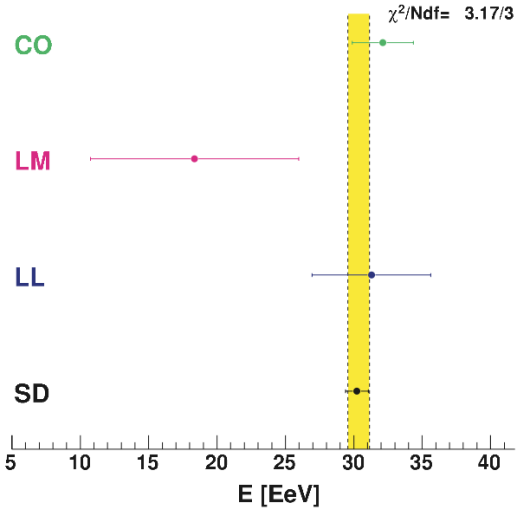
Figure 3.6: Hybrid events seen by Pierre Auger Observatory.

3.5 Astro findings of Pierre Auger Collaboration

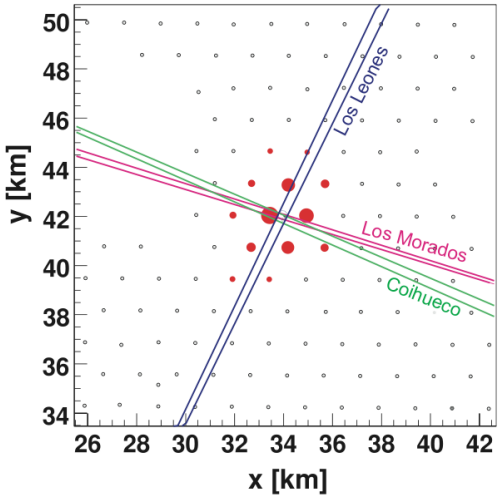
Several articles published by Pierre Auger Collaboration reveal some new findings in the area of astroparticle physics. One of the main findings is the correlation of ultra-high energy cosmic rays with the positions of the near by active galactic nuclei [1]. This provides the experimental proof of the theory that AGNs are very important sources of those UHECRs.

In the correlation analysis, the events are chosen in the period since 1 January 2004 to 31 August 2007. It contains 81 events with reconstructed energies above 40 EeV and zenith angles smaller than 60°. Higher the energy the better, for it would be less curved by the background cosmic magnetic field. It would arrive at the Earth with smaller deviation in the original direction. Besides, the higher energy leads to better measurement. Consequently, the AGNs taken in consideration should not be too faraway, the GZK effect would play a role and the particles would have a low possibility to reach Earth if they are with high energy. The catalogue of quasars and active nuclei by Véron-Cetty and Véron (V-C) is used [28]. The cut of redshift of the AGNs is $z \leq 0.024$, corresponding to distances D smaller than 100 Mpc, and 694 AGNs in total under this condition.

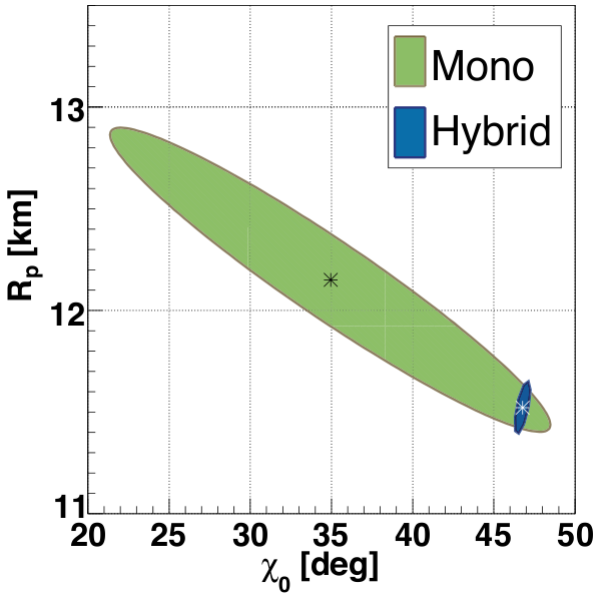
For consideration on the measurement quantity, only the events with strict criteria with regard to the quality of the reconstructions of their energy and direction are considered. The selection of those events is done via a quality trigger [29]. This trigger requires that



(a) The shower energies reconstructed by three eyes and the surface array of event 3351072. The yellow band represents a weighted average of the sub-detectors with the statistical uncertainty σ_E.



(b) Two dimensional plot of the surface array of event 3351072. Stations used for SD reconstruction are shown in red colors. Two lines from each FD with a successful FD reconstruction show the $\pm 1\sigma$ projections of the SDPs on the array plane.



(c) Solution for the axis for mono and hybrid reconstruction (1σ accuracy).

Figure 3.7: Hybrid events seen by Pierre Auger Observatory.

- The detector with the highest signal must be surrounded by five active nearest neighbors.
- The reconstructed shower core be inside an active equilateral triangle of detectors.

This represents an efficient quality cut while guaranteeing that no crucial information is missed for the shower reconstruction.

3.5.1 Search method

The main framework of the correlation test is to search a binomial possibility P in equation 3.1

$$P = \sum_{j=k}^{N} C_j^N p^j (1-p)^{N-j} \tag{3.1}$$

which denotes the probability P for a set of N events from an isotropic flux to contain k or more events at a maximum angular distance Φ from any member of a collection of candidate point sources. The p in the equation is the fraction of the sky (weighted by the exposure[1]) defined by the regions at angular separation less than Φ from the selected sources. The optimal cut for energy $E_{\rm th}$, Φ and the redshift z were unknown before the search work. During the search work, each parameter would be increased a little bit in a search loop, with a certain energy cut $E_{\rm th}$, we get N selected events; with redshift cut z, we determine the adapted AGNs; with the Φ, we scan all over the adapted AGNs with the selected events to determine the k, and can determine the p; finally the binomial probability P is know by equation 3.1. In this process, a function $P(E_{\rm th}, \Phi, z)$ is constructed, and the minimum value can be found. The data between 1 January 2004 and 26 May 2006, which is called period I, is scaned and the minimum value is found when $E_{\rm th} = 56$ EeV, $\Phi = 3.1°$, $z = 0.018 (D \leq 75$ Mpc). With this cut criteria, there are 15 events adapted and 12 of them are determined to be correlated. The $p = 0.21$ and the probability that this configuration would occur by chance if the flux were isotropic is $P = 1.755 \times 10^{-6}$ [30]. This denotes that the source positions of UHECRs are not isotropic and has a high correlation with AGNs. One example of the sky map is shown in Fig. 3.8

3.5.2 Challenge facing the detection

The observation during a certain period after 26 May 2006, which is the last day for the data set determine the cut criteria, is consistent with the theory of correlation between UHECRs and the AGNs. This period is from 27 May 2006 through 31 August 2007. It is called period II. There are 13 events with energy above 56 EeV, of which 8

[1] The definition of exposure is given in section 3.5.3.

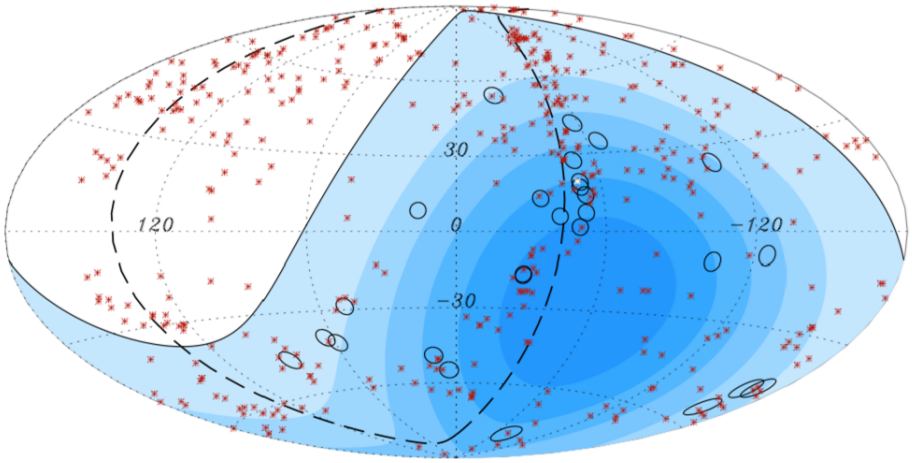


Figure 3.8: Aitoff projection of the celestial sphere in galactic coordinates with circles of 3.2° centred at the arrival directions of 27 cosmic rays detected by the Pierre Auger Observatory with reconstructed energies $E > 57$ EeV. The positions of the 442 AGN (292 within the field of view of the Observatory) with redshift $z \leq 0.017 (D < 71$ Mpc) from the 12th edition of the catalogue of quasars and active nuclei [31] are indicated by asterisks. The solid line draws the border of the field of view for the southern site of the Observatory (with zenith angles smaller than 60°). The dashed line is, for reference, the super-galactic plane. Darker colour indicates larger relative exposure[2]. Each coloured band has equal integrated exposure. Centaurus A, one of the closest AGN, is marked in white.

have arrival directions closer than 3.1∘ from the positions of AGN less than 75 Mpc away. The probability is $P = 1.74 \times 10^{-3}$.

A calibration was applied to the reconstruction then. Directions and energies of the previous captured events were measured again with the new calibration. It did not include large changes in the measurement results. A new cut criteria is determined[3], with $E_{th} = 55$ EeV, $\Phi = 3.1°, z = 0.018$.

However, during a long period from 1 September 2007 through 31 March 2009 (period II), only 8 out of total 31 events are found to be correlating with nearby AGNs. The P-value is as large as 0.31. This statistic supports the isotropic distribution hypothesis more. The signals detected in this period ring the alarm on the Pierre Auger Observatory research work: is the theory publish by this group should be reconsidered? or some problems of the Pierre Auger Observatory detectors were induced without being noticed? or the reconstruction has not high efficiency as demanded? or some problems else?

Of course there exists some possibility that this "unexpected" statistic result occurs without any problem happens. However, a cross check is needed to make sure the statistic (8 out of 31 total events correlated) occurs with an acceptable probability under the condition that everything is ok. Or find some other evidences illustrate problem existing.

3.5.3 One way to do the cross check

In this period, the negative thing is that only 8 out of 31 total events are determined to be correlated. It seems too rare of them. One way to do the check is to exam whether the correlation signals arrive with constant flux. This is expected if the "true flux" of these events (as arriving at Earth) is constant, and if the observation conditions are stable. Good result indicates the correctness of the two hypotheses, bad result indicates the "true flux" is not constant or points to problems in the observation conditions, or both. Here, the constant flux rate is not determined by time, for the observatory is in construction during those periods. The signals are expected to come at a faster rate when the observatory gets more efficiency because of construction accomplishment. The efficient parameter for the determination would be *exposure* of the Pierre Auger Observatory.

The integrated exposure represents the effective observation time of the Pierre Auger Observatory detectors. The surface detector array has full acceptance of events with energy above 3 EeV [24]. Above this energy the direction efficiency is lager than 99% and it is nearly independent of the direction of the shower axis defined by zenith angle (θ) with respect to the local vertical an azimuth (ϕ) with respect to South. Thus, above that energy the instantaneous instrument aperture as a function of zenith angle

[3]It is not determined in the way as described in previous content, which scans all over the events with AGNs and get the $E_{\rm th}$ ϕ z of the minimum P. It just make sure there are 27 total events pass the energy cut in the period I+II. As the result, the new energy cut is 55 EeV, and $\Phi = 3.1°$ $z = 0.018$ are unchanged.

is given by

$$A(t) = \int n(t) a_0 \cos\theta d\Omega dt \tag{3.2}$$

where $n(t)$ is the number of active cell of SD at time t, $a_0 \cos\theta$ is the surface of a unitary cell under the incidence zenith angle θ. And one cell is consist of several adjacent SD stations.

Doing the test for period III only is not the correct way, for we are checking if the correlation signals come at a constant rate in the period II+III, not just to see if it is constant in period III.

Chapter 4

Methods for Hypothesis Testing

Statistical decision problems in which there are just two possible actions constitute an important class called *hypothesis testing* problems. The possible states of nature are called *hypotheses* about nature; each individual state is termed a *simple* hypothesis. A simple hypothesis, then, is a complete specification of a probability distribution of the population on which observations are obtained for inference. The "hypothesis" is that this particular distribution is the correct one. A set of several states of nature, or the "hypothesis" that the actual state of nature is one of those in the set, is called a *composite hypothesis* [32]. Of course, we aim to see whether the correlation signal is a constant or not, so this would be just a simple hypothesis.

In this chapter, we will talk about the hypothesis testing methods.

4.1 Testing hypothesis

The process of standard hypothesis testing usually involves the following four steps [33]:

1. Formulate the null hypothesis H_0 (commonly, that the observations are the result of pure chance) and the alternative hypothesis H_α (commonly, that the observations show a real effect combined with a component of chance variation).

2. Identify a test statistic that can be used to assess the truth of the null hypothesis.

3. Compute the P-value, which is the probability that a test statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true, The smaller the P-value, the stronger the evidence against the null hypothesis.

4. Compare the P-value to an acceptable significance value α (sometimes called an *alpha* value). If $P \leq \alpha$, that the observed effect is statistically significant, the null hypothesis is ruled out, and the alternative hypothesis is valid.

Back to our research, we are testing whether the correlated events arrive with constant flux or not. We can formulate it as null hypothesis and alternative hypothesis as follow:

$$\begin{cases} H_0 & : & nullhypothesis \\ H_1 & : & alternativehypothesis \end{cases} \tag{4.1}$$

Of course, the null hypothesis is constant flux and alternative hypothesis is not constant flux in our case. After formulating the hypothesis, we should choose a statistical test to assess the truth of the hypothesis. There are several different test methods, and they have different advantages. Some powerful test method was chosen in this work. This test would be described in detail in the following sections. We will get the confidence level that the Pierre Auger Observatory signal sample consistent with the null hypothesis by the end.

4.2 Different testing methods

Goodness-of-fit tests are used to test the hypothesis that nature follows a certain law when the alternative hypothesis is the general one that nature does not.

In the case of a discrete random variable X with a finite number of possible values $x_1, \ldots, x_k$ with corresponding probability $p_1, \ldots, p_k$, the null hypothesis is

$$H_0 : \ p_1 = \pi_1, p_2 = \pi_2, \ldots, and \ p_k = \pi_k \tag{4.2}$$

where $pi_1, \ldots, \pi_k$ are specified numbers on the interval $[0, 1]$ whose sum is 1. The basis for testing H_0 is a random sample of n observations on X, usually presented in a tabulation such as this:

Value	x_1	x_2	$\cdots$	x_k
Frequency	f_1	f_2	$\cdots$	f_k

4.2.1 The classical χ^2-Square test

It is natural to consider the differences $(f_i - n\pi_i)$ as related to the goodness of fit of the observed freuencies f_i to the expected frequencies $n\pi_i$. If these differences are larger than sampling fluctuations would ordinarily produce, there would be reason to reject the π's as the true cell probabilities. Pearson introduced in 1900 the following measure, which is large when the differences $(f_i - n\pi_i)$ are large:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - n\pi_i)^2}{n\pi_i} \tag{4.3}$$

Aside from intuitive arguments that can be and have been proposed for this statistic, one of its virtues is that its asymptotic distribution is know under H_0. This asymptotic distribution, in fact, is identical with the χ^2-square distribution with $k-1$ degrees of freedom [34].

The χ^2-square test requires discrete distribution sample and it is not sensitive when the entries are not much in each bin. We will skip the details of this test and talk about Kolmogrov-Smirnov test, which is adopted here.

4.2.2 Other tests

The χ^2-square test is not the only tool for goodness-of-fit test. Some other tests such as Kolmogorov-Smirnov test can do the same task. The *Kolmogorov-Smirnov test (K-S test)* is a form of minimum distance estimation used as a nonparametric test of equality of one-dimensional probability distributions. It is useful to compare a sample with a reference probability distribution (one-sample K-S test), or to compare two samples (two-sample K-S test).The null distribution of this statistic is calculated under the null hypothesis that the samples are drawn from the same distribution (in the two-sample case) or that the sample is drawn from the reference distribution (in the one-sample case). In each case, the distributions considered under the null hypothesis are continuous distributions, and *exposure* value of the cosmic ray signal just full fill this requirement. Furthermore, this method has several advantages over χ^2-square test [35]:

> 1. It treats the individual observations separately, and no information is lost because of binning.
>
> 2. It works for small samples; for very small samples it is the only alternative. For intermediate sample sizes it is more powerful.
>
> 3. Note that as described here, the Kolmogorov-Smirnov test is non-directional or two-tailed, as is the chi-square test. However, a method of finding probabilities for the one-tailed test does exist (Birnbaum & Tingey 1951; Goodman 1954), giving the Kolmogorov-Smirnov test yet another advantage over the chi-square test.

With the fact that the signal of ultra-high energy cosmic rays measured by Pierre Auger Observatory is really scarce, the 2nd advantage mentioned above illustrates

that Kolmogorov-Smirnov test is much more powerful than chi-square test in our case.

Of course there exists other methods to do the test, they can also be applied and the results can be compared in further steps.

4.3 Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test is the method we adopted, and the reason is explained well in last section. In this section, we will look into the details of this method such as the Kolmogorov-Smirnov statistic and Kolmogorov distribution, which are the basic concepts furthermore the steps to apply it.

4.3.1 Kolmogrov-Smirnov statistic

Kolmogorov-Smirnov statistic is a very important value that decide the goodness-of-fit test, we need it for later on processes. Firstly we should know the empirical distribution function $F_n(x)$ for n independent and *identically-distributed (iid)* random variables X_j, which is defined as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \leq x} \tag{4.4}$$

where $I_{X_i \leq x}$ is the indicator function , equal to 1 if $X_i \leq x$ and equal to 0 otherwise, or represented as follow

$$I_{X_i}(x) = \begin{cases} 1 & : \quad X_i \leq x \\ 0 & : \quad X_i > x \end{cases} \tag{4.5}$$

This empirical function actually is the cumulative possibility function of the sample, and it should be normalised, for it is easy to see that $F_n(x)$ is from 0 to 1.

let us look at a simple example: assume we have a sample of data

X_i	X_1	X_2	X_3	X_4	X_5
	0.11	0.34	0.45	0.67	0.88

and our task is to test whether this sample of data is evenly distributed in the interval $[0, 1]$ or not, so the function $F_n(x)$ would be

$$F_n(x) = \begin{cases} 0 & : \quad 0.00 \leq x < 0.11 \\ 0.2 & : \quad 0.11 \leq x < 0.34 \\ 0.4 & : \quad 0.34 \leq x < 0.45 \\ 0.6 & : \quad 0.45 \leq x < 0.67 \\ 0.8 & : \quad 0.67 \leq x < 0.88 \\ 1 & : \quad 0.88 \leq x \leq 1.00 \end{cases} \tag{4.6}$$

and the function shape is shown in Fig. 4.1.

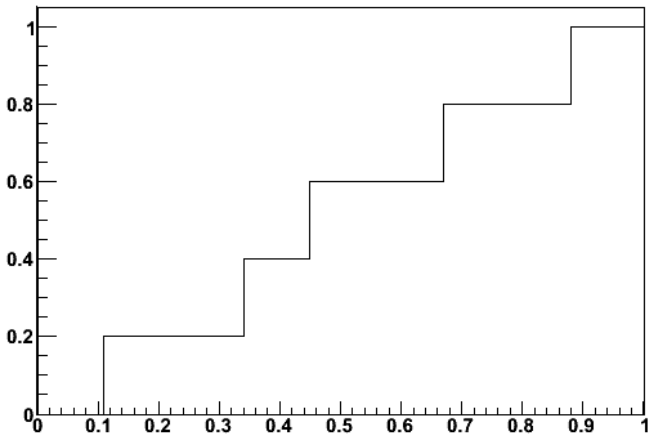


Figure 4.1: The empirical possibility function of the sample in our example. It is a step function that arises by 0.2 at each point X_i, and reaches to 1 finally.

Let $F(x)$ be our hypothesis, a cumulative possibility function of flat distribution, should be a straight line arise up from 0 to 1. And the *Kolmogorov-Smirnov statistic* for our sample $F_n(x)$ and hypothesis $F(x)$ is

$$D_n = \sup_x |F_n(x) - F(x)| \tag{4.7}$$

where sup S is the supremum of set S. This D_n here actually is the maximum deviation between $F_n(x)$ and $F(x)$, which would be 0.15 in our case, shows in the following Fig. 4.2

By the *Glivenko-Cantelli theorem*, if the sample comes from distribution $F(x)$, then D_n converges to 0 for $n \to \infty$ [36]. Kolmogorov strengthened this result, by effectively providing the rate of this convergence (see below).

4.3.2 Kolmogorov distribution

The *Kolmogorov distribution* is the distribution of the random variable

$$K = \sup_{t\in[0,1]} |B(t)| \tag{4.8}$$

where $B(t)$ is the *Brownian bridge*. This Brownian bridge is a continuous-time stochastic process whose probability distribution is the conditional probability distribution of a *Wiener process* $W(t)$ (a mathematical model of Brownian motion) given the condition that $B(0) = B(1) = 0$. The expected value of the bridge is zero, with variance $t(1-t)$, implying that the most uncertainty is in the middle of the bridge, with zero uncertainty at the nodes [37]. An example of the Brownian bridge is given in the Fig. 4.3.

The cumulative distribution function ok K is given by

$$Pr(K \leq x) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2x^2} = \frac{\sqrt{2\pi}}{x}\sum_{i=1}^{\infty}e^{-(2i-1)^2\pi^2/(8x^2)} \tag{4.9}$$

where K is a function of D_n and n

$$K = K(D_n, n) \tag{4.10}$$

This is the probability that the sample function concentrate in the shadow area in Fig. 4.4. However, we aim at get the confidence level for the interval $x \in [K, \infty]$, which means we are trying to get the probability that the sample exit out the shadow area, and this probability should be

$$Pr(K > x) = 1 - Pr(K \leq x) = 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2x^2} = 1 - \frac{\sqrt{2\pi}}{x}\sum_{i=1}^{\infty}e^{-(2i-1)^2\pi^2/(8x^2)} \tag{4.11}$$

and the shape of the function is shown in Fig. 4.5.
From 0 to 0.5, the confidence level is nearly 1, it drops fast down to 0 in the interval $[0.5, 1.7]$. This means the probability that K get as big as 1.7 is very low. Such big K rejects the null hypothesis surely.

4.3.3 Kolmogorov-Smirnov test

Under null hypothesis that the sample comes from the hypothesized distribution $F(x)$,

$$\sqrt{n}D_n \xrightarrow{n\to\infty} \sup_t |B(F(t))| \tag{4.12}$$

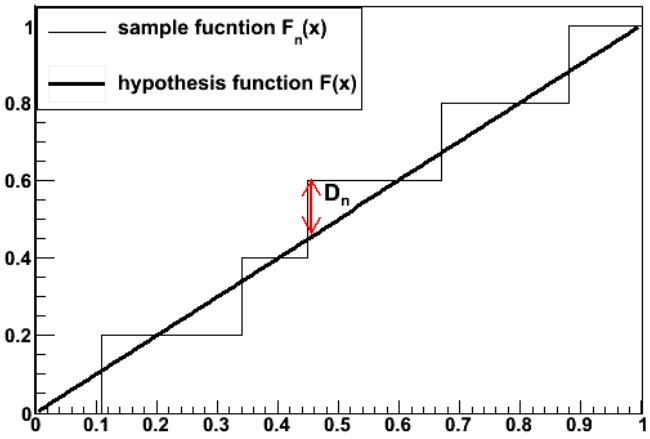


Figure 4.2: The sample function with the hypothesis function in the example. The D_n is the maximum deviation of the two functions. D_n is 0.15 in this case.

Figure 4.3: An example of Brownian bridge

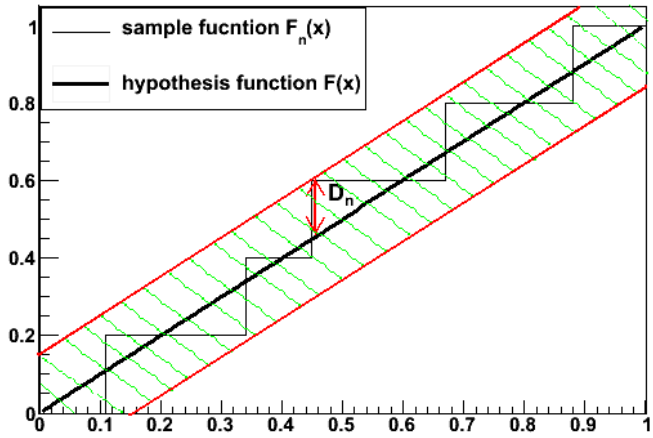


Figure 4.4: The two red lines above and below the hypothesis line are the lines with deviation of $\pm D_n$, and the shadow area is the space that $deviation < D_n$.

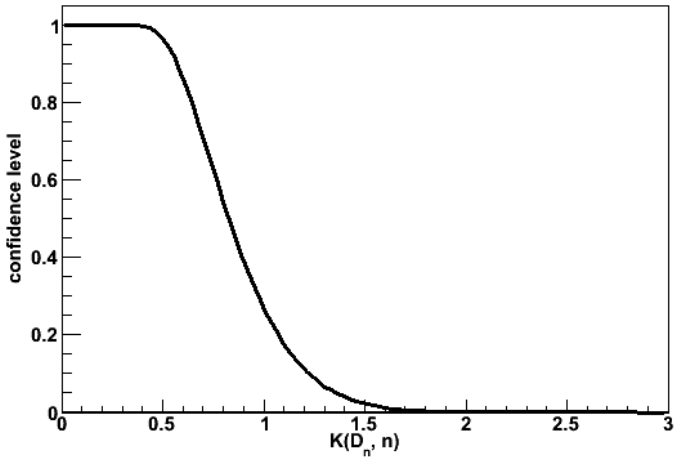


Figure 4.5: The Kolmogorov probability function. One must note that the y entry is not the probability with respect to K. It does not mean that the probability of K equals to 0.3 e.g. is nearly 1. By contraries, it means that the probability of $K > 0.3$ is nearly 1. In other words, it tells us how much the probability is to get a K as worse as the case. For example, when $K = 1.7$, y entry is nearly 0, this means the probability to get a K as large as 1.7 is nearly 0.

in distribution, where $B(t)$ is the Brownian bridge.

If F is continuous then under the null hypothesis $\sqrt{n}D_n$ converges to the Kolmogorov distribution, which does not depend on F. this result may also be know as the Kolmogorov theorem. The Kolmogorov-Smirnov test is constructed by using the critical values of the Kolmogorov distribution. The null hypothesis is rejected at level α if

$$\sqrt{n}D_n > K_\alpha, \tag{4.13}$$

where K is found from

$$Pr(K \leq K_\alpha) = 1 - \alpha \tag{4.14}$$

In the confidence level calculation (with confidence interval $x \in [K, \infty]$), it is equivalent to say that

$$K(D_n, n) = \sqrt{n}D_n \tag{4.15}$$

The asymptotic power of this test if 1. If the form or parameters of $F(x)$ are determined from the data X_j, the critical values determined in this way are invalid. In such cases, Monte Carlo or other methods may be required, but tables have been prepared for some cases.

4.3.4 Two-sample Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test may also be used to test whether two underlying one-dimensional probability distributions differ. In this case, the Kolmogorov-Smirnov statistic is

$$D_{n,n'} = \sup_x |F_n(x) - F_{n'}(x)| \tag{4.16}$$

and the null hypothesis is rejected at level α if

$$\sqrt{\frac{nn'}{n+n'}}D_{n,n'} > K_\alpha \tag{4.17}$$

which is equivalent to

$$\sqrt{\frac{nn'}{n+n'}}D_{n,n'} = K_\alpha \tag{4.18}$$

in confidence level calculation.

Note that the two-sample test checks whether the two data samples comes from the same distribution. This does not specify what that common distribution is (e.g. normal or not normal). Actually, if we treat the second sample as our hypothesis,

and let n' be infinity (for hypothesis is supported by infinite events), we can find that $\sqrt{\frac{nn'}{n+n'}}D_{n,n'} \xrightarrow{n'\to\infty} \sqrt{n}D_n$, which illustrate that the two-sample Kolmogorov-Smirnov test fails to one-sample Kolmogorov-Smirnov test. The standard Kolmogorov-Smirnov test can be treated as one special case of two-sample Kolmogorov-Smirnov test from this sense of view.

4.3.5 Steps to apply Kolmogorov-Smirnov test

The basic concepts about Kolmogorov-Smirnov test are given in previous sections, and now we summarise the steps to apply this test:

1. Prepare the sample empirical function, which is cumulative probability function, and normalise it. This function is $F_n(x)$ in our case.

2. Prepare the null hypothesis probability function, which is the normalised cumulative probability function of hypothesis distribution. We name it $F(x)$ in the previous part. In two-sample Kolmogorov-Smirnov test, this would be the second sample empirical function $F_{n'}(x)$.

3. Get the maximum deviation between this two functions, that is $D_n(x)$ by equation 4.7, or $D_{n,n'}(x)$ by equation 4.16 in case of two-sample K-S test.

4. By $D_n(x)$ and n, we get $K(D_n, n)$ in equation 4.15. For two-sample K-S test, use the equation 4.17.

5. Apply the Kolmogorov distribution probability function (equation 4.11) with $D_n(x)$ or $D_{n,n'}(x)$, we will get the confidence level for confidence interval $K \in [K_n, \infty]$ (or $K \in [K_{n,n'}, \infty]$ for two-sample K-S test).

In our case, we are testing the sample with hypothesis, so only a one-sample Kolmogorov-Smirnov test is applied, the two-sample test can be ignored.

Chapter 5

Testing Signal from Pierre Auger Observatory

The basic concepts and the steps of Kolmogorov-Smirnov test is given in last chapter, with a simple example given to help understanding the process. We can apply this method to the ultra-hight energy cosmic ray signal from Pierre Auger Observatory now.

Recall the content of previous chapters, we consider the same data as used in the recent update presented at the ICRC 2009 [38]. Specifically, an event is termed "correlating" for $\psi_{\max} = 3.1°$, $z_{\max} = 0.018$, and $E_{\rm th} = 55$ EeV, and all other events above $E_{\rm th}$ are called "non-correlating." The data are reconstructed with Herald v4r6p2f. We exclude the exploratory phase and regard the time period from 28 May 2006, the starting date of the prescription [39], to 31 March 2009, the termination date for data to enter the ICRC 2009 paper [38] (this time period corresponds to period II + III of Ref. [38]).

There are 17 correlating events in this period (as well as 27 non-correlating events $\Rightarrow$ total of 44 events). For each correlating event $i = 1...17$, we obtain the respective exposure value ϵ_i' of its detection from a web-interface ("exculator") provided by the acceptance group [40] choosing the "5T5" option. The exposure values are then taken relative to the starting date (28 May 2006), i.e. $\epsilon_i = \epsilon_i' - \epsilon_{\rm 28May06}$. The values ϵ_i of the events are listed in Tab. 5. The exposure value of the starting date (28 May 2006) is $\epsilon_{\rm start} = 0$ by definition, and of the ending date (31 March 2009) it is $\epsilon_{\rm end} = 12643$ km^2 sr yr.[1]

The selected events with their exposure are shown in Tab. 5.

5.1 Testing the correlation signal

Our trial task is to test the correlation signal with the hyphothesis, the normalized, cumulative step function of the data is plotted in Fig. 5.1 as a function of exposure between $[\epsilon_{\rm start} = 0, \epsilon_{\rm end}]$. That is, the step function starts at (0,0), increases by 1/17 at ϵ_i when a correlating event occured and ends at ($\epsilon_{\rm end}$,1).

We now test the hypothesis that the flux of correlating events is constant. Translated to the plot, this hypothesis is represented by the straight line between (0,0) and ($\epsilon_{\rm end}$,1). Then looks for the largest difference $D_{\rm n}$ between observation and hypothesis. The difference is maximal after the 9th correlating event at $\epsilon_9 = 4222$ km^2 sr yr with $D_{\rm n} = 9/17 - 4222/12643 \simeq 0.1955$.

[1] A simple estimate of the flux of correlating events then gives $\varphi_{\rm corr} \simeq 17/12643$ km^{-2} sr^{-1} yr^{-1} $\simeq 1.3 \times 10^{-3}$ km^{-2} sr^{-1} yr^{-1}.

Table 5.1: Event data: (i) running number j=1...44, (ii) Auger-ID, (iii) is the event a correlating one? (iv) exposure value ϵ_j (relative to 28 May 2006).

j	Auger-ID	correlating ?	ϵ_j (km^2 sr yr)
1	200618504666	yes	302
2	200629604873	yes	1222
3	200629900147	yes	1248
4	200701303199	no	1946
5	200705100383	yes	2309
6	200706903088	yes	2485
7	200708401685	yes	2637
8	200714500527	yes	3291
9	200718601091	no	3775
10	200719304931	yes	3865
11	200722101059	yes	4222
12	200723401180	no	4374
13	200723502725	no	4387
14	200729501686	no	5196
15	200734303548	no	5811
16	200734500536	no	5838
17	200801300140	no	6194
18	200801801764	no	6262
19	200803602961	no	6482
20	200805102810	no	6650
21	200805201552	no	6663
22	200808703016	yes	7122
23	200811801298	no	7551
24	200819202778	yes	8617
25	200820502900	no	8811
26	200826403841	no	9737
27	200826803082	no	9801
28	200828201372	yes	10010
29	200829605137	no	10236
30	200832203299	no	10651
31	200832805160	no	10741
32	200833700850	no	10884
33	200836200439	yes	11245
34	200900704718	no	11375
35	200903000345	no	11714
36	200903200633	no	11744
37	200903500506	no	11789
38	200903903934	yes	11851
39	200904702497	no	11976
40	200905104659	yes	12033
41	200907801387	yes	12456
42	200907802217	no	12456
43	200908004754	yes	12485
44	200908004974	no	12485

The KS probability $P_{\rm KS}$ to obtain values larger than $D_{\rm n}$ from statistical fluctuations can be determined by producing a large number of sets of 17 events each, picked at random according to the constant-flux hypothesis. This probability can be got by equation 4.11 if we have $K(D_n, n)$. Note that we are not going to use the equation 4.15 to get $K(D_n, n)$, but use the equation 5.1 instead[2]

$$K(D_n, n) = (\sqrt{n} + 0.12 + 0.11/\sqrt{n})D_{\rm n} \tag{5.1}$$

We obtain $P_{\rm KS} \simeq 47.6\%$ as the result. The equation 5.1 turns out to be more accurate with the simulated results, and it is given by [41]. For a cross-check, use the equation

[2]It is more accurate for small n, for large n it approaches equation 4.15

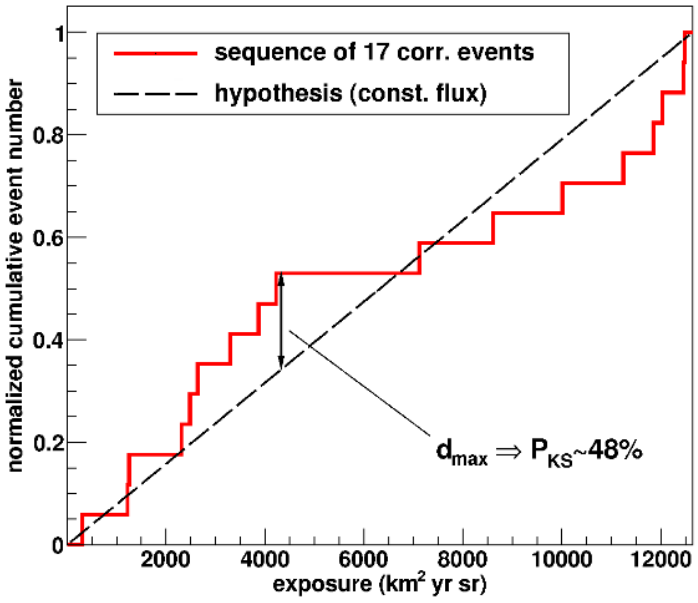


Figure 5.1: Normalized cumulative step function vs. exposure between 28 May 2006 and 31 March 2009 for the 17 correlating events. The straight line in each plot refers to the expectation in case of a constant flux.. The difference is maximal after the 9th correlating event at $\epsilon_9 = 4222$ km^2 sr yr with $D_\mathrm{n} =\simeq 0.1955$.

below:

$$P_\mathrm{approx}(x) = 2\sum_{j=1}^{\infty}(-1)^{j-1}e^{-2j^2x^2},\ x = (\sqrt{n} + 0.12 + 0.11/\sqrt{n})D_\mathrm{n} \tag{5.2}$$

where $n = 17$ and $D_\mathrm{n} \simeq 0.1955$.

There is an uncertainty connected to the exposure values. Due to the precise determination of the arrival times, and due to the relatively straightforward determination of the exposure of the array above threshold, uncertainties related to the exposure are generally expected to be small. Moreover, a feature of the KS test is that it is invariant under reparametrizations of the x-axis (exposure axis) such as applying a scaling factor or adding a constant term, all of which leaves d_max unchanged; thus, the effective uncertainty is likely to be smaller (and of somewhat different nature) than the 3% uncertainty generally quoted for the exposure determination [42]. The procedure [40] we used to obtain ϵ_i works on a daily basis (i.e. only full days are taken into account as the time of observation). However, the corresponding error due to the daily increment of exposure is below $\sim$20 km^2 sr yr. To estimate the effect of exposure uncertainties on P_KS, let us assume that d_max occured at a time

where the exposure was effectively off by $\Delta\epsilon \simeq 100$ km^2 sr yr (this would correspond to a neglected, or incorrectly added, exposure period of roughly one week). Then, $\Delta d_{\max} \simeq \Delta\epsilon/\epsilon_{\rm end} \simeq 0.008$, and $P_{\rm KS}(d_{\max} \pm \Delta d_{\max}) \simeq 43\%$ and 55%.

Thus, *the observation fits well to the hypothesis of a constant flux of correlating events.*

5.2 The non-correlation and total signal test

The same exercise was repeated for the non-correlating events and for all events (correlating and non-correlating events together; the all-event sample of course depends on the other two samples), they are listed in Tab. 5. Fig. 5.2 shows the empirical probability functions and D_n of the two samples. The KS probabilities are 2.7% for the non-correlating events and 4.8% for all events. One caveat related to the interpretation of the KS probabilities resulting from these subsequent analyses (after having analysed the correlating events) is that a "trial factor" needs to be kept in mind when analysing more than one data sample: when performing several trials (= analysing many samples) it is more likely to pick at least one fluctuation (= one small KS probability). Specifically, the probability to pick at least one KS probability value of $\leq$2.7% when performing two trials (sampling from the hypothesis to be tested) is $\sim$5.3% (and it is $\sim$9.4% for a value of $\leq$4.8%).

5.3 Simulated result

In the research work, we used Monte-Carlo method to simulate the stochastic process under the null hypothesis, which tells that the correlation signal comes at a constant rate (also the non-correlation and total signals).

In the simulation, we used the Kolmogorov-Smirnov statistic D_n, which is 0.1955 in case of correlation signal, 0.2744 in case of non-correlation signal and 0.2020 in case of all signal. We simulated for 1 million samples, counted out the number of the samples N that get a maximum deviation larger or equal to D_n, and get the probability by

$$P_{simu} = \frac{N}{1000,000} \tag{5.3}$$

We find that except there is a little difference between the results of correlation signal, the non-correlating and "all" signals are fitting with each other very well.

For comparation, see in the Tab. 5.2

5.4 Discussion

The value for the K-S probability of constant flux testing of correlating events is large. The signal rate deviation appearing between period 2 and 3 of correlating events is surely acceptable under the hypothesis of constant flux. This conclusion is,

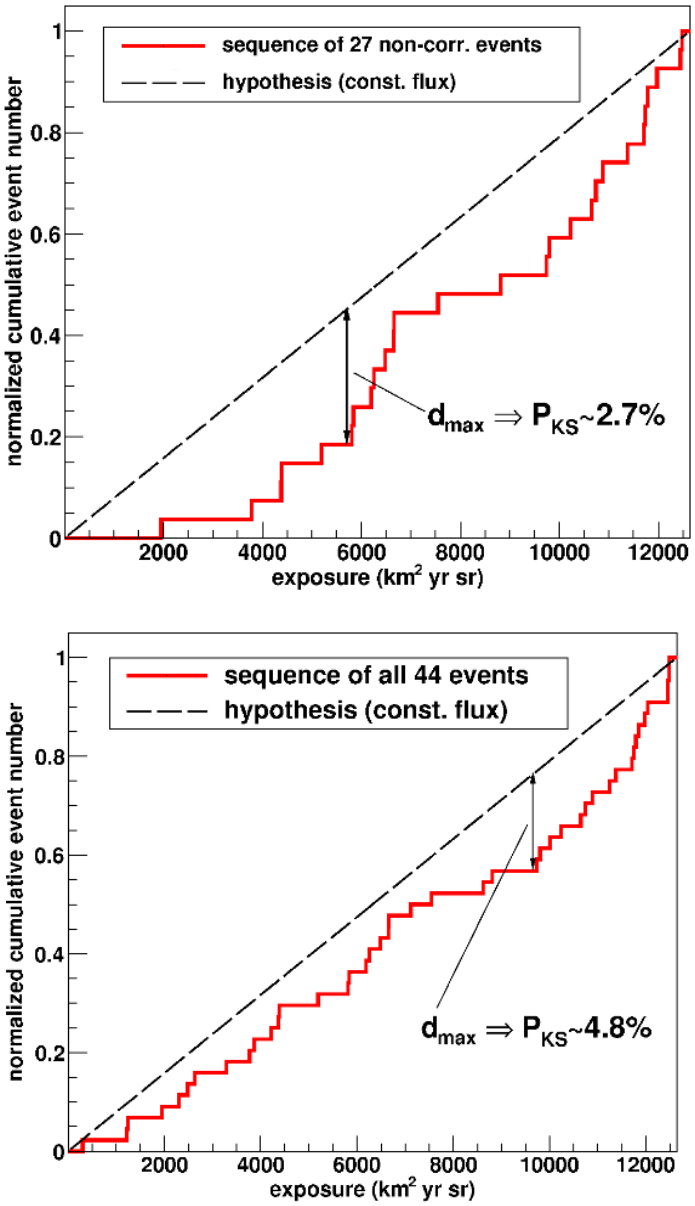


Figure 5.2: Normalized cumulative step function vs. exposure between 28 May 2006 and 31 March 2009 for the 27 non-correlating events (top), and all 44 events (bottom). The straight line in each plot refers to the expectation in case of a constant flux.

of course, limited by the small number of events. Still, It may put certain constraints on some scenarios of a possible extreme change of observation conditions. The KS probabilities for the non-correlating and for all events are smaller, particularly for the non-correlating events, with values at the few-percent level. As discussed above, a trial factor should be kept in mind when interpreting these values. We conclude that a rejection of the constant-flux hypothesis is not justified, but the relatively small probabilities motivate to "keep an eye" on the question whether the fluxes of non-correlating (and all) events are constant as expected. The approach introduced in this note can be used to monitor the behaviour.

There exists some possible speculation of a scenario of changing observation conditions while leaves the correlating events untouched but rather affects the non-correlating events. For instance, assume that the reconstructed energy scale of the selected events are slightly overestimated at later exposures, we can exclude some non-correlating events out by decreasing the energies of them. This will surely increase the testing result of the non-correlating events, while leaves the testing result of correlating evnets untouched. Related investigations are on going.

Comparing to other investigations of various aspects of the time evolution of the correlation signal (see e.g. [43]), our approach differs by using the quantity "exposure" as the effective detector time (instead of, e.g., total number of events); focusing just on one event sample (correlating events instead of ratio between correlating and non-correlating events); using a standard KS test (possible since exposure is a continuous observable); or using the full time period (instead of dividing the data set).

Table 5.2: Results from the KS test: (i) class of events, (ii) number of events N, (iii) maximum difference $d_{\max}$ between data and constant flux hypothesis, (iv) probability P_{KS} of MC simulation $d > D_{\mathrm{n}}$ (uncertainty due to MC statistics $\Delta P_{\mathrm{simu}} < 0.1\%$), (v) probability P_{approx} using the approximation formula Eq. (5.2), (vi) probability values when varying D_{n} by ± 0.01. The probability values resulting from the secondary analyses of the event classes "non-correlating" and "all" are put in brackets to indicate that they need to be interpreted with care (see text).

class of events	N	D_{n}	P_{simu}	P_{approx}	$P_{\mathrm{simu}}(D_{\mathrm{n}} \pm 0.01)$
correlating	17	0.1955	**47.6%**	49%	43%, 55%
non-correlating	27	0.2744	(2.7%)	(2.7%)	(2.1%, 3.5%)
all	44	0.2020	(4.8%)	(4.8%)	(3.5%, 6.4%)

Bibliography

[1] The Pierre Auger Collaboration. *Astropart. Phys.* 29:188-204, 2008.

[2] R.G. Harrison and D.B. Stephenson, Detection of a galactic cosmic ray influence on clouds, *Geophysical Research Abstracts*, Vol. 8, 07661, 2006 SRef-ID: 1607-7962/gra/EGU06-A-07661.

[3] J. C. Street and E.C. Stevenson. New evidence for the existance of a particle intermediate between th proton and electron. *Physical Revies*, 52:1003, 1937.

[4] W. Heitler, *Quantum Theory of Radiation.* Oxford University Press, 1944.

[5] Antonucci, R. (1993). "Unified Models for Active Galactic Nuclei and Quasars". *Annual Reviews in Astronomy and Astrophysics* 31 (1): 473–521.

[6] Urry, P.; Paolo Padovani (1995). "Unified schemes for radioloud AGN". *Publications of the Astronomical Society of the Pacific* 107: 803–845.

[7] HiRes Collaboration, D. R. Bergman, "Observation of the GZK cutoff using the HiRes detector," astro-ph/0609453.

[8] Daniel Kümpel. Geometry Reconstruction of Fluorescence Detectors Revisited. Diplome thesis.

[9] J. Bluemer, "Das nödliche Pierre Auger-Observatorium." DPG Frühjahrstagung, Heidelberg, 2007.

[10] Auger Collaboration, "The Pierre Auger Observatory design report,". Second Edition, 1997, (http://www.auger.org/technical info/design report.html).

[11] Čerenkov, P.A., "Visible Emission of Clean Liquids by Action of Radiation", Doklady Akad. Nauk SSSR 2 (1934) 451.

[12] Z. Szadkowski, K. H. Becker, and K.-H. Kampert, "Development of a new first level trigger for the surface array in the Pierre Auger Observatory based on the Cyclone Altera FPGA," *Nucl. Instrum. Meth.* A545 (2005) 793–802.

[13] Pierre Auger Collaboration, T. Suomijarvi, "The surface detectors of the Pierre Auger Observatory," Nucl. Phys. Proc. Suppl. 136 (2004) 393–398.

[14] Pierre Auger Collaboration, X. Bertou, "Performance of the Pierre Auger Observatory surface array," astro-ph/0508466.

[15] K.-H. Kampert, "The Pierre Auger observatory: Status and prospects," Nucl. Phys. Proc. Suppl. 151 (2006) 393–400, astro-ph/0501074.

[16] F. Kakimoto et al., "A Measurement of the air fluorescence yield," Nucl. Instrum. Meth. A372 (1996) 527–533.

[17] Pierre Auger Collaboration, R. Sato and C. O. Escobar, "The performance of the corrector lenses for the Auger fluorescence detector,". Presented at 29th International Cosmic Ray Conference (ICRC 2005), Pune, India, 3-11 Aug 2005.

[18] Pierre Auger Collaboration, J. A. Bellido, "Performance of the fluorescence detectors of the Pierre Auger Observatory," astro-ph/0508389.

[19] Pierre Auger Collaboration, C. Aramo et al., "Optical relative calibration and stability monitoring for the Auger fluorescence detector," astro-ph/0507577.

[20] Auger Collaboration, M. A. Mostafa, "Atmospheric monitoring for the Pierre Auger fluorescence detector," astro-ph/0308442.

[21] Pierre Auger Collaboration, F. Arqueros et al., "The central laser facility at the Pierre Auger observatory," JINST 1 (2006) P11003, astro-ph/0507334.

[22] Pierre Auger Collaboration, P. Allison et al., "Timing calibration and synchronization of surface and fluorescence detectors of the Pierre Auger Observatory,". Presented at 29th International Cosmic Ray Conference (ICRC 2005), Pune, India, 3-11 Aug 2005.

[23] Internet: "http://augersw1.physics.utah.edu/clf/, 06/27/07".

[24] M. Ave, Pierre Auger Collaboration. Astrophys, in: Proceedings of the 30th International Cosmic Ray Conference, Mérida, México, 2007. <arXiv:0709.2125>.

[25] C. Bonifazi, A. Letessier-Selvon, E. M. Santos, Astropart. Phys. 28 (2008) 523.

[26] C. Bonifazi,Pierre Auger Collaboration, in: Proceedings of the 29th International Cosmic Ray Conference Pune, India, vol. 7, 2005, p. 17.

[27] M. Roth, Pierre Auger Collaboration, in Proceedings of the 30th International Cosmic Ray Conference, Mérida, México, 2007. <arXiv:0706.2096>.

[28] M. -P. Véron, Astron. *Astrophys.* 22, 425 (1984).

[29] D. Allard et al. Pierre Auger Collaboration, Astrophysics.

[30] The Pierre Auger Collaboration. *Science* 318, 938 (2007).

[31] M.-P. Véron-Cetty, P. Véron, Astron. *Astrophys.* 455 (2006) 733.

[32] B. W. Lindgren. *Statistical Theory*, second edition, page 292.

[33] Weisstein, Eric W. "Hypothesis Testing." From MathWorld–A Wolfram Web Resource. "http://mathworld.wolfram.com/HypothesisTesting.html".

[34] B. W. Lindgren. *Statistical Theory*, second edition, page 325.

[35] J. V. Wall. From "http://nedwww.ipac.caltech.edu/level5/Wall2/Wal_contents.html".

[36] Shorak, G. R., Wellner J.A. (1986) *Empirical Processes with Applications to Statistics*, Wiley. ISBN 0-471-86725-X.

[37] Glasserman, Paul. Monte Carlo Methods in Financial Engineering, ISBN 0-387-00451-3, Springer-Verlag New York, 2004.

[38] D. Hague for the Pierre Auger Collaboration, 31th ICRC, Lodz (2009), available via <arXiv:0906.2347>.

[39] C. Covault and A. Letessier-Selvon for the AAARG, GAP-2006-096.

[40] Exculator, http://ipnweb.in2p3.fr/∼auger/AugerProtected/formulaire1.html.

[41] M.A. Stephens, Journ. Royal Stat. Soc. B **32**, 115 (1970), as cited in W.H. Press et al., *Numerical Recipes in C++*, Cambridge University Press (2002), 628ff.

[42] Pierre Auger Collaboration, *Trigger and Exposure of the Surface Detector Array of the Pierre Auger Observatory*, submitted to NIM A (2009); P.L. Ghia and I. Lhenry-Yvon, GAP-2009-092.

[43] For instance: F. Kuehn and P. Lebrun, GAP-2008-166; P. Lebrun and F. Kuehn, GAP-2009-030; P. Kasper, GAP-2009-089; H. Glass, GAP-2009-096; notes sent to pointsource task list, e.g. J. de Mello Neto and E.M. Santos (25 March 2009), P. Sommers (8 April 2009), P. Lebrun and F. Kuehn (13 July 2009).

Acknowledgements

I am thankful for support from many people. My parents supported my living in Germany for the first year. Prof. Dr. Zoltan Fodor and Prof. Dr. Karl-Heinz Kampert provided me a job to earn money for my living for the second year. The Wuppertal University exempted my tuition fees and gave me a scholarship during my thesis time. Great thanks for them giving me the finacial supports for my study and living abroad! Without their help, I would not finish my study for the master's degree.

Thanks to Prof. Dr. Francesco Knechtli. I gain a lot of knowledge from his lectures. And thanks to Prof. Dr. Karl-Heinz Kampert again. He inspired me and gave me a great insight to astro-particle physics. He offered me with the topic of this thesis, gave me continuous support even in busy times. I am indebted to Prof. Dr. Markus Risse. He gave me talks on the initial ideas with some important suggestions. Afterwards, a great amount of questions asked to him were imediately answered with patience. Special thanks to Dipl. -Phys. Nils Nierstenhöfer. He taught me a lot of programming knowledge in the beginning of my thesis, and kept to have meeting with me and Dr. Markus Risse discussing issues on the topics. Thanks to M.Sc. Pietro Oliva volunteering to help me. Thanks to Dr. Julian Rautenberg, and Dipl. -Phys. Daniel Fuhrmann. They all provided some precious help.

Heartful thanks to the Astroparticle Physics group of Wuppertal University. They gave me a warm welcome and I feel the friendly atmosphere when I was working and studying with them. Special thanks to the kind lady Ingrid Schaarwächter. She gave a lot of help in office problems. An additional thanks to Dipl. -Phys. Karl-Heinz Becker for his help of my computer when it went wrong.

I want to express my gratitude to my family and friends who concerned about me and kept encourage me to overcome the problems.

Thanks to Germany, thanks to University Wuppertal, thanks to the group. This would be one part of my most precious memory of all my life!